The Public Option: a Non-regulatory Alternative to Network Neutrality

Richard T. B. Ma
Advanced Digital Sciences Center, Illinois at
Singapore and National University of Singapore

Vishal Misra
Columbia University

ABSTRACT

Network neutrality and the role of regulation on the Internet have been heavily debated in recent times. Amongst the various definitions of network neutrality, we focus on the one which *prohibits paid prioritization of content*. We develop a model of the Internet ecosystem in terms of three primary players: consumers, ISPs and content providers. We analyze this issue from the point of view of the consumer, and target the desired system state that maximizes *consumer surplus*.

By analyzing the various structures of an ISP market, we obtain different conclusions on the desirability of regulation. We also introduce the notion of a *Public Option ISP*, an ISP that carries traffic in a network neutral manner. We find (i) in a monopolistic scenario, network neutral regulations might benefit consumers; however, the introduction of a Public Option ISP is even better, as it aligns the interests of the monopolistic ISP with the consumer surplus and (ii) in an oligopolistic scenario, the presence of a Public Option ISP is again preferable to network neutral regulations, although the presence of competing non-neutral ISPs provides the most desirable situation for the consumers.

Lastly, the ISP survivability is an orthogonal direction towards the debate. Nevertheless, our findings reveal that even ISPs can survive, network neutrality might still not be needed.

1. INTRODUCTION

Since around 2005, network neutrality has been a hotly debated topic amongst law and policy makers. The core debate has centered around the argument whether ISPs should be allowed to provide service differentiation and/or user discrimination, with the notion of "user" being either content providers (CPs) or consumers. Proponents of network neutrality, mostly the CPs, have argued that the Internet has been

"neutral" since its inception and that has been a critical factor in the innovation and rapid growth that has happened on it. Opponents of network neutrality, mostly the ISPs, claim that without some sort of service differentiation, ISPs will lose the incentive to invest in the networks and the end user experience will suffer. Both camps implicitly or explicitly claim that their approach is beneficial for consumers. A recent Federal Communications Commission (FCC) vote [1] in the US has sided with the proponents, although the ruling leaves some room for service differentiation in wireless networks. The controversy rages on though with corporations like Verizon filing lawsuits challenging the ruling and a "toll-tax" dispute between Level3/Netflix and Comcast being cast as a network neutrality issue.

We study the issue explicitly from the consumer's point of view under both monopolistic and oligopolistic scenarios. A lot of arguments for as well as against network neutrality live in an idealized world where economies of scale do not exist and monopolies cannot emerge, and therefore perfect competition solves all problems. We believe reality is more nuanced and hence we examine monopolistic scenarios too. The ISP survivability is an equally important but orthogonal aspect of the debate. We focus on the extra ISP profit generated by prioritizing content. Nevertheless, our conclusion on the unnecessity of network neutrality do not rely on the survivability of ISPs: if an ISP cannot survive, network neutrality should not be imposed; even if an ISP can survive, we show that network neutrality is still not needed.

We use game theoretic analyses and focus on the consumer surplus, i.e. the consumer utility derived from the network minus their access costs. We model the user demand for content and the rate allocation mechanism of the network. The interplay between the two determines the rate equilibrium for traffic flows. Our model of price discrimination is for the ISPs to offer two classes of service to CPs. The ISP divides its capacity into a premium and an ordinary class, and CPs get charged extra for sending traffic in the premium class, and more details are presented in Section 3.1. We then identify and analyze the strategic games played between ISPs, CPs and consumers in Section 3 for a monopolistic scenario and in Section 4 for oligopolistic scenarios. In Section 4.1, we introduce the notion of a *Public Option*

ISP which is neutral to all CPs. The Public Option ISP can be implemented by processes like local loop unbundling [3] in a monopolistic market and either government or a private organization could run the ISP and still be profitable [12]. Under this framework, our findings include:

- The impact of network neutrality on consumer surplus depends on the nature of competition at the ISP level. Concretely, a neutral network might be beneficial for consumers under a monopolistic regime (Section 3), whereas a non-neutral network is advantageous for consumers under oligopolistic scenarios (Section 4).

- Introducing a Public Option ISP is advantageous for consumers. In a monopolistic situation, the Public Option ISP offers the best scenario for consumers (Theorem 5), followed by network neutral regulations, and an unregulated market being the worst.

