UC Irvine
UC Irvine Electronic Theses and Dissertations

Title

Modeling Internet Service Provider (ISP) Tier Design and Impact of Data Caps

Permalink

https://escholarship.org/uc/item/096857ds

Author

Dai, Wei

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Modeling Internet Service Provider (ISP) Tier Design and Impact of Data Caps

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Networked Systems

by

Wei Dai

Dissertation Committee:
Professor Scott Jordan, Chair
Professor Amelia C. Regan
Associate Professor Vidyanand Choudhary

2015

DEDICATION

To

my parents and wife with love

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

My PhD graduate study at UC Irvine is the most memorable experience in my life. It is my fortune to meet and work with so many smart people.

First, I would like to thank my PhD advisor Professor Scott Jordan for his great support and guidance through these five years. Without him continuing to serve as my PhD advisor after my initial advisor left UC Irvine, I cannot successfully finish my PhD research. I learned a lot from him, not only on how to solve challenging academic problems, but also on how to choose the right problem and direction in the first place. His devotion to research always inspires me, even after my graduation.

Secondly, I would like to thank my PhD committee members, Prof. Amelia Regan and Professor Vidyanand Choudhary. Their provided valuable insights into how to model ISP service tier design during my PhD topic defense and the afterward discussions. I would also like to thank Professor Nalini Venkatasubramanian and Professor Athina Markopoulou, who are on my PhD advancement committee. I learned a lot on distributed systems and algorithm design by taking their courses.

Thirdly, I would like to thank all my friends met at UC Irvine, Xiaobin Liu, Chao Yang, Zhijing Qin, Ye Zhao, Qiyuan Tang, Tongzhou Wang, Akihiro Enomoto and many others. Thanks for their help and making my life fun.

Last, but not the least, I would like to thank my wife Wei Zhang and my parents for their love during my PhD study. Without their support and encouragement, I cannot successfully finish this dissertation in the United States, far away from my hometown.

CURRICULUM VITAE

Wei Dai

2009~2015	**PhD in Networked Systems** <u>Advisor:</u> Prof. Scott Jordan <u>GPA:</u> 3.99/4.0	**University of California, Irvine** *Department of Computer Science* *Irvine, CA, US*
2006~2009	**MS in Electromagnetic Fields and Microware Technology** <u>Advisor:</u> Prof. Guiling Wu, Prof. Jianping Chen <u>GPA:</u> 3.69/4.0	**Shanghai Jiao Tong University** *Department of Electronic Engineering* *Shanghai, China*
2002~2006	**BS in Information Engineering** <u>GPA:</u> 3.41/4.0	**Shanghai Jiao Tong University** *Department of Electronic Engineering* *Shanghai, China*

RESEARCH EXPERIENCE

Sep. 09~Jun. 11   **University of California, Irvine**
Irvine, CA: *Research Assistant/Teaching Assistant*

Sep. 07~Mar. 09   **Shanghai Jiao Tong University**
Shanghai, China: *Research Assistant*

WORK EXPERIENCE

Jun. 14~Sep. 14   **Arista Networks**
Santa Clara, CA: *Software Intern*

Jun. 11~Sep. 11   **Siemens PLM Software**
Cypress, CA: *Software Intern*

FIELD OF STUDY

Network economics, Internet policies and network performance optimization.

PUBLICATIONS

1. <u>Wei Dai</u>, Ji Won Baek and Scott Jordan. "The Impact of Data Caps on ISP Competition," *5th International Conference on Game Theory for Networks (GameNets)*, Beijing, China, Nov. 2014.

2. <u>Wei Dai</u>, Scott Jordan. "Design and Impact of Data Caps", *IEEE Global Communications Conference (Globecom)*, Atlanta, GA, US, Dec. 2013.

3. <u>Wei Dai</u>, Scott Jordan. "How Do ISP Data Caps Affect Subscribers?" *Telecommunications Policy Research Conference (TPRC)*, Virginia, US, Sep. 2013.

4. <u>Wei Dai</u>, Scott Jordan. "Modeling ISP Tier Design", *25th International Teletraffic Congress (ITC)*, Shanghai, China, Sep. 2013.

5. <u>Wei Dai</u>, Guiling Wu, etc. "Offset Time Configuration in Optical Burst Switching Ring Networks", *IEEE/OSA Journal of Lightwave Technology*, vol. 27, no. 19, Oct. 2009.

6. <u>Wei Dai</u>, Guiling Wu, etc. "Traffic Tree Based Burst Scheduling in Optical Burst Switching Network", *IEEE Communications Letters*, vol. 12, no. 8, Aug. 2008.

7. <u>Wei Dai</u>, Guiling Wu, etc. "Resource Scheduling in Centralized OBS Based Grids", *OSA Journal of Optical Networking*, vol. 7, no. 2, Feb. 2008.

8. Tairan Zhang, <u>Wei Dai</u>, etc. "A Dual Price-based Congestion Control Mechanism for Optical Burst Switching Networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 32, no. 14, July 2015.

9. Guiling Wu, <u>Wei Dai</u>, etc. "A maximum-efficiency-first multi-path route selection strategy for optical burst switching networks," *Optik - International Journal for Light and Electron Optics*, vol. 125, no. 10, pp. 2229-2233, May 2014.

10. Yi Shen, <u>Wei Dai</u>, etc. "A Matlab toolbox for the efficient estimation of the psychometric function using the Updated Maximum-Likelihood adaptive procedure," *Behavior Research Methods*, DOI 10.3758/s13428-014-0450-6, Mar. 2014.

11. Yahong Yang, Guiling Wu, <u>Wei Dai</u>, etc. "Joint congestion control and processor allocation for task scheduling in grid over OBS networks," *Expert Systems with Applications*, vol. 38, issue 7, July 2011.

12. Yahong Yang, Guiling Wu, Jianping Chen, <u>Wei Dai</u>. "Multi-objective optimization based on ant colony optimization in grid over optical burst switching networks," *Expert*

Systems with Applications, vol. 37, issue 2, Mar. 2010.

13. <u>Wei Dai</u>, Scott Jordan. "ISP Service Tier Design," *IEEE/ACM Transactions on Networking*, under revision.

14. <u>Wei Dai</u>, Scott Jordan. "The Effect of Data Caps upon ISP Service Tier Design and Users," *ACM Transactions on Internet Technology*, under revision.

ABSTRACT OF THE DISSERTATION

Modeling Internet Service Provider (ISP) Tier Design and Impact of Data Caps

By

Wei Dai

Doctor of Philosophy in Networked Systems

University of California, Irvine, 2015

Professor Scott Jordan, Chair

In this dissertation, we focus on the design of tiered pricing plans offered by Internet Service Providers (ISPs). We also analyze the impact of data caps on the pricing plans, users, ISP profits and social welfare, by considering ISP monopoly and ISP duopoly, respectively.

The initial work is about modeling ISP service tier design without data caps. Web browsing and video streaming are considered as the two dominant Internet applications. We propose a novel set of utility functions that depend on a user's willingness to pay for each application, the performance of each application, and the time devoted to each application. For a monopoly provider, the demand function for each tier is derived as a function of tier price and performance. We first give conditions for the tier rates, tier prices, and network capacity that maximize Internet Service Provider profit, defined as subscription revenue minus capacity cost. We then show how an Internet Service Provider may simplify tier and capacity design, by allowing their engineering department to set network capacity, their marketing department to set tier prices, and both to jointly set tier rates.

The next work is focused on how ISPs choose data caps and the resulting impact on users. We propose a data cap model by extending the ISP tier design model. A monopoly ISP is presumed to maximize its profit by controlling tier prices, tier rates, data caps and overage charges. We show how users fall into five categories: non-Internet subscribers, basic tier subscribers, premium tier subscribers unaffected by a data cap, premium tier subscribers who are capped but do not choose to exceed the cap, and premium tier subscribers who exceed the cap and pay overage charges. When data caps are used for profit maximization, we find that the monopoly ISP has the incentive to keep the basic tier price and basic tier rate unchanged, to increase the premium tier rate, and to reduce the premium tier price. The ISP also has the incentive to set smaller caps and higher overage charges than when caps are used only to ensure that heavy users pay for their usage.

Finally, we analyze the impact of data caps on ISP duopoly competition. In the duopoly competition model, users seek to maximize their surplus by making their ISP subscription choices and by controlling the time devoted to Internet activities. ISPs seek to maximize their profits by competing through their tier prices, tier rates, network capacities, data caps and overage charges. We illustrate how users' utilities are affected by data caps, and the resulting impact upon ISP market shares. We show that both ISPs have an incentive to use data caps and overage charges to ensure that heavy users pay at least an amount equal to the cost of their usage. The initial incentives for both ISPs to update their tier prices and tier rates given the newly added data caps are also predicted under different scenarios. The final Nash equilibrium with data caps for profit maximization is analyzed through simulation, and compared to the Nash equilibrium without data caps. The corresponding changes in ISP profits, user subscription choices, and user surpluses are illustrated.

CHAPTER 1: INTRODUCTION

SUBCHAPTER 1.1: Motivation and Challenge

The service tiers offered by Internet Service Providers (ISPs) are differentiated by their monthly price and their maximum downstream transmission rate. In recent years, ISPs have begun to market these tiers on the basis of the dominant applications of each tier's intended subscribers. Most ISPs now offer a basic tier which they market as good for web browsing and email, an intermediate tier which they market as also good for file sharing, and a higher tier which they market as also good for video streaming, see e.g. [1][2].

Although these service tiers used to be charged by a flat-rate Internet plan independent of data consumption, it has become common for wireless Internet Service Providers in the United States to place caps on the monthly usage of cellular data plans. Some wireline ISPs have also started placing data 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 from 50GB to 500GB per month [3]. 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.

It is important to investigate the ISP tier design problem. The tier offerings have significant influence on the Internet. They affect both application development and application use. Residential Internet use in countries with high tier rates is substantially different than in countries with low tier rates. Mobile Internet use in countries with flat rate pricing is substantially different than in countries with usage based pricing. Internet use in turn

affects what types of applications are developed. Thus, both technical and economic issues have a great impact on the Internet's development. And yet, the research literature provides little guidance as to why ISPs offer the tiers they do. Most existing Internet pricing models focus either on economic issues or on technical performance issues, but usually not both.

It is also necessary to analyze the data cap issue, because there is a vigorous debate over the use of caps. ISPs commonly claim that caps benefit most users. They cite statistics [4] 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 [5]. 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 [6]. They further claim that caps affect only a small percentage of heavy users [7], 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 will benefit a broader set of users [8].

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 [9]. They further claim that caps encourage a climate of scarcity, and that ISPs can increase their profit through the use of caps principally because of a lack of consumer choice in broadband providers [10]. 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 [11].

The debate has entered the public arena. Some public interest groups have called for government oversight [12]; in particular some have asked the US Federal Communications Commission (FCC) to investigate AT&T's broadband data caps [13] and its sponsored data agreement [14]. A US 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 [15]. The FCC Open Internet Advisory Committee examined policy issues in data caps; it noted that data caps are a source of concern where there are no or few substitutes for Internet access, but the lack of data about user behavior in the marketplace makes it difficult to draw conclusions about the role of data caps in competition [16].

In sum, we seek to answer the following questions related to the design of ISP service tiers and data caps:

1. How does an ISP determine the optimal number of tiers in a tiered Internet pricing plan?

2. How does an ISP determine the optimal rate and price of each tier, and when it should upgrade the underlying network capacity?

3. How may ISP engineering and marketing departments cooperate with each other to find near-optimal tier prices, tier rates and network capacity in a simplified design?

4. How does an ISP set data cap and overage charge in each tier, in the ISP monopoly case?

5. What is the impact of data cap on pricing plans and subscribers, in the ISP monopoly case?

6. How does the use of data caps affect the competition between ISPs? What is the effect of competition upon the adoption and use of data caps?

SUBCHAPTER 1.2: Background and Related Work

A. Literature on ISP Service Tier Design

A number of papers investigate the ISP tier design problem using economic models. A survey of a portion of this research literature can be found in Viswanathan and Anandalingam [17], including price discrimination, bundling and versioning. The literature on price discrimination topic is relevant to service tier design. Price discrimination potentially allows a provider to charge different prices based on users' willingness-to-pay (first degree), usage (second degree), or performance (third degree) [18]. In this context, ISP service tier design may be considered as third degree price discrimination. However, ISP service tier design is more complex than traditional third degree price discrimination models, as the ISP is offering a platform that supports multiple applications, each of which has a different mapping of capacity to performance. Furthermore, application performance affects usage, which in turn affects willingness-to-pay, thereby conflating first, second, and third degree price discrimination.

Another set of papers investigates the ISP tier design problem by analyzing the relationship between users, ISPs, and content providers, see e.g. [19][20]. Two-sided models, which take into account payments between users, ISPs, and content providers, are typically central to

the analysis. However, Internet architecture and topology are typically abstracted into a simple connectivity model. Applications are rarely modeled with respect to either traffic or utility. Similarly, congestion and traffic management are rarely considered.

A third set of papers investigates the ISP tier design problem using traditional performance models. The focus is on usage based pricing, see e.g. [21][22][23]. Traffic management models which take into account user traffic and congestion control are typically central to the analysis. However, economic aspects are typically abstracted into the revenue generated by the usage based pricing, and tier choice is rarely modeled.

A fourth set of papers analyzes ISP tier design for monopolies. He and Walrand [24] consider Internet service offered at multiple quality levels and prices, and obtain equilibrium results by modeling a game between users maximizing surplus under fixed prices; they also consider revenue maximization for a monopoly ISP. Lv and Rouskas [25][26] also consider Internet service offered at multiple quality levels and prices, but assume that users choose the highest tier they can afford; they propose an algorithm that attempts to maximize a monopoly ISP's profit under a fixed cost per user by controlling quality levels and prices. These papers model user willingness-to-pay solely as a function of an aggregated service quality (e.g. bandwidth) and sometimes a user dependent parameter. In the proposed algorithms, the number of the service levels (tiers) is assumed to be a fixed small number.

The last set of papers considers cooperation and competition between multiple ISPs. Although they can answer some ISP tier design questions, modeling ISP tier design is not the focus of these papers. Gibbens et. al. [27] consider competition between two ISPs that

may offer multiple service tiers by setting tier capacity and price. They find that although a monopolist may form multiple service tiers, two competing ISPs will not. Shetty et. al. [28] extend this model by modeling the cost of network capacity, and find that two service classes are now optimal.

B. Literature on Data caps

There is little related academic literature on data caps. Some papers seek to address the impact of data caps. Sen et al. [29] 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 main focus of this paper. Light users and heavy users are not differentiated in the model. Nor do they consider how data caps can affect network congestion or application performance, which is the reason that ISPs give for adopting data caps. Minne [30] 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 [31] evaluates the merits of data caps and other usage-based pricing strategies. They argue 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. [32] 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. [33] 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. Some papers approach the topic using mathematical modeling. Although Nabipay et al. [34] does not directly model data caps, it models a monopoly service provider with negligible marginal costs, and shows that flat rate pricing typically maximizes profit even in the presence of a small number of heavy users. Wei and Jordan [35] model the use of data caps by a monopoly ISP, and analyzes the resulting change on user surpluses, ISP profit and social welfare. Finally, there is a related literature on time-dependent Internet pricing, see [36].

There is even a smaller academic literature that examines the how caps affect, or affected by, the amount of competition between ISPs. Economides and Hermalin [37] model a monopoly ISP that imposes download limits on content consumers, and argues that the adoption of data caps may intensify competition among content providers, thereby generating greater surplus for consumers; however, the model does not directly model competition among multiple ISPs. In contrast, Waterman et al. [32] expresses concerns that data caps may be anticompetitive behavior in the online television market; however, data caps are not the focus of the paper and hence they are not modeled.

Some work can provide insights into how to model data caps, although each approach has some limitations. A set of papers investigate user decisions between multiple Internet technologies. Joseph et al. [38] and Sen et al. [39] 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. [40] 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. 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 consumption, 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.

Another set of papers investigates user choices between differentiated information goods by using existing decision models, statistical models or pricing models. Eikebrokk and Sorebo [41] 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. [42] use statistical models to analyze the impact of a variety of socio-economic 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 [43] 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.

A final set of papers investigates flat-rate, usage-based (including data caps) and time-dependent pricing plans. Sen et al. provide a survey on time-dependent Internet pricing plans [44]. 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. [45] investigate the design of time-dependent pricing plans. Although the method used to analyze the impact of time-dependent pricing on user behaviors is useful for data caps analysis, data caps are not the focus of that paper.

SUBCHAPTER 1.3: Contributions and Dissertation Structure

The major contributions of this dissertation are:

1. 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. (CHAPTER 2)

2. We introduce a basic model to investigate how an ISP determines the optimal number of tiers in a tiered Internet pricing plan. Closed form expressions are derived to show the condition under which an ISP will offer more than a single tier in this basic model. (CHAPTER 2)

3. We introduce an extended model to investigate how an ISP determines the optimal rate and price of each tier, and when it should upgrade the underlying network capacity. These models may be useful to ISPs, networking researchers and Internet policymakers. (CHAPTER 2)

4. We present what we believe is the first model in the academic literature of how an ISP may set data caps and overage charges. (CHAPTER 3)

5. We analyze the impact of caps on pricing plans and subscribers, by characterizing how tier rates and tier prices change in the presence of data caps, and which users benefit from or are hurt by caps. (CHAPTER 3)

6. We propose an ISP duopoly model to investigate how two ISPs compete with each other through tier prices, tier rates, overage charges and data caps. (CHAPTER 4)

7. We analyze the effect of competition upon the adoption of data caps, and how the use of data caps affects competition between ISPs. The impact of data caps under duopoly competition upon social welfare is also investigated. (CHAPTER 4)

This dissertation is organized as follows. In CHAPTER 2, we first propose a basic model for ISP tier design. Conditions under which an ISP will offer multiple tiers are analyzed. We then extend the basic model by considering more general user utility functions and more complex ISP traffic management. In the extended model, we derive the user demand for each tier and the density function of different users' willingness-to-pay in the market. We further explain how an ISP may simplify tier and capacity design, by decomposing the network capacity and tier design problem into three sub-problems for the ISP engineering and marketing departments. We finally present numerical results that illustrate the

variation on the design with key parameters, as well as the magnitude of the decrease in profit resulting from such a simplified design.

In CHPATER 3, we model the impact of data caps by incorporating data caps into the extended ISP tier design model in CHAPTER 2. 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. 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. The impact of caps, overage charges, and the corresponding tiered pricing plans on various subscribers are also analyzed. Numerical results are presented to illustrate how the tier rate, tier price, cap, and overage charges vary with the standard deviation in Internet usage amongst subscribers.

In CHAPTER 4, ISP duopoly competition is formulated as a non-cooperative game, where each ISP seeks to maximize its profit. We derive closed from conditions for natural ISP monopoly and ISP duopoly. We also partition users into three groups: users who are not capped, users who are capped without paying an overage charge, and users who are capped and paying an overage charge. The market share of each ISP and the Nash equilibrium of the ISP duopoly competition are derived from the model. We finally analyze the impact of data caps on the pricing plan, ISP profits, users and social welfare under ISP duopoly competition, through multiple case analyses and simulation.

In CHAPTER 5, we draw conclusions and discuss possible future work related to this dissertation.

CHAPTER 2:   ISP Service Tier Design

In this chapter, we decompose user willingness-to-pay for Internet access into the willingness to pay for two major set of applications: web browsing and video streaming, where two sets of novel utility functions are proposed for ISP service tier design. Rather than assuming a fixed number of tiers, we introduce a basic model to investigate how an ISP determines the optimal number of tiers in a tiered Internet pricing plan. Closed form expressions are derived to show the condition under which an ISP will offer more than a single tier in this basic model.

We also introduce an extended model to investigate how an ISP determines the optimal rate and price of each tier, and when it should upgrade the underlying network capacity. Compared to existing models, rather than modeling user willingness-to-pay solely as a function of bandwidth, we extend it as a joint function of bandwidth, performance, and the time devoted to each application. Rather than assuming a sunk cost or a fixed cost per user, we consider a general cost as a function of network capacity. Rather than considering only the cost of the tier in user surplus, we also consider the value a user places on time. The proposed utility function in the extended model is general enough to be applied to models with or without competition between ISPs.

Two interconnected problems separated by time scale are considered in the extended model. On a time scale of days, broadband Internet subscribers choose how much time to devote to web browsing and video streaming. On a time scale of months, ISPs choose what tiers to offer, and potential broadband Internet users choose tiers. The dependences of ISP profit on tier prices, tier rates and network capacity are derived from the extended model.

We also show how ISP engineering and marketing departments may cooperate with each other to find near-optimal tier prices, tier rates and network capacity in a simplified design. The magnitude of the decrease in profit resulting from such a simplified design is analyzed by changing certain key parameters in the simulation.

These models may be useful to ISPs, networking researchers and Internet policymakers. Although ISP service tier design is proprietary, we suspect that ISPs may improve upon their service tier designs using some of the ideas presented here. In particular, the models presented here suggest that while portions of the service tier design may be the sole domain of ISP engineering and marketing departments, some elements of the design are more effectively designed jointly by these departments. The models may also be of interest to networking researchers. Many networking research problems are affected by ISP service tier design. However as such designs are proprietary, the models presented here may serve as a public domain substitute for research purposes. Finally, these models may be of interest to Internet policymakers. In particular, the current debate over the use of data caps can be explored using such service tier models.

SUBCHAPTER 2.1: A Basic ISP Service Tier Model

The dominant applications on North American fixed access broadband Internet access networks, as measured by download traffic volume, are real-time entertainment, web browsing, and peer-to-peer (p2p) file sharing, which together account for approximately 85% [4]. Real-time entertainment traffic consists almost exclusively of video streaming. For the purposes of analysis, we split p2p into two subsets: p2p streaming, which we aggregate with other video streaming [46], and p2p file sharing, which we aggregate with web

browsing [46]. 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 paper on two applications: web browsing and video streaming.

In the basic model, we model user utility on both sets of Internet applications solely as functions of throughput performance. Users are characterized by their willingness-to-pay for each application. An ISP seeks to maximize its profit by setting the optimal number of tiers, as well as the rate and price in each tier.

A. Modeling Users

User i's willingness to pay for Internet (W_i) is composed of the willingness to pay for web browsing (W_i^b) and video streaming (W_i^s):

$$W_i = W_i^b + W_i^s$$

Most networking papers classify web browsing as an elastic application and model utility as an increasing concave function of throughput, see e.g. [47][48]. For the sake of simplicity, we propose to model user i's satisfaction with the quality of web browsing as a step function $Q^b\left(x_i^b\right)$ of throughput x_i^b:

$$Q^b\left(x_i^b\right) = \begin{cases} 0, & x_i^b < \overline{x}^b \\ 1, & x_i^b \geq \overline{x}^b \end{cases} \tag{2.1}$$

15

Shapes other than a step function will be considered in the extended model presented in SUBCHAPTER 2.3. Different users in the market will place different values upon web browsing. Thus, user i's willingness to pay for web browsing can be expressed as:

$$W_i^b = w_i^b Q^b \left(x_i^b \right)$$

where $w_i^b > 0$ is the amount of money user i is willing to pay for web browsing when the quality of web browsing $Q^b \left(x_i^b \right) = 1$.

Video streaming is commonly classified as a semi-elastic application, and its utility is modeled by a sigmoid function of throughput [49]. Similar to (2.1), we propose to model user i's satisfaction with the quality of video streaming as a step function $Q^s \left(x_i^s \right)$ of video streaming throughput x_i^s, where the throughput threshold is $\overline{x}^s$. This step function will be replaced by a sigmoid function in the extended model below. Thus, user i's willingness to pay for video streaming can be expressed as:

$$W_i^s = w_i^s Q^s \left(x_i^s \right)$$

where $w_i^s > 0$ is the amount of money user i is willing to pay for video streaming when $Q^s \left(x_i^s \right) = 1$.

The two throughput thresholds are presumed to satisfy $\overline{x}^s > \overline{x}^b$, since video streaming consumes more bandwidth than web browsing.

16

Different users in the market place different values on web browsing and video streaming. To model the market, denote the density function of users' willingness to pay for web browsing by $f_{w^b}\left(w^b\right)$. We assume $f_{w^b}\left(w^b\right) > 0$ on w^b>0. It is very likely that w^b and w^s will be positively correlated, because users who spend more time on web browsing are more likely to watch more videos online. In the basic model, we assume that there exists a fixed relationship between w^b and w^s for the users in the market:

$$w^s = g\left(w^b\right) \tag{2.2}$$

where $g()$ is a twice *continuously* differentiable monotonically increasing function such that g(0)=0. More general forms of correlation will be considered in the extended model.

B. Modeling an ISP

Instead of using priority queue based quality of service technology, we consider traditional tiered services, where a cap is added to the network download rate performance. The throughput of each tier is thus constrained by the tier rate, which is defined as the maximum allowed network download rate, while all tiers are experiencing the same load of the underlying network:

$$x_i^b = \min\left(X_{T_i}, TCP^b\left(\rho\right)\right), \quad x_i^s = \min\left(X_{T_i}, TCP^s\left(\rho\right)\right) \tag{2.3}$$

where T_i is user i's tier choice, X_{T_i} is the corresponding tier rate, ρ is the traffic load of the underlying network, and $TCP^b(\rho)$ and $TCP^s(\rho)$ are the maximum throughputs allowed by the TCP or TCP-friendly protocols adopted in web browsing and video streaming respectively [50][51]. The maximum throughput functions are complex; what matters here

17

is simply that $TCP^b(\rho)$ and $TCP^s(\rho)$ are non-increasing functions of the traffic load ρ. Since the quality functions $Q^b\left(x_i^b\right)$ and $Q^s\left(x_i^s\right)$ are assumed to be step functions of throughput, the ISP thus only need maintain the traffic load below certain thresholds. To maintain the throughput of web browsing (resp. video streaming) above its threshold, the network load ρ must be kept below $\rho^b = TCP^{b-1}\left(\overline{x}^b\right)$ (resp. $\rho^s = TCP^{s-1}\left(\overline{x}^s\right)$).

Denote the network capacity by μ. Denote the cost per month for the network capacity by a linear function $C(\mu) = p^\mu \mu + K$, where p^μ is the marginal network capacity cost and K is the fixed network cost. Denote the proportion of the month users devote to web browsing (resp. video streaming) by t^b (resp. t^s)[1]. (Later in the paper, we will model the times devoted by individual users.)

An ISP has the incentive to offer users one or two tiers, with prices, tier rates and network capacity determined to maximize the ISP's profit:

$$\max_{P_1,P_2,X_1,X_2,\mu} Profit = P_1 N_1 + P_2 N_2 - p^\mu \mu - K \tag{2.4}$$

where N_i denotes the number of users who subscribe to tier i and P_i denotes the price of tier i. (We order the tiers so that $X_2 \geq X_1$ and correspondingly $P_2 \geq P_1$; if an ISP offers users only one tier, this is denoted by $P_2 = P_1$ and $X_2 = X_1$.) When an ISP chooses the network capacity μ and the tier rates (X_1, X_2), it determines the network load ρ through:

$$\rho = \left(\left(N_1 + N_2\right)\overline{x}^b t^b + N_2 \overline{x}^s t^s\right)\Big/\mu$$

---

1 For example, if 20 hours are devoted to web browsing per month, then $t^b = 20/(30\times24)$.

SUBCHAPTER 2.2: One Tier Or Two Tiers?

In this section, we seek to determine under what conditions an ISP has the incentive to offer more than one tier for profit maximization.

A. General Results

We first seek to determine whether an ISP will set service tier rates high enough so that video streaming applications have sufficient performance. Denote the cumulative density function of users' willingness to pay for web browsing by $F_{w^b}(w) = \int_0^w f_{w^b}(w')dw'$.

Theorem 2.1: Assume an ISP seeks to maximize its profit as in (2.4). The ISP will set tier 1 rate $X_1 = \overline{x}^b$. If $\rho^b \leq \rho^s$, or if $\rho^b > \rho^s$ and

$$p^\mu \leq \max_{\Delta P > 0} P^\mu(\Delta P)$$

$$\text{where } P^\mu(\Delta P) \triangleq \Delta P \rho^s \bigg/ \left(1 + \frac{1}{F_{w^b}\left(g^{-1}(\Delta P)\right)} \frac{\overline{x}^b t^b}{\overline{x}^s t^s}\left(1 - \frac{\rho^s}{\rho^b}\right)\right) \overline{x}^s t^s \tag{2.5}$$

the ISP will set tier 2 rate $X_2 = \overline{x}^s$, and set network capacity

$$\mu = \left((N_1 + N_2)\overline{x}^b t^b + N_2 \overline{x}^s t^s\right) \bigg/ \min\{\rho^b, \rho^s\}.$$

Proof: Users' willingness-to-pay for web browsing (resp. streaming) is insensitive to throughputs above $\overline{x}^b$ (resp. $\overline{x}^s$). It follows that an ISP will set the tier 1 rate $X_1 = \overline{x}^b$, and if it offers two tiers, will set tier 2 rate $X_2 = \overline{x}^s$.

When $\rho^b \leq \rho^s$, it is straightforward to show that an ISP maximizes profit by offering two tiers, maintaining network load $\rho = \rho^b$, resulting in $\mu = \left((N_1 + N_2) \bar{x}^{-b} t^b + N_2 \bar{x}^{-s} t^s \right) \Big/ \min \{ \rho^b, \rho^s \}$.

When $\rho^b > \rho^s$, first suppose an ISP maintains network load $\rho = \rho^b$. Denote the corresponding profit-maximizing tier 1 price by P_{10}. In this case, video streaming applications do not have sufficient performance since $Q^s(x_i^s) = 0$, and thus the ISP offers a single tier.

Now consider the alternative, that the ISP reduces the traffic load ρ^b from ρ^b to ρ^s and offers tier 2 with rate $X_2 = \bar{x}^s$, at a price $P_{20} = P_{10} + \Delta P_0$, where $\Delta P_0 = \arg \max_{\Delta P > 0} P^\mu(\Delta P)$. The change in profit resulting from the offering of a second tier can be shown to be:

$$\Delta Profit = N_2 \left(\Delta P_0 - \frac{p^\mu}{\rho^s} \bar{x}^s t^s \right) - (N_1 + N_2) \left(\frac{p^\mu}{\rho^s} \bar{x}^{-b} t^b - \frac{p^\mu}{\rho^b} \bar{x}^{-b} t^b \right),$$

where the first term is the extra profit that the ISP earns from users who upgrade from tier 1 to tier 2, and the second term is the extra cost for network capacity to reduce the traffic load from ρ^b to ρ^s. If (2.5) holds, then $\Delta Profit > 0$, and $\mu = \left((N_1 + N_2) \bar{x}^{-b} t^b + N_2 \bar{x}^{-s} t^s \right) \Big/ \rho^s$.

∎

Theorem 2.1 indicates that ISP will set service tier rates high enough so that video streaming applications have sufficient performance if the marginal network capacity cost is smaller than a certain threshold. We will verify that (2.5) likely holds in today's Internet later in the section.

20

We then seek to answer whether an ISP will offer two differentiated tiers under the assumption that (2.5) is true (i.e. ISP will provide video streaming service). According to Theorem 2.1, since $Q^b\left(x_i^b\right)$ and $Q^s\left(x_i^s\right)$ are assumed to be step functions, ISPs are constrained to provide web browsing and video streaming with throughput performance $\overline{x}^b$ and $\overline{x}^s$, respectively. Thus, user i's tier choice, T_i, depends on the prices of each tier and her willingness-to-pay for each tier:

$$T_i = \begin{cases} 2, & \text{if } w_i^b + w_i^s - P_2 > \max\left(w_i^b - P_1, 0\right) \\ 1, & \text{if } w_i^b - P_1 > \max\left(w_i^b + w_i^s - P_2, 0\right) \\ 0, & \text{otherwise} \end{cases} \tag{2.6}$$

Our first result characterizes the number of subscribers to tier 1, N_1.

Lemma 2.1: $N_1 > 0$ iff $P_2 - P_1 > g(P_1)$, $N_1 = 0$ iff $P_2 - P_1 \leq g(P_1)$.

Proof: Let $h\left(w_i^b\right) = w_i^b + g\left(w_i^b\right)$. Then $P_2 - P_1 > g(P_1)$ iff

$$g^{-1}\left(P_2 - P_1\right) > P_1 \Leftrightarrow g^{-1}\left(P_2 - P_1\right) + g\left(g^{-1}\left(P_2 - P_1\right)\right) > P_2$$
$$\Leftrightarrow h\left(g^{-1}\left(P_2 - P_1\right)\right) > P_2 \Leftrightarrow g^{-1}\left(P_2 - P_1\right) > h^{-1}\left(P_2\right)$$

From (2.2) and (2.6), user i subscribes to tier 1 iff $P_1 < w_i^b < g^{-1}\left(P_2 - P_1\right)$ and subscribes to tier 2 iff $w_i^b > \max\left(g^{-1}\left(P_2 - P_1\right), h^{-1}\left(P_2\right)\right)$.

Thus, if $P_2 - P_1 > g(P_1)$, user i will not subscribe to the Internet iff $w_i^b < P_1$, will subscribe to tier 1 iff $P_1 < w_i^b < g^{-1}\left(P_2 - P_1\right)$, and will subscribe to tier 2 iff $w_i^b > g^{-1}\left(P_2 - P_1\right)$ and $w_i^b > h^{-1}\left(P_2\right)$. It follows that $N_1 > 0$ iff $P_2 - P_1 > g(P_1)$. Similarly, if $P_2 - P_1 \leq g(P_1)$, then

21

$g^{-1}(P_2 - P_1) \leq h^{-1}(P_2)$. As a result, user i will not subscribe to the Internet iff $w_i^b \leq h^{-1}(P_2)$, and will subscribe to tier 2 iff $w_i^b > h^{-1}(P_2)$. It follows that $N_1=0$ iff $P_2-P_1 \leq g(P_1)$.

∎

According to Lemma 2.1, no users will subscribe to tier 1 (i.e. $N_1=0$) when $P_2-P_1 \leq g(P_1)$. In this case, an ISP has the incentive to offer only tier 2 to users. Otherwise, an ISP has the incentive to offer both tier 1 and tier 2 to users. So, Lemma 2.1 can be used to provide a sufficient condition for an ISP to offer two tiers in the case in which

$$\left(F_{w^b}(w)-1\right)f_{w^b}'(w)-2f_{w^b}^2(w) < 0 \text{ for all } w \tag{2.7}$$

The density function $f_{w^b}(w^b)$ is often assumed to follow a Pareto distribution [52]; when the mean value is finite, then it can be easily shown that (2.7) is satisfied. It can also be easily shown that (2.7) is satisfied if $f_{w^b}(w^b)$ is exponentially distributed or uniformly distributed.

Theorem 2.2: If the density function of users' willingness-to-pay for web browsing $f_{w^b}(w^b)$ satisfies (2.7), marginal network cost p^μ satisfies (2.5), and if an ISP seeks to maximize its profit in (2.4), then the ISP will offer two different tiers if

$$\left(w^b - \frac{p^\mu \overline{x}^b t^b}{\rho^s}\right) g'(w^b) > g(w^b) - \frac{p^\mu \overline{x}^s t^s}{\rho^s} \text{ for all } w^b \tag{2.8}$$

The ISP will offer only one tier if

$$\left(w^b - \frac{p^\mu \overline{x}^b t^b}{\rho^s} \right) g'\left(w^b\right) \leq g\left(w^b\right) - \frac{p^\mu \overline{x}^s t^s}{\rho^s} \quad \text{for all } w^b \tag{2.9}$$

Proof: see Appendix A.

The sufficient condition (2.8) depends on the correlation between users' willingness to pay for web browsing and streaming, described by $g()$. When $g()$ is linear, we can obtain a simpler condition:

Corollary 2.1: If $w^s = g\left(w^b\right) = \beta w^b$, if the density function of users' willingness-to-pay for web browsing $f_{w^b}\left(w^b\right)$ satisfies (2.7), marginal network cost pµ satisfies (2.5), and if an ISP seeks to maximize its profit in (2.4), then the ISP will offer two different tiers if

$$\overline{x}^s t^s / \overline{x}^b t^b > \beta \quad \Leftrightarrow \quad w^b / \overline{x}^b t^b > \beta w^b / \overline{x}^s t^s \tag{2.10}$$

Proof: If $w^s = g(w^b) = \beta w^b$, then (2.8) reduces to (2.10).

The sufficient condition (2.10) has a nice interpretation: $\overline{x}^b t^b$ (resp. $\overline{x}^s t^s$) can be interpreted as the expected number of bits consumed by web browsing (resp. video streaming). Thus, $w^b / \overline{x}^b t^b$ (resp. $\beta w^b / \overline{x}^s t^s$) can be interpreted as the value user places per bit on web browsing (resp. video streaming) traffic. Corollary 2.1 thus states that when willingness-to-pay for video streaming is linearly proportional to willingness-to-pay for web browsing, an ISP will offer two tiers if users are willing to pay more per bit for web browsing than for video streaming.

An alternative sufficient condition can be derived for the case in which network capacity is cheap. If $g()$ is strictly convex, define w_0^b as the fixed point solution of equation $w^b / \overline{x}^{\,b} t^b = g\left(w^b\right) / \overline{x}^{\,s} t^s$ in w^b if $g'(0) < \overline{x}^{\,s} t^s / \overline{x}^{\,b} t^b$; and set w_0^b to zero if $g'(0) \geq \overline{x}^{\,s} t^s / \overline{x}^{\,b} t^b$.

Corollary 2.2: If $g()$ is strictly convex, if the density function of users' willingness-to-pay for web browsing $f_{w^b}\left(w^b\right)$ satisfies (2.7), marginal network cost p^μ satisfies (2.5), and if an ISP seeks to maximize its profit in (2.4), then the ISP will offer two different tiers if the cost of network capacity $p^\mu \leq w_0^b / \overline{x}^{\,b} t^b$.

Proof: Equation (2.8) can be rewritten as:

$$w^b g'\left(w^b\right) - g\left(w^b\right) > \frac{p^\mu}{\rho^s}\left(g'\left(w^b\right)\overline{x}^{\,b} t^b - \overline{x}^{\,s} t^s\right) \text{ for all } w^b \tag{2.11}$$

We consider three cases depending on the magnitude of $\overline{x}^{\,s} t^s / \overline{x}^{\,b} t^b$.

If $g'(0) < \overline{x}^{\,s} t^s / \overline{x}^{\,b} t^b < g'(\infty)$, then there exists $w_1^b = g'^{-1}\left(\overline{x}^{\,s} t^s / \overline{x}^{\,b} t^b\right)$. If $g()$ is strictly convex, then $g'(w^b)$ is an increasing function of w^b, and hence $g'\left(w^b\right)\overline{x}^{\,b} t^b - \overline{x}^{\,s} t^s \leq 0$ on $0 < w^b \leq w_1^b$ and $g'\left(w^b\right)\overline{x}^{\,b} t^b - \overline{x}^{\,s} t^s > 0$ on $w^b > w_1^b$. If $g'\left(w^b\right)\overline{x}^{\,b} t^b - \overline{x}^{\,s} t^s \leq 0$, then (2.11) is satisfied because the left side is positive (since convexity of $g()$ implies that $w^b g'(w^b) - g(w^b) > 0$ for all w^b) and the right side is negative. If $g'\left(w^b\right)\overline{x}^{\,b} t^b - \overline{x}^{\,s} t^s \geq 0$, then (2.11) is satisfied if

$$p^\mu < \min_{w^b > w_1^b} \rho^s \left(w^b g'\left(w^b\right) - g\left(w^b\right)\right) / \left(g'\left(w^b\right)\overline{x}^{\,b} t^b - \overline{x}^{\,s} t^s\right) \tag{2.12}$$

The right side of (2.12) is minimized when $w^b = w_0^b$, because

$$\partial \frac{\rho^s \left(w^b g'\left(w^b \right) - g\left(w^b \right) \right)}{g'\left(w^b \right) \overline{x}^b t^b - \overline{x}^s t^s} \Bigg/ \partial w^b \Bigg|_{w^b = w_0^b} = 0 \Leftrightarrow \frac{w_0^b}{\overline{x}^b t^b} = \frac{g\left(w_0^b \right)}{\overline{x}^s t^s}$$

and the second derivative is positive. After replacing $w^b = w_0^b$ in (2.12), (2.11) is satisfied if

$p^\mu \leq w_0^b \big/ \overline{x}^b t^b$.

If $\overline{x}^s t^s \big/ \overline{x}^b t^b \geq g'(\infty)$, then $g'\left(w^b \right) \overline{x}^b t^b - \overline{x}^s t^s \leq 0$ for all w^b. Thus (2.11) is satisfied because the

left side is positive and the right side is negative.

If $\overline{x}^s t^s \big/ \overline{x}^b t^b \leq g'(0)$, then $w_0^b = 0$ and $g'\left(w^b \right) \overline{x}^b t^b - \overline{x}^s t^s \geq 0$ for all w^b. The convexity of $g()$

implies that $w^b g'(w^b) - g(w^b) > 0$ for all w^b. Thus (2.11) is satisfied when $p^\mu = 0$ because the

left side is positive and the right side is zero.

∎

B. Today's Internet

In this section, we seek to answer whether ISPs have the incentive to offer more than one

tier in the real world, by analyzing the existing Internet use statistics. We first verify (2.5),

to see whether an ISP will set service tier rates high enough so that video streaming

applications have sufficient performance. According to [4], in 2011 web browsing

accounted for approximately 17% of aggregate total traffic and video streaming for

approximately 50% of aggregate total traffic. Thus, $N_2 \overline{x}^s t^s \big/ \left(N_1 + N_2 \right) \overline{x}^b t^b \approx 50/17$.

According to [53], users on average spend t^s =15 hours/month on video streaming, and are willing to pay ΔP = \$20/month to upgrade to the premium tier for video streaming [54]. The throughput threshold for video streaming is around $\overline{x}^s$ = 10Mbps [55]. A popular dimensioning rule of thumb is to maintain the network load near ρ^s = 0.7; we will explore this further in SUBCHAPTER 2.5. Thus the sufficient condition (2.5) is p^μ < \$50/Mbps/month. According to [56], a typical value for marginal network capacity cost is p^μ = \$10/Mbps/month, which indicates that (2.5) is indeed satisfied in the real world.

We then check whether ISP will offer two tiers, if both web browsing and videos streaming services are provided. We conjecture that condition (2.7) is satisfied in the real world, since it is satisfied by Pareto, Exponential, and Uniform distributions.