- In oligopolistic situations, the Public Option ISP is still preferable to network neutral regulations; however, since the incentive for an ISP to gain market share is aligned with maximizing consumer surplus (Theorem 6), no regulation is needed to protect the consumers.

- Under an oligopolistic competition, any ISP's optimal pricing and service differentiation strategy, whether network neutral or not, will be close to the one that maximizes consumer surplus (Theorem 6 and Corollary 1). Moreover, under a probable equilibrium where ISPs use homogenous strategies, their market shares will be proportional to their capacities (Lemma 4), which implies that ISPs do have incentives to invest and expand capacity so as to increase their market shares.

Our paper sheds new light on the network neutrality debate and concretely identifies where and how regulation can help. Additionally, our identification of the Public Option ISP is especially important as it provides a solution that combines the best of both worlds, protecting consumer interests without enforcing strict regulations on all ISPs. We start with describing our model in the next section.

2. THREE-PARTY ECOSYSTEM MODEL

We consider a model of the Internet with three parties: 1) CPs, 2) ISPs and 3) consumers. We focus on a fixed consumer group in a targeted geographic region. We denote M as the number of consumers in the region[1]. Each consumer subscribes to an Internet access service via an ISP. We consider the scenarios where one monopolistic ISP I or a set $\mathcal{I}$ of competing oligopolistic ISPs provide the Internet access for the consumers. We denote $\mathcal{N}$ as the set of

---

[1]M can also be interpreted as the *average* or *peak* number of consumers accessing the Internet simultaneously, which will scale with the total number of actual consumers in a region. This does not change the nature of any of the results we describe subsequently, but gives a more realistic interpretation of the rate equilibrium.

CPs from which the consumers request content. We define $N = |\mathcal{N}|$ as the number of CPs. Our model does not include the backbone ISPs for two reasons. First, the bottleneck of the Internet is often at the last-mile connection towards the consumers [9], both wired and wireless. We focus on the regional or so-called *eyeball ISPs* that provide the bottleneck last-mile towards the consumers. Second, the recent concern on network neutrality manifests itself in the cases where the last-mile ISPs, e.g. France Telecom, Telecom Italia and Vodafone, intended to differentiate services and charge CPs, e.g. Apple and Google, for service fees [5].

We denote μ as the last-mile bottleneck capacity towards the consumers in the region. Figure 1 depicts the contention
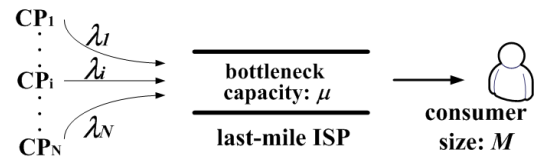

Figure 1: Contention at the last-mile bottleneck link.

at the bottleneck among different flows from the CPs. Given a set $\mathcal{N}$ of CPs, a group of M consumers and a link with capacity μ, we denote the system as a triple $(M, \mu, \mathcal{N})$. We denote λ_i as the aggregate throughput rate from CP i to the consumers. Because consumers initiate downloads and retrieve content from the CPs, we first model the consumer's demand so as to characterize the CPs' throughput rates λ_is.

2.1 Consumer Throughput Demand

We denote $\hat{\theta}_i$ as the *unconstrained throughput* for a typical user of CP i. For instance, the unconstrained throughput for the highest quality Netflix streaming movie is about 5 Mbps [4], and given an average query page of 20 KB and an average query response time of .25 seconds [2], the unconstrained throughput for a Google search is about 600 Kbps, or just over $1/10^{th}$ of Netflix. We denote $\alpha_i \in (0, 1]$ as the percentage of consumers that ever access CP i's content, which models the popularity of the content of CP i. We define $\hat{\lambda}_i = \alpha_i M \hat{\theta}_i$ as the unconstrained throughput of CP i. Without contention, CP i's throughput λ_i equals $\hat{\lambda}_i$. However, when the capacity μ cannot support the unconstrained throughput from all CPs, i.e. $\mu < \sum_{i \in \mathcal{N}} \hat{\lambda}_i$, two things will happen: 1) a typical user of CP i obtains throughput $\theta_i < \hat{\theta}_i$ from CP i, and 2) some active users might stop downloading content from CP i when θ_i goes below certain threshold, e.g. users of streaming content like Netflix. We denote θ_i as the *achievable throughput* for the consumers downloading content from CP i. We define a demand function $d_i(\theta_i)$ which represents the percentage of consumers that still demand content from CP i under the achievable throughput θ_i.