In the absence of studies regarding the correlation between users' willingness to pay for web browsing and streaming, we first conjecture that $g()$ is linear, and thus investigate whether (2.10) holds.

We start by comparing data usage for web browsing and video streaming. According to [4], $N_2\overline{x}^s t^s \big/ \left(N_1 + N_2\right)\overline{x}^b t^b \approx 50/17$. Thus $\overline{x}^s t^s \big/ \overline{x}^b t^b \geq 50/17$, considering $N_1 + N_2 \geq N_2$. We next evaluate the ratio of users' willingness-to-pay for video streaming to willingness-to-pay for web browsing. According to [54], in 2010 typical households in the US were willing to pay approximately \$60/month for basic Internet services like web browsing, and an extra \$20/month to upgrade the tier rate for services like video streaming. Thus, $w^s/w^b \approx 20/60$.

Finally the condition (2.10) asks us to compare these two ratios: $\overline{x}^s t^s \big/ \overline{x}^b t^b \geq 50/17 > w^s/w^b \approx 20/60$. Thus, users are willing to pay more per bit for web

browsing than for video streaming, and thus ISPs have the incentive to offer more than one tier.

An alternative conjecture is that $g()$ is not linear but convex, i.e. that heavy users place more relative value on video streaming over web browsing (i.e. larger w^s/w^b) than do light users. For the heavy users, twenty percent of the subscribers worldwide are consuming eighty percent of the bandwidth [57]. Since most of the traffic is from video streaming [4], that means some heavy users place substantial value on both video streaming and web browsing. In contrast, 28.4% of web users do not watch online video content in a typical week [53]. That means light Internet users are spending much less time on video streaming, and they may not care about video streaming services at all (i.e. w^s/w^b and $g'(0)$ close to zero). Thus, one could conjecture that heavy users indeed place relatively more value on video streaming over web browsing than do light users. It is thus likely that condition "$g()$ is strictly convex" and "$g'(0) < \overline{x}^s t^s \big/ \overline{x}^b t^b$" are satisfied, and thus by Corollary 2 that ISPs with low capacity costs (i.e. $p^\mu \leq w_0^b \big/ \overline{x}^b t^b$) will have the incentive to offer more than one tier.

We used existing Internet statistics to show that ISPs indeed have the incentive to offer more than one tier for the purpose of profit maximization. In the following section, we will analyze how ISPs may set the rate and price in each tier and upgrade the network capacity.

SUBCHAPTER 2.3: An Extended Model

In this subchapter, we formulate a more complex model that extends users' willingness-to-pay for Internet applications as joint functions of bandwidth, performance, and the devoted time. The application quality functions are generalized; Internet users are differentiated by their interest level placed on each application and their value placed on leisure time; and users' times devoted to each Internet application are generalized as functions of qualities, users' interest levels and users' value placed on leisure time. We consider two interconnected problems separated by time scale. On a time scale of days, broadband Internet subscribers choose how much time to devote to Internet applications. On a time scale of months, ISPs choose what tiers to offer, and potential broadband Internet users choose what tier to subscribe to. The interaction between the two time scales will be illustrated in the next section.

A. Short Term Model

As discussed above, the current networking literature models user utility as a function of an aggregated service quality, but does not consider the time users devote to each application. Based on common observations, the devoted time depends on the service quality and user characteristics, e.g. how users value applications and their time. The service quality also depends on the devoted time, because more time means more injected traffic, which can affect the network performance in return. We thus propose a novel set of utility functions that can capture the interaction between the devoted time and service quality. In the first subsection, we consider user utility. In the second subsection, we define user willingness-to-pay by considering both utility and a user's valuation of time.

Web browsing utility is commonly modeled as an increasing concave function of throughput [47]. However, users' utilities also depends on how much web browsing they do [48]. Define t_i^b as the time (in seconds per month) that user i devotes to web browsing, consisting of the time spent reading web pages, t_i^r, 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 time devoted to web browsing, the performance of web browsing, and a user's relative utility for web browsing. Utility is an increasing concave function $V^b\left(t_i^r\right)$ of the time devoted to it [58], independent of the user [24]. 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 web pages, rather than direct observation of throughput [48], 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 below. 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 is an open question to be validated by empirical results collected from real world experiments. Here, we model user i's utility for web browsing (in dollars per month) as the product:

$$U_i^b = v_i^b V^b \left(t_i^b r_i^b \right) \tag{2.13}$$

We similarly 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. Utility is an increasing concave function $V^s\left(t_i^s\right)$ of the time devoted to it [58], independent of the user [24]. 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 experienced by video streaming applications [22], normalized so that $Q^s\left(\infty\right)=1$. User i's utility for video streaming relative to other users is modeled using a scale factor v_i^s. Again, the interaction between these factors remains an open question to be validated; we model user i's utility for video streaming (in dollars per month) as the product:

$$U_i^s = v_i^s V^s\left(t_i^s\right)Q^s\left(x_i^s\right) \tag{2.14}$$

Users' willingness-to-pay for web browsing and for video streaming also depends on their incomes. The scale factors v_i^b and v_i^s should both be increasing with income. However the time devoted these activities 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, e.g. the minimum wage user i is willing to accept. We model user i's willingness-to-pay for web browsing and video steaming, respectively, in dollars per month as:

$$W_i^b = U_i^b - p_i^t t_i^b, \quad W_i^s = U_i^s - p_i^t t_i^s \tag{2.15}$$

Users will maximize their willingness-to-pay by controlling the time devoted to each application:

$$\max_{t_i^b, t_i^s} W_i\left(t_i^b, t_i^s\right) = W_i^b + W_i^s = U_i^b + U_i^s - p_i^t\left(t_i^b + t_i^s\right) \qquad (2.16)$$

The times that maximize user willingness-to-pay will thus satisfy:

$$\begin{cases} \partial v_i^b V^b\left(r_i^b t_i^b\right) \big/ \partial t_i^b = p_i^t \\ \partial v_i^s V^s\left(t_i^s\right) Q^s\left(x_i^s\right) \big/ \partial t_i^s = p_i^t \end{cases} \Rightarrow \begin{cases} t_i^b = V^{b\prime -1}\left(p^t \big/ v_i^b r_i^b\right) \big/ r_i^b \\ t_i^s = V^{s\prime -1}\left(p^t \big/ v_i^s Q^s\left(x_i^s\right)\right) \end{cases} \qquad (2.17)$$

We have thus incorporated both performance and economic factors of the two dominant applications.

B. Long Term Model

On a time scale of months, ISPs make decisions about what tiers to offer, and potential broadband Internet users make decisions about what tier to subscribe to. Although most ISPs offer several tiers, we focus here on only two tiers: a basic tier marketed to users primarily interested in email and web browsing, and a higher tier marketed to users also interested in video streaming according to the results obtained from the basic model in SUBCHAPTER 2.2.

Similarly to (2.4), in the basic model, ISPs are presumed to seek to maximize their profit:

$$\max_{P_1, P_2, X_1, X_2, \mu} P_1 N_1 + P_2 N_2 - C(\mu)$$

On a time scale of days, broadband Internet subscribers choose how much time to devote to Internet applications, presumably by evaluating their willingness-to-pay based on the utility accrued through their use of these applications. However, the tier design determines the performance of the applications, which in turn affects user decisions about the time

spent on the applications. Performance and time will further affect user decisions about tier subscription in return, as illustrated in Figure 2.1. We thus investigate these relationships in SUBCHAPTER 2.4.
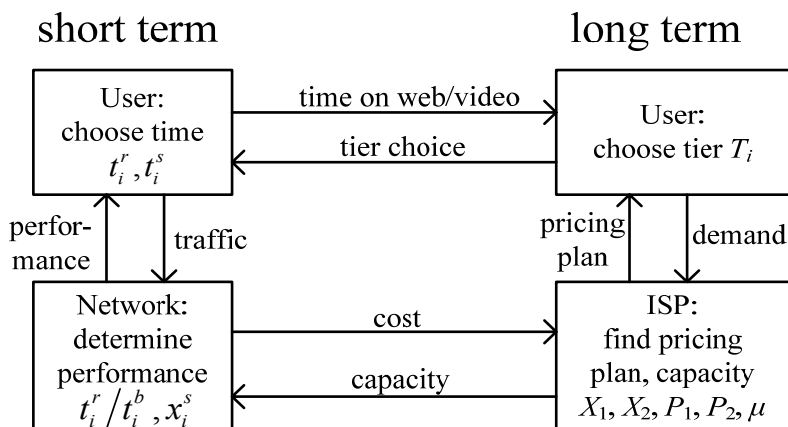
Figure 2.1 Relationships among an ISP and its subscriber

We focus on the bottleneck link within the access network. Denote by λ (in bits per month) the total downstream traffic for subscribers within the access network. This aggregate downstream traffic is simply the sum of demands by each user:

$$\lambda = \sum_i \left(t_i^b r_i^b \frac{L}{M} + x_i^s t_i^s \right) \tag{2.18}$$

where L is the average size (in bits) of a web page and M is the average time (in seconds) spent on reading a web page. 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. 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 web pages. The ratio of time spent reading web pages to time spent web browsing, $r_i^b = t_i^r / t_i^b$, can be derived from a TCP latency model [50]; 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:

$$r_i^b = t_i^r / t_i^b = TCP^b\left(d, l, X_j\right) \tag{2.19}$$

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 [50].

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 and the throughput can be derived from similar TCP throughput models [51]. Similar to (2.3), we again denote it as the minimum of TCP^s and the user's tier rate X_j:

$$x_i^s = \min\left(X_j, TCP^s\left(\rho\right)\right) \tag{2.20}$$

SUBCHAPTER 2.4: Demand Function and Density Function

In the United States and many other countries, it is common that only one or two ISPs offer wireline broadband service [59]. In SUBCHAPTER 2.4 and SUBCHAPTER 2.5, we consider one ISP that monopolizes the market. Since the current academic literature similarly analyzes a monopoly provider, and here our goal is to extend those models by considering two classes of applications and the time that users devote to each, a monopoly model is a reasonable starting point.

To derive the monopolist's demand function that expresses the dependence of user tier subscriptions upon prices and performance, we presume that 1) in the short term model users choose the time spent on web browsing and video streaming by maximizing surplus, and 2) in the long term users choose whether to subscribe to broadband Internet access and if so which tier to subscribe to. Users' time devoted to each application derived from the short term model in (2.16) is substituted into the long term model to get users' willingness-to-pay in (2.15).

Denote user i's willingness-to-pay if they have subscribed to tier j by $W_i\left(t_i^b, t_i^s \mid j\right)$. Denote the ratio of time spent reading web pages by users in tier j by $r^{b,j}$, and the throughput of video streaming by users in tier j by $x^{s,j}$. Using (2.13), (2.14), (2.16), and (2.17), $W_i\left(t_i^b, t_i^s \mid j\right)$ can be expressed as a function of $\left(v_i^b, v_i^s, p_i^t\right)$:

$$W_i\left(v_i^b, v_i^s, p_i^t \mid j\right) = v_i^b V^b\left(r^{b,j} t_i^b\right) - p_i^t t_i^b + v_i^s V^s\left(t_i^s\right) Q^s\left(x^{s,j}\right) - p_i^t t_i^s$$

where t_i^b and t_i^s can be obtained from (2.17) given performance $r^{b,j}$ and $x^{s,j}$ in tier j.

Denote user *i*'s tier choice by $T_i = 0,1,2$, where $T_i = 0$ means that user *i* chooses not to subscribe. The values of $\left(v_i^b, v_i^s, p_i^t \right)$ determine a user's choice of tiers, as shown in Figure 2.2. User *i* will choose tier T_i iff:

$$T_i = \arg \max_j \left[W_i \left(v_i^b, v_i^s, p_i^t \mid j \right) - P_j \right] \qquad (2.21)$$
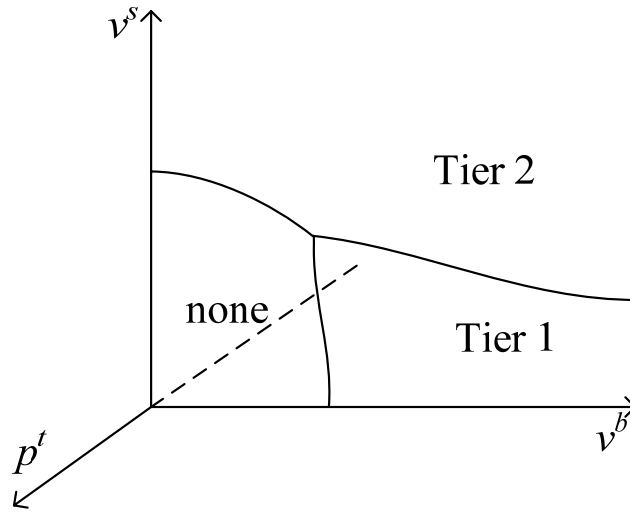
Figure 2.2 Users subscribe to tier 1 service, tier 2 service or nothing.

Denote the total number of users in the market by N_{total}. Denote the set of users that subscribe to tier *j* by $S_j = \left\{ \left(v_i^b, v_i^s, p_i^t \right) \text{ such that } T_i = j \right\}$, and the number of users that subscribe to tier *j* by $N_j = |S_j|$. Denote the distribution in the market of users' relative willingness-to-pay for web browsing, video streaming and their opportunity cost of time by a density function $f(v^b, v^s, p^t)$.

Marketing information like the density function $f(v^b, v^s, p^t)$ is important, but may be difficult to collect. Often, an ISP will conduct trials in small portions of their service area to try out new pricing plans, e.g. different tier prices [60][61]. Such trials can help an ISP estimate the

demand function. In Appendix B, we discuss how an ISP may estimate the density function $f(v^b, v^s, p^t)$ from such a trial.

The demand function for each tier is given by:

$$N_j = N_{total} \int_{(v^b, v^s, p^t) \in S_j} f\left(v^b, v^s, p^t\right) dv^b dv^s dp^t \tag{2.22}$$

According to (2.18), the aggregate traffic in the network is:

$$\lambda = \sum_j \int_{(v^b, v^s, p^t) \in S_j} N_{total} f\left(v^b, v^s, p^t\right) \left(x^{s,j} t^{s,j} + \frac{t^{b,j} r^{b,j} L}{M}\right) dv^b dv^s dp^t \tag{2.23}$$

where $t^{b,j}$ (resp. $t^{s,j}$) is the time a user in tier j with (v^b, v^s, p^t) spends on web browsing (resp. video streaming), which can be obtained from (2.17). Note that the performance $r^{b,1}$, $r^{b,2}$, $x^{s,1}$, and $x^{s,2}$ of each tier depends on the tier rates X_1, X_2, network loss l and network delay d. Furthermore, the loss l and delay d depend on the traffic rate λ using the M/M/1/K network model. Thus performance $r^{b,1}$, $r^{b,2}$, $x^{s,1}$, and $x^{s,2}$ can be expressed as functions of λ, and (2.23) is thus a nonlinear fixed point equation in λ.

Although multiple solutions from (2.23) may exist due to arbitrary density function $f(v^b, v^s, p^t)$ and performance functions TCP^b, TCP^s, we can show that there exists at most one solution if 1) $r^{b,1}=r^{b,2}$ and 2) $x^{s,1}=X_1$. Both conditions can be justified by Conjecture B in SUBCHAPTER 2.5. It is also possible that no solution can be solved from (2.23), which will lead to oscillations in user subscription choices. For example, two identical users with large v^s exist in the market, while the network capacity μ is only enough for one user to use video streaming in tier 2. When none or only one user subscribes to tier 2, network load is low

and $Q^s(x^s)$ is good enough. User who is still in tier 1 will have the incentive to upgrade from tier 1 to tier 2. When both users subscribe to tier 2, the network load is large and $Q^s(x^s)$ is poor. As a result, both users have the incentives to downgrade from tier 2 to tier 1. However, such subscription oscillation will not happen in the following ISP tier design problem, where an ISP can add network capacity to accommodate extra data usage.

SUBCHAPTER 2.5: ISP Service Tier Design

In the previous two subchapters, we introduced utility functions for web browsing and video streaming, and derived user demand for each tier. In this section, we seek to understand how an ISP may design a tiered pricing plan and bottleneck network capacity. ISP methods for tier design are proprietary; however, an understanding of how an ISP may approach tier design is essential for networking research. Our model can provide insight into this problem, by naturally decomposing the network capacity and tier design problem into three sub-problems for the ISP engineering and marketing departments.

Given the density function $f(v^b, v^s, p^t)$, the relative value functions $V^s(t^s)$, $V^b(t^b)$, the video streaming performance function $Q^s(x^s)$, and an accurate network model, an ISP could calculate the optimal tiered pricing plan P_1, P_2, X_1, X_2 and network capacity μ so as to maximize its profit, denoted by $Profit=P_1N_1+P_2N_2-C(\mu)$. The first order conditions for optimality are:

$$\left(\frac{\partial Profit}{\partial P_1}, \frac{\partial Profit}{\partial P_2}, \frac{\partial Profit}{\partial X_1}, \frac{\partial Profit}{\partial X_2}, \frac{\partial Profit}{\partial \rho} \right) = (0, 0, 0, 0, 0) \qquad (2.24)$$

However, it is difficult for an ISP to directly calculate the optimal pricing plan and network capacity from(2.24). First, an ISP may not have complete knowledge of all of the required functions. Second, an ISP may find it challenging to instill the required cooperation between its engineering department, which is traditionally focused on network architecture and performance, and its marketing department, which is traditionally focused on pricing and demand. Thus, it is natural for an ISP to attempt to decompose the task of profit maximization between its engineering and marketing departments.

A. Determination of Network Capacity

An ISP's engineering department typically has the primary responsibility for determining network capacity. While we are not privy to proprietary information about the operation of ISPs, our understanding is that many use a dimensioning rule of thumb: a capacity upgrade is scheduled when the load on a network link exceeds a threshold[2], here denoted by ρ^{th}. Thus, given network traffic λ during the peak time period, an ISP's engineering department may invest so that network capacity μ satisfies:

$$\rho = \lambda/\mu \leq \rho^{th} \tag{2.25}$$

We ask here whether such a rule of thumb applied to the capacity μ of the bottleneck link effectively maximizes profit. The optimal choice for ρ would result in:

$$\frac{\partial Profit}{\partial \rho} = \frac{\partial \sum_{j=1,2} P_j N_j}{\partial r^{b,1}} \frac{\partial r^{b,1}}{\partial \rho} + \frac{\partial \sum_{j=1,2} P_j N_j}{\partial r^{b,2}} \frac{\partial r^{b,2}}{\partial \rho}$$
$$+ \frac{\partial \sum_{j=1,2} P_j N_j}{\partial x^{s,1}} \frac{\partial x^{s,1}}{\partial \rho} + \frac{\partial \sum_{j=1,2} P_j N_j}{\partial x^{s,2}} \frac{\partial x^{s,2}}{\partial \rho} - p^\mu \left(\frac{1}{\rho} \frac{\partial \lambda}{\partial \rho} - \frac{\lambda}{\rho^2} \right) = 0 \tag{2.26}$$

---

[2] A commonly discussed choice for ρ^{th} is 0.7.

Web browsing performance in both tiers, $r^{b,1}$ and $r^{b,2}$, deteriorates with network load ρ. Similarly, video streaming performance in tier 2, $x^{s,2}$, deteriorates with network load ρ. However, video streaming performance in tier 1, $x^{s,1}$, would likely not change with network load ρ, since it would likely be constrained by tier rate X_1. As a result, any increase in load ρ will result in users spending less time on both applications, and the total traffic λ will fall. Thus:

$$\frac{\partial r^{b,1}}{\partial \rho} \leq 0, \quad \frac{\partial r^{b,2}}{\partial \rho} \leq 0, \quad \frac{\partial x^{s,1}}{\partial \rho} \approx 0, \quad \frac{\partial x^{s,2}}{\partial \rho} \leq 0, \quad \frac{\partial \lambda}{\partial \rho} \leq 0$$

The magnitude of these terms, however, depends on the load ρ. The dimensioning rule of thumb was based on observations that web browsing performance is good when loads are below a threshold, but begins to deteriorate quickly at loads above that threshold. With the increasing popularity of video streaming, ISPs seem to be using a similar rule of thumb for video streaming, but with a lower threshold. Thus, we conjecture that use of the dimensioning rule of thumb results in $\partial r^{b,1}/\partial\rho \approx 0$, $\partial r^{b,2}/\partial\rho \approx 0$, $\partial x^{s,2}/\partial\rho \approx 0$ when $\rho < \rho^{th}$, and $\partial r^{b,1}/\partial\rho < 0$, $\partial r^{b,2}/\partial\rho < 0$, $\partial x^{s,2}/\partial\rho < 0$ when $\rho > \rho^{th}$. The last term in (2.26) is a large positive number when $\rho < \rho^{th}$, and is a small positive number when $\rho > \rho^{th}$. Thus $\partial Profit/\partial\rho$ is a large positive number when $\rho < \rho^{th}$, is near 0 when $\rho \approx \rho^{th}$, and is negative when $\rho > \rho^{th}$.

Thus, it appears to be near optimal for an ISP to maintain a network load ρ slightly smaller than ρ^{th}. We expect that the amount of sub-optimality will depend on the choice of the threshold ρ^{th} and upon how quickly the performance of web browsing and video streaming change when the load exceeds the threshold. We will investigate this in SUBCHAPTER 2.6.

B. Determination of Tier Price

An ISP's marketing department typically has the primary responsibility for determining tier prices. While we are not privy to proprietary information about how they approach this task, we expect that they attempt to maximize profit. We presume here that the marketing department takes into account the engineering department's dimensioning rule of thumb, namely they assume that $\mu = \lambda/\rho^{th}$. Given this dependence, the optimal choices for P_1 and P_2 would result in:

$$\frac{\partial Profit}{\partial P_j} = N_j + P_1 \frac{\partial N_1}{\partial P_j} + P_2 \frac{\partial N_2}{\partial P_j} - \frac{p^\mu}{\rho^{th}} \frac{\partial \lambda}{\partial P_j} = 0 \qquad (2.27)$$

Denote by P_{21} the difference between P_2 and P_1, i.e. $P_2 - P_1$. Denote by $S_{j,k}$ the set of users with equal surplus in tier j and tier k (i.e. the users in Figure 2.2 on the boundary between two regions).

Conjecture A: $|S_{0,2}| \ll |S_{1,2}|$.

Conjecture A is based on the common observation that most marginal users in tier 2 prefer tier 1 to no Internet subscription.

Under Conjecture A, the total number of Internet users only depends on P_1, which gives $\partial(N_1+N_2)/\partial P_{21}=0$; the number of users in tier 2 only depends on price difference P_{21}, which gives $\partial N_2/\partial P_1=0$. Tier price design problem in (2.27) can thus be further decomposed into two sub-problems:

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

$$\frac{\partial Profit}{\partial P_{21}} = N_2 + P_{21} \frac{\partial N_2}{\partial P_{21}} - \frac{p^\mu}{\rho^{th}} \frac{\partial \lambda}{\partial P_{21}} = 0$$

As a result, tier 1 price P_1 is designed to segment all users into non-Internet users and Internet users for ISP profit maximization. The difference in tier prices P_{21} is designed to further segment Internet users into lighter users who devote less time to video streaming and heavier users who devote more time to video streaming, so as to further increase the ISP profit.

We then discuss how ISP marketing department may collect the marketing information required in (2.27). If the ISP has estimated the market density $f(v^b, v^s, p^t)$, then it can estimate the sensitivities of demands with prices $\{\partial N_1/\partial P_j, \partial N_2/\partial P_j\}$ from the demand functions in (2.22). In this case, it will likely consider the performance of web browsing and video streaming $\{r^{b,1}, r^{b,2}, x^{s,1}, x^{s,2}\}$ as fixed, i.e. estimate $\{dN_1/dP_j, dN_2/dP_j\}$ instead of $\{\partial N_1/\partial P_j, \partial N_2/\partial P_j\}$, since the dimensioning rule of thumb should keep the network load constant. Alternatively, we observe that some ISPs directly estimate these sensitivities from market surveys and/or pricing tests.

The last term in (2.27) is the impact of the tier prices on the network cost. Similarly, if the ISP has estimated the market density $f(v^b, v^s, p^t)$ and knows the time that users devote to web browsing and video streaming, then it can estimate the sensitivities of traffic with prices $\{\partial \lambda/\partial P_j\}$ from (2.23), now holding both the performance of web browsing and video streaming and the time devoted to each $\{t^{b,j}, t^{s,j}\}$ as fixed. Alternatively, if ISP directly

estimates the sensitivities of demands with prices, it may also directly estimate the sensitivities of traffic with prices.

Thus, the marketing department may attempt to maximize profit by selecting tier prices $\{P_1, P_2\}$ using (2.27). However, these prices will not be optimal since, through its reliance on the dimensioning rule of thumb, it presumes that optimal performance does not vary with price. We will investigate the amount of sub-optimality in SUBCHAPTER 2.6.

C. Determination of Tier Rates

We have presumed above that an ISP's engineering department is tasked with determining network capacity and that an ISP's marketing department is tasked with determining tier prices. The remaining task is that of determining tier rates $\{X_1, X_2\}$. We are unsure of how most ISPs handle this task. Tier rates affect the performance of web browsing and video streaming, and thus affect users' willingness-to-pay through (2.16). This in turn affects both the demand for each tier through (2.22) and the network traffic through (2.23). We conjecture that ISPs thus must involve both their engineering and marketing departments in this task.

Choosing tier rates to satisfy $\partial Profit/\partial X_1 = 0$ and $\partial Profit/\partial X_2 = 0$ appears to us to be too complex of a task to be undertaken directly by an ISP. So, we seek to first map web browsing and video streaming to separate tier rate ranges by looking into the relationship between tier rates and application performances. The rate value in each tier is thus chosen separately from the corresponding range. We address determination of the rate values in the following two subsections.

C.1 Determination of Tier 1 Rate

Given the use of a dimensioning rule of thumb, the choice of X_1 should have little effect upon the performance of web browsing and video streaming in tier 2. Similarly, the choice of X_2 should have little effect upon the performance of web browsing and video streaming in tier 1. Thus, we assume that:

$$\frac{\partial r^{b,1}}{\partial X_2} \approx 0, \quad \frac{\partial x^{s,1}}{\partial X_2} \approx 0, \quad \frac{\partial r^{b,2}}{\partial X_1} \approx 0, \quad \frac{\partial x^{s,2}}{\partial X_1} \approx 0$$

The partial derivative of profit with respect to tier 1 rate can then be simplified to:

$$\frac{\partial Profit}{\partial X_1} = P_1 \left(\frac{\partial N_1}{\partial r^{b,1}} \frac{\partial r^{b,1}}{\partial X_1} + \frac{\partial N_1}{\partial Q^s\left(x^{s,1}\right)} \frac{\partial Q^s\left(x^{s,1}\right)}{\partial X_1} \right) + P_2 \left(\frac{\partial N_2}{\partial r^{b,1}} \frac{\partial r^{b,1}}{\partial X_1} + \frac{\partial N_2}{\partial Q^s\left(x^{s,1}\right)} \frac{\partial Q^s\left(x^{s,1}\right)}{\partial X_1} \right) - \frac{p^\mu}{\rho^{th}} \frac{\partial \lambda}{\partial X_1}$$

and the partial derivative of λ with respect to tier 1 rate can be simplified to:

$$\frac{\partial \lambda}{\partial X_1} = \frac{\partial \lambda}{\partial r^{b,1}} \frac{\partial r^{b,1}}{\partial X_1} + \frac{\partial \lambda}{\partial Q^s\left(x^{s,1}\right)} \frac{\partial Q^s\left(x^{s,1}\right)}{\partial X_1}$$

The throughput of video streaming in tier 1, $x^{s,1}$, is very likely to be constrained by tier rate X_1, leading to $x^{s,1} = X_1$. The quality of web browsing, $r^{b,1}$, is an increasing concave function of X_1, while the quality of video streaming $Q^s(x^{s,1})$ is a sigmoid function of X_1. On this basis, we make the following conjecture:

Conjecture B: There exists an interval $X^0 \leq X_1 \leq X^1$, where the quality of web browsing r^b is very good, but the quality of video streaming $Q^s(x^s)$ is not desirable.

Conjecture B is based on the common observation that the minimum required tier rate X_1 for decent video streaming is larger than that of web browsing, i.e. X_0. According to Figure 2.5, Q^s has two flat portions. The initial flat portion $(X^1 \geq X_1)$ corresponds to poor video streaming performance under a low tier rate, where $\partial Q^s(x^{s,1})/\partial X_1 \approx 0$. Similarly, r^b also has a flat portion $(X_1 \geq X^0)$ corresponding to good web browsing, where $\partial r^{b,1}/\partial X_1 \approx 0$.

Thus, we can make the following approximations, when $X^1 \geq X_1 \geq X^0$:

$$\frac{\partial r^{b,1}}{\partial X_1} \approx 0, \quad \frac{\partial Q^s\left(x^{s,1}\right)}{\partial X_1} \approx 0 \Rightarrow \frac{\partial Profit}{\partial X_1} \approx 0$$

Thus, any choice of X_1 within $X^1 \geq X_1 \geq X^0$ can approximately maximize profit. One reasonable choice for X_1 is:

$$X_1 = Win/RTT \tag{2.28}$$

where *Win* is the maximum TCP receive window size and *RTT* denotes a typical round trip time.

The determination of tier 1 rate can thus be accomplished entirely by the engineering department. The amount of sub-optimality introduced by these approximations will largely depend upon the shape of the functions $r^{b,1}(X_1)$ and $Q^s(X_1)$, which we will investigate in SUBCHAPTER 2.6.

C.2 Determination of Tier 2 Rate

Tier 2 rate X_2 should be chosen from range $X_2 \geq X^1$, which is good enough for both web browsing and video streaming according to Conjecture B. However, the determination of X_2 is more complex, and we believe it will involve both the engineering and marketing departments. Using the approximations given in the previous subsection, the partial derivative of profit with respect to tier 2 rate can be simplified to:

$$\frac{\partial Profit}{\partial X_2} = P_1 \frac{\partial N_1}{\partial X_2} + P_2 \frac{\partial N_2}{\partial X_2} - \frac{p^\mu}{\rho^{th}} \frac{\partial \lambda}{\partial X_2}$$

We presume here that determination of tier rates occurs after network capacity and tier prices have been determined as outlined above. The throughput of video streaming in tier 2, $x^{s,2}$, is very likely to be constrained by tier rate X_2, leading to $x^{s,2} = X_2$.

The partial derivatives of demand and traffic to tier 2 rate, however, depend on many factors. We propose one additional conjecture solely to simplify estimation of $\partial Profit / \partial X_2$:

Conjecture C: $p_i^t = p^t \; \forall i$.

Conjecture C assumes that all users place the same value p^t on their time. We let p^t be the average value among all p_i^t, when calculating the near-optimal tier 2 rate.

Theorem 2.3: Based on conjectures A-C, $\partial Profit / \partial X_2$ can be approximated as follows:

$$\frac{\partial Profit}{\partial X_2} \approx N_2 v_{mar}^{s,2} V^s \left(t_{mar}^{s,2} \right) Q^{s\prime} \left(X_2 \right) - \frac{p^\mu N_2}{\rho^{th}} \bar{t}^{s,2} + X_2 \frac{p^\mu N_2 Q^{s\prime} \left(X_2 \right) V^{s\prime} \left(\bar{t}^{s,2} \right)}{\rho^{th} Q^s \left(X_2 \right) V^{s\prime\prime} \left(\bar{t}^{s,2} \right)} \qquad (2.29)$$

where $\bar{t}^{s,2} = E\left(t_i^s \mid T_i = 2\right)$ is the average amount of time users in tier 2 spend on video streaming, $t_{mar}^{s,2} = E\left(t_i^s \mid i \in S_{1,2}\right)$ is the average time users indifferent to tiers 1 and 2 spend on video streaming, and $v_{mar}^{s,2} = E\left(v_i^s \mid i \in S_{1,2}\right)$ is the average relative value placed on video streaming by users indifferent to tiers 1 and 2.

Proof: See Appendix C.

The first term in (2.29) can be interpreted as the marginal revenue produced by an increase in tier 2 rate, if the price of tier 2 is simultaneously increased by the amount that leaves the number of subscribers to tier 2 unchanged. The second term can be interpreted as the marginal cost for capacity produced by an increase in tier 2 rate required accommodating the increased transmission rate for video streaming. The third term can be interpreted as the marginal cost for capacity produced by an increase in tier 2 rate required accommodating the increased time spent on video streaming due to an increase in quality of video streaming.

The determination of tier 2 rate can be calculated from (2.29) by setting $\partial Profit/\partial X_2$ equal to zero. The engineering department would likely have knowledge of $Q^s, dQ^s/dX_2, p^\mu$, and $\bar{t}^{s,2}$, the marketing department would likely have knowledge of p^t, V^s and its derivatives, and both departments must cooperate to estimate $\partial Profit/\partial X_2$. The amount of sub-optimality introduced by these approximations will largely depend upon the validity of Conjecture C, which we will investigate in SUBCHAPTER 2.6.

The information flow and roles of the engineering and marketing departments is illustrated in Figure 2.3, where the complex version of the ISP profit maximization problem is decomposed into three sub-problems for the ISP departments respectively, using our proposed scheme.
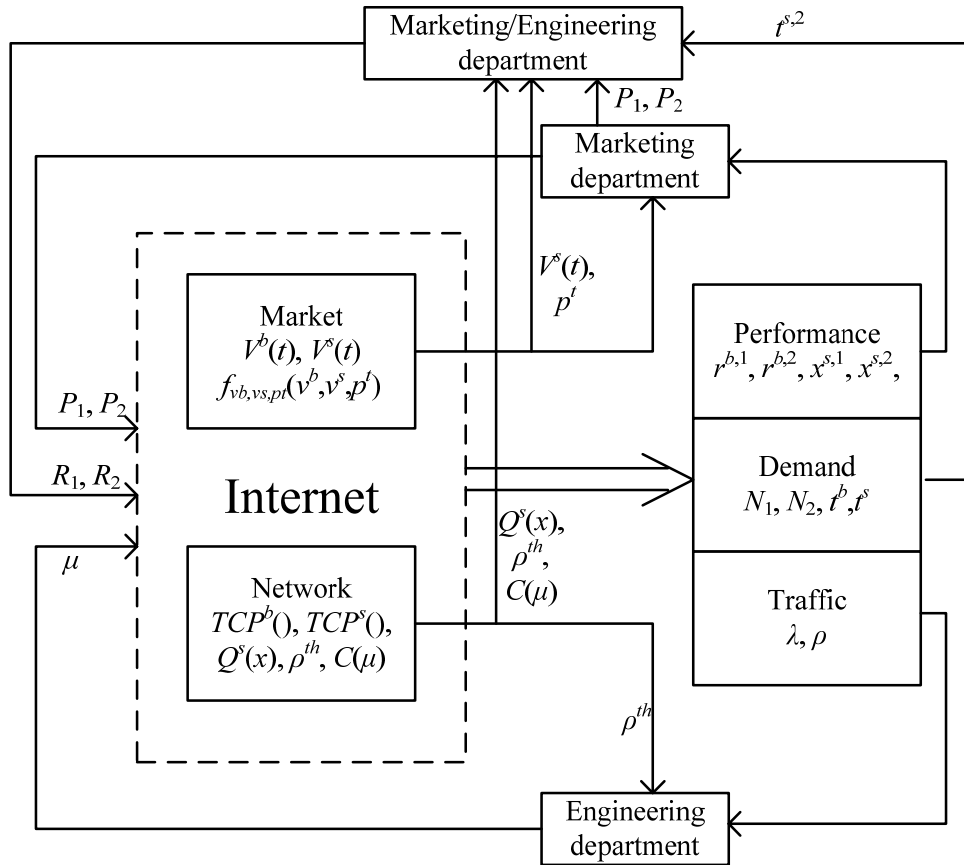
Figure 2.3 ISP profit maximization scheme designed for ISP departments.

SUBCHAPTER 2.6: Numerical Results

In this subchapter, we explore the magnitude of the decrease in profit resulting from the various sources of sub-optimality discussed in the previous section, and the variation of the design with key parameters.

A. Magnitude of Sub-optimality

The use of a dimensioning rule of thumb, based on the presumption of a threshold ρ^{th}, may cause significant sub-optimality. To investigate this, parameters are set as follows: L =750KB [62]; 10 concurrent TCP connections for web browsing; TCP packet size = 512B; RTT =100ms; M/M/1/K service rate = 600Mbps and buffer size = 25, 50, or 100 packets; video streaming service based on TCP with $Q^s(x^s)$ as in [55].
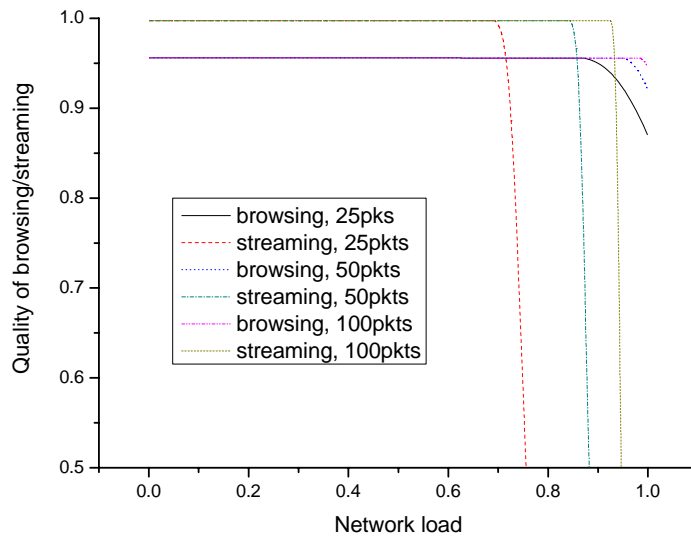
Figure 2.4 Dependence of performance r^b and Q^s upon network load ρ.

Figure 2.4 shows the performance of web browsing and video streaming as a function of network load. For a 50 packet buffer, there is a fairly steep decline in performance of web browsing when $\rho > 0.97$, and in the performance of video streaming when $\rho > 0.87$. Figure 2.5 shows the performance of web browsing and video streaming as a function of access tier rate for a 50 packet buffer when $\rho = 0.7$. The performance of web browsing is a concave function of the tier rate, and is fairly constant for $X > 2.5$Mbps. The performance of video streaming is a sigmoid function of the tier rate; it is fairly constant at poor performance when $X < 3$Mbps, rises quickly for 3Mbps$< X < 18$Mbps, and is fairly constant at high performance when $X > 18$Mbps. Thus, although both web browsing and video streaming performance experience a sharp threshold with respect to load, there is much slower change in video streaming performance with respect to tier rate.
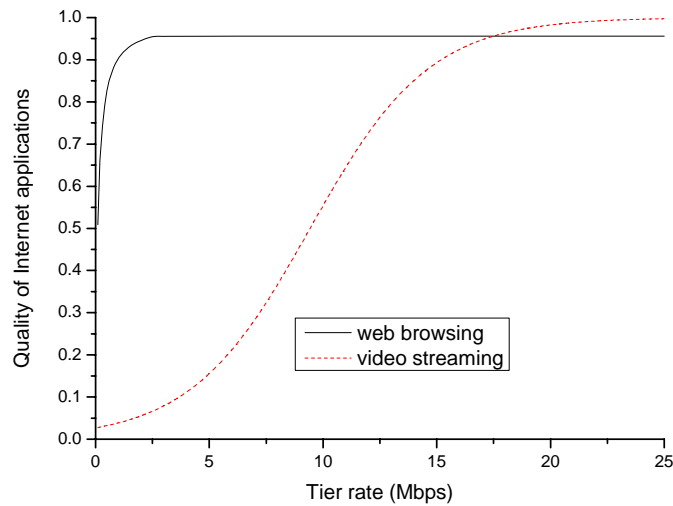
Figure 2.5 The dependence of performance r^b and Q^s on the tier rate.

To gauge the magnitude of the decrease in profit resulting from the simplified design, we compare the optimal choices of capacity, tier prices, and tier rates to those chosen using the

simplified scheme under the following parameters: $V^b(t^b) = \log(a_b t^b + 1)$ with $a_b = 0.0026$, so that a user with t^b=50 hours/month [63] is willing to pay $60 for tier 1 [54]; $V^s(t^s) = \log(a_s t^s + 1)$, with $a_s = 0.0093$, so that a user with t^s=15 hours/month [53] is willing to pay an additional $20 to move from tier 1 to tier 2 [54]; N=20000; $(v^b, v^s, p^t) \sim$ multivariate lognormal with $(v^b/p^t, v^s/p^t)$ independent of p^t, $f_p(p^t)$ given by 2009 US household income [64], $f(v^b/p^t)$ given by [60], $f(v^s/p^t)$ given by [53], and the correlation between v^b/p^t and $v^s/p^t = 0.5$; $p^\mu = \$10/\text{Mbps/month}$ [56]; peak traffic = 1.55 times average traffic [4]; buffer = 50 packets; $\rho^{th} = 0.7$.

Table 2.1 Optimal and simplified design comparison

	Optimal	Simplified
Tier 1 price P_1	$65	$68
Tier 2 price P_2	$80	$84
Tier 1 rate X_1	2.5Mbps	2.5Mbps
Tier 2 rate X_2	22.2Mbps	20Mbps
Users in tier 1 N_1	3326	3250
Users in tier 2 N_2	3471	3336
Capacity μ	6.41Gbps	6.31Gbps
Profit	$388280	$371600

Table 2.1 presents the parameters and profits of the optimal monopoly design (2.24) and the simplified design (2.25), (2.27) (2.28) and (2.29). Both the optimal and simplified design results are numerically obtained using a gradient descent algorithm. In the optimal design, all the marketing and network information are used to calculate the gradient in terms of each variable (i.e. tier rates, tier prices and capacity) during each iteration. In the simplified design, only the information in Figure 2.3 is used to calculate the gradients based on Conjectures A-C. Multiple initial conditions were used in case local maxima exist[3]. The

---

[3] The existence of local maxima depend on the shape of the density function $f(v^b, v^s, p^t)$, the video streaming quality function $Q^s(x^s)$, and of the web browsing quality function $TCP^b(d, l, X_j)$.

simplified near-optimal designs match fairly close to the optimal design, resulting in only 4.3% less profit. The tier 2 rate is smaller than the optimal rate because the marginal revenue produced by an increase in the tier 2 rate (i.e. the first item in (2.29)) is underestimated by Conjecture C. As a result, $dProfit/dX_2$ is negative and ISPs have the incentive to reduce the tier 2 rate.

Figure 2.6 shows the proportion of the optimal profit that the simplified design achieves under different load thresholds ρth. We observe the proportion increases with ρ^{th} until ρ^{th} = 0.85, when the simplified scheme achieves 98.5% of optimal, and then falls quickly after that. The optimal value of ρth = 0.85 corresponds to the load threshold for video streaming as seen in Figure 2.4.
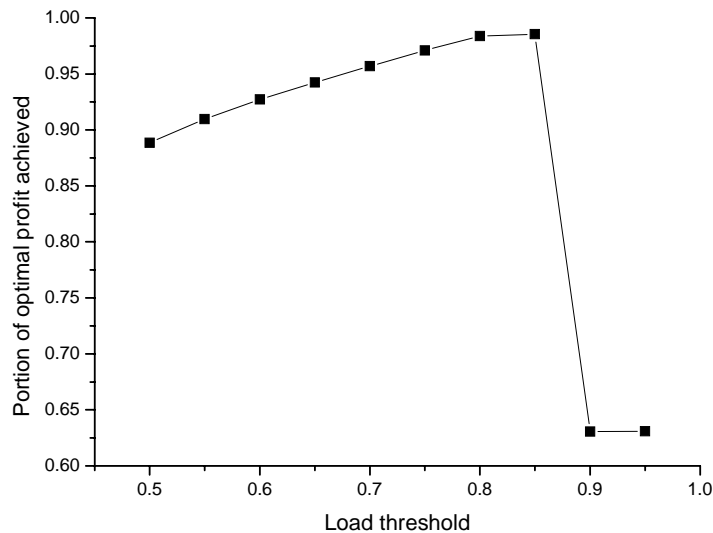
Figure 2.6 Sub-optimal profit over optimal profit under different ρ^{th}.

The dimensioning rule of thumb is the largest source of sub-optimality in the numerical results in this section, and the choice of the threshold is the most significant factor. The

determination of tier prices contributes additional sub-optimality through its reliance on the dimensioning rule of thumb, which presumes that optimal performance does not vary with price. The determination of tier rates contributes additional sub-optimality through approximations, which depend upon the shape of the functions $r^{b,1}(X_1)$ and $Q^s(X_1)$ and upon the validity of Conjecture C. In numerical results, these contributions are minor.

B. Variation of the Design with Key Parameters

In this final subsection, we explore the variation of the simplified design with key parameters. Figure 2.7 shows the dependence of profit upon the marginal network cost p^μ. Unsurprisingly, the cost of capacity decreases and profit increases as marginal network cost decreases.

The impact upon the demand for each tier is complex. First, consider the impact of p^μ on tier prices. The marketing department considers whether to increase or decrease P_1 in response. If it increases P_1, this will result in users in $S_{0,1}$ dropping service, with a small decrease in traffic, and users in $S_{1,2}$ upgrading from tier 1 to tier 2, with a substantial increase in traffic. As a result, $\partial\lambda/\partial P_1 > 0$ and when p^μ decreases, $\partial Profit/\partial P_1$ becomes positive from (2.27). Thus, the ISP will increase P_1 to earn more profit. The marketing department will then consider whether to modify P_2. If it increases P_2, this will result in users in $S_{1,2}$ downgrading from tier 2 to tier 1, with a substantial decrease in traffic. As a result, $\partial\lambda/\partial P_2 < 0$, and when p^μ decreases, $\partial Profit/\partial P_2$ becomes negative from (2.27). Thus, the ISP will decrease P_2 to earn more profit.

Next consider the impact upon tier rates. Tier 1 rate is set by the engineering department using (2.28) which does not depend upon p^μ. The engineering and marketing departments jointly use (2.29) to set tier 2 rate; decreasing p^μ makes $\partial Profit/\partial X_2$ positive, and thus the ISP will increase tier rate X_2.
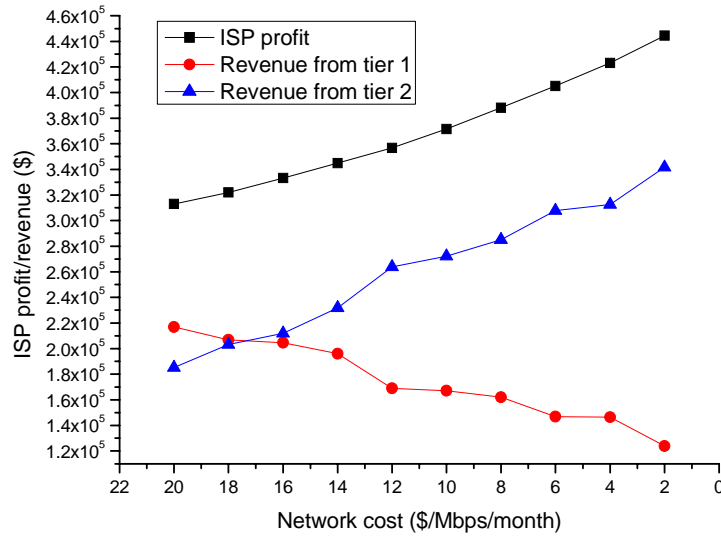
Figure 2.7 The dependence of profit upon the marginal network cost p^μ.
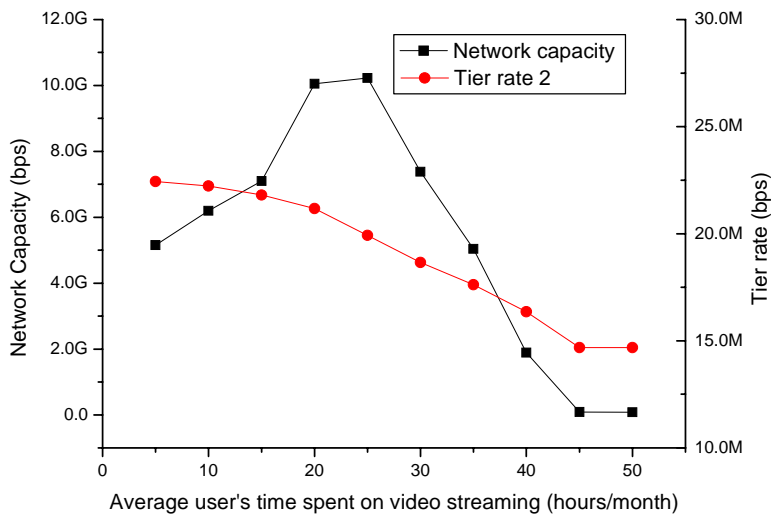
Figure 2.8 Network capacity and tier 2 rate vs. video streaming time.

Since the price of tier 2 has decreased while tier 2 rate has increased, the demand for tier 2 will increase. The increase in tier 2 demand outweighs the decrease in tier 2 price, and thus revenue from tier 2 increases. Similarly, the price of tier 1 has increased, causing users to upgrade to tier 2 and causing revenue from tier 1 to decrease.

Finally, we explore the effect of increasing video streaming popularity. To investigate this, we simultaneously increase v_b and decrease the parameter a_s in $V^s(t^s)$, so that the average user's time spent on video streaming increases but their willingness-to-pay for streaming remains unchanged. Figure 2.8 shows the network capacity and tier 2 rate as a function of the average user's time spent on video streaming.

As users devote more time to video streaming, $\partial Profit / \partial X_2$ becomes negative, and thus the engineering and marketing departments will jointly reduce tier rate X_2 using (2.29). This will cause some users to downgrade from tier 2 to tier 1.

The effect on traffic is more complex. For small increases in the average streaming time, the increase in video streaming time by those who remain in tier 2 outweighs the very small reduction in tier 2 subscriptions and performance, and hence results in an increase in traffic. As a result, the engineering department increases capacity μ according to (2.25). However, for larger increases in video streaming time, the tier 2 subscriptions and performance begin to drop quickly, outweighing the increase in video streaming time by those who remain in tier 2, and hence resulting in a decrease in traffic, and thus a decrease in capacity.

CHAPTER 3:   Effect of Data Caps on ISP Tier Design and Users

In this chapter, we evaluate the impact of data caps upon subscribers and ISPs by extending the ISP tier design model in CHAPTER 2. Compared with existing models, our 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 of network performance. This additional detail is a critical factor to be considered 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.

SUBCHAPTER 3.1: Cap Model Formulation

In this subchapter, 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 time scale. On a time scale of days, broadband Internet subscribers choose how much time to devote to Internet applications. On a time scale of months, subscribers choose what tier to subscribe to, and ISPs choose tier rates, tier prices, data caps, overage charges and network capacity.

A. Short Term Model

We model user i's willingness-to-pay for web browsing and video steaming, (in dollars per month) the same as the user willingness-to-pay in the extended ISP tier design model (see SUBCHAPTER 2.3):

$$W_i^b = U_i^b - p_i^t t_i^b, \quad W_i^s = U_i^s - p_i^t t_i^s \tag{3.1}$$

Most ISPs have started designing and marketing tiers on the basis of the applications they are intended for. 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 price of the basic tier and the premium tier by P_1 (i.e. tier1 price in CHAPTER 2) and P_2 (i.e. tier 2 price in CHAPTER 2) respectively. Denote the cap (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 web page and M is the average time (in seconds) spent on reading a web page. 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 time devoted to web browsing and video streaming so as to maximize surplus, S_i, defined as the difference between willingness-to-pay and cost:

$$\max_{t_i^b, t_i^s} S_i = W_i^b + W_i^s - p^o \max\left(0, x_i^b t_i^b + x_i^s t_i^s - C_{T_i}\right) - P_{T_i} \tag{3.2}$$

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

$$\partial v_i^b V^b\left(r_i^b t_i^b\right)\big/\partial t_i^b = p_i^t, \quad \partial v_i^s V^s\left(t_i^s\right)Q^s\left(x_i^s\right)\big/\partial t_i^s = p_i^t \tag{3.3}$$

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:

$$p_i^t < \partial v_i^s V^s\left(t_i^s\right)Q^s\left(x_i^s\right)\big/\partial t_i^s < p_i^t + p^o x_i^s, \quad p_i^t < \partial v_i^b V^b\left(r_i^b t_i^b\right)\big/\partial t_i^b < p_i^t + p^o x_i^b \tag{3.4}$$

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

$$\partial v_i^b V^b\left(r_i^b t_i^b\right)\big/\partial t_i^b = p_i^t + p^o x_i^b, \quad \partial v_i^s V^s\left(t_i^s\right)Q^s\left(x_i^s\right)\big/\partial t_i^s = p_i^t + p^o x_i^s \tag{3.5}$$

57

B. Long Term Model

In the long term, say on a time scale 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 iff:

$$T_i = \arg\max_j \left[W_i\left(t_i^b, t_i^s \mid j\right) - P_j - p^o \max\left(0, x_i^b t_i^b + x_i^s t_i^s - C_j\right) \right] \tag{3.6}$$

where, $W_i\left(t_i^b, t_i^s \mid j\right)$ is user i's willingness to pay for tier j; the amount of time user i devoted to web browsing (i.e. t_i^b) and videos streaming (i.e. t_i^s) can be derived from (3.2).

The total number of the 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 . The proposed user utility functions are general enough to be applied to models with or without competition between ISPs. In the remainder of this chapter, we consider one ISP that monopolizes the market. ISP duopoly competition will be considered in CHAPTER 4. A monopoly ISP is presumed to maximize its profit from video streaming by controlling the parameters P, X, C, p^o and μ:

$$\max_{P_1, P_2, X_1, X_2, C_1, C_2, p^o, \mu} P_1 N_1 + P_2 N_2 + p^o O - K(\mu) - k(N_1 + N_2) \tag{3.7}$$

where $O = \sum_i \max\left(0, x_i^b t_i^b + x_i^s t_i^s - C_{T_i}\right)$ is the total amount of data above the cap; $K(\mu)$ is the

ISP's variable cost per month for capacity required to accommodate the total user demand;

k is the ISP's fixed cost per subscriber per month.

The relationship between traffic and performance is the same as the traffic demand function and the application performance functions in the extended ISP tier design model, see (2.18), (2.19) and (2.20) in SUBCHAPTER 2.3.

SUBCHAPTER 3.2: 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}, i.e. $\mu = \lambda/\rho^{th}$. The ISP will set tier rates X_1 and X_2 higher than the achievable web browsing throughput over TCP, i.e. $X_2 > X_1 > X_0$, but no higher than the achievable video streaming throughput over TCP, i.e. $x_i^s = X_{T_i}$.

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 [65]. 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, i.e.

$$W_i^{b,1} + W_i^{s,1} - P_1 - p^o \max\left(0, x^{b,1} t_i^b + x^{s,1} t_i^s - C_1\right) > 0, \quad \text{for all } i : T_i \neq 0$$

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, i.e. a user's choice of time devoted to web browsing and video streaming so as to maximize surplus changes from (3.2) to:

$$\max_{t_i^b, t_i^s} S_i \Rightarrow \begin{cases} \max_{t_i^b} W_i^b \\ \max_{t_i^s} W_i^s - p^o \max\left(0, x_i^b t_i^b + x_i^s t_i^s - C_{T_i}\right) \end{cases} \tag{3.8}$$

For uncapped users, the marginal utility from web browsing or video streaming still satisfies (3.3). However, for users who are capped but not paying an overage charge, the marginal utility changes from (3.4) to:

$$p_i^t = \partial v_i^b V^b \left(r_i^b t_i^b\right) \Big/ \partial t_i^b, \quad p_i^t < \partial v_i^s V^s \left(t_i^s\right) Q^s \left(x_i^s\right) \Big/ \partial t_i^s < p_i^t + p^o x_i^s \tag{3.9}$$

And for users who are capped and paying an overage charge, the marginal utility changes from (3.5) to:

$$\partial v_i^b V^b \left(r_i^b t_i^b \right) / \partial t_i^b = p_i^t, \quad \partial v_i^s V^s \left(t_i^s \right) Q^s \left(x_i^s \right) / \partial t_i^s = p_i^t + p^o x_i^s \tag{3.10}$$

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

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 effect 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 [3][4]. 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 the user tier choice in (3.6) simplifies to:

$$T_i = \begin{cases} 2, & W_i^{s,2} - W_i^{s,1} > P_{21} + p^o \max\left(0, x_i^b t_i^b + X_2 t_i^s - C_2\right) \\ 1, & \text{otherwise} \\ 0, & W_i^{b,1} + W_i^{s,1} - P_1 > 0 \end{cases} \tag{3.11}$$

where $P_{21} = P_2 - P_1$, and t_i^b and t_i^s can be calculated from (3.8).

Proof: Using assumptions A, C and D, user i's willingness to pay for web browsing is the same in both tiers, i.e. $W_i^{b,1} = W_i^{b,2}$, 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 (3.6) simplifies to (3.11)

∎

SUBCHAPTER 3.3: 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 (3.11) 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:

$$W_i^{s,2} - W_i^{s,1} = P_{21} + p^o \max\left(0, x_i^b t_i^b + X_2 t_i^s - C_2\right) \tag{3.12}$$

Marginal premium subscribers who are not capped satisfy (3.3) and (3.12). Denote $v^{s1,u}\left(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b\right)$ as the solution to the fixed point equation in v_i^s resulting from

(3.3) and (3.12). There is a unique solution for v_i^s because by (3.1) and (3.3) $d\left(W_i^{s,2} - W_i^{s,1}\right)\Big/dv_i^s > 0$ for $v_i^s > 0$, which makes the left side of (3.12) 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 (3.9) and (3.12); denote $v^{s1,c}\left(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b\right)$ as the solution to the fixed point from these equations. Finally, marginal premium subscribers who are capped and paying overage charges satisfy (3.10) and (3.12); denote $v^{s1,o}\left(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b\right)$ as the solution to the fixed point from these equations.

Thus the marginal premium subscribers lie on the curve:

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

We can use this curve to partition all premium tier subscribers. According to (3.3), premium tier subscribers who are not capped satisfy:

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

Denote the set of such subscribers:

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

According to (3.9), premium tier subscribers who are capped but not paying an overage charge satisfy:

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

Denote the set of such subscribers:

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

Finally, according to (3.10), premium tier subscribers who are capped and paying overage charges satisfy:

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

Denote the set of such subscribers:

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

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 (3.11) 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:

$$W_i^{b,1} + W_i^{s,1} - P_1 = 0 \tag{3.13}$$

Denote $v^{s0}\left(P_1, X_1, p_i^t, v_i^b\right)$ as the solution to the fixed point equation in v_i^s resulting from (3.3) and (3.13). Non-subscribers, denoted by G_n, place a smaller value on video streaming than do marginal basic subscribers:

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

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:

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

These five sets define a partition of Internet subscribers on the basis of $\left(v_i^s, v_i^b, p_i^t\right)$. However, it is more revealing to use $\left(v_i^s / p_i^t, v_i^b / p_i^t, p_i^t\right)$ 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 3.1, 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.
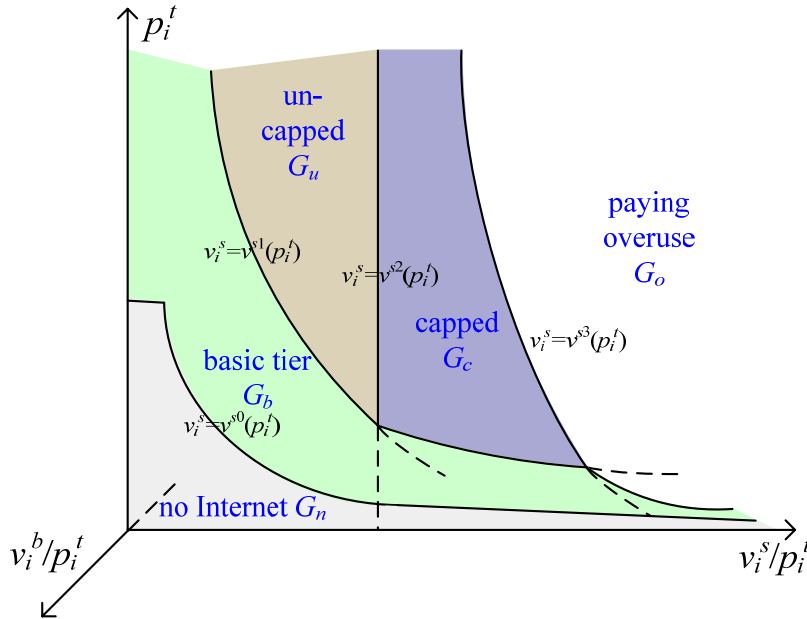
Figure 3.1 Partition of Internet subscribers.

SUBCHAPTER 3.4: Impact of Cap upon Pricing Plan

In this subchapter, 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.

A. Optimal Pricing Plan without Data Caps

We start by characterizing the rate, price and network capacity, without data caps, that maximize an ISP's profit. If assumptions A and B hold, the ISP profit maximization problem in (3.7) without data caps (i.e., $p^o = 0$ or $C_1 = C_2 = \infty$) can be reformulated as:

$$\max_{P_1, P_{21}, X_1, X_2} Profit_0 = (P_1 - k)(N_1 + N_2) + P_{21} N_2 - K(\lambda / \rho^{th}) \tag{3.14}$$

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

Theorem 3.2: If assumptions A and B hold, the first order optimality conditions for the profit maximization problem without data caps in (3.14) satisfy:

$$\begin{aligned}
\frac{\partial Profit_0}{\partial P_1} &= N_1 + N_2 + (P_1 - k)\frac{\partial(N_1 + N_2)}{\partial P_1} - \frac{p^\mu}{\rho^{th}}\frac{\partial \lambda}{\partial P_1} = 0 \\
\frac{\partial Profit_0}{\partial X_1} &= (P_1 - k)\frac{\partial(N_1 + N_2)}{\partial X_1} + P_{21}\frac{\partial N_2}{\partial X_1} - \frac{p^\mu}{\rho^{th}}\frac{\partial \lambda}{\partial X_1} = 0 \\
\frac{\partial Profit_0}{\partial P_{21}} &= N_2 + P_{21}\frac{\partial N_2}{\partial P_{21}} - \frac{p^\mu}{\rho^{th}}\frac{\partial \lambda}{\partial P_{21}} = 0 \\
\frac{\partial Profit_0}{\partial X_2} &= P_{21}\frac{\partial N_2}{\partial X_2} - \frac{p^\mu}{\rho^{th}}\frac{\partial \lambda}{\partial X_2} = 0
\end{aligned} \tag{3.15}$$

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

Proof: The proof is omitted, as the partial derivatives follow directly from earlier equations.

B. 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 [8]. In the next subsection, we consider the case in which an ISP uses caps to maximize profit.

Suppose that an ISP imputes a cost to user i equal to $p^\mu \left(t_i^s X_2 + t_i^b x_i^b \right) \Big/ \rho^{th}$, 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 $\left(t_i^s X_2 + t_i^b x_i^b \right) \Big/ \rho^{th}$ 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 2.2, we presume in this subsection that the goal of the

ISP is to set a data cap C_2 and overage charge p^o so that:

$$P_1 + P_{21} + p^o \left(t_i^s X_2 + t_i^b x_i^b - C_2 \right) - p^\mu \left(t_i^s X_2 + t_i^b x_i^b \right) \Big/ \rho^{th} - k \geq 0 \quad \text{for all } i : T_i = 2$$

Denote $t_{\max}^{b,1}$ and $t_{\max}^{s,1}$ by the maximum amount of time users in the basic tier spent on web

browsing and video streaming, respectively. Denote x^b by the throughput of web browsing

of all users, since $X_2 > X_1 > X_0$ by assumption A. Considering $x^{b,1} t_i^b + x^{s,1} t_i^s < C_1, \quad \forall i : T_i = 1$ from

assumption D, we examine a simple method of achieving this goal: $p^o = p^\mu / \rho^{th}$,

$C_1 = \left(P_1 - k \right) \big/ p^o$ and $C_2 = P_{21} / p^o + x^b t_{\max}^{b,1} + X_1 t_{\max}^{s,1}$ and. We henceforth refer to this choice of

cap C_1, 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

$P_2 + p^o \left(t_i^s X_2 + t_i^b x^b - C_2 \right)$, which is larger than or equal to the user's imputed cost, i.e.

$p^\mu \left(t_i^s X_2 + t_i^b x^b \right) \Big/ \rho^{th} + k$.

The impact of such a cap on premium tier subscribers is illustrated in Figure 3.2. Under the

optimal pricing plan without data caps from (3.14), 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 user cap is added.
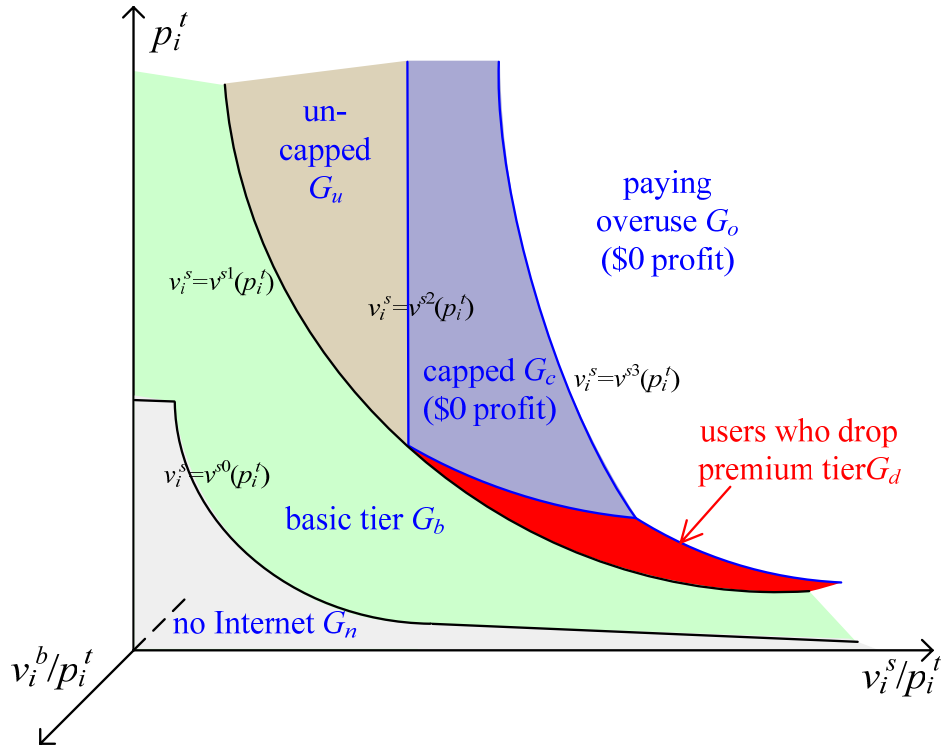
Figure 3.2 The impact of a cap designed to ensure that heavy users pay an amount equal to the cost of their usage.

C. 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 subsection does not maximize profit, 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 (3.7). Thus, the ISP profit maximization problem with cap in (3.7) can be reduced to:

$$\max_{P_1, P_{21}, X_1, X_2 C_2, p^o} Profit = (P_1 - k)(N_1 + N_2) + P_{21}N_2 + p^o O - K(\lambda/\rho^{th}) \qquad (3.16)$$

We henceforth refer the optimal tier rate, tier price, cap, and overage charge in (3.16) as the *profit-maximizing cap*.

Theorem 3.3: If assumptions A-D hold, then the partial derivatives of *Profit* defined in (3.16) with respect to P_1, X_1, P_{21}, X_2, C_2 and p^o can be expressed as:

$$
\begin{cases}
\dfrac{\partial Profit}{\partial P_1} = N_1 + N_2 + (P_1 - k)\dfrac{\partial(N_1 + N_2)}{\partial P_1} - \dfrac{p^\mu}{\rho^{th}}\dfrac{\partial \lambda}{\partial P_1} \\[2mm]
\dfrac{\partial Profit}{\partial X_1} = (P_1 - k)\dfrac{\partial(N_1 + N_2)}{\partial X_1} + P_{21}\dfrac{\partial N_2}{\partial X_1} + p^o\dfrac{\partial O}{\partial X_1} - \dfrac{p^\mu}{\rho^{th}}\dfrac{\partial \lambda}{\partial X_1} \\[2mm]
\dfrac{\partial Profit}{\partial P_{21}} = N_2 + P_{21}\dfrac{\partial N_2}{\partial P_{21}} + p^o\dfrac{\partial O}{\partial P_{21}} - \dfrac{p^\mu}{\rho^{th}}\dfrac{\partial \lambda}{\partial P_{21}} \\[2mm]
\dfrac{\partial Profit}{\partial X_2} = P_{21}\dfrac{\partial N_2}{\partial X_2} + p^o\dfrac{\partial O}{\partial X_2} - \dfrac{p^\mu}{\rho^{th}}\dfrac{\partial \lambda}{\partial X_2} \\[2mm]
\dfrac{\partial Profit}{\partial C_2} = P_{21}\dfrac{\partial N_2}{\partial C_2} + p^o\dfrac{\partial O}{\partial C_2} - \dfrac{p^\mu}{\rho^{th}}\dfrac{\partial \lambda}{\partial C_2} \\[2mm]
\dfrac{\partial Profit}{\partial p^o} = P_{21}\dfrac{\partial N_2}{\partial p^o} + O + p^o\dfrac{\partial O}{\partial p^o} - \dfrac{p^\mu}{\rho^{th}}\dfrac{\partial \lambda}{\partial p^o}
\end{cases}
\qquad (3.17)
$$

Proof: According (3.11) in theorem 3.1, N_2 only depends on $P_2 - P_1$ (or P_{21}), X_1, X_2, C_2 and p^o; $N_1 + N_2$ only depends on P_1, X_1. Thus, $\partial N_2/\partial P_1 = 0$, $\partial \lambda_2/\partial P_1 = 0$, $\partial(N_1 + N_2)/\partial X_2 = 0$, $\partial(N_1 + N_2)/\partial P_{21} = 0$, and $\partial(N_1 + N_2)/\partial C_2 = 0$, $\partial(N_1 + N_2)/\partial p^o = 0$. Thus, (3.17) can be obtained by replacing the partial derivatives of the ISP profit with cap defined in (3.16).

∎

The optimal pricing plans with and without caps can be numerically calculated from (3.16) and (3.14) 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 (3.16) to those in the heavy users cap:

Theorem 3.4: 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 3.2), and data caps and overage charges are set using the heavy users cap (i.e. $C_2 = P_{21}/p^o + x^b t_{max}^{b,1} + X_1 t_{max}^{s,1}$ and $p^o = p^\mu/\rho^{th}$), then:

$$\partial Profit/\partial C_2 \leq 0, \quad \partial Profit/\partial p^o \geq 0$$

Proof: see Appendix.

Based on theorem 3.4, we can summarize how ISPs might change the cap parameters starting from the heavy users cap presented in the previous subsection: 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 $f\left(v^b/p^t, v^s/p^t, p^t\right)$ 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 3.2. 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 noted above, 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 $f\left(v_i^b/p_i^t, v_i^s/p_i^t, p_i^t\right)$ with the value placed on time p_i^t. Household income in the United States can be approximated by a lognormal distribution [64]. Users with low income fall into the increasing portion of the lognormal distribution, i.e. the user density function $f\left(v_i^b/p_i^t, v_i^s/p_i^t, p_i^t\right)$ is an increasing function of value placed on time p_i^t on this set:

Assumption E: $\forall$ user i and i' from 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

$$f\left(v_i^b/p_i^t, v_i^s/p_i^t, p_i^t\right) > f\left(v_{i'}^b/p_{i'}^t, v_{i'}^s/p_{i'}^t, p_{i'}^t\right).$$

We now compare the tier rates and tier prices for an ISP that users data caps to maximize profit to those for an ISP that does not use data caps. We already know from Theorem 3.4 that a profit-maximizing ISP will set $p^o \geq p^\mu/\rho^{th}$ and $C_2 < P_{21}/p^o + x^b t_{max}^{b,1} + X_1 t_{max}^{s,1}$ 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 3.5: If assumptions A-E hold, tier rates and tier prices are set to maximize profit without caps (i.e. P_1, P_{21}, X_1, and X_2 are set using theorem 3.2), $C_2 \leq P_{21}/p^o$, and $p^o \geq p^\mu/\rho^{th}$, then:

$$\frac{\partial Profit}{\partial P_1} = 0, \frac{\partial Profit}{\partial X_1} \leq 0, \frac{\partial Profit}{\partial P_{21}} \leq 0, \frac{\partial Profit}{\partial X_2} \geq 0$$

Proof: see Appendix.

The theorem only applies when $C_2 \leq P_{21}/p^o$; 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 than an ISP that does not use caps or that uses caps only to ensure that heavy users pay for their usage.

SUBCHAPTER 3.5: Impact of Data Caps upon Users

In this subchapter, 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 subchapter.

A. Impact of Data Caps upon User Tier Choices

In Figure 3.3, 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, i.e. users who are indifferent between the basic tier and no Internet subscription, which form the boundary between the yellow and light grey areas in the figure. The impact of profit-maximizing data caps upon these subscribers is fairly straightforward: by theorem 3.5, 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.
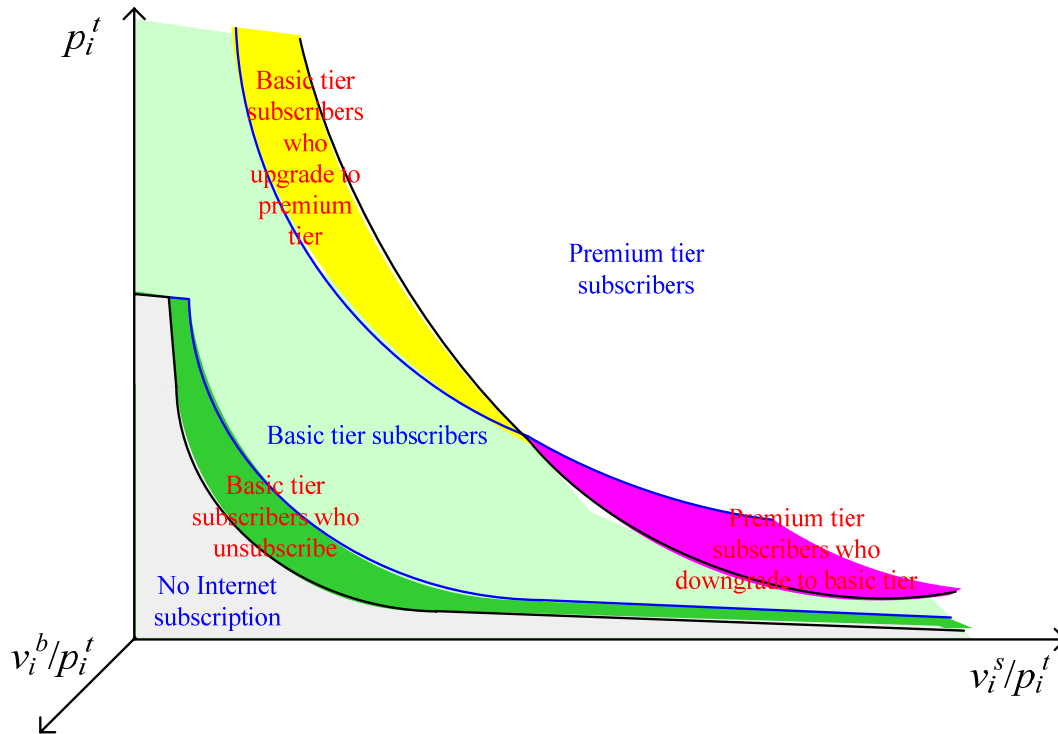
Figure 3.3 Comparing service tier choices when a profit-maximizing cap is implemented.

We next consider marginal premium subscribers under a pricing plan without caps, i.e. 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 3.5, 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 3.1: 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 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 but not paying overage charges 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 do before running into the cap. From equations (3.8)-(3.10), 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.

B. Impact of Data Caps upon User Surplus

We now turn to the change of user surplus when data caps are adopted. In Figure 3.4, we illustrate the change in user surplus when a profit-maximizing data cap is implemented.
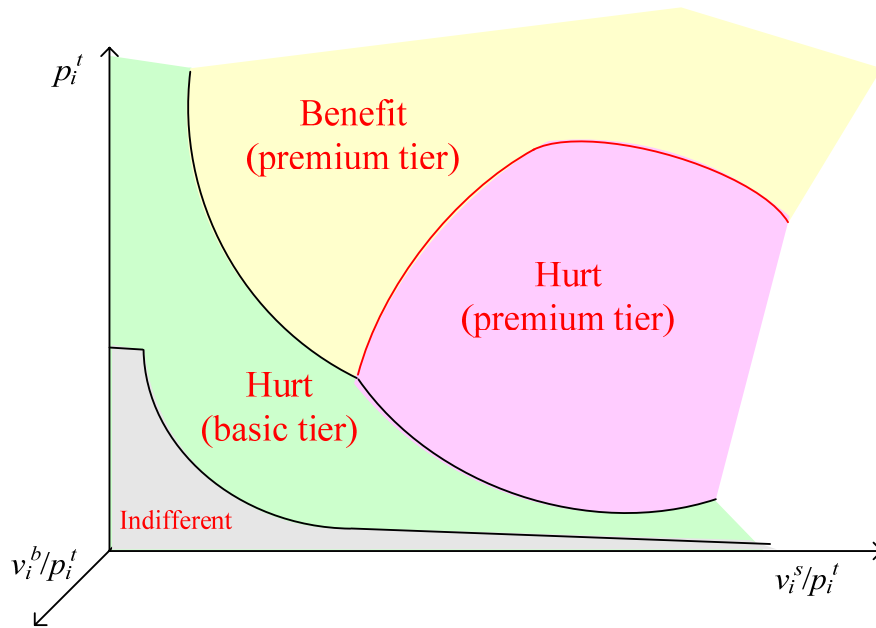
Figure 3.4 Comparing user surplus when a profit-maximizing cap is implemented.

According to the above analysis of the impact of data caps upon user tier choices, users who place low values on video streaming (i.e. v^s/p^t) or time (i.e. p^t) (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 3.5. Premium subscribers with moderate to high valuations on video streaming and moderate 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 the optimal data cap; the effect of the cap and overage charges outweigh the reduction in the 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, i.e. a subscriber on the boundary between the white and dark yellow regions in Figure 3.3. Denote by 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 3.3, 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 3.5, $\Delta X_1 < 0$, $\Delta X_2 > 0$, 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 discussed above, the users in the dark yellow region in Figure 3.3 upgrade from the basic to 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, i.e. a subscriber on the boundary between the light green and dark yellow regions in Figure 3.3.

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 3.6: 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, $\Delta S\left(p^t,v^b,v^s\right)>0, \ \forall v^s : h_2\left(p^t,v^b\right)<v^s<v^{th}$ and $\Delta S\left(p^t,v^b,v^s\right)\leq 0, \ \forall v^s : v^s \geq v^{th}$.

Proof: Proof is trivial.

Figure 3.5 User surplus under the pricing plan without caps and the pricing plan with absolute caps, i.e. $p^o = \infty$.

Theorem 3.6 is illustrated in Figure 3.5. 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 exists 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 [66], and a concave polynomial function $V^s(t)=ct^k$ ($0<t<t^{max}$, $0<k<1$) which produces a constant elasticity demand curve [66].

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:

Theorem 3.7: If $V^s(t)=at^2+bt$ ($0<t <-b/2a=t^{max}$, $a<0$, $b>0$):

1) For a fixed (p^t, v^b), if $\Delta S(p^t, v^b, v^{s4}) \geq 0$, then $\Delta S\left(p^t, v^b, v^s\right) \geq 0$, $\forall v^s : v^s \geq h_2\left(p^t, v^b\right)$.

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 $v^{th,2}$ ($v^{th,1} \leq v^{th,2}$). Furthermore, $\Delta S\left(p^t, v^b, v^s\right) \geq 0$, $\forall v^s : h_2\left(p^t, v^b\right) \leq v^s \leq v^{th,1}$, $\Delta S\left(p^t, v^b, v^s\right) \leq 0$, $\forall v^s : v^{th,1} \leq v^s \leq v^{th,2}$, and $\Delta S\left(p^t, v^b, v^s\right) \geq 0$, $\forall v^s : v^s \geq v^{th,2}$.

Proof: see Appendix.
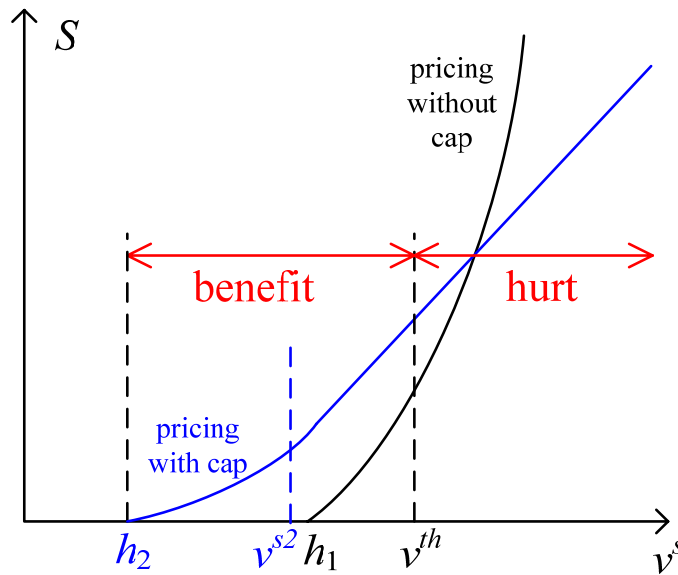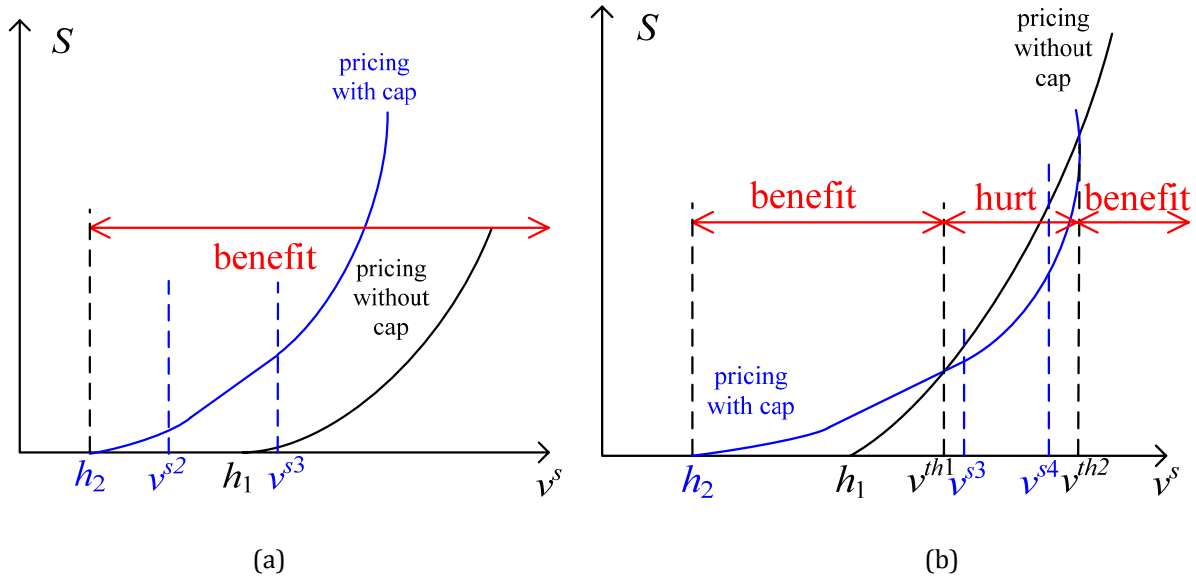

Figure 3.6 User surplus under the pricing plan without data caps and the pricing plan with data caps.

Theorem 3.7 is illustrated in Figure 3.6. The first case occurs when the cross section lies above the maximum point on the boundary, i.e. entirely within the light yellow region of Figure 3.4. In Figure 3.5(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, i.e. it crosses from the light yellow region to the light pink region and back into the light yellow region. In Figure 3.5(b), at the point v^{s4}, $\Delta S(p^t, v^b, v^{s4}) < 0$, i.e. 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 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 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 3.8: If $V^s(t)=ct^k$ ($0<t<t^{max}$, $0<k<1$):

1) For a fixed (p^t,v^b), if $\left(\dfrac{p^t+p^o\left(X_2+\Delta X_2\right)}{p^t}\right)^k \leq \dfrac{Q^s\left(X_2+\Delta X_2\right)}{Q^s\left(X_2\right)}$, then

$\Delta S\left(p^t,v^b,v^s\right)\geq 0,\ \forall v^s:v^s\geq h_2\left(p^t,v^b\right)$

2) For a fixed (p^t,v^b), if $\left(\dfrac{p^t+p^o\left(X_2+\Delta X_2\right)}{p^t}\right)^k > \dfrac{Q^s\left(X_2+\Delta X_2\right)}{Q^s\left(X_2\right)}$, then there exists a unique

root in v^s of $\Delta S(p^t,v^b,v^s)$, denoted v^{th}.