Assumption 1. *For any CP i, the demand $d_i(\cdot)$ is a non-negative, continuous and non-decreasing function defined on the domain of $[0, \hat{\theta}_i]$, and satisfies $d_i(\hat{\theta}_i) = 1$.*

We define the aggregate throughput of a set $\mathcal{N}$ of CPs as $\lambda_{\mathcal{N}} = \sum_{i \in \mathcal{N}} \lambda_i$, where each CP i's rate λ_i is defined as:

$$\lambda_i(\theta_i) = \alpha_i M d_i(\theta_i)\theta_i. \quad (1)$$

2.2 Rate Allocation Mechanism

When multiple flows share the same bottleneck link, they compete for capacity. The rates allocated to the flows depend on the rate allocation mechanism being used in the system.

Definition 1. *A* rate allocation mechanism *is a function that maps any fixed demand profile* $\{d_i : i \in \mathcal{N}\}^2$ *to an achievable throughput profile* $\{\theta_i : i \in \mathcal{N}\}$.

A rate allocation mechanism can be a flow control mechanism, e.g. CBR and VBR mechanisms, under which the bottleneck link decides the rates for each flow in a centralized manner, or a window-based end-to-end congestion control mechanism, e.g. TCP, under which each flow maintains a sliding window and adapts the its size based on implicit feedback from the network, e.g. the round-trip delay. We consider generic rate allocation mechanisms and assume that the resulting rate allocation obeys the physical constraints of the system and satisfies some desirable properties.

Axiom 1. $\theta_i \leq \hat{\theta}_i$ *for all* $i \in \mathcal{N}$.

Axiom 2. $\lambda_{\mathcal{N}} = \min\left\{\mu, \sum_{i \in \mathcal{N}} \hat{\lambda}_i\right\}$.

The above axioms characterize the feasibility of an allocation: the aggregate rate cannot exceed the capacity and the individual rate cannot exceed unconstrained throughput. It also characterizes a *work-conserving* property: if congestion can be alleviated without increasing capacity μ, the allocation would do so by fully utilizing the capacity.

Axiom 3 (MONOTONICITY). *A rate allocation is* monotonic *if for any* $M > 0$ *and capacity* $\mu_1 < \mu_2$, *the achievable throughput* θ_i *for any* $i \in \mathcal{N}$ *satisfies*

$$\theta_i(M, \mu_2, \mathcal{N}) \geq \theta_i(M, \mu_1, \mathcal{N}).$$

The Monotonicity axiom implies that if a flow gets an achievable throughput in one system, it will get at least that amount of throughput under a less congested system.

Axiom 4 (INDEPENDENCE OF SCALE). *A rate allocation is* independent of scale *if for any* $\xi > 0$, *the achievable throughput* θ_i *for any* $i \in \mathcal{N}$ *satisfies*

$$\theta_i(M, \mu, \mathcal{N}) = \theta_i(\xi M, \xi \mu, \mathcal{N}).$$

The Independence of Scale axiom states that if the capacity scales at the same rate as the consumer size, each flow's achievable throughput θ_i remains the same.

Assumption 2. *The network system implements a rate allocation mechanism that satisfies Axiom 1 to 4.*

---

²Without a bracket, we use d_i as a fixed demand.

2.3 Rate Equilibrium

The demand functions map the achievable throughput to a level of demand; the rate allocation mechanisms map fixed demands to achievable throughput. The interplay between a rate allocation mechanism and the demand functions determines the system rate equilibrium as the following theorem.

Theorem 1. *A system* $(M, \mu, \mathcal{N})$ *has a unique rate equilibrium* $\{\theta_i : i \in \mathcal{N}\}$ *under Assumption 1 and Axiom 1 to 3.*

We define $\nu = \mu/M$ as the per capita capacity of the system. By using Axiom 4, we further characterize the rate equilibrium $\{\theta_i : i \in \mathcal{N}\}$ as follows.

Lemma 1. *Under Assumption 1 and 2, for all* $i \in \mathcal{N}$, θ_i *in equilibrium can be expressed as* $\theta_i(M, \mu, \mathcal{N}) = \theta_i(\nu, \mathcal{N})$, *which is a non-decreasing and continuous function in* ν.

Lemma 1 states that when ν increases, users' achievable throughput θ_i would not be worse off for any CP i.