Furthermore, $\Delta S\left(p^t,v^b,v^s\right)>0,\ \forall v^s:h_2\left(p^t,v^b\right)<v^s<v^{th}$, and $\Delta S\left(p^t,v^b,v^s\right)\leq 0,\ \forall v^s:v^s\geq v^{th}$.

Proof: The proof is similar to that of theorem 3.7, and is omitted.

The first case occurs when the cross section lies above the boundary, i.e. entirely within the light yellow region of Figure 3.4. The second case occurs when the cross section lies crosses the boundary, i.e. it crosses 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.

SUBCHAPTER 3.6: Numerical Results

In this subchapter, 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 [63][64][67], and with a constant elasticity demand curve given by $V^b(t) = ct^k$ with parameters set to match the web browsing statistics in [63]. 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 [53][64][67], and with a constant elasticity demand curve given by $V^s(t) = ct^k$ with parameters set to match the video streaming statistics in [53]. Video streaming performance $Q^s(x)$ as a function of throughput is taken from [55]. The load threshold $\rho^{th} = 0.7$ [65]. The marginal price per unit capacity p^μ = \$10/Mbps/month [56].

In Figure 3.7, 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 rapidly with the proportion of heavy users, due to both the higher willingness-to-pay of heavy users and the

much higher usage (and therefore cost) of heavy users. When profit-maximizing data caps are used, the premium tier price P_2 decreases substantially, as predicted by the analytical results above. The prices both with and without caps would fall if there are competing ISPs. The basic tier price P_1 remains approximately unchanged after profit-maximizing data caps are added, as predicted by the analytical results above.
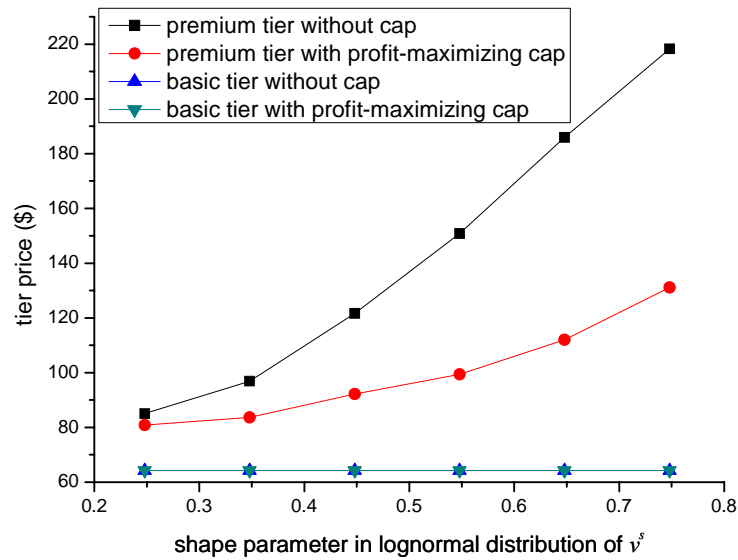
Figure 3.7 Tier price versus proportion of heavy users.

In Figure 3.8, 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 above. However, the basic tier rate X_1 only decreases slightly after profit-maximizing data caps are added. This occurs because the basic tier rate X_1 is already good enough for web browsing but far not good enough for video streaming,

and thus changing X_1 only slightly changes the quality of web browsing and video streaming.
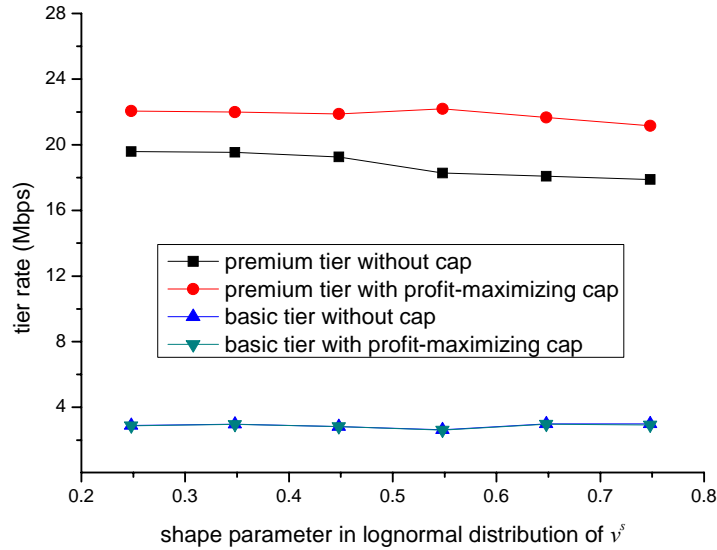
Figure 3.8 Tier rate versus proportion of heavy users.

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 3.9 and Figure 3.10, we illustrate the caps and overage charges under the profit-maximizing cap and the less aggressive heavy users cap discussed above. Under the heavy user 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 quite high. Current wireline ISP caps are roughly in this 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 [3].
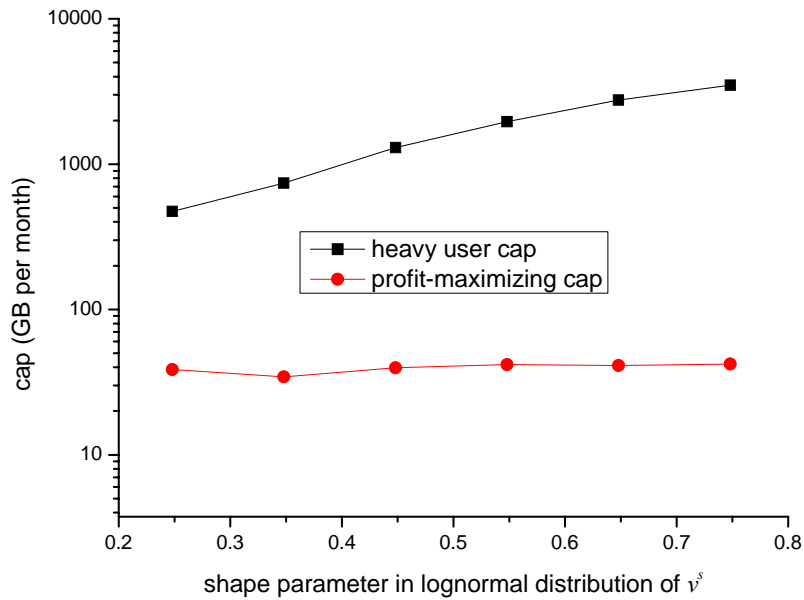
Figure 3.9 Data caps versus proportion of heavy users.
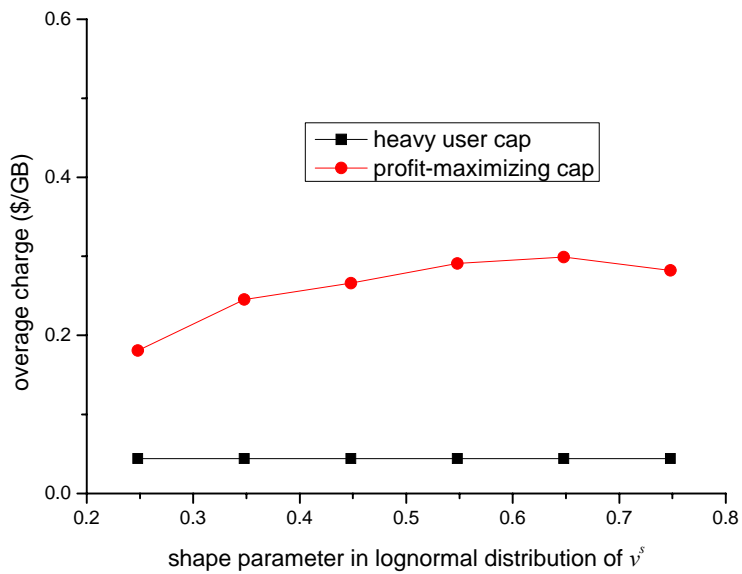
Figure 3.10 Overage charges versus proportion of heavy users.

To understand the impact upon different types of users, in Figure 3.11, we investigate which users are better off with and without profit-maximizing data caps, as a function of

$(v_i^s / p_i^t , p_i^t)$, when σ=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 rates. 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, while the basic tier price remains the same.
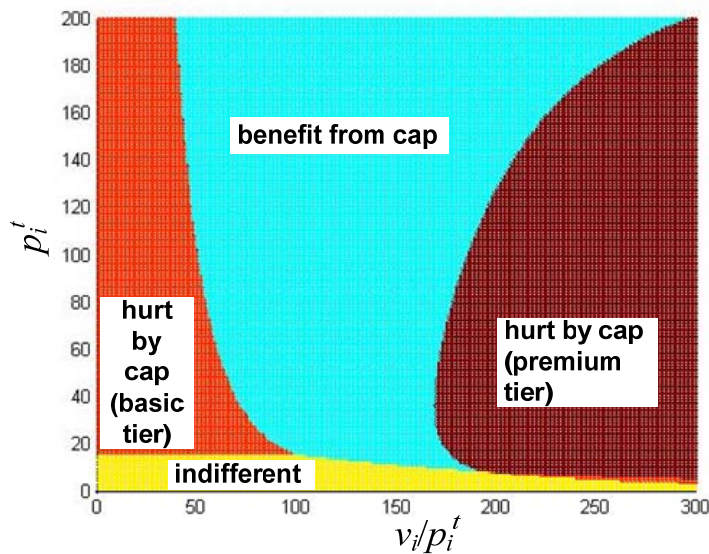
Figure 3.11 User surplus under profit-maximizing caps.

In Figure 3.12, we give the ISP profit, user surplus (defined as $\sum_i S_i$), 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

87

further increases substantially. The increase reflects the new overage charges, minus some 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.
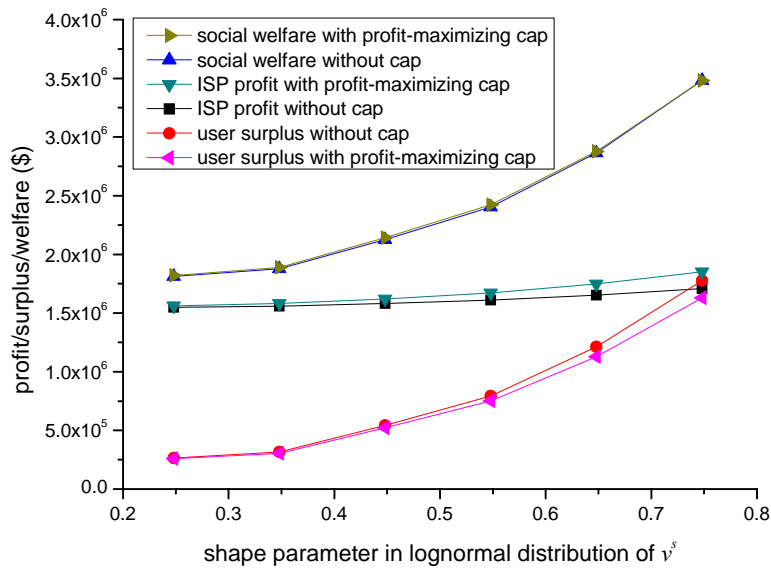
Figure 3.12 ISP profit, user surplus, and social welfare versus proportion of heavy users.

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 3.13, 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. Of greater interest is that 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.
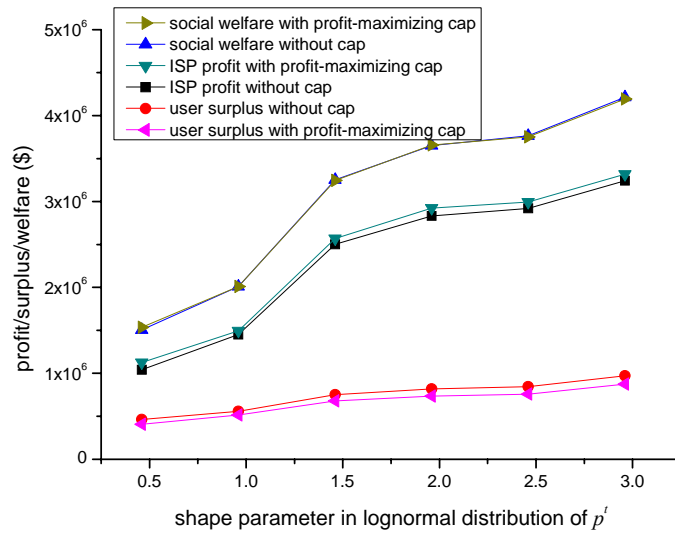
Figure 3.13 ISP profit, user surplus, and social welfare versus the proportion of wealthy users.

CHAPTER 4: Impact of Data Caps on ISP Duopoly Competition

In this chapter, we are interested in the effect of competition upon the adoption and use of data caps. We are interested both in whether competition decreases the prevalence of data caps (or increases the size of the cap), and in how the use of data caps affects competition between ISPs.

In the United States, it is common that a user has the choice between two broadband ISPs: a cable company using cable modems to provide broadband access and a telephone company using Digital Subscriber Line (DSL) modems to provide broadband access. Thus we will restrict our analysis to duopoly competition. The transmission rates of cable and DSL modems depend on the technology deployed by the ISP. It is common that cable modems offer higher transmission rates than DSL modems. Some researchers are concerned with the effects of this disparity in transmission rates upon ISP competition [68]. We are thus also concerned with how the differences between cable and DSL providers affect the use of data caps. If data caps are used principally to recover the costs of heavy users, one might expect that each provider sets data caps according to the costs imposed by its heavy users, whereas if data caps are used principally to extract user surplus, one might expect that duopoly providers will set data caps according to the surplus of its subscribers. In either case, since cable and DSL providers attract different sets of subscribers, one should expect that they will set different data caps.

To address these concerns, we consider one cable ISP competing against one DSL ISP. Based on the differences in technology, we associate different fixed costs, marginal access costs per user, and marginal costs per unit capacity to the two ISPs. They offer broadband

Internet services that differ by service tier rates, prices, and potentially caps on the data received per month and overage charges. We consider two sets of Internet applications: low bandwidth applications (e.g. browsing, email) and high bandwidth applications (e.g. video streaming). Application utilities are represented as functions of users' relative interest placed on the Internet, the time devoted to each application, and the performance of each application.

Each ISP seeks to maximize profit by setting network capacity, service rates, prices, caps and overage charges. We show that the two ISPs each obtain a positive market share under certain conditions on each ISP's fixed cost and on the convexity of each ISP's profit function. Presuming that the DSL ISP has a lower variable access cost than does the cable ISP, we illustrate how a DSL ISP and a cable ISP compete and converge to serving lighter users and heavier users, respectively.

We then compare the profit-maximizing tier rates, tier prices, network capacity, and data caps to those used if an ISP institutes caps only to ensure that heavy users pay an amount equal to the cost of their usage (called a heavy-users cap). We show that after profit-maximizing caps are added into the pricing plans, a cable ISP who is serving heavier users may have the same incentive as in the monopoly case to increase the premium tier rate, reduce the premium tier price, reduce data caps below the heavy-users cap, and increase overage charges above those in the heavy-users cap. In contrast, the DSL ISP may 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 under data caps, the cable ISP's profit increases.

SUBCHAPTER 4.1: ISP Duopoly Model Formulation

In this subchapter, we introduce a new set of user utility functions for low bandwidth applications and high bandwidth applications, by simplifying the utility functions used in CHAPTER 2 and CHAPTER 3. On a time scale of days, Internet users choose how much time to devote to Internet applications. On a time scale of months, users choose what ISP to subscribe to, and two ISPs compete with each other by setting tier prices, tier rates, overage charges and data caps.

A. Short term model: user utility

We consider users differentiated by their interest placed on the Internet. Denote user i's interest level placed on the Internet by θ_i, normalized between 0 and 1.

We consider two sets of Internet applications -- low bandwidth applications (e.g. browsing, email) and high bandwidth applications (e.g. video streaming) -- and we model user utility separately for each set. We presume that the user utility (in \$/month) associated with low bandwidth applications depends solely on the user interest level θ_i, i.e. $U^b(\theta_i) = \gamma + \delta\theta_i$, where $\gamma > 0$ is the utility placed on low bandwidth applications by users with the lowest interest level ($\theta_i = 0$), and $\delta > 0$ is fixed. In contrast, for high bandwidth applications, we posit that user i's utility (in \$/month) should be a function not only of the user interest level θ_i, but also of the application service tier rate and of the amount of time that the user

will devote to the applications. Denote the service tier transmission rate (in bps[4]) of ISP j by X_j. Denote the amount of time (in seconds/month) that user i will devote to high bandwidth applications by t_i. Denote user i's utility associated with high bandwidth applications, when subscribed to ISP j, by $U^d(t_i, X_j, \theta_i)$. (The notation is summarized in Table 4.1)

Table 4.1 List of notations used in the ISP duopoly model

Notation	Interpretation
θ_i	user i's interest level placed on the Internet
$f(\theta)$	density function of user type in the market
X_j	service tier transmission rate of ISP j (in bps)
$Q(X_j)$	quality of high bandwidth applications in ISP j
P_j	service tier price of ISP j (in \$/month)
C_j	data cap of ISP j (in bytes/month)
p_i^o	overage charge of ISP j (in \$/byte)
t_i	time that user i will devote to high bandwidth applications (in seconds/month)
$U^b()$	user i's utility associated with low bandwidth applications (in \$/month)
$U^d()$	user i's utility associated with high bandwidth applications (in \$/month)
$U()$	user i's utility associated with all Internet applications (in \$/month)
T_i	subscription choice of user i
N_j	number of users subscribing to ISP j
Z_j	fixed cost of ISP j (in \$/month)
k_i	access cost per subscriber to ISP j (in \$/subscriber/month)
μ_j	capacity of the bottleneck link in ISP j (in bps)
p_j^μ	marginal cost for network capacity in ISP j (in \$/Mbps/month)
$K_j()$	cost of network capacity in ISP j (in \$/month)
π_j	profit of ISP j (in \$/month)

Since users place a value on leisure time, it is commonly presumed that $U^d(t, X_j, \theta_i)$ is a concave function of t with a peak when t is equal to the maximum amount of time that user i will devote to high bandwidth applications. We thus model user i's marginal utility by:

$$U_t^d\left(t, X_j, \theta_i\right) \triangleq \partial U^d\left(t, X_j, \theta_i\right)\big/\partial t = A_i g\left(t/B_i\right) \tag{4.1}$$

---

[4] We use *bps* to denote bits per second and *Mbps* to denote megabits per second.

where $g(t)$ is a decreasing function, normalized so that $g(0)=1$ and $g(1)=0$, and where A_i and B_i are scaling factors for the marginal utility and the amount of time devoted to the high bandwidth applications, respectively. The scaling factor A_i can be interpreted as user i's marginal utility over time for high bandwidth applications when $t=0$. The scaling factor B_i can be interpreted as the optimal amount of time user i spends on high bandwidth applications. We are only interested in the case $t/B_i<1$, as a user would not spend more than B_i on high bandwidth applications even in the absence of data caps.

Both scaling factors likely depend on the user interest θ_i; here we presume that both marginal utility A_i and time B_i on high bandwidth applications are linearly correlated with θ_i. Both scaling factors also likely depend on the service tier rate X_j since the performance or "quality" of high bandwidth allocations are sensitive to X_j. Denote the quality of high bandwidth applications by a function $Q(X_j)$ of service tier rate X_j [55]. It is less known how marginal utility and time depend on service quality. We expect that time devoted to high bandwidth applications depends jointly on user interest and application quality. Based on utility models in [35][65][69], we conjecture that both marginal utility A_i and time B_i of high bandwidth applications are linearly correlated with $\theta_i Q(X_j)$. These assumptions can be summarized as:

Assumption F: A_i and B_i are both linearly proportional to $\theta_i Q(X_j)$:

$$A_i = \alpha \theta_i Q\left(X_j\right), \quad B_i = A_i/\beta$$

where α and β are fixed ratios to be fitted to Internet statistics.

The dependence of a user's marginal utility upon the time devoted to high bandwidth applications is unknown. Here we assume that it is linearly decreasing:

Assumption G: $g(t) = 1 - t$, where $t \geq 0$.

A subscriber's usage demand in terms of time devoted to high bandwidth applications, under Assumption G, is a linear demand function of the overage charge. Linear demand is a common assumption in economic models when the precise form of the demand is unknown.

Based on assumptions F & G, if a user subscribes to ISP j, then that user's utility function and marginal utility function of all Internet applications can be expressed as a function of time t_i, service rate X_j and user type θ_i:

$$
\begin{aligned}
U\left(t_i, X_j, \theta_i\right) &= U_i^b + U_i^d = \gamma + \delta\theta_i + \alpha\theta_i Q\left(X_j\right)t_i - \beta t_i^2 / 2 \\
U_t\left(t_i, X_j, \theta_i\right) &= \alpha\theta_i Q\left(X_j\right) - \beta t_i
\end{aligned}
\tag{4.2}
$$

Similar to CHAPTER 2 and CHAPTER 3, suppose that ISP j charges P_j (in \$/month) for broadband Internet access at rate X_j and charges overage fees of p_j^o (in \$/byte) for downloads above a data cap of C_j (in bytes/month). The monthly usage of low bandwidth applications is typically negligible relative to that of high bandwidth applications [4], and thus user i's overage fees are approximately $p_j^o \max(0, X_j t_i - C_j)$. Each month, user i is assumed to choose the time devoted to the Internet, so as to maximize her surplus, $S_{i,j}$, defined as the difference between utility and cost:

$$
\max_{t_i} S_{i,j} = U\left(t_i, X_j, \theta_i\right) - p_j^o \max\left(0, X_j t_i - C_j\right) - P_j
\tag{4.3}
$$

B. Long term model: user subscription

We consider one DSL ISP denoted by j=1 and one cable ISP denoted by j=2. As noted above, it is common that the cable ISP will offer a substantially higher service tier rate than the DSL ISP, and thus we assume that $X_2 > X_1$. On a time scale of years, we presume that a user chooses an ISP (or not to subscribe) by maximizing long term user surplus. Denote user i's ISP choice by $T_i = j \in \{0,1,2\}$, where T_i=0 means that user i chooses not to subscribe to either ISP (in which case we set X_0=0 and P_0=0). Then:

$$T_i = \arg \max_{j \in \{0,1,2\}} \left[U\left(t_{i,j}^*, X_j, \theta_i\right) - p_j^o \max\left(0, X_j t_{i,j}^* - C_j\right) - P_j \right]$$
$$\text{where} \quad t_{i,j}^* = \arg \max_{t_i} S_{i,j}$$

(4.4)

Denote the total number of users in the market by N. Denote the set of users who subscribe to ISP j by $\mathbb{N}_j = \{\text{user } i \mid T_i = j\}$. Thus the demands for both ISPs are functions of the prices, rates, caps and overage charges:

$$N_j\left(\overline{P}, \overline{X}, \overline{C}, \overrightarrow{p^o}\right) = \left|\mathbb{N}_j\right|$$

where vectors $\overline{P}, \overline{X}, \overline{C}, \overrightarrow{p^o}$ are the prices, rates, caps and overage charges set by the DSL ISP and the cable ISP.

C. ISP duopoly competition game

ISP j's revenue (in \$/month) can be easily obtained from the demand functions:

$$Rev_j = P_j N_j + p_j^o O_j$$

where $O_j = \sum_{i \in \{\text{user } i | T_i = j\}} \max \left(0, X_j t_{i,j}^* - C_j\right)$ is the total amount of data above the cap.

An ISP's network cost is complex. For purposes of this analysis, we model the network cost (in \$/month) as comprised of three parts:

$$Cost_j = Z_j + k_j N_j + K_j \left(\mu_j\right)$$

All three terms reflect the cost (in \$/month) for building and maintaining the ISP's network. The first term Z_j is the portion of the cost that does not depend either on the number of subscribers or the traffic; we refer to this as the fixed cost. The second term $k_j N_j$ is the portion of the cost that depends on the number of subscribers; we refer to this as the variable access cost. The third term K_j is the portion of the cost that depends on the capacity of the network; we refer to this as the variable capacity cost. We represent this latter cost as a function of the capacity μ_j of the bottleneck link in the ISP's network. It has been shown that an ISP can achieve near-optimal profit by setting network capacity so that the network load remains below a threshold ρ^{th}, i.e. $\mu_j = \lambda_j/\rho^{th}$, where λ_j is the total downstream traffic (in bits/month) on the bottleneck link within the access network of ISP j [65]. As discussed above, we only consider high bandwidth applications when calculating the total downstream traffic, i.e. $\lambda_j = \sum_{i \in \{\text{user } i, T_i = j\}} X_j t_{i,j}^*$. This approach reduces the ISP's profit maximization problem to:

$$\max_{P_j, X_j, C_j, p_j^o} \pi_j = Rev_j - Cost_j = P_j N_j + p_j^o O_j - K_j \left(\lambda_j / \rho^{th} \right) - k_j N_j - Z_j \qquad (4.5)$$

On a time scale of months to years, the DSL ISP and the cable ISP compete by offering their Internet access services in a non-cooperative fashion, each trying to maximize its own profit defined in (4.5) by controlling its strategy vector $E_j = (P_j, X_j, C_j, p_j^o)$. Users respond with subscription choices according to (4.4). A Nash equilibrium is established if no ISP can unilaterally improve its profit by selecting an alternative strategy vector E_j. The Nash equilibrium of this non-cooperative game can thus be defined as follows:

Definition 4.1: Nash Equilibrium in ISP Duopoly Game. A pair of ISP strategy vectors (E_1^*, E_2^*), is said to constitute a Nash equilibrium iff:

$$\begin{cases} \pi_1 \left(E_1^*, E_2^* \right) \ge \pi_1 \left(E_1, E_2^* \right), \forall E_1 \in \mathbb{R}_+^4 \\ \pi_2 \left(E_1^*, E_2^* \right) \ge \pi_2 \left(E_1^*, E_2 \right), \forall E_2 \in \mathbb{R}_+^4 \end{cases}$$

The conditions for the existence of Nash equilibria in ISP duopoly game are illustrated in Appendix G.

SUBCHAPTER 4.2: ISP Competition Model Analysis

In the first subsection, we obtain specific conditions for the existence of Nash equilibria by simplifying the duopoly model, and use these conditions to determine whether the market results in a natural monopoly or duopoly. We then consider the use of such a pricing plan with data caps in a duopoly, based on the assumption that the conditions for natural monopoly do not hold. In the second subsection, we examine the time that each user will devote to high bandwidth applications. In the third subsection, we show how the market is segmented by the two ISPs.

A. Natural monopoly

In this subsection, we simplify the model so as to obtain specific conditions for the existence of Nash equilibria. These conditions are then used to determine whether the market results in a natural monopoly.

In this simplified model, we do not consider data caps, i.e. $C_1=C_2=\infty$. We assume that the service tier rates X_1 and X_2 have been chosen and fixed. We ignore the ISP variable capacity costs, i.e. $K_1(\mu_1)=K_2(\mu_2)=0$. We presume that the utility placed on low bandwidth applications by users with the lowest interest level $\gamma \geq P_1$ so that all users subscribe to broadband Internet access. Using (4.2)-(4.3), it can be easily shown that in the absence of data caps, the utility of high bandwidth applications of a user of type θ is $U^d(\theta)=\alpha^2\theta^2 Q\left(X_j\right)^2 \big/ 2\beta$. Denote the type of the user in the market with the lowest (resp. highest) interest level by $\theta_{\min}$ (resp. $\theta_{\max}$). We also assume that $U^d(\theta)$ is uniformly

distributed between $U^d(\theta_{\min})$ and $U^d(\theta_{\max})$. Under these conditions, we can give specific conditions for the existence and uniqueness of a Nash equilibrium:

Theorem 4.1: If assumptions F & G hold, users seek to maximize their surplus $S_{i,j}$ by controlling time t_i in (4.3), and ISPs seek to maximize profit (4.5) by controlling prices P_j, then there exists a unique Nash equilibrium to the simplified ISP duopoly game, in which both ISPs are earning positive profits, if and only if:

$$
\begin{cases}
2\theta_{\min}^2 - \theta_{\max}^2 < \dfrac{2\beta(k_2 - k_1)}{\alpha^2\left(Q(X_2)^2 - Q(X_1)^2\right)} < 2\theta_{\max}^2 - \theta_{\min}^2 \\[4mm]
Z_1 < \dfrac{\delta}{9}\left(\dfrac{\alpha^2\left(Q(X_2)^2 - Q(X_1)^2\right)}{2\beta}\left(\theta_{\max}^4 - 6\theta_{\max}^2\theta_{\min}^2 + 8\theta_{\min}^4\right) + \dfrac{2\beta(k_2 - k_1)^2}{\alpha^2\left(Q(X_2)^2 - Q(X_1)^2\right)} + (k_2 - k_1)\left(2\theta_{\max}^2 - 6\theta_{\min}^2\right)\right) \\[4mm]
Z_2 < \dfrac{\delta}{9}\left(\dfrac{\alpha^2\left(Q(X_2)^2 - Q(X_1)^2\right)}{2\beta}\left(4\theta_{\max}^4 - \theta_{\min}^4\right) + \dfrac{2\beta(k_2 - k_1)^2}{\alpha^2\left(Q(X_2)^2 - Q(X_1)^2\right)} - 4(k_2 - k_1)\theta_{\max}^2\right)
\end{cases}
$$

Proof: See Appendix H.

The theorem can be used to determine whether the market results in a natural monopoly, where for purposes of this discussion we say that a natural monopoly occurs if under the Nash equilibrium one of the ISPs is earning negative profit:

Corollary 4.1: If assumptions F & G hold, users seek to maximize their surplus $S_{i,j}$ by controlling time t_i in (4.3), ISPs seek to maximize profit (4.5) by controlling prices P_j, and $Z_1 = Z_2 = 0$, then:

1) If $\dfrac{k_2 - k_1}{Q(X_2)^2 - Q(X_1)^2} \geq \dfrac{\alpha^2\left(2\theta_{\max}^2 - \theta_{\min}^2\right)}{2\beta}$, the DSL ISP will monopolize the market.

2) If $\dfrac{k_2 - k_1}{Q(X_2)^2 - Q(X_1)^2} \leq \dfrac{\alpha^2 \left(2\theta_{\min}^2 - \theta_{\max}^2\right)}{2\beta}$, the cable ISP will monopolize the market.

3) Otherwise, the DSL ISP and the cable ISP will share the market.

Proof: If $2\theta_{\min}^2 - \theta_{\max}^2 < \dfrac{2\beta\left(k_2 - k_1\right)}{\alpha^2 \left(Q(X_2)^2 - Q(X_1)^2\right)} < 2\theta_{\max}^2 - \theta_{\min}^2$, we can prove that both the DSL

ISP and the cable ISP share the market according to Theorem 4.1 under the condition that

$Z_1 = Z_2 = 0$.

If $\dfrac{2\beta\left(k_2 - k_1\right)}{\alpha^2 \left(Q(X_2)^2 - Q(X_1)^2\right)} \geq 2\theta_{\max}^2 - \theta_{\min}^2$, the DSL ISP and the cable ISP will not share the

market because the condition in Theorem 4.1 is not satisfied. The DSL ISP will monopolize

the market by setting its tier price $P_1 = k_2 - \alpha^2 \left(Q(X_2)^2 - Q(X_1)^2\right)\theta_{\max}^2 \big/ 2\beta > k_1$, where no

users will subscribe to the cable ISP even when the cable ISP sets its tier price $P_2 = k_2$.

If $\dfrac{2\beta\left(k_2 - k_1\right)}{\alpha^2 \left(Q(X_2)^2 - Q(X_1)^2\right)} \leq 2\theta_{\min}^2 - \theta_{\max}^2$, the cable ISP will monopolize the market by setting

its tier price $P_2 = k_1 + \alpha^2 \left(Q(X_2)^2 - Q(X_1)^2\right)\theta_{\min}^2 \big/ 2\beta > k_2$, where no users will subscribe to

the DSL ISP even when the DSL ISP sets its tier price $P_1 = k_1$.

■

Whether a natural monopoly occurs thus depends on the parameters α and β that

characterize user utility of high bandwidth applications, the ISP variable access cost per

subscriber k_j, and the range of user interest $[\theta_{\min}, \theta_{\max}]$. It is reasonable to presume that the

better technology (here, the cable ISP) has both higher variable access cost per subscriber, i.e. $k_2 > k_1$, and higher quality, i.e. $Q(X_2) > Q(X_1)$. Case 1 thus states that if the ratio of the incremental cost $k_2 - k_1$ to the incremental squared quality $Q(X_2)^2 - Q(X_1)^2$ is above a certain threshold, then the cable ISP is not competitive and eventually will leave the market, resulting in the DSL ISP being a natural monopoly. Case 2 states that if the same ratio is below a lower threshold, then the DSL ISP is not competitive and eventually will leave the market, resulting in the cable ISP being a natural monopoly. Case 3 states that if the same ratio is between the two thresholds, then both ISPs will earn non-negative profit under the Nash equilibrium and thus remain in the market. Finally, the DSL ISP is more likely to monopolize the market when users place less value on their marginal utility over time (i.e. smaller α) or spend less time on high bandwidth applications (i.e. larger β). The cable ISP is more likely to monopolize the market when users place more value on their marginal utility over time (i.e. larger α) or spend more time on high bandwidth applications (i.e. smaller β).

B. User time devoted to high bandwidth applications

Users may devote less time to high bandwidth applications in the presence of data caps. The amount of time devoted affects the usage charge, and potentially affects the user choice amongst the ISPs. The following theorem partitions users into three groups: users who are not capped, users who are capped without paying an overage charge, and users who are capped and paying an overage charge.

Theorem 4.2: If assumptions F & G hold, and users seek to maximize their surplus $S_{i,j}$ by controlling time t_i in (4.3), then there exist thresholds $\vartheta_j^{(1)} = \beta C_j \big/ \alpha Q(X_j) X_j$ and $\vartheta_j^{(2)} = \left(\beta C_j + p_j^o X_j^2\right)\big/ \alpha Q(X_j) X_j$, such that:

1) A user with $0 \le \theta_i < \vartheta_j^{(1)}$ who subscribes to ISP j will devote time $t_i = \alpha \theta_i Q(X_j) \big/ \beta$ to high bandwidth applications, will have usage below ISP j's data cap, and will thus not pay an overage charge.

2) A user with $\vartheta_j^{(1)} \le \theta_i \le \vartheta_j^{(2)}$ who subscribes to ISP j will devote time $t_i = C_j / X_j$ to high bandwidth applications, will have usage equal to ISP j's data cap, and will thus not pay an overage charge.

3) A user with $\vartheta_j^{(2)} < \theta_i \le 1$ who subscribes to ISP j will devote time $t_i = \left(\alpha \theta_i Q(X_j) - p_j^o X_j\right) \big/ \beta$ to high bandwidth applications, will have usage above ISP j's data cap, and will thus pay an overage charge.

Proof: Focus on user i who subscribes to ISP j. If $0 \le \theta_i < \vartheta_j^{(1)}$, her marginal utility is smaller than zero when $t_i = C_j/X_j$, i.e. $d U(t, X_j, \theta_i) \big/ dt \big|_{t = C_j / X_j} < 0$. Thus, user i consumes data $t_i X_j = \alpha \theta_i Q(X_j) X_j \big/ \beta < C_j$, and will not be capped.

If $\vartheta_j^{(1)} \le \theta_i \le \vartheta_j^{(2)}$, her marginal utility is non-negative but not larger than $p_j^o X_j$ when $t_i = C_j/X_j$, i.e. $0 \le d U(t, X_j, \theta_i) \big/ dt \big|_{t = C_j / X_j} \le p_j^o X_j$. Thus, user i consumes data $t_i X_j = C_j$, and will be capped without paying an overage charge.

If $\vartheta_j^{(2)} < \theta_i \leq 1$, her marginal utility is larger than $p_j^o X_j$, i.e. $dU(t, X_j, \theta_i)/dt\big|_{t=C_j/X_j} > p_j^o X_j$.

Thus, user i consumes data $t_i X_j = \left(\alpha \theta_i Q(X_j) - p_j^o X_j\right) X_j/\beta > C_j$, and will pay an overage charge.

■

C. Product differentiation and heterogeneous users

In this subsection, we use Theorem 4.2 to determine how the market is segmented by the two ISPs. The results will illustrate how ISP choices of tier prices and rates will affect market shares.

As above, we presume that the cable ISP's technology results in a higher tier rate, i.e. $X_2 > X_1$, and thus a higher quality, i.e. $Q(X_2) > Q(X_1)$. Correspondingly, we presume that the cable ISP has a higher variable access cost per subscriber, i.e. $k_2 > k_1$, and will set a higher service tier price, i.e. $P_2 > P_1$. In economic terms, the two ISPs thus compete using vertical services differentiation[5].

Recall that user interest in Internet access is characterized by the user type $0 \leq \theta_i \leq 1$. Denote the density (resp. cumulative distribution) function of user type by $f(\theta)$ (resp. $F(\theta)$). Users with higher interest levels have higher utilities at the same level of usage. Since two ISPs are differentiated by quality, one would normally expect that users with higher interest levels are more likely to subscribe to the cable ISP. Denote the minimum user type that

---

[5] The model also bears some resemblance to a Hotelling model [29]. However, in a classical Hotelling model, some users prefer each product, whereas here we presume that all users would prefer the higher tier rate if it were offered at the same price as the lower tier rate. More importantly, in our model ISPs choose both tier prices and tier quality as strategic variables.

purchases broadband Internet access (henceforth called a marginal user of the DSL ISP) by $\theta^{(1)}$. Denotes the minimum user type that purchases broadband Internet access from the cable ISP (henceforth called a marginal user of the cable ISP) by $\theta^{(2)}$.

We investigate how the market is segmented by the two ISPs, and how data caps may alter the market segmentation. We start with the case in which there are no data caps, and determine the market share N_j of each ISP:

Theorem 4.3: If assumptions F & G hold, there are no data caps i.e. $C_1=C_2=\infty$, and users seek to maximize their surplus $S_{i,j}$ by controlling time t_i in (4.3), then there exist thresholds

$$\theta^{(1,1)} = \left(\sqrt{\beta^2\delta^2 + 2\beta(P_1 - \gamma)\alpha^2 Q(X_1)^2} - \delta\beta \right) \Big/ \alpha^2 Q(X_1)^2,$$

$$\theta^{(1,2)} = \left(\sqrt{\beta^2\delta^2 + 2\beta(P_2 - \gamma)\alpha^2 Q(X_2)^2} - \delta\beta \right) \Big/ \alpha^2 Q(X_2)^2, \text{ and}$$

$$\theta^{(2)} = \sqrt{2\beta(P_2 - P_1)} \Big/ \alpha\sqrt{Q(X_2)^2 - Q(X_1)^2} \text{ such that:}$$

1) If $P_1 \big/ Q(X_1)^2 \leq P_2 \big/ Q(X_2)^2$, users within $0 \leq \theta < \max\left(\theta^{(1,1)}, 0\right)$ will not subscribe to either ISP, users within $\max\left(\theta^{(1,1)}, 0\right) \leq \theta < \min\left(\theta^{(2)}, 1\right)$ will subscribe to the DSL ISP, and users within $\min\left(\theta^{(2)}, 1\right) \leq \theta \leq 1$ will subscribe to the cable ISP. Thus, $N_1 = N\left(F\left(\theta^{(2)}\right) - F\left(\theta^{(1,1)}\right)\right), N_2 = N\left(1 - F\left(\theta^{(2)}\right)\right)$.

2) If $P_1 \big/ Q(X_1)^2 > P_2 \big/ Q(X_2)^2$, users within $0 \leq \theta < \max\left(\theta^{(1,2)}, 0\right)$ will not subscribe to either ISP, no users will subscribe to the DSL ISP, and users within $\max\left(\theta^{(1,2)}, 0\right) \leq \theta \leq 1$ will subscribe to the cable ISP. Thus, $N_1 = 0, N_2 = N\left(1 - F\left(\theta^{(1,2)}\right)\right)$.

Proof: In the absence of data caps, a user i within $\theta_i \geq \dfrac{\sqrt{2\beta(P_2 - P_1)}}{\alpha\sqrt{Q(X_2)^2 - Q(X_1)^2}}$ prefers the

cable ISP to the DSL ISP, since $U(t^*_{i,2}, X_2, \theta_i) - P_2 \geq U(t^*_{i,1}, X_1, \theta_i) - P_1$. A user i within

$\theta_i \geq \left(\sqrt{\beta^2\delta^2 + 2\beta(P_1 - \gamma)\alpha^2 Q(X_1)^2} - \delta\beta\right)\Big/\alpha^2 Q(X_1)^2$ prefers the DSL ISP to no broadband

Internet subscription, since $U(t^*_{i,1}, X_1, \theta_i) - P_1 \geq 0$. A user i within

$\theta_i \geq \left(\sqrt{\beta^2\delta^2 + 2\beta(P_2 - \gamma)\alpha^2 Q(X_2)^2} - \delta\beta\right)\Big/\alpha^2 Q(X_2)^2$ prefers the cable ISP to no broadband

Internet subscription, since $U(t^*_{i,2}, X_2, \theta_i) - P_2 \geq 0$. Thus, theorem 4.3 follows directly from

the above user subscription choice.

■

As expected, the cable ISP obtains a higher market share when it sets a lower service tier

price P_2 or a higher service tier rate X_2. The market share of each ISP is also related to the

user utility function. The cable ISP obtains a larger market share when users place more

value on their marginal utility over time (i.e. larger α) or spend more time on high

bandwidth applications (i.e. smaller β).