We define the consumer surplus as $CS = \sum_{i \in \mathcal{N}} \phi_i \lambda_i$, where ϕ_i denotes the per unit traffic surplus (utility minus cost) that the consumers obtain by receiving content from CP i. This surplus can be derived from communicating with friends, e.g. Skype, watching movies, e.g. Netflix, and obtaining information, e.g. Google. Notice that the single parameter ϕ_i is a linear model for user surplus, which might over/under-estimate the real utility; however, we can always adjust the demand functions $d_i(\cdot)$s, which still hold the monotonicity property in Assumption 1, to compensate the difference between the assumed linear utility and the real non-linear utility. Although these utility-adjusted demand functions would modify the real rate equilibrium, they do not invalidate our theoretical results at a macroscopic level. We denote Φ as the per capita consumer surplus defined as

$$\Phi = \frac{CS}{M} = \frac{1}{M} \sum_{i \in \mathcal{N}} \phi_i \lambda_i(\theta_i) = \sum_{i \in \mathcal{N}} \phi_i \alpha_i d_i(\theta_i)\theta_i. \quad (2)$$

Theorem 2. *Under Assumption 1 and 2, the per capita consumer surplus* Φ *can be expressed as* $\Phi(M, \mu, \mathcal{N}) = \Phi(\nu, \mathcal{N})$, *which is non-decreasing function in* ν. *In particular, it strictly increases in* $\nu \in [0, \sum_{i \in \mathcal{N}} \alpha_i \hat{\theta}_i]$.

Theorem 2 states that the per capita consumer surplus will strictly increase with the system per capita capacity ν, unless it is already maximized when unconstrained throughput is obtained. Notice that it does not depend on the values of ϕ_is, but relies on the monotonic traffic demand (Assumption 1) and the monotonic (Axiom 3) and work-conserving (Axiom 2) properties of the rate allocation mechanism.

2.4 Examples and Illustrations

2.4.1 Demand as a function of throughput sensitivity

Users often have different demand patterns for different CPs. For example, the demand for real-time applications decreases dramatically when their throughput drops below

certain threshold where performance cannot be tolerated by users. We can characterize this throughput sensitivity by a positive parameter β_i and consider the demand function

$$d_i(\theta_i) = e^{-\beta_i\left(\frac{\hat{\theta}_i}{\theta_i}-1\right)} = e^{-\beta_i\left(\frac{1}{\omega_i}-1\right)}, \qquad (3)$$

where we define $\omega_i = \theta_i/\hat{\theta}_i$ as the percentage of unconstrained throughput achieved for CP i. The user demand decays exponentially with the level of congestion (measured by $\frac{\hat{\theta}_i-\theta_i}{\theta_i}$, the ratio of unsatisfied and achieved throughput) scaled by β_i. This demand function distinguishes the CPs via their throughput sensitivity β_i: larger β_i indicates higher sensitivity to throughput for CP i. Figure 2 illustrates the
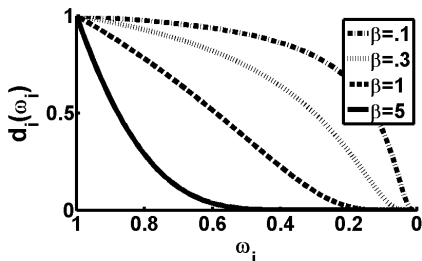

Figure 2: Demand function $d_i(\omega_i)$.

demand functions with various values of β_i. To normalize $\hat{\theta}_i$, we plot d_i against ω_i instead of θ_i. We observe that when throughput drops linearly, the demand drops sharply for large β_i, e.g. when $\beta_i = 5$, the demand is halved with a 10% drop in throughput from $\hat{\theta}_i$. Large β_is can be used to model CPs that have stringent throughput requirements, e.g. Netflix; while, small β_is can be used to model CPs that are less sensitive to throughput, e.g. a Google search query.

2.4.2 *End-to-end congestion control mechanisms*

Due to the end-to-end design principle of the Internet, congestion control has been implemented by window-based protocols, i.e. TCP and its variations. Mo and Walrand [21] showed that a class of α-proportional fair solutions[3] can be implemented by window-based end-to-end protocols. Among the class of α-proportional fair solutions, the max-min fair allocation, a special case with $\alpha = \infty$, is the result of the AIMD mechanism of TCP [7]. Differing round trip times, receiver window sizes and loss rates can result in different bandwidths, but to a first approximation, TCP provides a max-min fair allocation of available bandwidth amongst flows. Although other protocols, e.g. UDP, coexist in the Internet, recent research [15] sees a growing concentration of application traffic, especially video, over TCP.

We illustrate the rate allocation under the max-min fair mechanism using an example of three CPs with demand functions of Equation (3) and parameters $(\alpha_1, \hat{\theta}_1, \beta_1) = (1, 1, 0.1)$, $(\alpha_2, \hat{\theta}_2, \beta_2) = (0.3, 10, 3)$ and $(\alpha_3, \hat{\theta}_3, \beta_3) = (0.5, 3, 5)$.

---

[3]Any α-proportional fair solution also satisfies Assumption 2.

CP 1 represents Google-type of CPs that are extensively accessed and less sensitive to throughput. CP 2 represents Netflix-type of CPs that are more throughput-sensitive and have high unconstrained throughput. CP 3 represents Skype-type of CPs that are extremely sensitive to throughput and have medium unconstrained throughput. Figure 3 illustrates
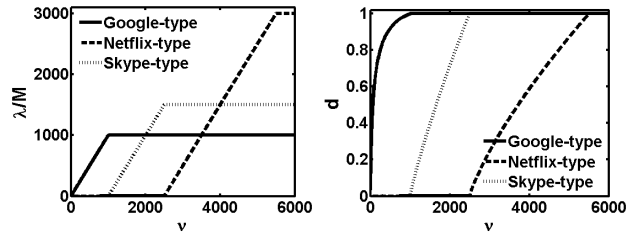

Figure 3: Throughput under max-min fair mechanism.

the rates and the corresponding demands of the three CPs under a max-min fair allocation mechanism. We vary the per capita capacity ν from 0 to $6,000$. We observe that when ν increases from zero, the demand for Google-type content increases first, followed by the demand for Skype-type content and, the demand for Netflix-type content being the last.

3. MONOPOLISTIC ISP ANALYSIS

In this section, we start with the scenario where the last-mile capacity is controlled by a single monopolistic ISP I. We analyze the ISP's strategy under which non-neutral service differentiation is allowed, and the corresponding best responses of the CPs. We derive the equilibria of the system and analyze the ISP's impact on the system congestion and the welfare of the consumers and the ISP itself.

3.1 Non-Neutral Service Differentiation

We assume that the monopolistic last-mile ISP I has a capacity of μ. This ISP can be a retail residential ISP, e.g. Comcast, and Time Warner Cable, or a mobile operator, e.g. Verizon and AT&T. Regardless being a wired or wireless provider, it serves as the last-mile service provider for the consumers. We assume that the ISP is allowed to allocate a fraction $\kappa \in [0, 1]$ of its capacity to serve premium CPs and charge them for an extra rate $c \in [0, \infty)$ (dollar per unit traffic), besides ordinary transit charges. For a wired ISP, κ can be interpreted as the percentage of capacity deployed for private peering that charge c per unit incoming traffic and $1 - \kappa$ can be interpreted as the percentage of capacity deployed for public peering where incoming traffic is charge-free. For a wireless ISP, κ can be interpreted as the percentage of capacity devoted for prioritized traffic. The pair of parameters (κ, c) can also be thought of a type of Paris Metro Pricing (PMP) [23, 24], where an ordinary and a premium class have capacities of $(1 - \kappa)\mu$ and $\kappa\mu$ and charge 0 and c respectively. In reality, content might be delegated via content distribution networks (CDNs), e.g. Akamai, or backbone ISPs, e.g. Level3 is a major tier-1 ISP that delivers Netflix traffic

towards regional ISPs. Therefore, the extra charge c might be imposed on the delivering ISP, e.g. Level3, and then be recouped from the CP, e.g. Netflix, by its delivering ISP. Our model does not assume any form of the implementation.