We now turn to the case when data caps are present. We first show that if the cable ISP's

data cap is large enough and/or its overage charge is low enough, then the market is

similarly segmented so that users with higher interest levels subscribe to the cable ISP:

Theorem 4.4: If assumptions F & G hold, users seek to maximize their surplus $S_{i,j}$ by

controlling time t_i in (4.3), and $\dfrac{C_2}{p_2^o} > \dfrac{X_2^2 Q(X_1)^2}{\beta\left(Q(X_2)^2 - Q(X_1)^2\right)}$, then there exists a threshold α^*

and a threshold $\theta^{(2)}$ such that for all $\alpha > \alpha^*$, users within $0 \le \theta < \theta^{(2)}$ prefer the DSL ISP to the

cable ISP and users within $\theta^{(2)} \le \theta < 1$ prefer the cable ISP to the DSL ISP.

Proof: See Appendix I.

However, such a market segmentation does not always hold. If the cable ISP's data cap is

too low and/or its overage charge is too high, then this may cause very heavy users to

switch from the cable ISP to the DSL ISP. An extreme case can illustrate this. In the case in

which the cable ISP sets an absolute data cap, i.e. $p_2^o = \infty$, we can show that the market

segments such that, while users with low interest levels subscribe to the DSL ISP and users

with moderate low interest level subscribe to the cable ISP (as before), the heaviest users

switch to the DSL ISP to avoid high overage charges:

Theorem 4.5: If assumptions F & G hold, users seek to maximize their surplus $S_{i,j}$ by

controlling time t_i in (4.3), $p_1^o < \infty$, $p_2^o = \infty$, and $C_2 > X_2 \sqrt{\dfrac{2(P_2 - P_1)Q(X_2)^2}{\beta\left(Q(X_2)^2 - Q(X_1)^2\right)}}$, then there

exists a threshold α^* and two thresholds $\theta^{(2)}$ and $\theta^{(3)}$ such that for all $\alpha > \alpha^*$, users within

$0 \le \theta < \theta^{(2)}$ and $\theta^{(3)} \le \theta < 1$ prefer the DSL ISP to the cable ISP and users within $\theta^{(2)} \le \theta < \theta^{(3)}$ prefer

the cable ISP to the DSL ISP.

Proof: See Appendix J.

SUBCHAPTER 4.3: Impact of Data Caps on ISPs

In SUBCHAPTER 4.2, we derived conditions that determine whether the market results in a natural monopoly. In this subchapter, we analyze the impact of data caps on an ISP duopoly, under the assumption that the market is not a natural monopoly. We wish to compare the service rates, service prices, and network capacity at the Nash equilibrium without caps to the same quantities at the Nash equilibrium with data caps.

The game proceeds in three stages. First, we allow the two ISPs to compete on service tier price and rate, with no data caps or overage charges. In the absence of data caps, the ISP's profit maximization problem in (4.5) can be reduced to:

$$\max_{P_j, X_j} \pi_j = P_j N_j - K_j \left(\lambda_j / \rho^{th} \right) - k_j N_j - Z_j \tag{4.6}$$

The two ISPs compete in a non-cooperative fashion in which each ISP j attempts to maximize its own profit by controlling its strategy vector $E_j = (P_j, X_j)$. The Nash equilibrium is characterized in Definition 4.1.

We then turn to the effect of allowing each ISP to adopt data caps. The second stage of the game starts with the initial condition set to be the Nash equilibrium of the first stage (without data caps). In the second stage, each ISP institutes caps only to ensure that heavy users pay an amount equal to the cost of their usage. This stage is interesting in its own right, as some ISPs claim this is the purpose of their data caps.

Denote the marginal cost for network capacity μ in ISP j by $p_j^{\mu} = dK_j(\mu_j)/d\mu_j$. In the second stage, we suppose that ISP j imputes a cost to user i equal to $k_j + p_j^{\mu} X_j t_i / \rho^{th}$, on the

basis that user i's usage is $X_j t_i$ and that this requires incremental capacity $X_j t_i / \rho^{th}$ at an incremental cost per unit capacity p_j^μ. Given the price P_j and rate X_j under the Nash equilibrium without data caps, we presume that the goal of ISP j in the second stage is to set a data cap C_j and overage charge p_j^o so that:

$$P_j + p_j^o \max\left(0, t_i X_j - C_j\right) \geq p_j^\mu t_i X_j / \rho^{th} + k_j \text{ for all } i$$

Similar to the heavy users cap in CHAPTER 3, we examine a simple method of achieving this goal: $p_j^o = p_j^\mu / \rho^{th}$ and $C_j = \left(P_j - k_j\right) / p_j^o$, which we still henceforth refer to as the *heavy-users cap*. Under this policy, subscribers to ISP j with usage greater than or equal to C_j will pay an amount $P_j + p_j^o \left(t_i X_j - C_j\right)$, which exactly equals their imputed cost, i.e. $p_j^\mu t_i X_j / \rho^{th} + k_j$.

Recall that in theorem 4.2 and 4.3, we denoted the type of the marginal users who are indifferent between ISP j and no Internet subscription by $\theta^{(1,j)}$, the type of the marginal users who are indifferent between DSL ISP and cable ISP by $\theta^{(2)}$, and the smallest type of users whose usage is equal to the heavy-users cap in ISP j by type $\vartheta_j^{(1)}$.

We first consider the effect of the DSL ISP's heavy-users cap. We distinguish between scenario a-1 in which no DSL ISP subscribers are impacted by its heavy-users cap, and scenario a-2 in which some are:

Scenario a-1: In the heavy-users cap Nash equilibrium, no DSL ISP subscribers are impacted by its heavy-users cap, i.e. $\theta^{(2)} < \vartheta_1^{(1)}$, as shown in Figure 4.1(a).

Scenario a-2: In the heavy-users cap Nash equilibrium, some DSL ISP subscribers are impacted by its heavy-users cap, i.e. $\theta^{(2)} \geq \vartheta_1^{(1)}$, as shown in Figure 4.1(b).
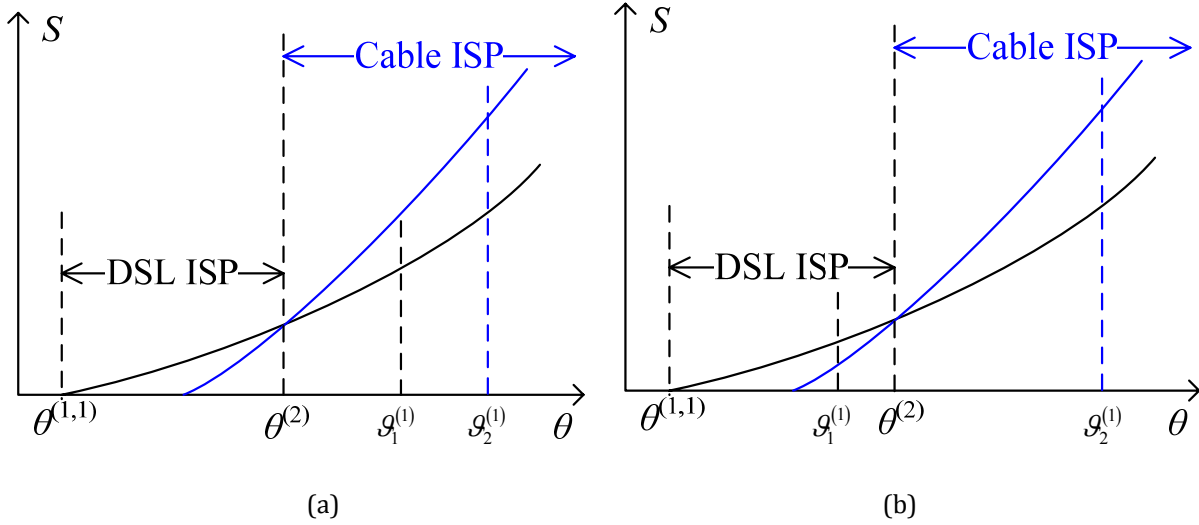
Figure 4.1 User surplus versus user type θ in the presence of data caps.

We next consider the effect of the cable ISP's heavy-users cap. According to theorems 4.4 and 4.5, if the cable ISP's data cap is too low and/or its overage charge is too high, then this may cause very heavy users to switch from the cable ISP to the DSL ISP. We distinguish between scenario b-1 in which the heaviest users subscribe to the cable ISP, and scenario b-2 in which the heaviest users subscribe to the DSL ISP:

Scenario b-1: In the heavy-users cap Nash equilibrium, there exists a threshold $\theta^{(2)}$ such that users with types $\max(0,\theta^{(1,1)}) \leq \theta < \theta^{(2)}$ subscribe to the DSL ISP and users with types $\theta^{(2)} \leq \theta < 1$ subscribe to the cable ISP, as shown in Figure 4.2(a).

Scenario b-2: In the heavy-users cap Nash equilibrium, there exists two thresholds $\theta^{(2)}$ and $\theta^{(3)}$ such that users with types $\max(0,\theta^{(1,1)}) \leq \theta < \theta^{(2)}$ and $\theta^{(3)} \leq \theta < 1$ subscribe to the DSL ISP and users with types $\theta^{(2)} \leq \theta < \theta^{(3)}$ subscribe to the cable ISP, as shown in Figure 4.2(b).
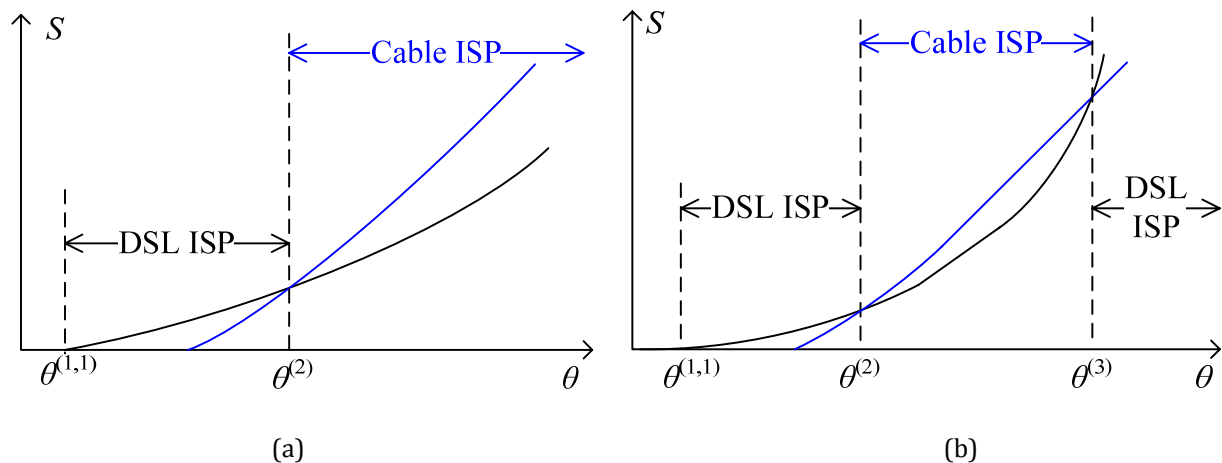
Figure 4.2 User surplus versus user type θ under heavy-users caps.

Any combination of scenarios a-1 and a-2 with scenarios b-1 and b-2 may occur. In SUBCHAPTER 4.4, we will present numerical results based on existing Internet statistics which show that the combination of scenario a-1 and scenario b-1 is most common, namely that the DSL ISP is unlikely to adopt a heavy-users cap and that the heaviest users subscribe to the cable ISP despite the overage charges.

Finally, we consider a third stage of the game in which both ISPs use data caps to maximize their profits. The pricing plans using a heavy-users cap do not maximize ISP profits, since the cap and overage charge were only intended to ensure that heavy users pay for their usage. In the third stage of the game, both ISPs set caps and overage charges to maximize their profits according to (4.5). Similar to CHPATER 3, we henceforth still refer to the optimal tier rate, tier price, cap, and overage charge at the profit-maximizing Nash equilibrium described in Definition 4.1 as the *profit-maximizing cap*.

Direct calculation of the profit-maximizing caps is difficult. Thus, we start at the heavy-users cap Nash equilibrium and examine the gradient of each ISP's profit with respect to tier prices, tier rates, caps, and overage charges. The next theorem analyzes the gradients

111

of profit with respect to each of these design parameters for each combination of scenarios. We presume that each ISP starts at its heavy-users cap Nash equilibrium. $E_j^{(0)}$ and then uses a gradient descent method to maximize its profit i.e. $E_j^{(n+1)} = E_j^{(n+1)} + \varepsilon \nabla \pi_j \left(E_j^{(n)} \right), n \geq 0$, where ε is the step size. Denote $\pi_j \left(E_j^{(n)} \right)$ by $\pi_j^{(n)}$. We require the following additional assumptions:

a. The step size ε is sufficiently small.

b. $\theta^{(2)} < \vartheta_2^{(1)}$, which means that marginal users indifferent between the DSL ISP and the cable ISP are not impacted by the cable ISP's heavy-users cap. This is to be expected, since otherwise the cable ISP is losing money on all of its subscribers if it does not adopt a cap.

c. Under scenario b-2, $\vartheta_1^{(1)} < \theta^{(3)}$, which means that all users with types $\theta^{(3)} \leq \theta < 1$ are impacted by the DSL ISP's heavy-users cap. We expect that scenario b-2 is rare, but if it does occur, it would be expected that the DSL ISP would set a heavy-users cap that would impact these users.

d. $Q'\left(X_j^{(0)} \right) \geq Q\left(X_j^{(0)} \right)\left(\vartheta_j^{(2)} - 2\vartheta_j^{(1)} \right) \Big/ X_j^{(0)} \vartheta_j^{(2)}$ for $j=1,2$, which means that users who are paying an overage charge to ISP j will consume more data if the tier rate X_j is increased, i.e. $\partial\left(X_j t_i \right) \big/ \partial X_j \Big|_{X_j = X_j^{(n)}} \geq 0$ for all θ_i within $\vartheta_j^{(2)} < \theta_i \leq 1$. This is to be expected, since an increase in the tier rate would increase the transmission rate of video streaming, which would in turn increase data consumption.

112

e. $f'\left(\theta^{(2)}\right) \geq Th$, where *Th* is given in the Appendix K. *Th* is usually a negative value, which means that the density function $f(\theta)$ should not decrease faster than a threshold *Th* when $\theta=\theta^{(2)}$. Based on the statistics of users' time devoted to the Internet applications [53], we can observe that $f(\theta)$ should be increasing with θ among light users (i.e. $f'(\theta)>0$), relatively flat among moderate users (i.e. $f'(\theta)\approx0$), and decreasing with θ among heavy users (i.e. $f'(\theta)<0$). We conjecture that marginal users with type $\theta^{(2)}$ should place moderate interest on the Internet, so that both ISPs have a positive market share. Thus, we have $f'(\theta^{(2)}) \approx 0 \geq Th$.

Theorem 4.6: Under additional assumptions (a)-(e) above:

$$\frac{\partial \pi_j^{(0)}}{\partial P_j} = 0, \frac{\partial \pi_j^{(0)}}{\partial X_j} = 0, \frac{\partial \pi_j^{(0)}}{\partial C_j} \leq 0, \frac{\partial \pi_j^{(0)}}{\partial p_j^o} \geq 0 \text{ for } j = 1,2$$

When scenarios a-1 and scenario b-1 occur, $\dfrac{\partial \pi_1^{(1)}}{\partial P_1} = 0, \dfrac{\partial \pi_1^{(1)}}{\partial X_1} = 0, \dfrac{\partial \pi_2^{(1)}}{\partial P_2} = 0, \dfrac{\partial \pi_2^{(1)}}{\partial X_2} \geq 0$

When scenarios a-1 and scenario b-2 occur, $\dfrac{\partial \pi_1^{(1)}}{\partial X_1} \geq 0, \dfrac{\partial \pi_2^{(1)}}{\partial X_2} \geq 0$

When scenarios a-2 and scenario b-1 occur, $\dfrac{\partial \pi_1^{(1)}}{\partial P_1} \leq 0, \dfrac{\partial \pi_1^{(1)}}{\partial X_1} \geq 0, \dfrac{\partial \pi_2^{(1)}}{\partial P_2} \geq 0$

When scenarios a-2 and scenario b-2 occur, $\dfrac{\partial \pi_1^{(1)}}{\partial X_1} \geq 0$

Proof: See Appendix K.

Theorem 4.6 thus predicts that if both ISPs start from the heavy-users cap Nash equilibrium, then both ISPs have the incentives to set overage charges p_j^o not smaller than p_j^μ / ρ^{th} and data caps C_j not larger than $\left(P_j - k_j \right) / p_j^o$, so as to extract profits from heavy users.

Initially, neither ISP has the incentive to change its tier price or tier rate, which are already optimized for profit maximization at the heavy-users cap Nash equilibrium. However, after overage charges and data caps are updated, ISPs may have the incentive to update their tier prices or tier rates. We consider these second order effects under each combination of scenarios.

If scenarios a-1 and b-1 occur (in which no DSL ISP subscribers are impacted by the DSL ISP's heavy-users cap and the heaviest users subscribe to the cable ISP), then $\partial \pi_1^{(0)} / \partial C_1 = 0, \partial \pi_1^{(0)} / \partial p_1^o = 0$, which means that the DSL ISP does not initially change its data cap or overage charge, since its subscribers are not affected by the heavy-users cap. The cable ISP does, however, decrease its data cap and increase its overage charge. As a result, after C_2 and p_2^o are updated, $\partial \pi_2^{(1)} / \partial X_2 \geq 0$, which means that the cable ISP will increase its tier rate X_2 to attract more subscribers, because the corresponding incremental cost to serve heavy users can now be more than compensated from overage charges. The cable ISP does not yet change its tier price, and the DSL ISP does not yet change either its tier price or tier rate. (They may both change these parameters in future iterations; both ISPs' actions at $n>1$ and the final profit-maximizing cap Nash equilibrium will be analyzed by simulation in SUBCHAPTER 4.4.)

If scenarios a-1 and b-2 occur (in which no DSL ISP subscribers are impacted by the DSL ISP's heavy-users cap and the heaviest users subscribe to the DSL ISP), then both ISPs initially decrease their data caps and increase their overage charges, since they are both serving heavy users. As a result, after these caps and overage charges are updated, $\partial \pi_j^{(1)}/\partial X_j \geq 0$ for both ISPs, which means that both ISPs will increase their tier rates to attract more subscribers, because for both ISPs the corresponding incremental cost to serve heavy users can now be more than compensated from overage charges. The second order effects on tier prices are indeterminate. An ISP may decrease its tier price to compensate for lost market share, but since both ISPs changed their caps simultaneously the change in market share is indeterminate.

If scenarios a-2 and b-1 occur (in which some DSL ISP subscribers are impacted by the DSL ISP's heavy-users cap and the heaviest users subscribe to the cable ISP), then both ISPs initially decrease their data caps and increase their overage charges, since they are both serving heavy users. After these caps and overage charges are updated, marginal users with type a little lower than $\theta^{(2)}$ switch from the DSL ISP to the cable ISP due to the DSL ISP's cap. As a result, $\partial \pi_1^{(1)}/\partial X_1 \geq 0$ and $\partial \pi_1^{(1)}/\partial P_1 \leq 0$, which means that the DSL ISP will increase its tier rate and decrease its tier price to attract more subscribers. The cable ISP will react to the DSL ISP's lower data cap by increasing its tier price. However, the change in the cable ISP's tier rate is now indeterminate; it may increase also its tier rate to attract more subscribers, but it might not need to due to its already higher market share.

Finally, if scenarios a-2 and b-2 occur (in which some DSL ISP subscribers are impacted by the DSL ISP's heavy-users cap and the heaviest users subscribe to the DSL ISP), then both

ISPs initially decrease their data caps and increase their overage charges, since they are both serving heavy users. As in the previous set of scenarios, after these caps and overage charges are updated, marginal users with type a little lower than $\theta^{(2)}$ switch from the DSL ISP to the cable ISP due to the DSL ISP's cap. As before, the DSL ISP will increase its tier rate to attract more subscribers. However, now the change in the DSL ISP's tier price is indeterminate, as are the changes in the cable ISP's tier rate and tier price, because the two ISP's are competing for marginal users on two fronts.

SUBCHAPTER 4.4: Numerical Results

In this subchapter, we numerically evaluate the impact of data caps on the pricing plans, users, ISP profits, and social welfare. We consider N=10,000 users in the market. To set the parameters of users' utility functions, we presume that the user type θ has a lognormal distribution (similar in shape to the distribution of household income), with mean and variance determined by fitting the statistics of household income [64], time devoted to video streaming [53], and willingness to pay for broadband Internet access [70]; this results in the median at θ=0.3. The parameters γ, δ for low bandwidth applications are also set by fitting these same statistics. User valuation of time is characterized by parameters α and β. We set α=$12.5/hour based on the distribution of household income (truncated at a maximum of $100,000/year) [64] and based on an estimate that users value leisure time at one fourth of their hourly wage [67]. We set β=0.125 hr²/$/month based on the maximum user income and the maximum amount of time users devote to video streaming [53]. The performance function $Q(x)$ is taken from [55].

The cost function of each ISP is comprised of three parts. The fixed costs of both ISPs are set to be equal: $Z_1=Z_2=\$10,000$/month. The variable access cost of the DSL ISP is set to $k_1=\$15$/user/month [71], and the variable access cost of the cable ISP is set to be \$20/user/month reflecting its more expensive per subscriber technology [72]. The variable capacity cost $K_j(\mu_j)$ is linearly proportional to capacity μ_j, where the marginal cost per unit capacity is set to \$10/Mbps/month for the DSL ISP and \$5/Mbps/month for the cable ISP (reflecting the different returns to scale of the two technologies) [56], and the load threshold $\rho^{th}=0.7$ [65]. In certain graphs, some of these parameters are modified as described below.

Unfortunately, these parameters choices are somewhat inconsistent with each other, as they are taken from studies in different years ranging from 2007 to 2012. In addition, good cost estimates are not available. As a consequence, the results below are presented to encourage an understanding of relative prices and rates, not as a prediction of absolute market prices.

In the first subsection, we compare pricing plans with and without data caps by showing how tier rates, tier prices, data caps and overage charges converge as the DSL ISP and cable ISP compete with each other. In the second subsection, we illustrate the impact of data caps on different set of users based on the previously obtained pricing plans. In the third subsection, we investigate the sensitivity to the marginal cost per unit capacity and to the mean and variance of user type.

A. The impact of data caps upon service plans and ISP profits

The simulation starts with the first stage of the game, in which the two ISPs compete on service tier price and rate to approach the Nash equilibrium without data caps. At iteration 1001, the second stage of the game begins when both ISPs add a heavy-users cap. The second stage continues with the ISPs competing on service tier price and rate to approach the heavy-users cap Nash equilibrium. At iteration 2001, the third stage of the game begins when both ISPs are now allowed to not only set service tier prices and rates but also to set data caps and overage charges to maximize profits. The third stage continues through iteration 6000 to approach the profit-maximizing cap Nash equilibrium.

Figure 4.3 illustrates the service tier prices and rates offered by each ISP. In the first stage of the game (iterations 1-1000), the tier prices and rates quickly converge in the absence of data caps. As expected, the cable ISP provides a higher tier rate than does the DSL ISP, due to its lower marginal cost per unit capacity. The corresponding tier price of the cable ISP is also higher than that of the DSL ISP[6]. As predicted by Theorem 4.3 case 2, light to moderate users subscribe to the DSL ISP, while moderate to heavy users subscribe to the cable ISP; we will examine user behavior in more detail below.

---

[6] We consider here only one tier offered by each service provider. In the United States, most broadband providers offer multiple tiers. The DSL ISP's service tier considered here should be thought of as the highest tier it offers, while the cable ISP's service tier considered here should be thought of as one of the lower tiers it offers. Consequently, the ISPs offer competing products, and without data caps there are only small differences in the offered tier rates and pries. The tier prices shown here are lower than that of most premium tiers [34][35].
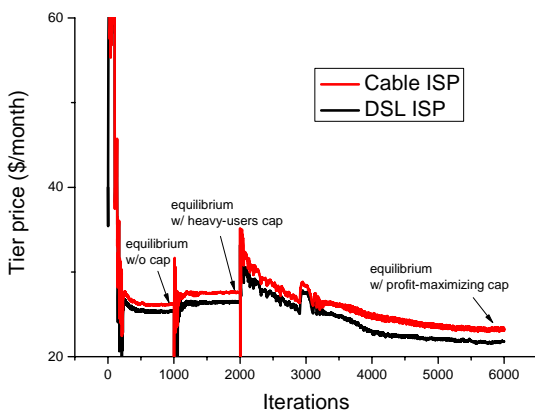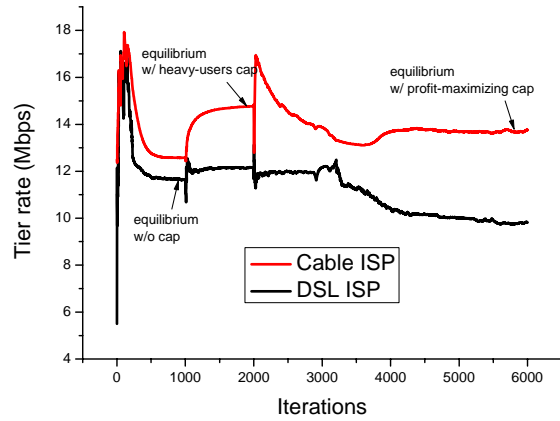
(a) Tier prices   (b) Tier rates

Figure 4.3 Tier prices and rates under duopoly competition

At iteration 1001, both ISPs adopt heavy-users caps. The overage charges are fixed at

$p_j^o = p_j^\mu / \rho^{th}$, and the data cap is a function $C_j = \left(P_j - k_j \right) / p_j^o$ of the tier price and the

overage charge. About 74 percent of the cable ISP's subscribers (18 percent of all users) are

affected by its heavy-users cap, but none of the DSL ISP's subscribers are affected by its

heavy-users cap[7]. In the second stage of the game (iterations 1001-2000), both ISPs

continue to compete on tier prices and rates to maximize profits. As a result, the cable ISP

increases its tier rate significantly, because it can now do so without losing money on the

heavier users. As discussed above, the DSL ISP's heavy-users cap is above the usage level of

any of its subscribers, meaning that the system is in scenario a-1. Consequently, the DSL

ISP's tier rate increases only slightly in an attempt to remain competitive with the cable

ISP. Both ISPs raise their tier prices slightly, because they can demand a higher price given

the higher tier rates. While the overage charges are fixed in this stage, the data caps vary

---

[7] In fixed broadband Internet service in the United States, we do not observe such a large percentage of users affected by data caps. This may reflect reluctance by cable ISPs to put in place data caps that affect a high percentage of its subscribers.

with the tier prices, as shown in Figure 4.4(a), and thus the data caps are also raised slightly. At iteration 2000 (close to the heavy-users cap Nash equilibrium), about 68 percent of the cable ISP's subscribers (16 percent of all users) are affected by its heavy-users cap, but none of the DSL ISP's subscribers are affected by its heavy-users cap.
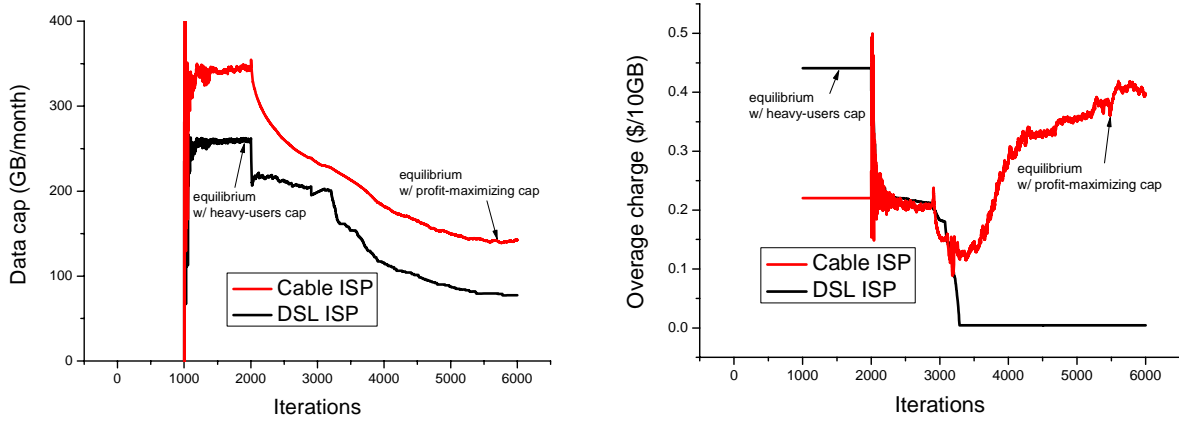
(a) Data caps   (b) Overage charge

Figure 4.4 Data caps and overage charges under duopoly competition

At iteration 2001, both ISPs are allowed to modify their data caps and overage charges to maximize their profits. In the third stage of the game (iterations 2001-6000), both ISPs compete on all four parts of their plans: tier prices, tier rates, caps, and overage prices. As we will see below, the heavy users subscribe to the cable ISP, and hence the system is in scenario b-1. Theorem 4.6 predicts that the cable ISP will decrease its data cap and increase its overage charge from the heavy-users cap to increase the profit that it earns from heavy users. Indeed, in Figure 4.4, we observe very large movements from iteration 2001 to iteration 2100, wherein the percentage of capped subscribers of the cable ISP quickly increases from 68% to 95% (from 16% to 23% of all users). Shortly after these initial movements, second order effects are observed. Theorem 4.6, for the combination of

scenarios a-1 and b-1, predicts that the cable ISP will increase it tier rate in response to its reduced data cap and increased overage charge, which we see as the quick jump in Figure 4.3(b), thereby attracting some of the DSL ISP's users. The tier price of the cable ISP is also pushed higher because of its higher tier rate; see the jump in Figure 4.3(a).

The DSL ISP, after initially increasing its tier price together with the cable ISP, starts to lose some of its subscribers. The DSL ISP reacts by starting a tier price war from iteration 2101 to iteration 2900. For the DSL ISP, lowering the tier price risks losing money on its heavier users, so it also decreases its data cap enough to collect a small amount of revenue from them. For the cable ISP, the reduction in tier price can be compensated by continued reductions in its data cap, resulting in increased revenue from overage charges. So, from iteration 2101 to iteration 2900, the percentage of capped subscribers of the cable ISP further increases from 95% to almost 100% (from 23% to 26% of all users). Whereas the cable ISP's initial response was to increase its tier rate, it now decreases its tier rate, since some lighter user who do not demand high tier rate begin to subscribe the cable ISP.

At approximately iteration 2900, when the DSL ISP's tier price is reduced below that in the heavy-users cap Nash equilibrium, it changes its strategy, and competes by decreasing its overage charge. It also briefly increases its tier price, before the tier price war continues. The cable ISP reacts by also briefly decreasing its overage charge and increasing its tier price. By iteration 3200, the DSL ISP has reduced its overage charge to zero, unable to extract much revenue from its relatively light users. (After iteration 3200, the DSL ISP's data cap no longer matters, since its subscribers are no longer charged for overage.)

After approximately iteration 3200, the cable ISP continues its reduction in its data cap and resumes increasing its overage charge, as it has a monopoly on the heaviest users. Almost all users in the cable ISP are capped, and the percentage of all Internet users who are capped increases, as the market share of the cable ISP increases. The DSL ISP keeps reducing its tier rate and tier price, in an attempt to attract even lighter users, who did not previously subscribe to broadband Internet access, so as to compensate for the lost market share. The tier prices of both ISPs stabilize.

At iteration 6000, the pricing plans have almost converged to the profit-maximizing Nash equilibrium. The tier prices at the profit-maximizing cap Nash equilibrium are substantially lower than either those at the Nash equilibrium without data caps or those at the heavy-users cap Nash equilibrium. The cable ISP's tier rate under profit-maximizing caps is higher than that without caps but lower than that under heavy-users caps. The DSL ISP's tier rate under profit-maximizing caps is lower than either that without caps or under heavy-users caps. The cable ISP's data cap is lower and its overage charge is higher than those under heavy-users caps. Almost all cable ISP subscribers (48 percent of all users) are capped by its profit-maximizing cap. The DSL ISP's data caps are not binding on its subscribers at either equilibrium.

The increased difference in tier rates between the two ISPs under profit-maximizing caps is consistent with economic theory. When prices are endogenously determined, economic theory predicts that firms will increase the differentiation of their products to reduce the intensity of price competition [73][74]. Here, under profit-maximizing caps, the price war causes the DSL ISP to decrease its tier price and rate, thereby attracting very light users

122

who didn't subscribe to the Internet when prices were higher. The cable ISP does not match the DSL ISP's reduction in tier rates, continuing to focus on moderate and heavy users, resulting in the greater differentiation.
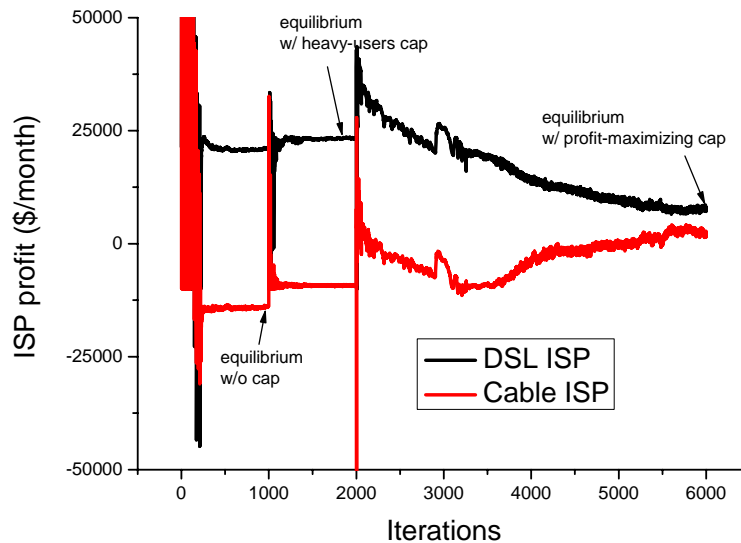
Figure 4.5 ISP profits under duopoly competition

We now turn to examining the resulting ISP profits, illustrated in Figure 4.5. At the Nash equilibrium without data caps, the two ISPs have a similar number of subscribers, with the DSL ISP serving light users and the cable ISP serving moderate and heavy users. However, the DSL ISP is earning a higher profit than is the cable ISP, because the cable ISP's slightly higher tier price does not compensate for the much higher capacity cost required to support the heavy users.

In the second stage of the game, at the heavy-users cap Nash equilibrium, the cable ISP's profit increases significantly, since it is no longer losing money for serving its heaviest users. In the third stage of the game, when the ISPs are allowed to adopt profit-maximizing

caps, both ISP's profits initially jump due to increased overage charges. However, when the price war ensues, both ISP's profits start to fall. After iteration 3200, the cable ISP has resumed its attempt to increase the profit earned by its heaviest subscribers, raising its profit. However the continuing price war decreases the DSL's profit. The cable ISP has an advantage, as the data cap and overage charge give it two degrees of the freedom that the DSL ISP can no longer use having given up on data caps.

These numerical results support the ISPs' claim that data caps result in lower tier prices than would be offered without caps. The results partially support some public interest groups' claim that ISPs can increase their profit by adopting data caps; we predict that cable ISPs can increase profit by adopting either a heavy-users cap or a profit-maximizing cap, but that DSL ISPs may see decreased profit from competition.

The results also shed some light onto the conflicting claims of ISPs that data caps increase the incentive for ISPs to add capacity to the network and of some public interest groups that the use of caps may decrease an ISP's incentive to add capacity. While heavy-users caps result in higher capacity and higher tier rates, profit-maximizing caps may result in lower DSL tier rates and in cable tier rates that are higher than those without caps but lower than those under heavy-users caps.

B. The impacts of data caps upon different types of users

In the previous subsection, we compared pricing plans with and without data caps. In this subsection, we illustrate the impact of these service plans on different sets of users. We focus on two cases: the Nash equilibrium when neither ISP uses data caps, and the Nash equilibrium when both ISPs use profit-maximizing caps. (For the remainder of this section, data caps refer to those under the Nash equilibrium when both ISPs use profit-maximizing caps.)
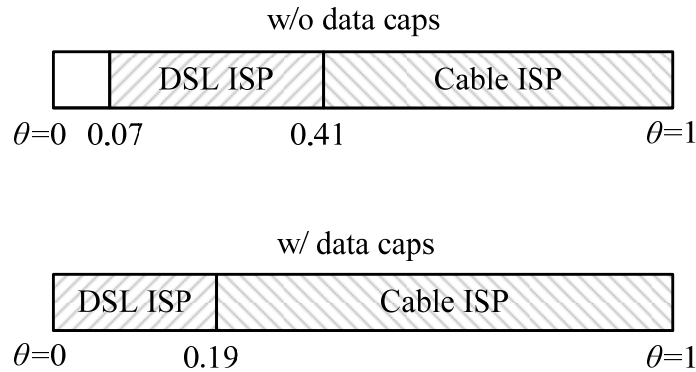
Figure 4.6 User subscription choice vs. user type, under service plans without and with data caps.

Figure 4.6 shows users' subscription choices in the absence and presence of data caps. For purposes of this discussion, we call users whose types satisfy $\theta<0.07$ light users, $0.07<\theta<0.41$ moderate users, and $\theta>0.41$ heavy users. As predicted by Theorem 4.4, under service plans without data caps light users do not subscribe to the Internet, moderate users subscribe to the DSL ISP, and heavy users subscribe to the cable ISP (which offers a higher tier rate and higher tier price than the DSL ISP). Under service plans with data caps, light users switch from no subscription to subscribing to the DSL ISP, due to the DSL ISP's reduced tier price. Some moderate users ($0.19<\theta<0.41$) upgrade from the DSL ISP to the

cable ISP, due to the cable ISP's reduced tier price and increased tier rate. The net result is that the DSL ISP loses market share (since its lost moderate users outnumber its new light users), and the cable ISP gains market share.

Figure 4.7(a) plots the ISP profit per user versus the user type, under service plans without and with data caps. Under service plans without data caps, since there is no usage-based pricing, the profit per subscriber is decreasing with θ. The DSL ISP earns profit from all of its light users, while the cable ISP loses money on its heaviest users (those with type $\theta>0.52$). Under service plans with data caps, the DSL ISP continues to earn profit from all of its light users. The cable ISP, having adopted usage-based pricing, now also earns profit from all of its users, with the profit per subscriber increasing with θ.

Figure 4.7(b) plots the density function of ISP profit (in $ per user type θ). Recall that the user type θ has a lognormal distribution, with the median at $\theta=0.3$. Under service plans without data caps, the ISPs earn almost all of their profit from moderate users, as these users outnumber the small number of heavy users on whom the cable ISP profits. Under service plans with data caps, although the profit per cable subscriber increases with θ, there are fewer users as θ increases, and thus the cable ISP earns similar total profit from each range of user type.

These results support ISPs' claims that without usage-based pricing moderate subscribers subsidize a small percentage of the heaviest users. However, whereas heavy-users caps remove this subsidy, profit-maximizing caps may result in a substantial proportion of the cable ISP's profit coming from overage charges. These results do not support ISPs' claims that caps affect will only a small percentage of heavy users. Instead, the results are partially

consistent with public interests groups' claim that data caps may eventually hurt a significant portion of users. ISPs can earn more profits by extracting higher surplus from moderate to heavy users under the data caps, since a usage-based pricing scheme makes users reveal how much they value the Internet service through overage charges.

(a)   (b)
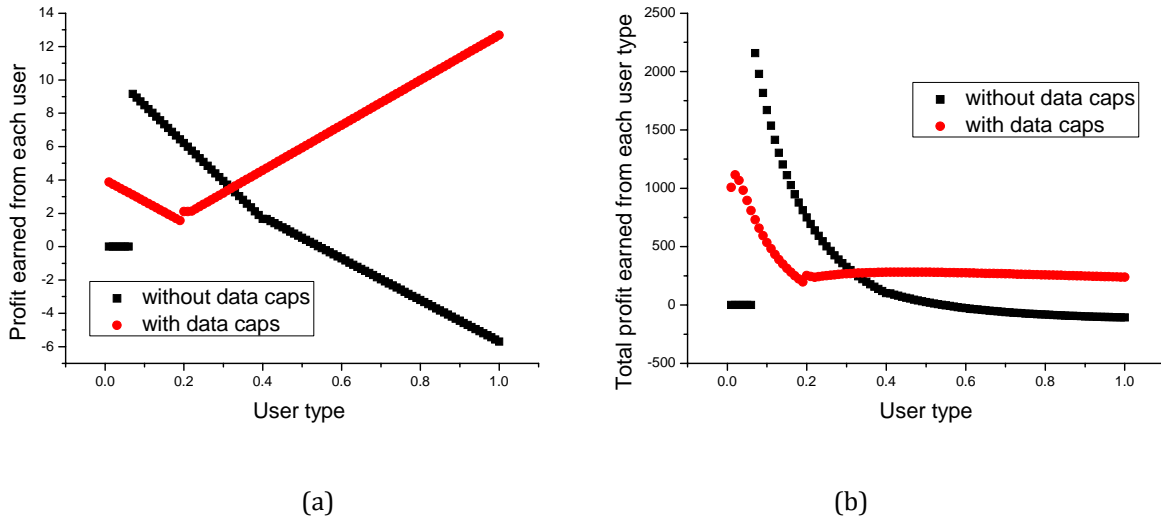
Figure 4.7 (a) ISP profit per user vs. user type. (b) ISP profit density function.

Figure 4.8(a) plots the per user surplus (on a log scale) versus the user type, under service plans without and with data caps. The adoption of data caps benefits light users, who switch from no subscription to a DSL ISP subscription, and thus now experience a small but positive surplus. The adoption also benefits moderate users ($0.07<\theta<0.19$) who subscribe to the DSL ISP under both sets of service plans; their surplus increases because the lower tier price more than compensates for the DSL ISP's lower service tier rate. The adoption benefits some moderate users who switch from the DSL ISP to the cable ISP ($0.19<\theta<0.26$); relatively light moderate users experience higher surplus because they receive a higher quality service at a lower price; other moderate users experience higher surplus because the increased service tier rate and decreased tier price can compensate for the extra

127

overage charges. However, the adoption of data caps hurts relatively heavy moderate users who switch from the DSL ISP to the cable ISP (0.26<θ<0.41), as overage charges outweigh the benefit from the increased tier rate and decreased tier price. Similarly, data caps hurt heavy users (0.41<θ) for the same reasons.