We denote $\mathcal{O}$ and $\mathcal{P}$ as the disjoint sets of CPs that join the ordinary and premium class respectively. We denote v_i as CP i's per unit traffic profit. This profit can be generated by advertising for media clients, e.g Google, by selling online, e.g. Amazon, or by providing online services, e.g. Netflix and other e-commerce. Our model does not assume how the profit is generated. We define each CP i's utility u_i as

$$u_i(\lambda_i) = \begin{cases} v_i\lambda_i & \text{if } i \in \mathcal{O}, \\ v_i\lambda_i - c\lambda_i & \text{if } i \in \mathcal{P}. \end{cases} \quad (4)$$

We define $IS = c\lambda_{\mathcal{P}}$ as the ISP surplus (extra revenue from CPs) and denote Ψ as the per capita ISP surplus defined as

$$\Psi = \frac{IS}{M} = \frac{c}{M}\lambda_{\mathcal{P}} = \frac{c}{M}\sum_{i\in\mathcal{P}}\lambda_i(\theta_i) = c\sum_{i\in\mathcal{P}}\alpha_i d_i(\theta_i)\theta_i.$$

Notice that our focus is the additional ISP revenue earned by providing a differentiated service. The ISP surplus does not reflect the ISP's normal operating costs or their core revenue from the subscription payments from their residential users.

3.2 Content Provider's Best Response

Given the ISP's decision κ and c, each CP chooses the service class, $\mathcal{O}$ or $\mathcal{P}$, to join. We denote ρ_i as the per capita throughput over CP i's user base, i.e. $\alpha_i M$ users, defined as

$$\rho_i(\nu, \mathcal{N}) = d_i(\theta_i(\nu, \mathcal{N}))\theta_i(\nu, \mathcal{N}). \quad (5)$$

Lemma 2. *Given a fixed set $\mathcal{O}$ of CPs in the ordinary class and a fixed set $\mathcal{P}$ of CPs in the premium class, a new CP i's optimal strategy is to join the premium service class, if*

$$(v_i - c)\,\rho_i\big(\kappa\nu, \mathcal{P}\cup\{i\}\big) \geq v_i\,\rho_i\big((1-\kappa)\nu, \mathcal{O}\cup\{i\}\big). \quad (6)$$

And with equality, both service classes gives the same utility.

Lemma 2 states that a CP will join the premium service class if that results higher profit, which is per-unit flow profit ($v_i - c$ for the premium class) multiplied by the per capita throughput ρ_i. The above decision is clear for a CP only if all other CPs have already made their choices. To treat all CPs equally, we model the decisions of all CPs as a simultaneous-move game as part of a two-stage game.

3.3 Two-Stage Strategic Game

We model the strategic behavior of the ISP and the CPs as a two-stage game, denoted as a quadruple $(M, \mu, \mathcal{N}, I)$.

1. *Players*: The monopolistic ISP I and the set of CPs $\mathcal{N}$.

2. *Strategies*: ISP I chooses a strategy $s_I = (\kappa, c)$. Each CP i chooses a binary strategy of whether to join the premium class. The CPs' strategy profile can be written as $s_{\mathcal{N}} = (\mathcal{O}, \mathcal{P})$, where $\mathcal{O}\cup\mathcal{P} = \mathcal{N}$ and $\mathcal{O}\cap\mathcal{P} = \emptyset$.

3. *Rules*: In the first stage, ISP I decides $s_I = (\kappa, c)$ and announces it to all the CPs. In the second stage, all the CPs make their binary decisions simultaneously and reach a joint decision $s_{\mathcal{N}} = (\mathcal{O}, \mathcal{P})$.

4. *Outcome*: The set $\mathcal{P}$ of the CPs shares a capacity of $\kappa\mu$ and the set $\mathcal{O}$ of the CPs shares a capacity of $(1-\kappa)\mu$. Each CP $i \in \mathcal{O}$ gets a rate λ_i in $(M, (1-\kappa)\mu, \mathcal{O})$ and each CP $j \in \mathcal{P}$ gets a rate λ_j in $(M, \kappa\mu, \mathcal{P})$.

5. *Payoffs*: Each CP i's payoff is defined by the utility $u_i(\lambda_i)$ in Equation (4). The ISP's payoff is its surplus $IS = c\lambda_{\mathcal{P}}$ received from the premium class.

Notice that as a consequence of service differentiation, the original system $(M, \mu, \mathcal{N})$ breaks into two independent sub-systems $(M, (1-\kappa)\mu, \mathcal{O})$ and $(M, \kappa\mu, \mathcal{P})$. In practice, if the premium service class is under utilized, i.e. $\lambda_{\mathcal{P}} < \kappa\mu$, and if the ISP uses a work-conserving mechanism so that the extra capacity $\kappa\mu - \lambda_{\mathcal{P}}$ in $\mathcal{P}$ would be used by ordinary class, then equivalently, we can think of the ISP's strategy as setting an effective κ that equals $1 - \lambda_{\mathcal{P}}/\mu$, or virtually restricting the domain of κ to be upper-bounded by some value less than 1. Effectively, it limits the level of service differentiations and avoids the ordinary class being made a damaged good [11].

If we regard the set of CPs as a single player that chooses a strategy $s_{\mathcal{N}}$, our two-stage game can be thought as a Stackelberg game [20], where the first-mover ISP can take the best-responses of the CPs into consideration and derive its optimal strategy s_I using *backward induction* [20]. Given any fixed strategy $s_I = (\kappa, c)$, the CPs derive their best strategies under a simultaneous-move game, denoted as $(M, \mu, \mathcal{N}, s_I)$. We denote $s_{\mathcal{N}}(M, \mu, \mathcal{N}, s_I) = (\mathcal{O}, \mathcal{P})$ as a strategy profile of the CPs under the game $(M, \mu, \mathcal{N}, s_I)$. Technically speaking, when $\kappa = 0$ or 1, there is only one service class. When $\kappa = 0$, we define the trivial strategy profile as $s_{\mathcal{N}} = (\mathcal{N}, \emptyset)$; when $\kappa = 1$, although there is not a physical ordinary class, we define the trivial strategy profile as $s_{\mathcal{N}} = (\mathcal{O}, \mathcal{N}\backslash\mathcal{O})$, with $\mathcal{O} = \{i : v_i \leq c, i \in \mathcal{N}\}$ which defines the set of CPs that cannot afford to join the premium class. Based on Lemma 2, we can define an equilibrium in the sense of a Nash or competitive equilibrium. To break a tie, we assume that a CP always chooses to join the ordinary service class when both classes provide the same utility.

Definition 2. *A strategy profile $s_{\mathcal{N}} = (\mathcal{O}, \mathcal{P})$ is a Nash equilibrium of a game $(M, \mu, \mathcal{N}, s_I)$, if*

$$\frac{v_i - c}{v_i} \begin{cases} \leq \rho_i\big((1-\kappa)\nu, \mathcal{O}\big)/\rho_i\big(\kappa\nu, \mathcal{P}\cup\{i\}\big) & \text{if } i \in \mathcal{O}, \\ > \rho_i\big((1-\kappa)\nu, \mathcal{O}\cup\{i\}\big)/\rho_i\big(\kappa\nu, \mathcal{P}\big) & \text{if } i \in \mathcal{P}. \end{cases}$$

3.4 Competitive Equilibrium

Notice that a CP's joining decision to a service class might increase the congestion level and reduce the throughput of flows of that service class; however, if the number of CPs in a service class is big and no single CP's traffic will dominate,

an additional CP i's effect will be marginal. Analogous to the *pricing-taking assumption* [20] in a competitive market, we make a *throughput-taking assumption* as follows.

Assumption 3. *Any CP $i \notin \mathcal{N}$ makes an estimate $\tilde{\rho}_i(\nu, \mathcal{N})$ on its ex-post per capita throughput $\rho_i(\nu, \mathcal{N} \cup \{i\})$ in the decision-making under a competitive equilibrium.*

Based on the above throughput-taking assumption, we can define a competitive equilibrium of the CPs as follows.

Definition 3. *A strategy profile $s_{\mathcal{N}} = (\mathcal{O}, \mathcal{P})$ is a competitive equilibrium of a game $(M, \mu, \mathcal{N}, s_I)$, if*

$$\frac{v_i - c}{v_i} \begin{cases} \leq \rho_i\big((1-\kappa)\nu, \mathcal{O}\big)/\tilde{\rho}_i\big(\kappa\nu, \mathcal{P}\big) & \text{if } i \in \mathcal{O}, \\ > \tilde{\rho}_i\big((1-\kappa)\nu, \mathcal{O}\big)/\rho_i\big(\kappa\nu, \mathcal{P}\big) & \text{if } i \in \mathcal{P}. \end{cases} \quad (7)$$

The competitive equilibrium depends on how each CP i calculates $\tilde{\rho}_i = d_i(\tilde{\theta}_i)\tilde{\theta}_i$, which boils down to an estimation of the ex-post throughput $\tilde{\theta}_i$. This estimation depends on the rate allocation mechanism being used. For example, under the max-min fair mechanism, CP i can expect an achievable throughput of $\theta_{\mathcal{N}} = \max\{\theta_j : j \in \mathcal{N}\}$. Thus, CP i can take this *throughput* as given and estimate that $\tilde{\theta}_i = \min\{\hat{\theta}_i, \theta_{\mathcal{N}}\}$. The competitive equilibrium under the throughput-taking assumption can be regarded as a special type of *congestion equilibrium* [18], where the throughput of the CPs indicates the level of congestion in system.