(a)   (b)
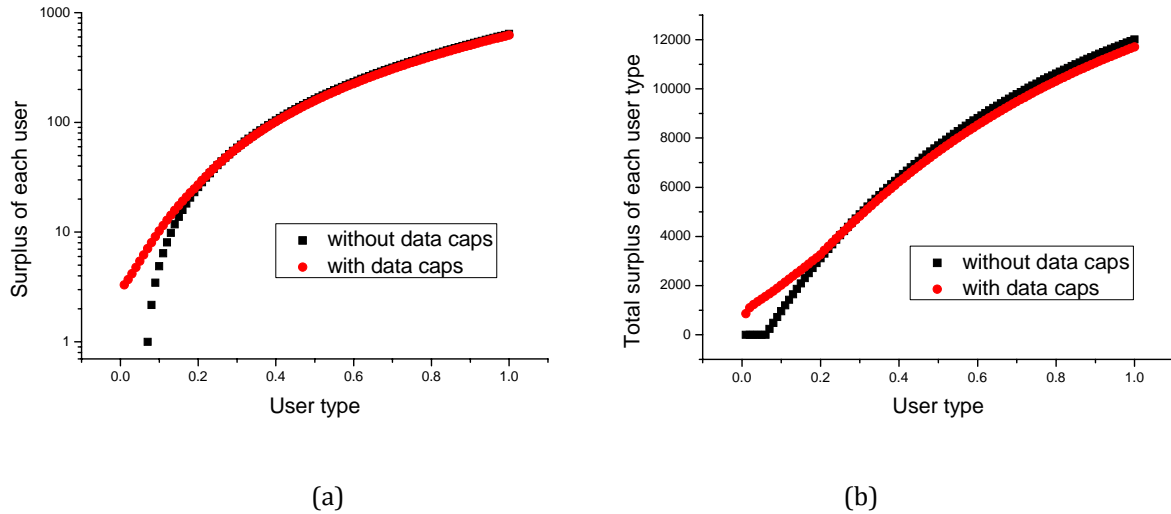
Figure 4.8 (a) Per user surplus vs. user type. (b) User surplus density function.

Figure 4.8(b) plots the density function of user surplus on a linear scale. Again we see the increase in user surplus for all light and some moderate users (θ<0.26) and the decrease in surplus for some moderate and all heavy users (0.26<θ). The two sets of users partially offset each other, and the resulting total user surplus may either decrease or increase depending on the shape of the distribution of user type. With the parameters used here, the total user surplus without data caps is slightly higher than that with data caps.

C. The effect of changes in cost factors and user distribution

In this subsection, we investigate the sensitivity of pricing plans using data caps to the marginal cost per unit capacity and to the mean and variance of user type.

We first change the marginal cost per unit capacity of the DSL ISP (i.e. p_1^μ), while fixing the marginal cost per unit capacity of the cable ISP at \$5/Mbps/month. As expected, both ISP's tier prices increase slightly and both ISP's tier rates decrease slightly as the marginal capacity cost of the DSL ISP increases. Figure 4.9(a) illustrates the market share of each ISP, and Figure 4.9(b) illustrates the resulting profit of each ISP, under service plans without and with data caps. As expected, as the marginal capacity cost of the DSL ISP increases, the DSL ISP loses market share to the cable ISP under either set of pricing plans. It is straightforward that the DSL ISP serves fewer users and earns less profit, while the cable ISP serves more users and earns more profit as the marginal capacity cost of the DSL ISP increases, since the cable ISP has an increasing advantage over the DSL ISP in terms of the capacity cost. We also observe that the cable ISP's profit increases dramatically with the marginal cost per unit capacity of the DSL ISP in the presence of data caps. When p_1^μ is low, the cable ISP's use of an overage charge is limited by the threat of its heavy subscribers to switch to the DSL ISP (which would switch the system from scenario b-1 to scenario b-2). However, as the DSL's capacity cost increases, this threat is blunted by the DSL's high cost to serve such heavy users.
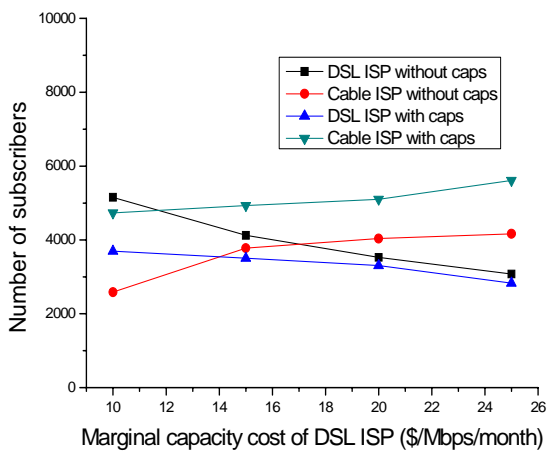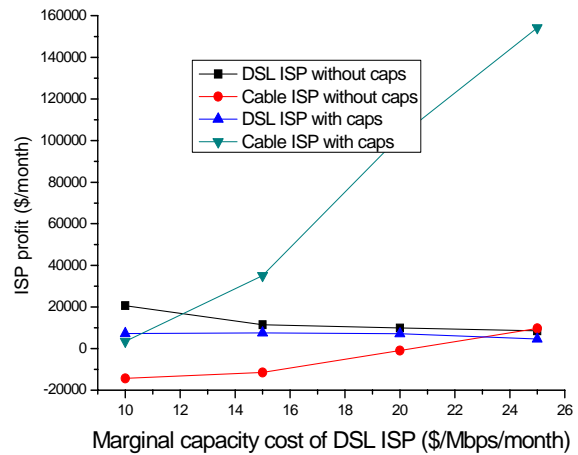
<div style="text-align:center">(a)   (b)</div>

Figure 4.9 Number of subscribers and ISP profit vs. marginal cost per unit capacity of the DSL ISP

Next we consider the effect of the mean and variance of user type. We vary the shape parameter[8] of the lognormal distribution; as the shape parameter increases, both the mean and variance of θ increase, but the median of θ is fixed. Thus the shape parameter can be interpreted as reflecting the proportion of heavy users. As the proportion of heavy users increases, both ISPs slightly reduce their tier rates, but the difference between their tier rates increases, reflecting increased market differentiation. In Figure 4.10, we plot the number of subscribers and ISP profits versus the shape parameter. In service plans without data caps, as the proportion of heavy users increases, the cable ISP increases its tier price to recover the increased capacity costs associated with a greater number of heavy users. The DSL ISP responds by also increasing its tier price, but not as much as the cable ISP does, to increase its profit. Both ISPs lose market share, since an increasing number of light users do not subscribe to broadband Internet access. Although the cable ISP is serving fewer subscribers, it earns slightly higher profit because it is now losing less money on

---

[8] The shape parameter is the standard deviation of $\ln\theta$, which is equal to 1.2 in Figure 4.3-Figure 4.9.

fewer heavy users at a higher tier price. In contrast, the DSL ISP earns slightly lower profit because of a decreasing number of subscribers. Under service plans with data caps, as the proportion of heavy users increases, the cable ISP slightly decreases its tier price and increases its tier rate, since it is now using the overage charge to extract some of the surplus of a greater number of heavy users. The DSL ISP, which is now serving light users who did not previously subscribe to the Internet, slightly increases its tier price. The DSL ISP also slightly increases its tier rate, in reaction to the cable ISP's increased tier rate. Thus, as the proportion of heavy users increases, the cable ISP loses market share but the DSL ISP's market share remains almost constant. Correspondingly, the cable ISP earns a lower profit due to the lower number of subscribers. In contrast, the DSL ISP earns more profit, due to its increased tier price.

(a)   (b)
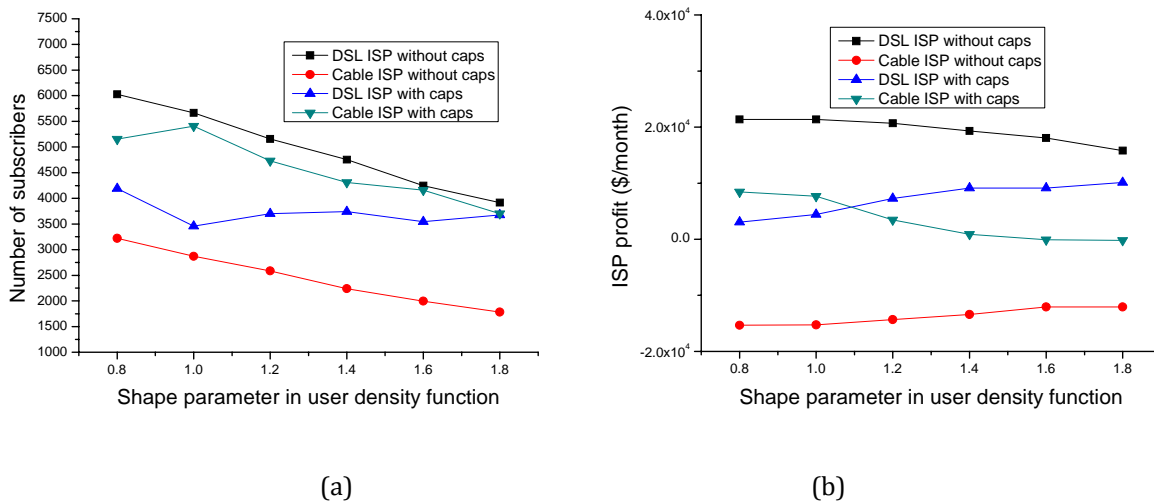
Figure 4.10 Number of subscribers (a) and ISP profits (b) vs. proportion of heavy users.

Although the simulation results above support our theoretic analysis, these results might not correspond to the current situation in the Internet market for several reasons. The model assumes competition between high quality and low quality services; in reality, each

ISP offers multiple tiers and the competition between them is more complex. The numerical results assume a lognormal distribution of user types; in reality, the distribution may have a different shape. Finally, the model does not consider the hesitation of a user to switch ISPs.

CHAPTER 5: Conclusions and Future Work

This dissertation investigates the design of ISP service tiers and the impact of data caps. For ISP tier design, we propose a basic model to obtain the condition under which ISPs may offer users multiple service tiers. Existing Internet usage statistics are used together with the basic model to explain why an ISP will offer multiple tiers for profit maximization. We also proposed a more complex model by extending the basic model to illustrate how ISPs may set tier prices, tier rates, and network capacity by considering both technical and economic issues. Web browsing and video streaming are modeled by utility functions depending on performance, devoted time, user's valuation of time and applications, instead of only an aggregated service quality (e.g. bandwidth). A general cost function depending on the network capacity is used, instead of assuming a fixed cost per user. On a time scale of days, users choose how much time to devote to applications based on the opportunity cost of their time. On a time scale of months, ISPs choose tier rates and prices, and users make subscriptions decisions. We then use our extended model to answer how ISPs may design tiered pricing plans, which is proprietary but important to networking research. Model analysis shows that the complex ISP profit maximization problem can be decomposed by the ISP, where the engineering department sets network capacity, the marketing department sets tier prices, and they jointly set tier rates. Numerical results are presented to illustrate the magnitude of the decrease in profit resulting from such a simplified design, and from duopoly competition between ISPs.

For the impact of data caps, we first consider a monopoly ISP that maximizes profit by setting tier prices, tier rates, network capacity, data caps, and overage charges. We show

that users with a small relative value on streaming and a large income subscribe to the premium tier but are not capped due to their low interest in streaming, users with a moderate relative value on streaming and moderate or high incomes subscribe to the premium tier and are capped, and that users with a 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 tier rate and decrease the tier price when data caps are used to maximize profit. The ISP will also 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 users benefit from data caps because of the increased tier rate and reduced tier price, while heavy users are hurt by the caps and overage charges. User welfare and social welfare may increase or decrease depending on the shape of the market density function and user utility function.

We then propose an ISP duopoly model to analyze the impact of data caps on ISP competition. Each ISP seeks to maximize its profit by setting tier prices, tier rates, network capacity, data caps, and overage charges. Based on the ISP service offerings, users choose which ISP to subscribe to (or not to subscribe) and how much time to devote to each application. In the absence of data caps, simplified network and user models show that the two ISPs each obtain a positive market share, when the incremental performance from DSL to cable is between two thresholds that depend on the incremental cost, the distribution of user type, and the correlation of user interest in the Internet with time devoted to the Internet. In the presence of data caps, we compare the tier rates, tier prices, network capacity, and data caps under profit-maximizing caps to those used if an ISP institutes caps only to ensure that heavy users pay an amount equal to the cost of their usage. The analysis

shows that ISPs will set higher overage charges and lower data caps under profit-maximizing caps than under heavy-users caps. ISPs have incentives to increase their tier rates under profit-maximizing caps under most scenarios. Numerical results based on existing Internet statistics are presented. In addition to verifying the results from the theoretical model analysis, we further show that (1) the DSL ISP is likely to maintain flat rate pricing while the cable ISP can increase profit by shifting to usage-based pricing, (2) these changes in pricing and tier rates can shift market share from the DSL ISP to the cable ISP, (3) the cable ISP's resulting profit may increase substantially, and (4) light to moderate users are likely to have higher surpluses due to reduced tier prices while moderate to heavy users are likely to have lower surpluses.

In the future, we would like to use our model to answer more complex Internet policy questions. For example, ISPs are beginning to provide integrated voice, video and data services, which compete with over-the-top content providers like Netflix and Hulu. ISPs are also capable of prioritizing one service over others by deploying Quality of Service (QoS) technology. Thus, QoS based pricing may greatly affect future Internet service competition.

Thus, it is important to investigate: 1) how to extend existing ISP tier design and data cap models to incorporate content providers and QoS; 2) based on the new model with content provider and QoS, answer Internet policy questions like: whether ISPs are willing to deploy QoS; whether ISPs should be allowed to use QoS solely for their own services; to whom should ISPs charge QoS, other content providers or end users; at what price should QoS be charged, whether the QoS price should be regulated or not.

REFERENCES

[1]   AT&T, "U-Verse pricing", http://www.att.com/u-verse/shop/, accessed Apr. 30, 2012.

[2]   Cox Communications, "Cox Internet pricing", http://ww2.cox.com/residential/internet.cox, accessed Apr. 30 2012.

[3]   Stacey Higginbotham, "Which ISPs are Capping Your Broadband, and Why?" *Gigaom.com*, Oct. 2012, available at http://gigaom.com/2013/11/15/data-cap-2013/

[4]   Sandvine Incorporated, "Global Internet Phenomena Report 2012," *Technical Report*, Fall 2012.

[5]   Advanced Communications Law &Policy Institute at New York Law School, "A Primer on Data Consumption: Trends& Emerging Business Models," *Technical Report*, Feb. 2012.

[6]   George S. Ford, "A Most Egregious Act? The Impact on Consumers of Usage-Based Pricing," *Phoenix Center*, May. 2012.

[7]   Richard Bennett, "Comcast Raises Invisible Data Cap," *The Information Technology & Innovation Foundation*, May. 2012.

[8]   Roger Yu, "Cable Companies Cap Data Use for Revenue," *USA Today*, available at http://www.usatoday.com/story/tech/2012/10/01/internet-data-cap/1595683/

[9]   Joel Kelsey, "Comcast Should Eliminate Punitive Data Caps Altogether," *Free Press*, May. 2012.

[10] Hibah Hussain, Danielle Kehl, Patrick Lucey, "The Destructive Power of Data Caps," *Free Press*, Dec. 2012.

[11] Hibah Hussain, Danielle Kehl, Benjamin Lennett, Patrick Lucey, "Capping the Nation's Broadband Future?" *New America Foundation's Open Technology Institute*, Dec. 2012.

[12] Andrew Odlyzko, Bill St. Arnaud, Erik Stallman and Michael Weinberg, "Know Your Limits Considering the Role of Data Caps and Usage Based Billing in Internet Access Service," *Public Knowledge*, May 2012.

[13] New America, "Letter to FCC on AT&A Data Caps," May 2011.

[14] Public Knowledge, "Letter to FCC on AT&T's new sponsored data scheme," Jan. 2014.

[15] US Congress. S. 3703 Data Cap Integrity Act, 112th Congress. *Senate Bill*, 2012.

[16] Economic Impacts of Open Internet Frameworks Working Group, "Policy Issues in Data Caps and Usage Based Pricing," *Open Internet Advisory Committee, Federal Communications Commission Annual Report*, Aug. 2013.

[17] Siva Viswanathan, G. Anandalingam, "Pricing strategies for Information goods," *Sadhana*, vol. 30, issue 2-3, pp. 257-274, April 2005.

[18] Robert Phillips. "Pricing and Revenue Optimization". *Stanford University Press*, 2005.

[19] A. Gold and C. Hogendorn, "Tipping in two-sided markets with asymmetric platforms," *Telecommunications Policy Research Conference* (*TPRC*), 2011.

[20] J. Musacchio and D. Kim, "Network platform competition in a two-sided market: Implications to the net neutrality issue," *Telecommunications Policy Research Conference* (*TPRC*), 2009.

[21] M. Chiang et al., "Pricing broadband: Survey and open problems," *Second International Conference on Ubiquitous and Future Networks* (*ICUFN*), 2010.

[22] P. Hande et al., "Pricing under constraints in access networks: Revenue maximization and congestion management," *IEEE International Conference on Computer Communications* (*INFOCOM*), Mar. 2010.

[23] S. Shakkottai et al., "The price of simplicity," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1269-1276, Sep. 2008.

[24] Linhai He and Jean Walrand, "Pricing differentiated Internet services," *IEEE International Conference on Computer Communications* (*INFOCOM*), Mar. 2005.

[25] Qian Lv and George N. Rouskas, "An Economic Model for Pricing Tiered Network Services," *IEEE International Conference on Communications* (*ICC*), 2009.

[26] Qian Lv and George N. Rouskas, "Internet Service Tiering as a Market Segmentation Strategy," *IEEE Global Communications Conference* (*GLOBECOM*), 2009.

[27] Richard Gibbens, Robin Mason, and Richard Steinberg, "Internet Service Classes Under Competition," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 12, pp. 2490-2498, Dec. 2000.

[28] Nikhil Shetty, Galina Schwartz, and Jean Walrand, "Internet QoS and Regulations," *IEEE/ACM Transaction on Networking*, vol. 18, issue 6, pp. 1725-1737, Dec. 2010.

[29] Soumya Sen, Carlee Joe-Wong, and Sangtae Ha, "The Economics of Shared Data Plans," *In Workshop on Information Technologies and Systems (WITS)*, 2012.

[30] Minne, Jacob Joseph Orion, "Data Caps: How ISPs are Stunting the Growth of Online Video Distributors and What Regulators Can Do About it," *Federal Communications Law Journal*, April, 2013. Available at http://dx.doi.org/10.2139/ssrn.2049174

[31] Daniel A Lyons, "The Impact of Data Caps and Other Forms of Usage-Based Pricing for Broadband Access," *Technical Report*, Mercatus Center at George Mason University.

[32] David Waterman, Ryland Sherman, Sung Wook Ji, "The Economics of Online Television: Revenue Models, Aggregation, and TV Everywhere," *Telecommunications Policy Research Conference (TPRC)*, Sep. 2012.

[33] Marshini Chetty, Richard Banks, A.J. Bernheim Brush, Jonathan Donner and Rebecca E. Grinter, "You're Capped! Understanding the Effects of Bandwidth Caps on Broadband Use in the Home," *ACM Conference on Human Factors in Computing Systems*, May 2012.

[34] Papak Nabipay, Andrew Odlyzko, Zhi-Li Zhang, "Flat Versus Metered Rates, Bundling, and Bandwidth Hogs," *Proceedings of NetEcon 11: 6th Workshop on the Economics of Networks, Systems, and Computation*, 2011.

[35] Wei Dai and Scott Jordan, "How do ISP Data Caps Affect Subscribers?" *Telecommunications Policy Research Conference (TPRC)*, Sep. 2013.

[36] Soumya Sen, Carlee Joe-Wong, Sangtae Ha, and Mung Chiang, "Incentivizing Time-Shifting of Data: A Survey of Time-Dependent Pricing for Internet Access," *IEEE Communications Magazine*, vol. 50 no. 11, Nov. 2012.

[37] Nicholas Economides and Benjamin E. Hermalin, "The Strategic Use of Download Limits by a Monopoly Platform," *NET Institute Working Paper*, Dec. 2013.

[38] Dilip Joseph, Nikhil Shetty, John Chuang, and Ion Stoica, "Modeling the Adoption of new Network Architectures," *In ACM CoNEXT Conference*, 2007.

[39] Soumya Sen, Youngmi Jin, Roch Guerin, and Kartik Hosanagar, "Modeling the Dynamics of Network Technology Adoption and the Role of Converters," *IEEE/ACM Transaction on Networking*, vol. 18, no.6, pp. 1793–1805, Dec. 2010.

[40] Carlee Joe-Wong, Soumya Seny, and Sangtae Hay, "Offering Supplementary Wireless Technologies: Adoption Behavior and Offloading Benefits," *IEEE International Conference on Computer Communications* (*INFOCOM*), 2013.

[41] Tom R. Eikebrokk and Oystein Sorebo, "Technology Acceptance in Situations with Alternative Technologies: An Empirical Evaluation of the Technology Acceptance Model in a Multiple-choice Situation," *In proceeding of the Seventh European Conference on Information Systems (ECIS)*, 1999.

[42] Anindya Chaudhuri, Kenneth S. Flamma, and John Horriganb, "An analysis of the determinants of internet access," *Telecommunications Policy*, vol. 29, no. 9, pp. 731–755, Nov. 2005.

[43] Hemant K. Bhargava and Vidyanand Choudhary, "Research Note-When Is Versioning Optimal for Information Goods?" *Management Science*, vol. 54, issue 5, pp. 1029–1035, May 2008.

[44] Soumya Sen, Carlee Joe-Wong, Sangtae Ha, Jasika Bawa, and Mung Chiang, "When the Price Is Right: Enabling Time-Dependent Pricing of Broadband Data," *In CHI13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013.

[45] Libin Jiang, Shyam Parekh, and Jean Walrand, "Time-Dependent Network Pricing and Bandwidth Trading," *In Network Operations and Management Symposium Workshops*, 2008.

[46] Yong Liu, Yang Guo, Chao Liang, "A survey on peer-to-peer video streaming systems," *Peer-to-Peer Networking and Applications*, vol. 1, no. 1, Mar. 2008.

[47] W. Wang, M. Palaniswami, and S. H. Low, "Application-oriented flow control: Fundamentals, algorithms and fairness," *IEEE/ACM Transaction on Networking*, vol. 14, no.6, pp. 1282–1291, Dec. 2006.

[48] O. Ormond, J. Murphy, and G.-M. Muntean, "Utility-based intelligent network selection in beyond 3G systems," *IEEE International Conference on Communications (ICC)*, 2006.

[49] P. Hande, Z. Shengyu, and C. Mung, "Distributed rate allocation for inelastic flows," *IEEE/ACM Transaction on Networking*, vol. 15, no.6, pp. 1240–1253, Dec. 2007.

[50] N. Cardwell, S. Savage, and T. Anderson, "Modeling TCP latency," *IEEE International Conference on Computer Communications* (*INFOCOM*), Mar. 2000.

[51] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP throughput: A simple model and its empirical validation," *ACM Special Interest Group on Data Communication* (*SIGCOMM*), 1998.

[52] Srinivasan Jagannathan , Kevin C. Almeroth "The dynamic of Price, Revenue, and System Utilization," *4th IFIP/IEEE International Conference on Management of Multimedia Networks and Services*, 2001.

[53] Burstmedia.com, "Online video content & advertising video preferences, habits and actions in Q4 2011," Oct. 2011, available at http://www.burstmedia.com/pdf/burst_media_online_insights_2011_11.pdf.

[54] G. Rosston, S. Savage, and D. Waldman, "Household demand for broadband Internet service," *Final report to the Broadband.gov Task Force Federal Communications Commission*, Feb. 3, 2010.

[55] S. Weber and V. Veeraraghavan, "Distributed algorithms for rate-adaptive media streams," *INFORMS Telecommunications Conference*, Dec. 2007.

[56] CCS Leeds Network Solutions, "UK leased lines pricing," http://www.uk-leased-line.co.uk/prices.html, accessed Apr.30 2012.

[57] MOTOROLA, "WHITE PAPER: Anticipating the Bandwidth Bottleneck," 2008.

[58] Hans U Gerber, Gerard Pafumi, "Utility Functions: From Risk Theory to Finance," *North American Actuarial Journal*, vol. 2, issue 3, Jul. 1998.

[59] Nicholas Economides, "Net Neutrality, Non-Discrimination and Digital Distribution of Content Through the Internet," *Journal of Law and Policy for the Information Society*, vol. 4, no. 2, 2008.

[60] R. Kumer and A. Tomkins. "A characterization of online browsing behavior," In *International World Wide Web Conference Committee (IW3C2)*, 2010.

[61] X. Wang and H. Schulzrinne. "Pricing network resources for adaptive applications," *IEEE/ACM Transaction on Networking*, vol. 14, no. 6, pp. 1282-1291, Dec. 2006.

[62] Website Optimization LLC, "Average web page size septuples since 2003," http://www.websiteoptimization.com/speed/tweak/averageweb-page/, accessed April 30th, 2012.

[63] Georgia Tech Research Corporation, "GVN's 10th WWW user survey," 1998, available at http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/graphs/use/q02.htm.

[64] United States Census Bureau, "Money income of households, 2009," 2009, available at http://www.census.gov/compendia/statab/2012/tables/12s0690.pdf.

[65] Wei Dai and Scott Jordan, "Modeling ISP Tier Design," *25th International Teletraffic Congress (ITC)*, Sep. 2013.

[66] David Besanko and Ronald R. Braeutigam, "Microeconomics," *John Wiley & Sons, Inc*, 2005.

[67] Frank J. Cesario, "Value of Time in Recreation Benefit Studies," *Land Econmics*, vol. 52 no. 1, pp. 32-41, Feb. 1976.

[68] Susan Crawford, "The Looming Cable Monopoly," *Yale Law & Policy Review*, vol. 29, pp. 34-40, Dec. 2010.

[69] Wei Dai, Scott Jordan. "Design and Impact of Data Caps", *IEEE Global Communications Conference (Globecom)*, Atlanta, GA, US, Dec. 2013.

[70] Bureau of Business & Economic Research at University of New Mexico, "Broadband Subscription and Internet Use in New Mexico," *The New Mexico Broadband Program, New Mexico Department of Information Technology*, June 2013.

[71] AT&T, "DSL High speed Internet pricing," http://www.att.com/shop/internet/internet-service.html, accessed Mar. 2014.

[72] Times Warner, "Cable High Speed Internet pricing" http://losangeles-offer.aiprx.timewarnercable.com/losangeles/twcInternet.php, accessed Mar. 2014.

[73] Shaked,A., and J.Sutton, "Relaxing Price Competition through Product Differentiation", *Review of Economic Studies*, vol. 49, issue 1, pp. 3-13, Jan. 1982.

[74] Jean Tirole, "The Theory of Industrial Organization," *the MIT Press*, 1988.

[75] Stella Dafermos, Anna Nagurney, "Oligopolistic and competitive behavior of spatially separated markets," *Regional Science and Urban Economics*, vo. 17, no. 2, pg. 245-254, 1987.

[76] Anna Nagurney, Ding Zhang, "Projected Dynamical Systems and Variational Inequalities with Applications," *International Series in Operations Research & Management Science (Book 2)*, Springer, Dec. 1995.

APPENDIX A: Proof of Theorem 2.2

For simplicity of notation, we henceforth represent w^b, $f_{w^b}\left(w^b\right)$, $F_{w^b}\left(w^b\right)$ and P_2–P_1 by w, $f(w)$, $F(w)$ and P_{21}, respectively.

We will first show that the optimal ISP profit in (2.4) can be achieved on the subset of $\{(P_1,\ P_2)\}$ such that P_2–$P_1 \geq g(P_1)$. Suppose that $(P_1{}^*,\ P_2{}^*)$ achieves the maximum in (2.4) and that $P_2{}^*$–$P_1{}^* < g(P_1{}^*)$. On P_2–$P_1 < g(P_1)$, the ISP profit in (2.4) is independent of P_1. Thus the same ISP profit is generated by $(P_1',\ P_2{}^*)$ where $P_2{}^*$–$P_1' = g(P_1')$.

On this subset, we can simplify the ISP profit maximization problem. On $\{(P_1,\ P_2): P_2$–$P_1 \geq g(P_1)\}$, N_1 and N_2 can be expressed as:

$$N_1 = N_{total}\left(F\left(g^{-1}\left(P_{21}\right)\right) - F\left(P_1\right)\right)$$
$$N_2 = N_{total}\left(1 - F\left(g^{-1}\left(P_{21}\right)\right)\right)$$

where N_{total} is the total number of users in the market. Recall that an ISP will set $X_1 = \overline{x}^b$ and $X_2 = \overline{x}^s$. We can express the capacity μ in terms of other variables by relating it to the network traffic λ. Recall that the ISP must keep the network load below $\rho^{min} = \min(\rho^b, \rho^s)$. Thus, if the network traffic rate is λ, network capacity must be at least $\mu = \lambda/\rho^{min}$. Hence,

$$\mu = N_{total}\,\frac{\left(F\left(g^{-1}\left(P_{21}\right)\right) - F\left(P_1\right)\right)\overline{x}^b t^b + \left(1 - F\left(g^{-1}\left(P_{21}\right)\right)\right)\left(\overline{x}^b t^b + \overline{x}^s t^s\right)}{\rho^{min}}$$

The ISP profit maximization problem in (2.4) reduces to:

$$\max_{P_1, P_{21}} N_{total} \left(P_1 + P_{21} - P_1 F\left(P_1\right) - P_{21} F\left(g^{-1}\left(P_{21}\right)\right) \right) - C(\mu)$$

$$\text{s.t. } P_1 - g^{-1}\left(P_{21}\right) \le 0 \tag{2.30}$$

It remains to characterize the solution to (2.30) and determine the number of tiers. The first order necessary optimality conditions are:

$$1 - F\left(P_1\right) - \left(P_1 - \frac{p^{\mu} \overline{x}^{-b} t^b}{\rho^{\min}} \right) f\left(P_1\right) - \kappa = 0$$

$$1 - F\left(g^{-1}\left(P_{21}\right)\right) - \left(P_{21} - \frac{p^{\mu} \overline{x}^{-s} t^s}{\rho^{\min}} \right) \frac{f\left(g^{-1}\left(P_{21}\right)\right)}{g'\left(g^{-1}\left(P_{21}\right)\right)} + \kappa \frac{1}{g'\left(g^{-1}\left(P_{21}\right)\right)} = 0$$

$$\kappa\left(P_1 - g^{-1}\left(P_{21}\right) \right) = 0$$

$$P_1 - g^{-1}\left(P_{21}\right) \le 0, \kappa \ge 0$$

where κ is the Lagrangian multiplier associated with the inequality constraint. Theorem 2.2 presents two sufficient conditions that depend on the sign of

$$\left(w - \frac{p^{\mu} \overline{x}^{-b} t^b}{\rho^{\min}} \right) g'(w) - \left(g(w) - \frac{p^{\mu} \overline{x}^{-s} t^s}{\rho^{\min}} \right).$$

We first consider the case when $\left(w - \frac{p^{\mu} \overline{x}^{-b} t^b}{\rho^{\min}} \right) g'(w) > g(w) - \frac{p^{\mu} \overline{x}^{-s} t^s}{\rho^{\min}}$ for all w. Setting $w = P_1$

gives:

$$\left(P_1 - \frac{p^{\mu} \overline{x}^{-b} t^b}{\rho^{\min}} \right) g'\left(P_1\right) > g\left(P_1\right) - \frac{p^{\mu} \overline{x}^{-s} t^s}{\rho^{\min}} \tag{2.31}$$

We will show by contradiction that the ISP will offer two tiers. Suppose that it is optimal for the ISP to offer one tier. Then by Lemma 2.1, $P_2 - P_1 \le g(P_1)$; since $g()$ is invertible and

monotonically increasing, this is equivalent to $P_1 - g^{-1}(P_{21}) \geq 0$. The first order necessary optimality conditions require $P_1 - g^{-1}(P_{21}) \leq 0$; thus it follows that $P_1 - g^{-1}(P_{21}) = 0$. Then, the first two conditions become:

$$1 - F(P_1) - \left(P_1 - \frac{p^\mu \bar{x}^{-b} t^b}{\rho^{\min}} \right) f(P_1) - \kappa = 0$$

$$1 - F(P_1) - \left(g(P_1) - \frac{p^\mu \bar{x}^{-s} t^s}{\rho^{\min}} \right) \frac{f(P_1)}{g'(P_1)} + \kappa \frac{1}{g'(P_1)} = 0$$

Substituting these conditions into (2.31) gives $\kappa < -\kappa/g'(P_1)$. Since $g'(P_1) > 0$, it follows that $\kappa < 0$, which contradicts the optimality condition $\kappa \geq 0$. By contradiction it follows that the ISP will offer two tiers.

Finally, we consider the case when $\left(w - \frac{p^\mu \bar{x}^{-b} t^b}{\rho^{\min}} \right) g'(w) \leq g(w) - \frac{p^\mu \bar{x}^{-s} t^s}{\rho^{\min}}$ for all w.

Setting $w = g^{-1}(P_{21})$ gives:

$$\frac{P_{21}}{g'\left(g^{-1}(P_{21})\right)} - g^{-1}(P_{21}) - \frac{p^\mu \bar{x}^{-s} t^s}{\rho^{\min} g'\left(g^{-1}(P_{21})\right)} \leq -\frac{p^\mu \bar{x}^{-b} t^b}{\rho^{\min}} \tag{2.32}$$

We will show by contradiction that the ISP will offer one tier. Suppose that it is optimal for the ISP to offer more than one tier. Then by Lemma 2.1, $P_2 - P_1 > g(P_1)$; since $g()$ is invertible and monotonically increasing, this is equivalent to $P_1 - g^{-1}(P_{21}) < 0$. The complementary slackness condition requires that $\kappa(P_1 - g^{-1}(P_{21})) = 0$, thus $\kappa = 0$. Then, the first two conditions become:

$$1 - F(P_1) - \left(P_1 - \frac{p^\mu \bar{x}^{-b} t^b}{\rho^{\min}}\right) f(P_1) = 0$$

$$1 - F\left(g^{-1}(P_{21})\right) - \left(P_{21} - \frac{p^\mu \bar{x}^{-s} t^s}{\rho^{\min}}\right) \frac{f\left(g^{-1}(P_{21})\right)}{g'\left(g^{-1}(P_{21})\right)} = 0$$

These two conditions have a similar form:

$$1 - F(x) - (x + y) f(x) = 0$$

where in the first condition $x = P_1$ and $y = -p^\mu \bar{x}^{-b} t^b / \rho^{\min}$ and in the second condition $x = g^-$

$^1(P_{21})$ and $y = \dfrac{P_{21}}{g'\left(g^{-1}(P_{21})\right)} - g^{-1}(P_{21}) - \dfrac{p^\mu \bar{x}^{-s} t^s}{\rho^{\min} g'\left(g^{-1}(P_{21})\right)}$.

In both, y can be written as: $y = (1 - F(x))/f(x) - x$.

The hypothesis of the theorem is that the density function of users' willingness-to-pay for web browsing $f(w)$ satisfies (2.7), i.e. $(F(w) - 1)f'(w) - 2f^2(w) < 0$ for all w. When this is true,

$$\frac{dy}{dx} = \left(\frac{F(x) f'(x) - f'(x) - 2f^2(x)}{f^2(x)}\right) < 0 \text{ for all } x$$

Thus by (2.32), $P_1 \geq g^{-1}(P_{21})$. This contradicts the constraint $P_1 - g^{-1}(P_{21}) < 0$. Thus the ISP will offer only one tier.

APPENDIX B: Estimation of Density Function in ISP Service Tier Design

In this section, we discuss how an ISP may estimate the density function $f(v^b, v^s, p^t)$ from a trial to try out new pricing plans in small portions of their service area.

Denote a user's relative value placed on video streaming by $v=v^s/p^t$. Denote the time a user devotes to video streaming, given relative value v and tier subscription j, by $t^{s,j}(v) = \max[0, V^{sr-1}(1/Q^s(X_j)v)]$. Denote the maximum user willingness to pay by

$$P_{\max} = \arg \max_{\text{user } i} W_i \left(v_i^b, v_i^s, p_i^t \mid T_i = 2 \right).$$

Suppose that an ISP tests all combinations of tier prices $(n_1\Delta P, n_2\Delta P)$, where $P_{\max}/\Delta P > n_2 > n_1 > 0$ are positive integers and ΔP is the increment between tested prices, and measures the number of users who subscribe to each tier and the amount of the time each user devotes to video streaming. Denote the number of users who subscribe to tier j and devote a time to video streaming in the interval $t^{s,j}(k\Delta v) < t^{s,j} \leq t^{s,j}((k+1)\Delta v)$ by:

$$N_{j,k}\left(n_1, n_2 \right) = \left| \left\{ i : T_i = j, t^{s,j}\left(k\Delta v \right) < t_i^s \leq t^{s,j}\left((k+1)\Delta v \right) \right\} \right|$$

Denote user's willingness-to-pay for browsing (resp. streaming) in tier j by:

$$W^{b,j}\left(v^b, p^t \right) \triangleq v^b V^b \left(t^{b,j} r^{b,j} \right) - t^{b,j} p^t$$
$$W^{s,j}\left(v^s, p^t \right) \triangleq v^s V^s \left(t^{s,j} \right) Q^s \left(x^{s,j} \right) - t^{s,j} p^t$$

and denote user's total willingness-to-pay for tier j by:

$$W^j\left(v^b, v^s, p^t \right) = W^{b,j}\left(v^b, p^t \right) + W^{s,j}\left(v^s, p^t \right) \tag{2.33}$$

As a result, users can be differentiated in the domain of (W^1, W^2, υ). When the ISP tests prices $(n_1\Delta P, n_2\Delta P)$, a user subscribes to tier 1 and devotes time $t^{s,1}$ to video streaming if and only if $W^1 > n_1\Delta P$ and $W^1 - n_1\Delta P > W^2 - n_2\Delta P$, and subscribes to tier 2 and devotes time ts,2 to video streaming if and only if $W^2 > n_2\Delta P$, $W^1 - n_1\Delta P < W^2 - n_2\Delta P$.

The joint density function of (W^1, W^2, υ) can be estimated by:

$$f_{W^1, W^2, \upsilon}\left(W^1, W^2, \upsilon\right) \approx \Delta N_k\left(n_1, n_2\right) / \Delta\upsilon\Delta P^2$$

where ΔN_k denotes the number of users measured in the region $\{n_1\Delta P < W^1 \le (n_1+1)\Delta P, n_2\Delta P < W^2 \le (n_2+1)\Delta P, k\Delta\upsilon < \upsilon \le (k+1)\Delta\upsilon\}$, which can be calculated by:

$$\Delta N_k\left(n_1, n_2\right) = N_{1,k}\left(n_1, n_2+1\right) + N_{2,k}\left(n_1, n_2+1\right) + N_{1,k}\left(n_1+1, n_2\right) + N_{2,k}\left(n_1+1, n_2\right)$$
$$- N_{1,k}\left(n_1, n_2\right) - N_{2,k}\left(n_1, n_2\right) - N_{1,k}\left(n_1+1, n_2+1\right) - N_{2,k}\left(n_1+1, n_2+1\right)$$

The remaining task is to estimate the density function $f(\upsilon^b, \upsilon^s, p^t)$ from the density function $f_{W^1, W^2, \upsilon}\left(W^1, W^2, \upsilon\right)$. This can be done via a change of variable in the subset of $(\upsilon^b, \upsilon^s, p^t)$ in which the transformation given by (2.33) and by $\upsilon = \upsilon^s/p^t$ is a one-to-one function. Assume that during the trial the ISP sets $X_1 = X^0$, $X_2 > X_1$, and maintains a constant network load ρ, thus keeping web browsing performance $r^{b,1}$, $r^{b,2}$ and video streaming performance $x^{s,1}$, $x^{s,2}$ constant. Then (2.33) gives:

$$W^2 - W^1 = W^{s,2}\left(\upsilon^s, p^t\right) - W^{s,1}\left(\upsilon^s, p^t\right) \Rightarrow p^t = \frac{W^2 - W^1}{\upsilon V^s\left(t^{s,2}(\upsilon)\right)Q^s\left(X_2\right) - \upsilon V^s\left(t^{s,1}(\upsilon)\right)Q^s\left(X_1\right) - t^{s,2}(\upsilon) + t^{s,1}(\upsilon)}$$

Denote $a_b = V^{b\prime}(0)$ and $a_s = V^{s\prime}(0)$. Thus, p^t is uniquely determined if $W^{s,2} > 0$ or $\upsilon^s > p^t/a_s Q^s(X_2)$. Then, υ^s is determined by $\upsilon^s = \upsilon p^t$. Finally, υ^b is determined by $W^2 = W^{b,2}(\upsilon^b, p^t) +$

$W^{s,2}(v^s, p^t)$ if $W^{b,2} > 0$ or $v^b > p^t/a_b r^{b,2}$. Thus in the region $\{v^b > p^t/a_b r^{b,2}, v^s > p^t/a_s Q^s(X_2)\}$, the transformation is a one-to-one function, and the density function $f_{v^b, v^s, p^t}\left(v^b, v^s, p^t\right)$ can be estimated by using the Jacobian of the transformation:

$$f_{v^b, v^s, p^t}\left(v^b, v^s, p^t\right) = \left|J\left(v^b, v^s, p^t\right)\right| \cdot$$
$$f_{W^1, W^2, v}\left(W^{b,1}\left(v^b, p^t\right) + W^{s,1}\left(v^s, p^t\right), W^{b,2}\left(v^b, p^t\right) + W^{s,2}\left(v^s, p^t\right), v^s/p^t\right)$$

where $J\left(v^b, v^s, p^t\right) = \begin{vmatrix} \dfrac{\partial W^{b,1}\left(v^b, p^t\right)}{\partial v^b} & \dfrac{\partial W^{s,1}\left(v^s, p^t\right)}{\partial v^s} & \dfrac{\partial W^1}{\partial p^t} \\[2ex] \dfrac{\partial W^{b,2}\left(v^b, p^t\right)}{\partial v^b} & \dfrac{\partial W^{s,2}\left(v^s, p^t\right)}{\partial v^s} & \dfrac{\partial W^2}{\partial p^t} \\[2ex] 0 & \dfrac{1}{p^t} & -\dfrac{v^s}{\left(p^t\right)^2} \end{vmatrix}$

The density function cannot be estimated in the region $\{v^s \leq p^t/a_s Q^s(X_2) \text{ or } v^b \leq p^t/a_b r^{b,2}\}$, since users in this region will not subscribe to the Internet even when prices are zero.

APPENDIX C: Proof of Theorem 2.3

Using Conjecture B:

$$\frac{\partial x^{s,1}}{\partial X_2} \approx 0, \frac{\partial r^{b,1}}{\partial X_2} \approx 0, \frac{\partial r^{b,2}}{\partial X_2} \approx 0, r^{b,1} \approx r^{b,2}, t_{mar}^{b,1} \approx t_{mar}^{b,2}, t_{mar}^{s,1} \approx 0$$

where $t_{mar}^{b,1}$ and $t_{mar}^{b,2}$ (resp. $t_{mar}^{s,1}$ and $t_{mar}^{s,2}$) are the average times users in $S_{1,2}$, and in tiers 1 and 2 respectively, spend on web browsing (resp. video streaming). Thus, when $x^{s,2}$ is constrained by tier rate X_2:

$$\frac{\partial \lambda_2}{\partial X_2} \approx N_2 \left(X_2 \frac{\partial \overline{t}^{s,2}}{\partial X_2} + \overline{t}^{s,2} \right) + \frac{\partial N_2}{\partial X_2} \left(\frac{t_{mar}^{b,2} r^{b,2} L}{M} + t_{mar}^{s,2} X_2 \right)$$

The first term gives the marginal traffic of current tier 2 users resulting from increased transmission rates and from additional time devoted to video streaming due to improved performance. The second term gives the marginal traffic due to new tier 2 subscribers.

By considering the performances in tier 1 as independent of tier rate X_2:

$$\frac{\partial \lambda_1}{\partial X_2} \approx \frac{\partial N_1}{\partial X_2} \left(\frac{t_{mar}^{b,1} r^{b,1} L}{M} + t_{mar}^{s,1} x^{s,1} \right) \approx \frac{\partial N_1}{\partial X_2} \frac{t_{mar}^{b,1} r^{b,1} L}{M}$$

Using Conjecture A, a marginal increase in X_2 primarily causes some marginal users to upgrade from tier 1 to tier 2, with the total number of subscribers remaining constant:

$$\frac{\partial (N_1 + N_2)}{\partial X_2} \approx 0 \Rightarrow \frac{\partial N_1}{\partial X_2} = -\frac{\partial N_2}{\partial X_2}$$

Thus, we have the following approximations for $\partial \lambda / \partial X_2$:

$$\frac{\partial \lambda}{\partial X_2} = \frac{\partial (\lambda_1 + \lambda_2)}{\partial X_2} \approx N_2 \left(X_2 \frac{\partial \bar{t}^{\,s,2}}{\partial X_2} + \bar{t}^{\,s,2} \right) + \frac{\partial N_2}{\partial X_2} t_{mar}^{s,2} X_2$$

It remains to find expressions for $\partial N_2 / \partial X_2$ and $\partial \bar{t}^{\,s,2} / \partial X_2$ to get an expression for $\partial Profit / \partial X_2$.

We presumed that tier prices have already been determined according to (2.27). Thus:

$$\frac{\partial Profit}{\partial P_2} = 0 \Rightarrow \frac{\partial N_2}{\partial P_2} = -N_2 \bigg/ \left(P_2 - P_1 - \frac{p^\mu X_2 t_{mar}^{s,2}}{\rho^{th}} \right)$$

An increase in tier 2 rate will increase such marginal users' willingness-to-pay for video streaming in tier 2. Thus:

$$\frac{\partial N_2}{\partial X_2} \approx \partial N_{total} \operatorname{Prob}\left(W^{s,2} > P_2 - P_1 \right) \big/ \partial X_2$$

$$= N_{total} \operatorname{Prob}\left(W^{s,2} + \Delta X_2 \, \partial W^{s,2} / \partial X_2 > P_2 - P_1 \right) \big/ \Delta X_2$$

$$= N_{total} \operatorname{Prob}\left(W^{s,2} + \Delta X_2 \, \partial W_{mar}^{s,2} / \partial X_2 > P_2 - P_1 \right) \big/ \Delta X_2$$

$$= \frac{N_{total} \operatorname{Prob}\left(W^{s,2} > P_2 - P_1 - \Delta X_2 \, \partial W_{mar}^{s,2} / \partial X_2 \right)}{\Delta X_2 \, \partial W_{mar}^{s,2} / \partial X_2} \frac{\partial W_{mar}^{s,2}}{\partial X_2}$$

$$= -\frac{\partial N_2}{\partial P_2} \frac{\partial W_{mar}^{s,2}}{\partial X_2} = -\frac{\partial N_2}{\partial P_2} v_{mar}^{s,2} V^s \left(t_{mar}^{s,2} \right) Q^{s\prime} (X_2)$$

Using Conjecture C, the average time users in tier 2 spend on video streaming can be estimated from (2.17) using:

$$v^{s,2} V^{s\prime} \left(\bar{t}^{\,s,2} \right) Q^s (X_2) = p^t \Rightarrow \bar{t}^{\,s,2} = V^{s\prime-1} \left(\frac{p^t}{v^{s,2} Q^s (X_2)} \right)$$

Thus, given p^t and $v^{s,2}$, $\partial \bar{t}^{\,s,2} / \partial X_2$ can be expressed as:

$$\frac{\partial \bar{t}^{s,2}}{\partial X_2} = \partial V^{s'-1}\left(\frac{p^t}{v^{s,2}Q^s(X_2)}\right)\Bigg/ \partial X_2 = -\frac{Q^{s'}(X_2)V^{s'}\left(\bar{t}^{s,2}\right)}{Q^s(X_2)V^{s''}\left(\bar{t}^{s,2}\right)}$$

Thus, we can derive the final expression for $\partial Profit/\partial X_2$ in (2.29) by replacing $\partial\lambda/\partial X_2$, $\partial N_2/\partial X_2$ and $\partial\bar{t}^{s,2}\big/\partial X_2$ in $\dfrac{\partial Profit}{\partial X_2} = P_1\dfrac{\partial N_1}{\partial X_2} + P_2\dfrac{\partial N_2}{\partial X_2} - \dfrac{p^\mu}{\rho^{th}}\dfrac{\partial\lambda}{\partial X_2}.$

APPENDIX D: Proof of Theorem 3.4

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

We first consider the overage charge p^o; (3.17) gives:

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

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 charge will switch between the basic tier and the premium tier. Denote these set of users by ΔG_o; thus:

$$\frac{\partial N_2}{\partial p^o} = \frac{\int_{\Delta G_o} f(\upsilon)d\upsilon}{\Delta p^o}$$

$$\frac{\partial O}{\partial p^o} = \frac{\int_{\Delta G_o} \left(X_2 t^{s,2} + x^b t^b - C_2 \right) f(\upsilon)d\upsilon}{\Delta p^o} + \int_{G_o}\frac{\partial t^{s,2}}{\partial p^o}X_2 f(\upsilon)d\upsilon$$

$$\frac{\partial \lambda^{s,1}}{\partial p^o} = -\frac{\int_{\Delta G_o} X_1 t^{s,1} f(\upsilon)d\upsilon}{\Delta p^o}$$

$$\frac{\partial \lambda^{s,2}}{\partial p^o} = \frac{\int_{\Delta G_o} X_2 t^{s,2} f(\upsilon)d\upsilon}{\Delta p^o} + \int_{G_o}\frac{\partial t^{s,2}}{\partial p^o}X_2 f(\upsilon)d\upsilon$$

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

$$\frac{\partial Profit}{\partial p^o} = \frac{\int_{\Delta G_o}\left(P_{21} - p^o\left(C_2 - x^b t^b \right) + p^o X_1 t^{s,1} \right) f(\upsilon)d\upsilon}{\Delta p^o} + \int_{G_o}\left(X_2 t^{s,2} + x^b t^b - C_2 \right) f(\upsilon)d\upsilon$$

The data cap is set to be $C_2 = P_{21}/p^o + x^b t_{\max}^{b,1} + X_1 t_{\max}^{s,1}$. Thus, we have:

$$P_{21} - p^o \left(C_2 - x^b t^b\right) + p^o X_1 t^{s,1} = p^o \left(x^b \left(t^b - t_{\max}^{b,1}\right) + X_1 \left(t^{s,1} - t_{\max}^{s,1}\right)\right) \leq 0$$

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 non-negative value. Marginal premium subscribers in G_o consume more data than the cap C_2 if they subscribe to the premium tier, and hence $X_2 t^{s,2} + x^b t^b - C_2 \geq 0$. Thus, the second term is also a non-negative value. So, $\partial Profit/\partial p^o \geq 0$.

We now turn to the cap C_2; (3.17) also gives:

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

Similarly, if the cap C_2 is changed by ΔC_2, some marginal premium subscribers who are capped will switch between the basic tier and the premium tier. Denote these set of users who by ΔG_c and ΔG_o, respectively. Thus,

$$\frac{\partial N_2}{\partial C_2} = \frac{\int_{\Delta G_c \cup \Delta G_o} f(\upsilon) d\upsilon}{\Delta C_2}$$

$$\frac{\partial O}{\partial C_2} = \frac{\int_{\Delta G_o} \left(X_2 t^{s,2} + x^b t^b - C_2\right) f(\upsilon) d\upsilon}{\Delta C_2} - \int_{G_o} f(\upsilon) d\upsilon$$

$$\frac{\partial \lambda^{s,1}}{\partial C_2} = -\frac{\int_{\Delta G_c \cup \Delta G_o} X_1 t^{s,1} f(\upsilon) d\upsilon}{\Delta C_2}$$

$$\frac{\partial \lambda^{s,2}}{\partial C_2} = \frac{\int_{\Delta G_c \cup \Delta G_o} X_2 t^{s,2} f(\upsilon) d\upsilon}{\Delta C_2}$$

Considering $p^o = p^\mu/\rho^{th}$ and $C_2 = P_{21}/p^o + x^b t_{\max}^{b,1} + X_1 t_{\max}^{s,1}$, we have:

$$\frac{\partial Profit}{\partial C_2} = \frac{\int_{\Delta G_c \cup \Delta G_o} p^o \left(x^b t^b - x^b t^{b,1}_{max} + X_1 t^{s,1} - X_1 t^{s,1}_{max} \right) f(\upsilon) d\upsilon}{\Delta C_2} - \int_{G_o} p^o f(\upsilon) d\upsilon$$

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

APPENDIX E: Proof of Theorem 3.5

We first consider the price of basic tier, i.e. P_1. According to Theorem 3.2, the optimal tiered pricing plan without data caps satisfies:

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

From (3.17) in Theorem 3.3, we have:

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

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 (3.11).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 the 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 3.3. Denote these set of users by ΔG_u and ΔG_d. Thus, Theorem 3.2 gives:

$$\frac{\partial Profit_0}{\partial P_{21}} = N_2 + P_{21}\frac{\partial N_2}{\partial P_{21}} - \frac{p^\mu}{\rho^{th}}\frac{\partial \lambda}{\partial P_{21}}$$

$$= \int_{\Delta G_u \cup \Delta G_c \cup \Delta G_o \cup \Delta G_d} f(\upsilon)d\upsilon + P_{21}\frac{\int_{\Delta G_u \cup \Delta G_d} f(\upsilon)d\upsilon}{\Delta P_{21}} - \frac{p^\mu}{\rho^{th}}\frac{\int_{\Delta G_u \cup \Delta G_d}(X_2 t^{s,2} - X_1 t^{s,1})f(\upsilon)d\upsilon}{\Delta P_{21}} = 0$$

In the presence of data caps, if the price P_{21} is changed by ΔP_{21}, some marginal premium subscribers in the 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 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 3.3. Denote these set of users by ΔG_u, ΔG_c, and ΔG_o. Thus, Theorem 3.3 gives:

$$\frac{\partial \text{Profit}}{\partial P_{21}} = N_2 + P_{21}\frac{\partial N_2}{\partial P_{21}} + p^o \frac{\partial O}{\partial P_{21}} - \frac{p^\mu}{\rho^{th}}\frac{\partial \lambda}{\partial P_{21}}$$

$$= \int_{\Delta G_u \cup \Delta G_c \cup \Delta G_o} f(\upsilon)\,d\upsilon + P_{21}\frac{\int_{\Delta G_u \cup \Delta G_c \cup \Delta G_o} f(\upsilon)\,d\upsilon}{\Delta P_{21}} + p^o \frac{\int_{\Delta G_o}\left(X_2 t^{s,2} + x^b t^b - C_2\right) f(\upsilon)\,d\upsilon}{\Delta P_{21}}$$

$$- \frac{p^\mu}{\rho^{th}}\frac{\int_{\Delta G_u \cup \Delta G_c \cup \Delta G_o}\left(X_2 t^{s,2} - X_1 t^{s,1}\right) f(\upsilon)\,d\upsilon}{\Delta P_{21}}$$

Thus, we have:

$$\frac{\partial\left(\text{Profit} - \text{Profit}_0\right)}{\partial P_{21}} = -\int_{G_d} f(\upsilon)\,d\upsilon - \frac{\int_{\Delta G_d}\left(P_{21} - p^\mu\left(X_2 t^{s,2} - X_1 t^{s,1}\right)/\rho^{th}\right) f(\upsilon)\,d\upsilon}{\Delta P_{21}}$$

$$+ \frac{\int_{\Delta G_c \cup \Delta G_o}\left(P_{21} - p^\mu\left(C_2 - x^b t^b - X_1 t^{s,1}\right)/\rho^{th}\right) f(\upsilon)\,d\upsilon}{\Delta P_{21}} + \left(p^o - \frac{p^\mu}{\rho^{th}}\right)\frac{\int_{\Delta G_o}\left(X_2 t^{s,2} + x^b t^b - C_2\right) f(\upsilon)\,d\upsilon}{\Delta P_{21}}$$

(3.18)

Obviously, the first term in (3.18) is non-positive. 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 $X_2 t^{s,2} + x^b t^b - C_2 \geq 0$. Thus, the last term in (3.18) is non-positive, since $p^o \geq p^\mu/\rho^{th}$. The sum of the second and third term in (3.18) can be expressed as:

$$-\int\limits_{\frac{v^b}{p^t}}\int\limits_{\frac{v^s}{p^t}} \left(\left(P_{21} - \frac{p^u\left(C_2 - x^b t^b - X_1 t^{s,1}\right)}{\rho^{th}} \right) \frac{f\left(v^b/p^t, v^s/p^t, p^{t,c}\right)}{\partial S^c/\partial p^t\big|_{p^t=p^{t,c}}} - \left(P_{21} - \frac{p^u\left(X_2 t^{s,2} - X_1 t^{s,1}\right)}{\rho^{th}} \right) \frac{f\left(v^b/p^t, v^s/p^t, p^{t,u}\right)}{\partial S^u/\partial p^t\big|_{p^t=p^{t,u}}} \right) d\frac{v^s}{p^t} d\frac{v^b}{p^t}$$

where $p^{t,u}$ and $p^{t,c}$ are the values placed on time by the uncapped and capped marginal premium subscribers with relative values v^b/p^t and v^s/p^t, respectively, which 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 (3.8).

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:

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

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

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

Marginal premium subscribers in G_o consume more data than the 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 (3.18) can be expressed as:

$$-\int\limits_{\frac{v^b}{p^t}}\int\limits_{\frac{v^s}{p^t}} \left(\begin{array}{l} \left(P_{21} - \dfrac{p^\mu \left(C_2 - x^b t^b - X_1 t^{s,1} \right)}{\rho^{th}} \right) \dfrac{f\left(v^b/p^t, v^s/p^t, p^{t,c} \right)}{\partial S^c / \partial p^t \big|_{p^t = p^{t,c}}} \\[3mm] - \left(P_{21} - \dfrac{p^\mu \left(C_2 - x^b t^b - X_1 t^{s,1} \right)}{\rho^{th}} \right) \dfrac{f\left(v^b/p^t, v^s/p^t, p^{t,u} \right)}{\partial S^u / \partial p^t \big|_{p^t = p^{t,u}}} \\[3mm] + \dfrac{p^\mu \left(X_2 t^{s,2} + x^b t^b - C_2 \right)}{\rho^{th}} \dfrac{f\left(v^b/p^t, v^s/p^t, p^{t,u} \right)}{\partial S^u / \partial p^t \big|_{p^t = p^{t,u}}} \end{array} \right) d\dfrac{v^s}{p^t} d\dfrac{v^b}{p^t} \le 0$$

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

We then turn to the basic tier rate X_1. In the absence of data caps, if the price X_1 is changed by ΔX_1, some marginal premium subscribers in the 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 the 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 3.2 gives:

$$\begin{aligned} \frac{\partial Profit_0}{\partial X_1} &= \left(P_1 - k \right) \frac{\partial \left(N_1 + N_2 \right)}{\partial X_1} + P_{21} \frac{\partial N_2}{\partial X_1} - \frac{p^\mu}{\rho^{th}} \frac{\partial \lambda}{\partial X_1} \\[3mm] &= \left(P_1 - k \right) \frac{\partial \left(N_1 + N_2 \right)}{\partial X_1} + P_{21} \frac{\int_{\Delta G_u \cup \Delta G_d} f(v) dv}{\Delta X_1} - \frac{p^\mu}{\rho^{th}} \frac{\partial \lambda \left(G_b \right)}{\partial X_1} \\[3mm] &\quad + \frac{p^\mu}{\rho^{th}} \frac{\int_{\Delta G_n \cup \Delta G_u \cup \Delta G_d} X_1 t^{s,1} f(v) dv}{\Delta X_1} - \frac{p^\mu}{\rho^{th}} \frac{\int_{\Delta G_u \cup \Delta G_d} X_2 t^{s,2} f(v) dv}{\Delta X_1} \\[3mm] &= 0 \end{aligned}$$

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 the 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 set of users by ΔG_u, ΔG_c, and ΔG_o. Thus, Theorem 3.3 gives:

$$\frac{\partial Profit}{\partial X_1} = \left(P_1 - k\right)\frac{\partial\left(N_1 + N_2\right)}{\partial X_1} + P_{21}\frac{\partial N_2}{\partial X_1} + p^o\frac{\partial O}{\partial X_1} - \frac{p^\mu}{\rho^{th}}\frac{\partial\lambda}{\partial X_1}$$

$$= \left(P_1 - k\right)\frac{\partial\left(N_1 + N_2\right)}{\partial X_1} + P_{21}\frac{\int_{\Delta G_u \cup \Delta G_c \cup \Delta G_o} f\left(\upsilon\right)d\upsilon}{\Delta X_1} - \frac{p^\mu}{\rho^{th}}\frac{\partial\left(\lambda\left(G_b\right) + \lambda\left(G_d\right)\right)}{\partial X_1}$$

$$+ \frac{p^\mu}{\rho^{th}}\frac{\int_{\Delta G_n \cup \Delta G_u \cup \Delta G_c \cup \Delta G_o} X_1 t^{s,1} f\left(\upsilon\right)d\upsilon}{\Delta X_1} - \frac{p^\mu}{\rho^{th}}\frac{\int_{\Delta G_u \cup \Delta G_c \cup \Delta G_o} X_2 t^{s,2} f\left(\upsilon\right)d\upsilon}{\Delta X_1}$$

$$+ p^o\frac{\int_{\Delta G_o}\left(X_2 t^{s,2} + x^b t^b - C_2\right)f\left(\upsilon\right)d\upsilon}{\Delta X_1}$$

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. $\partial\,\lambda(G_b)/\partial P_1$ is also a function of X_1 and P_1, according to Theorem 3.1. Thus, we have:

$$\frac{\partial\left(Profit - Profit_0\right)}{\partial X_1}$$

$$= -\frac{p^\mu}{\rho^{th}}\frac{\partial\lambda\left(G_d\right)}{\partial X_1} + \frac{\int_{\Delta G_c \cup \Delta G_o}\left(P_{21} - p^\mu\left(C_2 - x^b t^b - X_1 t^{s,1}\right)\big/\rho^{th}\right)f\left(\upsilon\right)d\upsilon}{\Delta X_1} \qquad (3.19)$$

$$- \frac{\int_{\Delta G_d}\left(P_{21} - p^\mu\left(X_2 t^{s,2} - X_1 t^{s,1}\right)\big/\rho^{th}\right)f\left(\upsilon\right)d\upsilon}{\Delta X_1} + \left(p^o - \frac{p^\mu}{\rho^{th}}\right)\frac{\int_{\Delta G_o}\left(X_2 t^{s,2} + x^b t^b - C_2\right)f\left(\upsilon\right)d\upsilon}{\Delta X_1}$$

Obviously, the first term in (3.19) is non-positive. As X_1 decreases, more users subscribe to the premium tier. Marginal premium subscribers in ΔG_o consume more data than the data 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 (3.19) is non-positive, since $p^o \geq p^\mu/\rho^{th}$. Similar to (3.18), we can prove the sum of the first and second term in (3.19) is also non-positive. So, $\partial(Profit - Profit_0)/\partial X_1 \leq 0$. 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 the rate X_2 is changed by ΔX_2, some marginal premium subscribers in the 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 3.2 gives:

$$
\begin{aligned}
\frac{\partial Profit_0}{\partial X_2} &= P_{21} \frac{\partial N_2}{\partial X_2} - \frac{p^\mu}{\rho^{th}} \frac{\partial \lambda}{\partial X_2} \\
&= P_{21} \frac{\int_{\Delta G_u \cup \Delta G_d} f(\upsilon) d\upsilon}{\Delta X_2} - \frac{p^\mu}{\rho^{th}} \frac{\int_{\Delta G_u \cup \Delta G_d} \left(X_2 t^{s,2} - X_1 t^{s,1}\right) f(\upsilon) d\upsilon}{\Delta X_2} \\
&\quad - \frac{p^\mu}{\rho^{th}} \int_{G_u \cup G_c \cup G_o \cup G_d} \left(t^{s,2} + X_2 \, \partial t^{s,2}/\partial X_2\right) f(\upsilon) d\upsilon
\end{aligned}
$$

In the presence of data caps, if the rate X_2 is changed by ΔX_2, some marginal premium subscribers in the 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 set of users by ΔG_u, ΔG_c, and ΔG_o. Thus, Theorem 3.3 gives:

$$
\begin{aligned}
\frac{\partial Profit}{\partial X_2} &= P_{21} \frac{\partial N_2}{\partial X_2} + p^o \frac{\partial O}{\partial X_2} - \frac{p^\mu}{\rho^{th}} \frac{\partial \lambda}{\partial X_2} \\
&= P_{21} \frac{\int_{\Delta G_u \cup \Delta G_c \cup \Delta G_o} f(\upsilon) d\upsilon}{\Delta X_2} + p^o \frac{\int_{\Delta G_o} \left(X_2 t^{s,2} + x^b t^b - C_2\right) f(\upsilon) d\upsilon}{\Delta X_2} \\
&\quad - \frac{p^\mu}{\rho^{th}} \frac{\int_{\Delta G_u \cup \Delta G_c \cup \Delta G_o} \left(X_2 t^{s,2} - X_1 t^{s,1}\right) f(\upsilon) d\upsilon}{\Delta X_2} - \frac{p^\mu}{\rho^{th}} \int_{G_u} \left(t^{s,2} + X_2 \frac{\partial t^{s,2}}{\partial X_2}\right) f(\upsilon) d\upsilon \\
&\quad + \left(p^o - \frac{p^\mu}{\rho^{th}}\right) \int_{G_o} \left(t^{s,2} + X_2 \, \partial t^{s,2}/\partial X_2\right) f(\upsilon) d\upsilon
\end{aligned}
$$

Thus, we have:

$$\frac{\partial\left(Profit - Profit_0\right)}{\partial X_2} = \frac{p^\mu}{\rho^{th}} \int_{G_c \cup G_o \cup G_d} \left(t^{s,2} + X_2\, \partial t^{s,2}/\partial X_2\right) f(\upsilon)\, d\upsilon$$

$$+ \frac{\int_{\Delta G_c \cup \Delta G_o} \left(P_{21} - p^\mu \left(C_2 - x^b t^b - X_1 t^{s,1}\right)\big/\rho^{th}\right) f(\upsilon)\, d\upsilon}{\Delta X_2}$$

$$- \frac{\int_{\Delta G_d} \left(P_{21} - p^\mu \left(X_2 t^{s,2} - X_1 t^{s,1}\right)\big/\rho^{th}\right) f(\upsilon)\, d\upsilon}{\Delta X_2} \qquad (3.20)$$

$$+ \left(p^o - \frac{p^\mu}{\rho^{th}}\right)\left(\int_{G_o} \left(t^{s,2} + X_2\, \partial t^{s,2}/\partial X_2\right) f(\upsilon)\, d\upsilon + \frac{\int_{\Delta G_o} \left(X_2 t^{s,2} + x^b t^b - C_2\right) f(\upsilon)\, d\upsilon}{\Delta X_2}\right)$$

In the absence of data caps, we can easily prove that $\partial t^{s,2}/\partial X_2 \geq 0$ from (3.3), since users will devote more time to video streaming when the tier rate X_2 improves. Thus, the first term in (3.20) is non-negative. As X_2 increases, 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 $X_2 t^{s,2} + x^b t^b - C_2 \geq 0$. Thus, the last term in (3.20) is non-negative, since $p^o \geq p^\mu/\rho^{th}$. Similar to (3.18), we can prove the sum of the second and third term in (3.20) is also non-negative. So, $\partial(Profit - Profit_0)/\partial X_2 \geq 0$. Considering $\partial Profit_0/\partial X_2 = 0$, it follows that $\partial Profit/\partial X_2 \geq 0$.

APPENDIX F: Proof of Theorem 3.7

If $\Delta S(p^t, v^b, v^{s4}) \geq 0$, we can easily show that $\Delta S\left(p^t, v^b, v^s\right) \geq 0, \ \forall v^s: v^s \geq h_2\left(p^t, v^b\right)$, since

$v^{s4} = \arg \min\limits_{v^s \in \{v^s \geq h_1(p^t, v^b)\}} \Delta S\left(p^t, v^b, v^s\right)$ and $\Delta S\left(p^t, v^b, v^s\right) \geq 0, \ \forall v^s: h_2\left(p^t, v^b\right) \leq v^s \leq h_1\left(p^t, v^b\right)$.

We then consider the case when $\Delta S(p^t, v^b, v^{s4}) < 0$. We can easily show that

$\Delta S\left(p^t, v^b, v^s\right) \geq 0, \ \forall v^s: h_2\left(p^t, v^b\right) < v^s < v^{s2}\left(C_2, X_2 + \Delta X, p^t\right)$, since $\Delta X > 0$, $\Delta P < 0$ and hence

uncapped premium users benefit from the increased premium tier rate and reduced

premium tier price. Denote the surplus of a user with v^s value placed on video streaming,

under a pricing plan with and without profit-maximizing caps, by $S_c(v^s)$ and $S_u(v^s)$

respectively. $S_c(v^s)$ corresponds to the blue curve in Figure 3.6, and we can show that $S_c(v^s)$

is a linear increasing function when premium users are capped but not paying an overage

charge, i.e. $v^{s2}\left(C_2, X_2 + \Delta X, p^t\right) < v^s < v^{s3}\left(C_2, X_2 + \Delta X, p^o, p^t\right)$.

If $S_c(v^{s3}) < S_u(v^{s3})$, there exists a unique solution denoted by $v^{th,1}$ to the equation $S_c(v^s) = S_u(v^s)$, $v^{s2} < v^s < v^{s3}$, and $\Delta S\left(p^t, v^b, v^s\right) \geq 0, \ \forall v^s: h_2\left(p^t, v^b\right) \leq v^s \leq v^{th,1}$. Considering $S_c(v^{s2}) \geq S_u(v^{s2})$,

$S_c(v^{s3}) < S_u(v^{s3})$, $S_u(v^s)$ is an increasing convex function for $v^s > h_1(p^t, v^b)$ and $S_c(v^s)$ is a linear

increasing function for $v^{s2} < v^s < v^{s3}$, we have $S_u'(v^{s3}) > S_c'(v^{s3})$, which gives the following

results by replacing $V^s(t) = at^2 + bt$:

$$\frac{\left(p^t + p^o\left(X_2 + \Delta X\right)^2\right)}{Q\left(X_2 + \Delta X\right)} > \frac{\left(p^t\right)^2}{Q\left(X_2\right)}$$

165

We thus can conclude that $S_c(v^s) - S_u(v^s)$ or $\Delta S(p^t,v^b,v^s)$ is a decreasing function of v^s when

$$v^{s3} < v^s < \sqrt{\left(\frac{\left(p^t+p^o\left(X_2+\Delta X\right)^2\right)}{Q\left(X_2+\Delta X\right)} - \frac{\left(p^t\right)^2}{Q\left(X_2\right)}\right) \Big/ b^2\left(Q\left(X_2+\Delta X\right)-Q\left(X_2\right)\right)}. $$
$S_c(v^s) - S_u(v^s)$ or $\Delta S(p^t,v^b,v^s)$ is an

increasing function of v^s when $v^s > \sqrt{\left(\frac{\left(p^t+p^o\left(X_2+\Delta X\right)^2\right)}{Q\left(X_2+\Delta X\right)} - \frac{\left(p^t\right)^2}{Q\left(X_2\right)}\right) \Big/ b^2\left(Q\left(X_2+\Delta X\right)-Q\left(X_2\right)\right)}.$

So, there exists a unique solution denoted by $v^{th,2}$ to the equation $S_c(v^s) = S_u(v^s)$, $v^s > v^{s3}$, and

$\Delta S\left(p^t,v^b,v^s\right) \leq 0$, $\forall v^s : v^{th,1} \leq v^s \leq v^{th,2}$, $\Delta S\left(p^t,v^b,v^s\right) \geq 0$, $\forall v^s : v^s \geq v^{th,2}$.

If $S_c(v^{s3}) \geq S_u(v^{s3})$, we can also prove that $\left(p^t + p^o\left(X_2+\Delta X\right)^2\right)\Big/Q\left(X_2+\Delta X\right) > \left(p^t\right)^2\Big/Q\left(X_2\right)$.

Otherwise, $\Delta S(p^t,v^b,v^s) = S_c(v^s) - S_u(v^s)$ is an increasing function of v^s for $v^s \geq v^{s3}$, which gives

$\Delta S(p^t,v^b,v^s) > 0$ for $v^s \geq h_1(p^t, v^b)$ and contradicts $\Delta S(p^t,v^b,v^{s4})<0$. We can thus show that the

minimum of $\Delta S(p^t,v^b,v^s)$ is achieved when:

$$v^{s4} = \sqrt{\left(\frac{\left(p^t+p^o\left(X_2+\Delta X\right)^2\right)}{Q\left(X_2+\Delta X\right)} - \frac{\left(p^t\right)^2}{Q\left(X_2\right)}\right) \Big/ b^2\left(Q\left(X_2+\Delta X\right)-Q\left(X_2\right)\right)}$$

Considering $\Delta S(p^t,v^b,v^{s4})<0$, $\Delta S(p^t,v^b,v^{s3})\geq 0$ and $\Delta S(p^t,v^b,\infty)\geq 0$, we can prove that there exists a

unique solution denoted by $v^{th,1}$ to the equation $S_c(v^s) = S_u(v^s)$, $v^{s3} < v^s < v^{s4}$, and

$\Delta S\left(p^t,v^b,v^s\right) \geq 0$, $\forall v^s : h_2\left(p^t,v^b\right) \leq v^s \leq v^{th,1}$. Similarly, we can also prove that there exists a

unique solution denoted by $v^{th,2}$ to the equation $S_c(v^s) = S_u(v^s)$, $v^s > v^{s4}$, and

$\Delta S\left(p^t,v^b,v^s\right) \leq 0$, $\forall v^s : v^{th,1} \leq v^s \leq v^{th,2}$, $\Delta S\left(p^t,v^b,v^s\right) \geq 0$, $\forall v^s : v^s \geq v^{th,2}$.

APPENDIX G: Conditions for the Existence of Nash Equilibria from Definition 4.1

Under certain concavity assumptions, the Nash equilibria can be characterized using this standard theorem.

Theorem 4.7: Assume for each ISP j the profit function $\pi_j(E_1, E_2)$ is continuously differentiable and concave with respect to the variables in $E_j \in \mathbb{R}_+^4$. Then $E^* = (E_1^*, E_2^*)$ is a Nash equilibrium in Definition 4.1 if and only if it satisfies the variational inequality:

$$-\sum_{j=1}^{2}\sum_{k=1}^{4}\frac{\partial \pi_j\left(E_1^*, E_2^*\right)}{\partial E_{j,k}}\left(E_{j,k} - E_{j,k}^*\right) \geq 0, \quad \forall E_j \in \mathbb{R}_+^4 \tag{4.7}$$

where $\mathbb{R}_+^4$ is a 4-dimensional vector space over positive real numbers, E_1 and E_2 are any strategy vectors, and $E_{j,k}$ denotes the kth element in E_j.

Proof: (4.7) follows directly from [75] Theorem 2.1.

We now consider the concavity assumptions in Theorem 4.7. The profit function $\pi_j(E_1, E_2)$ is not concave with respect to the variables $\{P_j, C_j, p_j^o\}$ because the limit as any of those variables go to infinity exists. However, it is reasonable to assume that there exists a strategy sub-space $\mathbb{C}_+^4 \subset \mathbb{R}_+^4$ such that $\pi_j(E_1, E_2)$ is concave with respect to the variables in E_j within $\mathbb{C}_+^4$. As a result, we can obtain a Nash equilibrium within the strategy sub-space $\mathbb{C}_+^4$, which is very likely to include all strategies actually pursued by ISPs. It is also likely that $\mathbb{C}_+^4$ is compact and convex. This leads to a sufficient condition for the existence of a Nash equilibrium:

Theorem 4.8: Assume $\pi_j \left(E_1, E_2 \right)$, $E_j \in \mathbb{C}_+^4$ is continuously differentiable and concave with respect to the variables in E_j, $\mathbb{C}_+^4$ is compact and convex, and $\vec{F}(E) \triangleq -\nabla \left(\pi_1, \pi_2 \right)$ is continuous on $\mathbb{C}_+^4$. Then there exists at least one Nash equilibrium in the ISP duopoly game.

Proof: Using [76] Theorem 2.1, it can easily be proven that (4.7) has at least one solution.

A common sufficient condition for the uniqueness of a Nash equilibrium is that $\vec{F}(E) \triangleq -\nabla \left(\pi_1, \pi_2 \right)$ is strongly monotone in $\mathbb{C}_+^4$, see e.g. [76] Theorem 2.2. However, since the distribution of the user type θ depends on user behavior and may be arbitrary, one should not generally expect that this strong monotonicity condition will hold.

∎

APPENDIX H: Proof of Theorem 4.1

Since data caps are absent by assumption in the simplified model, users' utility can be expressed from (4.2) using assumptions F & G as:

$$U(X_j,\theta) = \gamma + \delta\theta + \alpha^2 Q(X_j)^2 \,\theta^2 \big/ 2\beta$$

Since $\gamma \geq P_1$, all users will at least subscribe to the DSL ISP. Denote marginal users who are indifferent between the DSL ISP and the cable ISP by user type θ_{mar}. Thus, users within $\theta_{\text{min}}^2 \leq \theta^2 < \theta_{\text{mar}}^2 \triangleq 2\beta(P_2 - P_1)\big/\alpha^2\left(Q(X_2)^2 - Q(X_1)^2\right)$ will subscribe to the DSL ISP, and users within $\theta_{\text{mar}}^2 \leq \theta^2 \leq \theta_{\text{max}}^2$ will subscribe to the cable ISP.

Since by assumption $K_1(\mu_1) = K_2(\mu_2) = 0$, $U^d(\theta)$ is uniformly distributed between $U^d(\theta_{\text{min}})$ and $U^d(\theta_{\text{max}})$, and X_1 and X_2 are fixed, the ISP profit maximization problem in (4.5) can be reduced to:

$$
\begin{cases}
\max\limits_{P_1} \pi_1 = N F\left(\theta_{\text{mar}}^2\right)\left(P_1 - k_1\right) - Z_1 \\
\max\limits_{P_2} \pi_2 = N\left(1 - F\left(\theta_{\text{mar}}^2\right)\right)\left(P_2 - k_2\right) - Z_2
\end{cases}
\tag{4.8}
$$

where $F(\theta) = \begin{cases} 1 & , \theta > \theta_{\text{max}} \\ \left(\theta^2 - \theta_{\text{min}}^2\right)\big/\left(\theta_{\text{max}}^2 - \theta_{\text{min}}^2\right), & \theta_{\text{min}} \leq \theta \leq \theta_{\text{max}} \\ 0 & , \theta < \theta_{\text{min}} \end{cases}$. Denote the tier prices of the DSL ISP

and the cable ISP in the equilibrium by P_1^* and P_2^*, respectively. If both ISPs earn positive profits in the equilibrium, the ISPs will each have a positive number of subscribers, i.e.

$\theta_{\min}^2 < \theta_{\mathrm{mar}}^2 = 2\beta\left(P_2^* - P_1^*\right) / \alpha^2\left(Q(X_2)^2 - Q(X_1)^2\right) < \theta_{\max}^2$; and the prices in the equilibrium can

be calculated from:

$$\begin{cases} P_1^* = \alpha^2\left(Q(X_2)^2 - Q(X_1)^2\right)\left(\theta_{\max}^2 - 2\theta_{\min}^2\right) / 6\beta + (2k_1 + k_2)/3 \\ P_2^* = \alpha^2\left(Q(X_2)^2 - Q(X_1)^2\right)\left(2\theta_{\max}^2 - \theta_{\min}^2\right) / 6\beta + (k_1 + 2k_2)/3 \end{cases} \qquad (4.9)$$

By replacing P_1^* and P_2^* in $\theta_{\min}^2 < \theta_{\mathrm{mar}}^2 < \theta_{\max}^2$, we have:

$$2\theta_{\min}^2 - \theta_{\max}^2 < \frac{2\beta\left(k_2 - k_1\right)}{\alpha^2\left(Q(X_2)^2 - Q(X_1)^2\right)} < 2\theta_{\max}^2 - \theta_{\min}^2$$

The conditions for ISP fixed costs (i.e. Z_1, Z_2) in Theorem 4.1 can be similarly proved by replacing P_1^* and P_2^* in $\pi_1 > 0$, $\pi_2 > 0$.

On the other hand, if $2\theta_{\min}^2 - \theta_{\max}^2 < \dfrac{2\beta\left(k_2 - k_1\right)}{\alpha^2\left(Q(X_2)^2 - Q(X_1)^2\right)} < 2\theta_{\max}^2 - \theta_{\min}^2$, we can prove that

the prices solved from (4.9) is the equilibrium for the ISP competition, because they satisfy

the ISP profit maximization condition in (4.8) and $\theta_{\min}^2 < \theta_{\mathrm{mar}}^2 < \theta_{\max}^2$. By considering the

conditions for Z_1, Z_2 in Theorem 4.1, we can thus prove that both ISPs are earning positive

profits in the equilibrium.

APPENDIX I: Proof of Theorem 4.4

Denote the optimal surplus of user i when subscribing to ISP j by $S_{i,j}^*(\theta_i) = \max_{t_i} S_{i,j}$.

According to user surplus in (4.3) and theorem 4.2, for a user i who subscribes to ISP j and who is not capped, $\partial S_{i,j}^* / \partial \theta_i = \alpha^2 Q(X_j)^2 \theta_i / \beta \triangleq S_\theta^{u,i,j}$. For a user i who subscribes to ISP j and who is capped but not paying an overage charge, $\partial S_{i,j}^* / \partial \theta_i = \alpha Q(X_j) C_j / X_j \triangleq S_\theta^{c,i,j}$. For a user i who subscribes to ISP j and who is capped and paying an overage charge, $\partial S_{i,j}^* / \partial \theta_i = \alpha Q(X_j)(\alpha\theta_i Q(X_j) - p_j^o X_j) / \beta \triangleq S_\theta^{o,i,j}$. Thus:

$$\begin{cases} S_\theta^{o,2} > S_\theta^{u,1} \Leftrightarrow \theta_i > p_2^o X_2 Q(X_2) / \alpha \left(Q(X_2)^2 - Q(X_1)^2 \right) \\ S_\theta^{c,2} > S_\theta^{u,1} \Leftrightarrow \theta_i < \beta C_2 Q(X_2) / \alpha X_2 Q(X_1)^2 \end{cases}$$

For a user i who is not capped by the cable ISP, i.e. $0 \leq \theta_i \leq \vartheta_2^{(1)}$, we can easily show that $S_\theta^{u,i,2} > S_\theta^{u,i,1}$ since $X_2 > X_1$. So, $\partial \left(S_{i,2}^* - S_{i,1}^* \right) / \partial \theta_i < 0$ for $0 \leq \theta_i \leq \vartheta_2^{(1)}$.

For a user who is capped but not paying an overage charge to the cable ISP, i.e. $\vartheta_2^{(1)} < \theta_i \leq \vartheta_2^{(2)}$, $\theta_i \leq \vartheta_2^{(2)} = \left(\beta C_2 + p_2^o X_2^2 \right) / \alpha Q(X_2) X_2$ and $\dfrac{C_2}{p_2^o} > \dfrac{X_2^2 Q(X_1)^2}{\beta \left(Q(X_2)^2 - Q(X_1)^2 \right)}$ gives $S_\theta^{c,i,2} > S_\theta^{u,i,1}$, or $\partial \left(S_{i,2}^* - S_{i,1}^* \right) / \partial \theta_i < 0$ for $\vartheta_2^{(1)} < \theta_i \leq \vartheta_2^{(2)}$.

For a user who is paying overage charge to the cable ISP, i.e. $\vartheta_2^{(2)} \leq \theta_i \leq 1$,

$$\theta_i \geq \vartheta_2^{(2)} = \left(\beta C_2 + p_2^o X_2^2\right)\Big/\alpha Q(X_2)X_2 \text{ and } \frac{C_2}{p_2^o} > \frac{X_2^2 Q(X_1)^2}{\beta\left(Q(X_2)^2 - Q(X_1)^2\right)} \text{ gives } S_\theta^{o,i,2} > S_\theta^{u,i,1}, \text{ or}$$

$$\partial\left(S_{i,2}^* - S_{i,1}^*\right)\Big/\partial\theta_i < 0 \text{ for } \vartheta_2^{(2)} \leq \theta_i \leq 1.$$

It follows that $\partial\left(S_{i,2}^* - S_{i,1}^*\right)\Big/\partial\theta_i < 0$ for all $0 \leq \theta_i \leq 1$. For users of type $\theta=1$, given

$$S_{i,1}^*(1) \leq \gamma + \delta + \frac{\left(\alpha Q(X_1)\right)^2}{2\beta} - P_1, \quad S_{i,2}^*(1) \geq \gamma + \delta + \frac{\left(\alpha Q(X_2) - p_2^o X_2\right)^2}{2\beta} - P_2 \text{ and } X_2 > X_1, \text{ there}$$

exists a threshold $\alpha^* > 0$, such that when $\alpha \geq \alpha^*$, $S_{i,2}^*(1) > S_{i,1}^*(1)$, i.e. the user prefers the

cable ISP. From $S_{i,1}^*(0) > S_{i,2}^*(0)$, $S_{i,2}\left(1, t_{i,2}^*\right) > S_{i,1}\left(1, t_{i,1}^*\right)$ and $\partial\left(S_{i,2}^* - S_{i,1}^*\right)\Big/\partial\theta_i < 0$ for all

$0 \leq \theta_i \leq 1$, there exists a threshold α^* and a threshold $\theta^{(2)}$ such that for all $\alpha > \alpha^*$, users within

$0 \leq \theta < \theta^{(2)}$ prefer the DSL ISP and users within $\theta^{(2)} \leq \theta < 1$ prefer the cable ISP.

APPENDIX J: Proof of Theorem 4.5

Based on user surplus in (4.3) and Theorem 4.2, $C_2 > X_2 \sqrt{\dfrac{2(P_2 - P_1)Q(X_2)^2}{\beta\left(Q(X_2)^2 - Q(X_1)^2\right)}}$ gives

$S_{i,2}^*\left(\vartheta_2^{(1)}\right) > S_{i,1}^*\left(\vartheta_2^{(1)}\right)$, i.e. users of type $\vartheta_2^{(1)}$ prefer the cable ISP. Thus, when

$\alpha \geq \alpha_1^* \triangleq \sqrt{\dfrac{2\beta(P_2 - P_1)}{Q(X_2)^2 - Q(X_1)^2}}$, there exists a threshold $0 < \theta^{(2)} < \vartheta_2^{(1)}$ such that users within

$0 \leq \theta < \theta^{(2)}$ prefer the DSL ISP and users within $\theta^{(2)} \leq \theta_i \leq \vartheta_2^{(1)}$ prefer the cable ISP.

Similarly, $p_2^o = \infty$ gives $S_{i,2}^*(1) = \gamma + \delta + \dfrac{\alpha Q(X_2)C_2}{X_2} - \dfrac{\beta C_2^2}{2X_2^2}$; and $p_1^o < \infty$ gives

$S_{i,1}^*(1) \geq \gamma + \delta + \dfrac{\left(\alpha Q(X_1) - p_1^o X_1\right)^2}{2\beta}$. Thus there exists a threshold $\alpha_2^* > 0$ such that when

$\alpha \geq \alpha_2^*$, $S_{i,1}^*(1) > S_{i,2}^*(1)$, i.e. a user of type $\theta = 1$ prefers the DSL ISP. From $S_{i,2}^*\left(\vartheta_2^{(1)}\right) > S_{i,1}^*\left(\vartheta_2^{(1)}\right)$

and $S_{i,1}^*(1) > S_{i,2}^*(1)$, there exists a threshold $\theta^{(3)}$ such that users within $\vartheta_2^{(1)} \leq \theta_i < \theta^{(3)}$ prefer

the cable ISP and users within $\theta^{(3)} \leq \theta < 1$ prefer the DSL ISP.

In summary, when $\alpha \geq \max\left(\alpha_1^*, \alpha_2^*\right) \triangleq \alpha^*$, there exists two thresholds $\theta^{(2)}$ and $\theta^{(3)}$ such that

users within $0 \leq \theta < \theta^{(2)}$ and $\theta^{(3)} \leq \theta < 1$ prefer the DSL ISP and users within $\theta^{(2)} \leq \theta < \theta^{(3)}$ prefer the

cable ISP.

APPENDIX K: Proof of Theorem 4.6

Denote by $Data(\theta, X_j)$ the data consumption by users of type θ who subscribe to ISP j in the presence of data caps. Thus,

$$Data\left(\theta, X_j\right) = \begin{cases} \alpha\theta Q\left(X_j\right)X_j \big/ \beta, & \text{if } 0 \leq \theta < \vartheta_j^{(1)} \\ C_j, & \text{if } \vartheta_j^{(1)} \leq \theta \leq \vartheta_j^{(2)} \\ \left(\alpha\theta Q\left(X_j\right) - p_j^o X_j\right)\big/\beta, & \text{if } \vartheta_j^{(1)} < \theta \leq 1 \end{cases}$$

Denote by Θ_j^u the set of user types who subscribe to ISP j and are not capped. Denote by Θ_j^c the set of user type who subscribe to ISP j and are capped but not paying overage charges. Denote by Θ_j^o the set of user type who subscribe to ISP j and are capped and are paying overage charges. Thus, profit in the presence of data caps is:

$$\pi_j = \int_{\theta \in \Theta_j^u} f\left(\theta\right)\left(P_j - k_j - p_j^\mu Data\left(\theta, X_j\right)\big/\rho^{th}\right)d\theta +$$

$$\int_{\theta \in \Theta_j^c \cup \Theta_j^o} f\left(\theta\right)\left(P_j + p_j^o\left(Data\left(\theta, X_j\right) - C_j\right) - k_j - p_j^\mu Data\left(\theta, X_j\right)\big/\rho^{th}\right)d\theta - Z_j$$

At the heavy-users cap Nash equilibrium $\left(E_1^{(0)}, E_2^{(0)}\right)$, the first order optimization conditions, $p_j^o = p_j^\mu \big/ \rho^{th}$, and $C_j = \left(P_j - k_j\right)\big/ p_j^o$ give:

$$\frac{\partial \pi_j^{(0)}}{\partial P_j} = 0, \quad \frac{\partial \pi_j^{(0)}}{\partial X_j} = 0,$$

$$\frac{\partial \pi_j^{(0)}}{\partial C_j} = -p_j^o \int_{\theta \in \Theta_j^c \cup \Theta_j^o} f\left(\theta\right)d\theta \leq 0, \quad \frac{\partial \pi_j^{(0)}}{\partial p_j^o} = \int_{\theta \in \Theta_j^o} f\left(\theta\right)\left(Data\left(\theta, X_j\right) - C_j\right)d\theta \geq 0$$

The expressions for $\partial \pi_j^{(1)}/\partial P_j$ and $\partial \pi_j^{(1)}/\partial X_j$ can thus be derived from $\partial \pi_j^{(0)}/\partial P_j = 0$ and $\partial \pi_j^{(0)}/\partial X_j = 0$ as follows, when both ISPs use the gradient descent method to maximize their profits:

$$\frac{\partial \pi_j^{(1)}}{\partial P_j} = \frac{\partial^2 \pi_j^{(0)}}{\partial C_j \partial P_j} \frac{\partial \pi_j^{(0)}}{\partial C_j} \varepsilon + \frac{\partial^2 \pi_j^{(0)}}{\partial p_j^o \partial P_j} \frac{\partial \pi_j^{(0)}}{\partial p_j^o} \varepsilon + \frac{\partial^2 \pi_j^{(0)}}{\partial C_{3-j} \partial P_j} \frac{\partial \pi_{3-j}^{(0)}}{\partial C_{3-j}} \varepsilon + \frac{\partial^2 \pi_j^{(0)}}{\partial p_{3-j}^o \partial P_j} \frac{\partial \pi_{3-j}^{(0)}}{\partial p_{3-j}^o} \varepsilon$$

$$\frac{\partial \pi_j^{(1)}}{\partial X_j} = \frac{\partial^2 \pi_j^{(0)}}{\partial C_j \partial X_j} \frac{\partial \pi_j^{(0)}}{\partial C_j} \varepsilon + \frac{\partial^2 \pi_j^{(0)}}{\partial p_j^o \partial X_j} \frac{\partial \pi_j^{(0)}}{\partial p_j^o} \varepsilon + \frac{\partial^2 \pi_j^{(0)}}{\partial C_{3-j} \partial X_j} \frac{\partial \pi_{3-j}^{(0)}}{\partial C_{3-j}} \varepsilon + \frac{\partial^2 \pi_j^{(0)}}{\partial p_{3-j}^o \partial X_j} \frac{\partial \pi_{3-j}^{(0)}}{\partial p_{3-j}^o} \varepsilon$$

(4.10)

Case 1: scenarios a-1 and scenario b-1

Under scenario a-1, no users in the DSL ISP are capped. Thus, we can easily prove:

$$\frac{\partial \pi_1^{(0)}}{\partial C_1} = 0, \quad \frac{\partial \pi_1^{(0)}}{\partial p_1^o} = 0, \quad \frac{\partial^2 \pi_1^{(0)}}{\partial p_2^o \partial P_1} = 0, \quad \frac{\partial^2 \pi_1^{(0)}}{\partial C_2 \partial P_1} = 0, \quad \frac{\partial^2 \pi_1^{(0)}}{\partial p_2^o \partial X_1} = 0, \quad \frac{\partial^2 \pi_1^{(0)}}{\partial C_2 \partial X_1} = 0$$

When scenario b-1 occurs, the cable ISP's profit can be expressed as follows, given $\theta^{(2)} < \vartheta_2^{(1)}$:

$$\pi_2^{(0)} = \int_{\theta=\theta^{(2)}}^{\vartheta_2^{(1)}} f(\theta)\left(P_2 - k_2 - \frac{p_2^\mu Data(\theta, X_2)}{\rho^{th}} \right) d\theta +$$

$$\int_{\theta=\vartheta_2^{(1)}}^{1} f(\theta)\left(P_2 + p_2^o \left(Data(\theta, X_2) - C_2 \right) - k_2 - \frac{p_2^\mu Data(\theta, X_2)}{\rho^{th}} \right) d\theta - Z_2$$

which gives,

$$\frac{\partial \pi_2^{(0)}}{\partial P_2} = \int_{\theta=\theta^{(2)}}^{1} f(\theta) d\theta - f(\theta^{(2)}) \frac{\partial \theta^{(2)}}{\partial P_2} \left(P_2 - k_2 - \frac{p_2^{\mu} Data(\theta^{(2)}, X_2)}{\rho^{th}} \right)$$

$$\frac{\partial \pi_2^{(0)}}{\partial X_2} = -f(\theta^{(2)}) \frac{\partial \theta^{(2)}}{\partial X_2} \left(P_2 - k_2 - \frac{p_2^{\mu} Data(\theta^{(2)}, X_2)}{\rho^{th}} \right) - \int_{\theta=\theta^{(2)}}^{\vartheta_2^{(1)}} \frac{f(\theta) p_2^{\mu}}{\rho^{th}} \frac{\partial Data(\theta, X_2)}{\partial X_2} d\theta$$

$$+ \int_{\theta=\vartheta_2^{(2)}}^{1} f(\theta) \left(p_2^{o} - \frac{p_2^{\mu}}{\rho^{th}} \right) \frac{\partial Data(\theta, X_2)}{\partial X_2} d\theta$$

Thus, $p_j^o = p_j^{\mu} / \rho^{th}$ and $C_j = (P_j - k_j)/p_j^o$ can give:

$$\frac{\partial^2 \pi_2^{(0)}}{\partial p_2^o \partial P_2} = 0, \quad \frac{\partial^2 \pi_2^{(0)}}{\partial C_2 \partial P_2} = 0, \quad \frac{\partial^2 \pi_2^{(0)}}{\partial C_2 \partial X_2} = -\frac{\partial \vartheta_2^{(1)}}{\partial C_2} \frac{f(\vartheta_2^{(1)}) p_2^{\mu}}{\rho^{th}} \frac{\partial Data(\theta, X_2)}{\partial X_2} \bigg|_{\theta=\vartheta_2^{(1)}} \leq 0,$$

$$\frac{\partial^2 \pi_2^{(0)}}{\partial p_2^o \partial X_2} = \int_{\theta=\vartheta_2^{(2)}}^{1} f(\theta) \frac{\partial Data(\theta, X_2)}{\partial X_2} d\theta$$

Since $Q'(X_2^{(0)}) \geq Q(X_2^{(0)})(\vartheta_2^{(2)} - 2\vartheta_2^{(1)})/X_2^{(0)} \vartheta_2^{(2)}$, i.e. $\partial Data(\theta, X_2)/\partial X_2 \geq 0$ for all θ within

$\vartheta_2^{(2)} < \theta \leq 1$, we have $\frac{\partial^2 \pi_2^{(0)}}{\partial p_2^o \partial X_2} \geq 0$. $\frac{\partial \pi_1^{(1)}}{\partial P_1} = 0, \frac{\partial \pi_1^{(1)}}{\partial X_1} = 0, \frac{\partial \pi_2^{(1)}}{\partial P_2} = 0, \frac{\partial \pi_2^{(1)}}{\partial X_2} \geq 0$ can thus be derived

from (4.10) by substituting $\frac{\partial^2 \pi_2^{(0)}}{\partial p_2^o \partial P_2} = 0, \frac{\partial^2 \pi_2^{(0)}}{\partial C_2 \partial P_2} = 0, \frac{\partial^2 \pi_2^{(0)}}{\partial C_2 \partial X_2} = 0, \frac{\partial^2 \pi_2^{(0)}}{\partial p_2^o \partial X_2} \geq 0$.

Case 2: scenarios a-1 and scenario b-2

In this case, users with type $\theta^{(1,1)} \leq \theta \leq \theta^{(2)}$ subscribe to the DSL ISP and are not capped, and

users with type $\theta \geq \theta^{(3)}$ subscribe to the DSL ISP and will be capped. Thus, the DSL ISP's

profit is:

$$\pi_1^{(0)} = \int_{\theta=\theta^{(1,1)}}^{\theta^{(2)}} f(\theta) \left(P_1 - k_1 - \frac{p_1^{\mu} Data(\theta, X_1)}{\rho^{th}} \right) d\theta +$$

$$\int_{\theta=\theta^{(3)}}^{1} f(\theta) \left(P_1 + p_1^{o} \left(Data(\theta, X_1) - C_1 \right) - k_1 - \frac{p_1^{\mu} Data(\theta, X_1)}{\rho^{th}} \right) d\theta - Z_1$$

which gives:

$$\frac{\partial \pi_1^{(0)}}{\partial P_1} = \int_{\theta=\theta^{(1,1)}}^{\theta^{(2)}} f(\theta) d\theta + f(\theta^{(2)}) \frac{\partial \theta^{(2)}}{\partial P_1} \left(P_1 - k_1 - \frac{p_1^{\mu} Data(\theta^{(2)}, X_1)}{\rho^{th}} \right)$$

$$- f(\theta^{(1,1)}) \frac{\partial \theta^{(1,1)}}{\partial P_1} \left(P_1 - k_1 - \frac{p_1^{\mu} Data(\theta^{(1,1)}, X_1)}{\rho^{th}} \right) + \int_{\theta=\theta^{(3)}}^{1} f(\theta) d\theta$$

$$- f(\theta^{(3)}) \frac{\partial \theta^{(3)}}{\partial P_1} \left(P_1 + p_1^{o} \left(Data(\theta^{(3)}, X_1) - C_1 \right) - k_1 - \frac{p_1^{\mu} Data(\theta^{(3)}, X_1)}{\rho^{th}} \right)$$

$$\frac{\partial \pi_1^{(0)}}{\partial X_1} = -\int_{\theta=\theta^{(1,1)}}^{\theta^{(2)}} f(\theta) \frac{p_1^{\mu}}{\rho^{th}} \frac{\partial Data(\theta, X_1)}{\partial X_1} d\theta + \int_{\theta=\theta^{(3)}}^{1} f(\theta) \left(p_1^{o} - \frac{p_1^{\mu}}{\rho^{th}} \right) \frac{\partial Data(\theta, X_1)}{\partial X_1} d\theta$$

$$+ f(\theta^{(2)}) \frac{\partial \theta^{(2)}}{\partial X_1} \left(P_1 - k_1 - \frac{p_1^{\mu} Data(\theta^{(2)}, X_1)}{\rho^{th}} \right)$$

$$- f(\theta^{(1,1)}) \frac{\partial \theta^{(1,1)}}{\partial X_1} \left(P_1 - k_1 - \frac{p_1^{\mu} Data(\theta^{(1,1)}, X_1)}{\rho^{th}} \right)$$

$$- f(\theta^{(3)}) \frac{\partial \theta^{(3)}}{\partial X_1} \left(P_1 + p_1^{o} \left(Data(\theta^{(3)}, X_1) - C_1 \right) - k_1 - \frac{p_1^{\mu} Data(\theta^{(3)}, X_1)}{\rho^{th}} \right)$$

Thus, $p_j^{o} = p_j^{\mu} / \rho^{th}$ and $C_j = \left(P_j - k_j \right) / p_j^{o}$ can give:

$$\frac{\partial^2 \pi_1^{(0)}}{\partial p_1^o \partial P_1} = -f\left(\theta^{(3)}\right)\frac{\partial \theta^{(3)}}{\partial P_1}\left(Data\left(\theta^{(3)}, X_1\right) - C_1\right) - f\left(\theta^{(3)}\right)\frac{\partial \theta^{(3)}}{\partial p_1^o} \leq 0$$

$$\frac{\partial^2 \pi_1^{(0)}}{\partial C_1 \partial P_1} = f\left(\theta^{(3)}\right)\frac{\partial \theta^{(3)}}{\partial P_1}p_1^o - f\left(\theta^{(3)}\right)\frac{\partial \theta^{(3)}}{\partial C_1} \geq 0$$

$$\frac{\partial^2 \pi_1^{(0)}}{\partial p_2^o \partial P_1} = -f\left(\theta^{(3)}\right)\frac{\partial \theta^{(3)}}{\partial p_2^o} \geq 0, \quad \frac{\partial^2 \pi_1^{(0)}}{\partial C_2 \partial P_1} = -f\left(\theta^{(3)}\right)\frac{\partial \theta^{(3)}}{\partial C_2} \leq 0$$

$$\frac{\partial^2 \pi_1^{(0)}}{\partial p_1^o \partial X_1} = \int_{\theta=\theta^{(3)}}^{1} f\left(\theta\right)\frac{\partial Data\left(\theta, X_1\right)}{\partial X_1}d\theta - f\left(\theta^{(3)}\right)\frac{\partial \theta^{(3)}}{\partial X_1}\left(Data\left(\theta^{(3)}, X_1\right) - C_1\right)$$

$$\frac{\partial^2 \pi_1^{(0)}}{\partial C_1 \partial X_1} = f\left(\theta^{(3)}\right)\frac{\partial \theta^{(3)}}{\partial X_1}p_1^o \leq 0, \quad \frac{\partial^2 \pi_1^{(0)}}{\partial p_2^o \partial X_1} = 0, \quad \frac{\partial^2 \pi_1^{(0)}}{\partial C_2 \partial X_1} = 0$$

Users with type $\theta^{(2)} \leq \theta \leq \theta^{(3)}$ subscribe to the cable ISP. Thus, the cable ISP's profit can be expressed as follows given $\theta^{(2)} < \vartheta_2^{(1)}$:

$$\pi_2^{(0)} = \int_{\theta=\theta^{(2)}}^{\vartheta_2^{(1)}} f\left(\theta\right)\left(P_2 - k_2 - \frac{p_2^\mu Data\left(\theta, X_2\right)}{\rho^{th}}\right)d\theta +$$

$$\int_{\theta=\vartheta_2^{(1)}}^{\theta^{(3)}} f\left(\theta\right)\left(P_2 + p_2^o\left(Data\left(\theta, X_2\right) - C_2\right) - k_2 - \frac{p_2^\mu Data\left(\theta, X_2\right)}{\rho^{th}}\right)d\theta - Z_2$$

which gives:

$$\frac{\partial \pi_2^{(0)}}{\partial P_2} = \int_{\theta=\theta^{(2)}}^{\theta^{(3)}} f\left(\theta\right)d\theta - f\left(\theta^{(2)}\right)\frac{\partial \theta^{(2)}}{\partial P_2}\left(P_2 - k_2 - \frac{p_2^\mu Data\left(\theta^{(2)}, X_2\right)}{\rho^{th}}\right)$$

$$+ f\left(\theta^{(3)}\right)\frac{\partial \theta^{(3)}}{\partial P_2}\left(P_2 + p_2^o\left(Data\left(\theta^{(3)}, X_2\right) - C_2\right) - k_2 - \frac{p_2^\mu Data\left(\theta^{(3)}, X_2\right)}{\rho^{th}}\right)$$

$$\frac{\partial \pi_2^{(0)}}{\partial X_2} = -\int_{\theta=\theta^{(2)}}^{\vartheta_2^{(1)}} \frac{f(\theta)\, p_2^{\mu}}{\rho^{th}} \frac{\partial Data(\theta, X_2)}{\partial X_2} d\theta + \int_{\theta=\vartheta_2^{(2)}}^{\theta^{(3)}} f(\theta)\left(p_2^{o} - \frac{p_2^{\mu}}{\rho^{th}}\right) \frac{\partial Data(\theta, X_2)}{\partial X_2} d\theta$$

$$- f(\theta^{(2)}) \frac{\partial \theta^{(2)}}{\partial X_2}\left(P_2 - k_2 - \frac{p_2^{\mu} Data(\theta^{(2)}, X_2)}{\rho^{th}}\right)$$

$$+ f(\theta^{(3)}) \frac{\partial \theta^{(3)}}{\partial X_2}\left(P_2 + p_2^{o}\left(Data(\theta^{(3)}, X_2) - C_2\right) - k_2 - \frac{p_2^{\mu} Data(\theta^{(3)}, X_2)}{\rho^{th}}\right)$$

Thus, $p_j^{o} = p_j^{\mu}/\rho^{th}$ and $C_j = \left(P_j - k_j\right)/p_j^{o}$ can give:

$$\frac{\partial^2 \pi_2^{(0)}}{\partial p_1^{o} \partial P_2} = f(\theta^{(3)}) \frac{\partial \theta^{(3)}}{\partial p_1^{o}} \geq 0, \quad \frac{\partial^2 \pi_2^{(0)}}{\partial C_1 \partial P_2} = f(\theta^{(3)}) \frac{\partial \theta^{(3)}}{\partial C_1} \leq 0$$

$$\frac{\partial^2 \pi_2^{(0)}}{\partial p_2^{o} \partial P_2} = f(\theta^{(3)}) \frac{\partial \theta^{(3)}}{\partial P_2}\left(Data(\theta^{(3)}, X_2) - C_2\right) + f(\theta^{(3)}) \frac{\partial \theta^{(3)}}{\partial p_2^{o}} \leq 0$$

$$\frac{\partial^2 \pi_2^{(0)}}{\partial C_2 \partial P_2} = -f(\theta^{(3)}) \frac{\partial \theta^{(3)}}{\partial P_2} p_2^{o} + f(\theta^{(3)}) \frac{\partial \theta^{(3)}}{\partial C_2} \geq 0$$

$$\frac{\partial^2 \pi_2^{(0)}}{\partial p_1^{o} \partial X_2} = 0, \quad \frac{\partial^2 \pi_2^{(0)}}{\partial C_1 \partial X_2} = 0, \quad \frac{\partial^2 \pi_2^{(0)}}{\partial C_2 \partial X_2} = -f(\theta^{(3)}) \frac{\partial \theta^{(3)}}{\partial X_2} p_2^{o} \leq 0$$

$$\frac{\partial^2 \pi_2^{(0)}}{\partial p_2^{o} \partial X_2} = \int_{\theta=\vartheta_2^{(2)}}^{\theta^{(3)}} f(\theta) \frac{\partial Data(\theta, X_2)}{\partial X_2} d\theta + f(\theta^{(3)}) \frac{\partial \theta^{(3)}}{\partial X_2}\left(Data(\theta^{(3)}, X_2) - C_2\right)$$

Since $Q'\left(X_j^{(0)}\right) \geq Q\left(X_j^{(0)}\right)\left(\vartheta_j^{(2)} - 2\vartheta_j^{(1)}\right)\Big/ X_j^{(0)} \vartheta_j^{(2)}$, i.e. $\partial Data(\theta, X_j)/\partial X_j \geq 0$ for all θ within

$\vartheta_j^{(2)} < \theta \leq 1$, we have $\dfrac{\partial^2 \pi_1^{(0)}}{\partial p_1^{o} \partial X_1} \geq 0, \dfrac{\partial^2 \pi_2^{(0)}}{\partial p_2^{o} \partial X_2} \geq 0$. $\dfrac{\partial \pi_1^{(1)}}{\partial X_1} \geq 0, \dfrac{\partial \pi_2^{(1)}}{\partial X_2} \geq 0$ can thus be derived from

(4.10) by substituting the above second order partial derivatives.

Case 3: scenarios a-2 and scenario b-1:

In this case, the DSL ISP's profit is:

$$\pi_1^{(0)} = \int_{\theta=\theta^{(1,1)}}^{\mathcal{G}_1^{(1)}} f(\theta) \left(P_1 - k_1 - \frac{p_1^{\mu} Data(\theta, X_1)}{\rho^{th}} \right) d\theta$$

, which gives:

$$+ \int_{\theta=\mathcal{G}_1^{(1)}}^{\theta^{(2)}} f(\theta) \left(P_1 + p_1^o \left(Data(\theta, X_1) - C_1 \right) - k_1 - \frac{p_1^{\mu} Data(\theta, X_1)}{\rho^{th}} \right) d\theta - Z_1$$

$$\frac{\partial \pi_1^{(0)}}{\partial P_1} = \int_{\theta=\theta^{(1,1)}}^{\theta^{(2)}} f(\theta) d\theta - f\left(\theta^{(1,1)}\right) \frac{\partial \theta^{(1,1)}}{\partial P_1} \left(P_1 - k_1 - \frac{p_1^{\mu} Data\left(\theta^{(1,1)}, X_1\right)}{\rho^{th}} \right)$$

$$+ f\left(\theta^{(2)}\right) \frac{\partial \theta^{(2)}}{\partial P_1} \left(P_1 + p_1^o \left(Data\left(\theta^{(2)}, X_1\right) - C_1 \right) - k_1 - \frac{p_1^{\mu} Data\left(\theta^{(2)}, X_1\right)}{\rho^{th}} \right)$$

$$\frac{\partial \pi_1^{(0)}}{\partial X_1} = -\int_{\theta=\theta^{(1,1)}}^{\mathcal{G}_1^{(1)}} \frac{f(\theta) p_1^{\mu}}{\rho^{th}} \frac{\partial Data(\theta, X_1)}{\partial X_1} d\theta + \int_{\theta=\mathcal{G}_1^{(2)}}^{\theta^{(2)}} f(\theta) \left(p_1^o - \frac{p_1^{\mu}}{\rho^{th}} \right) \frac{\partial Data(\theta, X_1)}{\partial X_1} d\theta$$

$$- f\left(\theta^{(1,1)}\right) \frac{\partial \theta^{(1,1)}}{\partial X_1} \left(P_1 - k_1 - \frac{p_1^{\mu} Data\left(\theta^{(1,1)}, X_1\right)}{\rho^{th}} \right)$$

$$+ f\left(\theta^{(2)}\right) \frac{\partial \theta^{(2)}}{\partial X_1} \left(P_1 + p_1^o \left(Data\left(\theta^{(2)}, X_1\right) - C_1 \right) - k_1 - \frac{p_1^{\mu} Data\left(\theta^{(2)}, X_1\right)}{\rho^{th}} \right)$$

Thus, $p_j^o = p_j^{\mu}/\rho^{th}$ and $C_j = \left(P_j - k_j\right)/p_j^o$ can give:

$$\frac{\partial^2 \pi_1^{(0)}}{\partial p_1^o \partial P_1} = f\left(\theta^{(2)}\right) \frac{\partial \theta^{(2)}}{\partial P_1} \left(Data\left(\theta^{(2)}, X_1\right) - C_1 \right) + f\left(\theta^{(2)}\right) \frac{\partial \theta^{(2)}}{\partial p_1^o} \le 0,$$

$$\frac{\partial^2 \pi_1^{(0)}}{\partial C_1 \partial P_1} = -f\left(\theta^{(2)}\right) \frac{\partial \theta^{(2)}}{\partial P_1} p_1^o + f\left(\theta^{(2)}\right) \frac{\partial \theta^{(2)}}{\partial C_1} \ge 0, \quad \frac{\partial^2 \pi_1^{(0)}}{\partial p_2^o \partial P_1} = 0, \quad \frac{\partial^2 \pi_1^{(0)}}{\partial C_2 \partial P_1} = 0$$

$$\frac{\partial^2 \pi_1^{(0)}}{\partial p_1^o \partial X_1} = \int_{\theta=\mathcal{G}_1^{(2)}}^{\theta^{(2)}} f(\theta) \frac{\partial Data(\theta, X_1)}{\partial X_1} d\theta + f\left(\theta^{(2)}\right) \frac{\partial \theta^{(2)}}{\partial X_1} \left(Data\left(\theta^{(2)}, X_1\right) - C_1 \right)$$

$$\frac{\partial^2 \pi_1^{(0)}}{\partial C_1 \partial X_1} = -f\left(\theta^{(2)}\right) \frac{\partial \theta^{(2)}}{\partial X_1} p_1^o \le 0, \quad \frac{\partial^2 \pi_1^{(0)}}{\partial p_2^o \partial X_1} = 0, \quad \frac{\partial^2 \pi_1^{(0)}}{\partial C_2 \partial X_1} = 0$$

When scenarios a-2 and b-1 occur, the cable ISP's profit is:

$$\pi_2^{(0)} = \int_{\theta=\theta^{(2)}}^{\vartheta_2^{(1)}} f\left(\theta\right)\left(P_2 - k_2 - \frac{p_2^{\mu} Data\left(\theta, X_2\right)}{\rho^{th}} \right) d\theta$$

$$+ \int_{\theta=\vartheta_2^{(1)}}^{1} f\left(\theta\right)\left(P_2 + p_2^{o}\left(Data\left(\theta, X_2\right) - C_2 \right) - k_2 - \frac{p_2^{\mu} Data\left(\theta, X_2\right)}{\rho^{th}} \right) d\theta - Z_2$$

, which

gives:

$$\frac{\partial \pi_2^{(0)}}{\partial P_2} = \int_{\theta=\theta^{(2)}}^{1} f\left(\theta\right) d\theta - f\left(\theta^{(2)}\right)\frac{\partial \theta^{(2)}}{\partial P_2}\left(P_2 - k_2 - \frac{p_2^{\mu} Data\left(\theta^{(2)}, X_2\right)}{\rho^{th}} \right)$$

$$\frac{\partial \pi_2^{(0)}}{\partial X_2} = -\int_{\theta=\theta^{(2)}}^{\vartheta_2^{(1)}} \frac{f\left(\theta\right) p_2^{\mu}}{\rho^{th}} \frac{\partial Data\left(\theta, X_2\right)}{\partial X_2} d\theta + \int_{\theta=\vartheta_2^{(2)}}^{1} f\left(\theta\right)\left(p_2^{o} - \frac{p_2^{\mu}}{\rho^{th}} \right)\frac{\partial Data\left(\theta, X_2\right)}{\partial X_2} d\theta$$

$$- f\left(\theta^{(2)}\right)\frac{\partial \theta^{(2)}}{\partial X_2}\left(P_2 - k_2 - \frac{p_2^{\mu} Data\left(\theta^{(2)}, X_2\right)}{\rho^{th}} \right)$$

Thus, $p_j^{o} = p_j^{\mu}/\rho^{th}$ and $C_j = \left(P_j - k_j\right)/p_j^{o}$ can give:

$$\frac{\partial^2 \pi_2^{(0)}}{\partial p_1^{o} \partial P_2} = -f\left(\theta^{(2)}\right)\frac{\partial \theta^{(2)}}{\partial p_1^{o}} - f\left(\theta^{(2)}\right)\frac{\partial^2 \theta^{(2)}}{\partial p_1^{o} \partial P_2}\left(P_2 - k_2 - \frac{p_2^{\mu} Data\left(\theta^{(2)}, X_2\right)}{\rho^{th}} \right)$$

$$- f'\left(\theta^{(2)}\right)\frac{\partial \theta^{(2)}}{\partial p_1^{o}}\frac{\partial \theta^{(2)}}{\partial P_2}\left(P_2 - k_2 - \frac{p_2^{\mu} Data\left(\theta^{(2)}, X_2\right)}{\rho^{th}} \right)$$

$$- f\left(\theta^{(2)}\right)\frac{\partial \theta^{(2)}}{\partial p_1^{o}}\frac{\partial \theta^{(2)}}{\partial P_2}\frac{p_2^{\mu}}{\rho^{th}}\frac{\partial Data\left(\theta, X_2\right)}{\partial \theta}\Bigg|_{\theta=\theta^{(2)}}$$

$$\frac{\partial^2 \pi_2^{(0)}}{\partial C_1 \partial P_2} = -f\left(\theta^{(2)}\right)\frac{\partial \theta^{(2)}}{\partial C_1} - f\left(\theta^{(2)}\right)\frac{\partial^2 \theta^{(2)}}{\partial C_1 \partial P_2}\left(P_2 - k_2 - \frac{p_2^{\mu} Data\left(\theta^{(2)}, X_2\right)}{\rho^{th}} \right)$$

$$- f'\left(\theta^{(2)}\right)\frac{\partial \theta^{(2)}}{\partial C_1}\frac{\partial \theta^{(2)}}{\partial P_2}\left(P_2 - k_2 - \frac{p_2^{\mu} Data\left(\theta^{(2)}, X_2\right)}{\rho^{th}} \right)$$

$$- f\left(\theta^{(2)}\right)\frac{\partial \theta^{(2)}}{\partial C_1}\frac{\partial \theta^{(2)}}{\partial P_2}\frac{p_2^{\mu}}{\rho^{th}}\frac{\partial Data\left(\theta, X_2\right)}{\partial \theta}\Bigg|_{\theta=\theta^{(2)}}$$

$$\frac{\partial^2 \pi_2^{(0)}}{\partial p_2^o \partial P_2} = 0, \quad \frac{\partial^2 \pi_2^{(0)}}{\partial C_2 \partial P_2} = 0$$

$$\frac{\partial^2 \pi_2^{(0)}}{\partial p_1^o \partial X_2} = \frac{f\left(\theta^{(2)}\right) p_2^\mu}{\rho^{th}} \frac{\partial Data\left(\theta^{(2)}, X_2\right)}{\partial X_2} \frac{\partial \theta^{(2)}}{\partial p_1^o} - f\left(\theta^{(2)}\right) \frac{\partial^2 \theta^{(2)}}{\partial p_1^o \partial X_2} \left(P_2 - k_2 - \frac{p_2^\mu Data\left(\theta^{(2)}, X_2\right)}{\rho^{th}} \right)$$

$$- f'\left(\theta^{(2)}\right) \frac{\partial \theta^{(2)}}{\partial p_1^o} \frac{\partial \theta^{(2)}}{\partial X_2} \left(P_2 - k_2 - \frac{p_2^\mu Data\left(\theta^{(2)}, X_2\right)}{\rho^{th}} \right)$$

$$- f\left(\theta^{(2)}\right) \frac{\partial \theta^{(2)}}{\partial p_1^o} \frac{\partial \theta^{(2)}}{\partial X_2} \frac{p_2^\mu}{\rho^{th}} \left. \frac{\partial Data\left(\theta, X_2\right)}{\partial \theta} \right|_{\theta=\theta^{(2)}}$$

$$\frac{\partial^2 \pi_2^{(0)}}{\partial C_1 \partial X_2} = \frac{f\left(\theta^{(2)}\right) p_2^\mu}{\rho^{th}} \frac{\partial Data\left(\theta^{(2)}, X_2\right)}{\partial X_2} \frac{\partial \theta^{(2)}}{\partial C_1} - f\left(\theta^{(2)}\right) \frac{\partial^2 \theta^{(2)}}{\partial C_1 \partial X_2} \left(P_2 - k_2 - \frac{p_2^\mu Data\left(\theta^{(2)}, X_2\right)}{\rho^{th}} \right)$$

$$- f'\left(\theta^{(2)}\right) \frac{\partial \theta^{(2)}}{\partial C_1} \frac{\partial \theta^{(2)}}{\partial X_2} \left(P_2 - k_2 - \frac{p_2^\mu Data\left(\theta^{(2)}, X_2\right)}{\rho^{th}} \right)$$

$$- f\left(\theta^{(2)}\right) \frac{\partial \theta^{(2)}}{\partial C_1} \frac{\partial \theta^{(2)}}{\partial X_2} \frac{p_2^\mu}{\rho^{th}} \left. \frac{\partial Data\left(\theta, X_2\right)}{\partial \theta} \right|_{\theta=\theta^{(2)}}$$

Denote $Th \triangleq \max\left(Th_1\left(E_1^{(0)}, E_2^{(0)}\right), Th_2\left(E_1^{(0)}, E_2^{(0)}\right) \right)$, where:

$$Th_1\left(E_1, E_2\right) = -\frac{f\left(\theta^{(2)}\right)}{\left(P_2 - D\right)\left(\partial \theta^{(2)} / \partial P_2\right)} - \frac{f\left(\theta^{(2)}\right)}{\partial \theta^{(2)} / \partial P_2} \min\left(\frac{\partial^2 \theta^{(2)} / \partial p_1^o \partial P_2}{\partial \theta^{(2)} / \partial p_1^o}, \frac{\partial^2 \theta^{(2)} / \partial C_1 \partial P_2}{\partial \theta^{(2)} / \partial C_1} \right)$$

$$+ \frac{f\left(\theta^{(2)}\right) p_2^\mu X_2 \alpha Q\left(X_2\right)}{\rho^{th}\left(P_2 - D\right)\beta}$$

$$Th_2\left(E_1, E_2\right) = \frac{f\left(\theta^{(2)}\right) p_2^\mu \theta^{(2)} \alpha\left(Q\left(X_2\right) + X_2 Q'\left(X_2\right)\right)}{\rho^{th}\left(P_2 - D\right)\left(\partial \theta^{(2)} / \partial X_2\right)\beta} + \frac{f\left(\theta^{(2)}\right) p_2^\mu X_2 \alpha Q\left(X_2\right)}{\rho^{th}\left(P_2 - D\right)\beta}$$

$$- \frac{f\left(\theta^{(2)}\right)}{\partial \theta^{(2)} / \partial X_2} \max\left(\frac{\partial^2 \theta^{(2)} / \partial p_1^o \partial X_2}{\partial \theta^{(2)} / \partial p_1^o}, \frac{\partial^2 \theta^{(2)} / \partial C_1 \partial X_2}{\partial \theta^{(2)} / \partial C_1} \right)$$

$$D = p_2^\mu X_2 \alpha \theta^{(2)} Q\left(X_2\right) / \rho^{th} \beta$$

$f'(\theta) \geq Th$ can give $\dfrac{\partial^2 \pi_2^{(0)}}{\partial p_1^o \partial P_2} \geq 0, \dfrac{\partial^2 \pi_2^{(0)}}{\partial C_1 \partial P_2} \leq 0$. As a result, since

$Q'\left(X_1^{(0)}\right) \geq Q\left(X_1^{(0)}\right)\left(\mathcal{G}_1^{(2)} - 2\mathcal{G}_1^{(1)}\right)\Big/ X_1^{(0)} \mathcal{G}_1^{(2)}$, we have $\dfrac{\partial^2 \pi_1^{(0)}}{\partial p_1^o \partial X_1} \geq 0, \dfrac{\partial^2 \pi_1^{(0)}}{\partial C_1 \partial X_1} \leq 0$. Since

$Q'\left(X_2^{(0)}\right) \geq Q\left(X_2^{(0)}\right)\left(\mathcal{G}_2^{(2)} - 2\mathcal{G}_2^{(1)}\right)\Big/ X_2^{(0)} \mathcal{G}_2^{(2)}$ and $f'(\theta) \geq Th$, we have $\dfrac{\partial^2 \pi_2^{(0)}}{\partial p_1^o \partial X_2} \leq 0, \dfrac{\partial^2 \pi_2^{(0)}}{\partial C_1 \partial X_2} \geq 0$.

$\dfrac{\partial \pi_1^{(1)}}{\partial P_1} \leq 0, \dfrac{\partial \pi_2^{(1)}}{\partial P_2} \geq 0, \dfrac{\partial \pi_1^{(1)}}{\partial X_1} \geq 0$ can thus be derived from (4.10) by substituting the above

second order partial derivatives.

Case 4: scenarios a-2 and scenario b-2:

In this case, $\dfrac{\partial^2 \pi_1^{(0)}}{\partial p_1^o \partial P_1} \leq 0$, $\dfrac{\partial^2 \pi_1^{(0)}}{\partial C_1 \partial P_1} \geq 0$, $\dfrac{\partial^2 \pi_1^{(0)}}{\partial p_2^o \partial P_1} \geq 0$, $\dfrac{\partial^2 \pi_1^{(0)}}{\partial C_2 \partial P_1} \leq 0$, $\dfrac{\partial^2 \pi_2^{(0)}}{\partial p_1^o \partial P_2} \geq 0$, $\dfrac{\partial^2 \pi_2^{(0)}}{\partial C_1 \partial P_2} \leq 0$,

$\dfrac{\partial^2 \pi_2^{(0)}}{\partial p_2^o \partial P_2} \leq 0$, $\dfrac{\partial^2 \pi_2^{(0)}}{\partial C_2 \partial P_2} \geq 0$ can be proved similarly to that of case 2 and case 3. $\dfrac{\partial^2 \pi_1^{(0)}}{\partial p_1^o \partial X_1} \geq 0$,

$\dfrac{\partial^2 \pi_1^{(0)}}{\partial C_1 \partial X_1} \leq 0$, $\dfrac{\partial^2 \pi_1^{(0)}}{\partial p_2^o \partial X_1} = 0$, $\dfrac{\partial^2 \pi_1^{(0)}}{\partial C_2 \partial X_1} = 0$, $\dfrac{\partial^2 \pi_2^{(0)}}{\partial p_1^o \partial X_2} \leq 0$, $\dfrac{\partial^2 \pi_2^{(0)}}{\partial C_1 \partial X_2} \geq 0$, $\dfrac{\partial^2 \pi_2^{(0)}}{\partial p_2^o \partial X_2} \geq 0$, $\dfrac{\partial^2 \pi_2^{(0)}}{\partial C_2 \partial X_2} \leq 0$ can

be proved similarly to that of case 3. $\dfrac{\partial^2 \pi_1^{(0)}}{\partial p_1^o \partial X_1} \geq 0$ can thus be derived from (4.10) by

substituting the above second order partial derivatives.