In practice, because CPs rarely know the characteristics of all other CPs, the *common knowledge* assumption [20] for reaching Nash equilibria might be questionable. Thus, we use competitive equilibria for numerical evaluations[4]. Although the throughput-taking assumption might not be valid if one of the CPs' has significant percentage of traffic, our results do not depend on the underlying equilibrium type, and apply for both equilibrium definitions. In the rest of the paper, unless we specifically indicate an equilibrium to be Nash (Definition 2) or competitive (Definition 3), we use the term *equilibrium* to indicate both.

Theorem 3. *If $s_{\mathcal{N}} = (\mathcal{O}, \mathcal{P})$ is an equilibrium of a game $(M, \mu, \mathcal{N}, s_I)$, it is also a same type of equilibrium (Nash or competitive) of a game $(\xi M, \xi \mu, \mathcal{N}, s_I)$ for any $\xi > 0$.*

Although a game $(M, \mu, \mathcal{N}, s_I)$ might have multiple equilibria, we do not assume that it reaches a particular equilibrium. However, to make equilibria under the same per capita capacity ν consistent, we make the following assumption.

Assumption 4. *If $s_{\mathcal{N}} = (\mathcal{O}, \mathcal{P})$ is a realized equilibrium of a game $(M, \mu, \mathcal{N}, s_I)$, then it is also the realized equilibrium of the linearly scaled game $(\xi M, \xi \mu, \mathcal{N}, s_I)$ for any $\xi > 0$.*

The above assumption implies that when the ISP scales its capacity μ linearly and smoothly with its consumer size M, the CPs will not diverge abruptly into another equilibrium, if there exists any. For the game $(M, \mu, \mathcal{N}, s_I)$ with strategy

$s_I = (\kappa, c)$ and equilibrium $s_{\mathcal{N}} = (\mathcal{O}, \mathcal{P})$, the per capita consumer surplus Φ is a function of ν, written as

$$\Phi(M, \mu, \mathcal{N}, s_I) = \Phi(\nu, \mathcal{N}, s_I) = \Phi((1-\kappa)\nu, \mathcal{O}) + \Phi(\kappa\nu, \mathcal{P}).$$

Under the above assumption, the per capita consumer and ISP surplus will remain the same in linearly scaled games $\{(\xi M, \xi \mu, \mathcal{N}, s_I) : \xi > 0\}$ in equilibrium.

Lemma 3. *Under Assumption 4, the per capita consumer surplus Φ satisfies*

$$\Phi(\nu, \mathcal{N}, s_I) = \Phi(\xi M, \xi \mu, \mathcal{N}, s_I), \quad \forall \xi > 0.$$

The above is true for the per capita ISP surplus Ψ as well.

3.5 Monopolistic ISP's Strategy

In order to increase surplus, the ISP's optimal strategy would encourage more CPs to join its premium service class.

Theorem 4. *In the game $(M, \mu, \mathcal{N}, I)$, for any $0 \leq c < 1$, strategy $s_I = (\kappa, c)$ is always dominated by $s_I^1 = (1, c)$. If $\lambda_{\mathcal{P}} < \min\{\mu, \sum_{v_i \geq c} \hat{\lambda}_i\}$, s_I is strictly dominated by s_I^1. $s_I = (\kappa, c)$ is also dominated by $s_I' = (\kappa', c)$ with $\kappa' > \kappa$, if equilibrium $(\mathcal{O}', \mathcal{P}')$ under s_I' satisfies $\mathcal{P} \subseteq \mathcal{P}'$.*

When the monopoly ISP increases κ, it improves the condition in the premium service class and in a new equilibrium, $\mathcal{P}'$ would attract more CPs to join than $\mathcal{P}$. Theorem 4 states that the ISP would have incentives to increase κ so as to maximize revenue. The effect of increasing κ is twofold: 1) more capacity is allocated to the premium class for sale, and 2) the reduced capacity in the ordinary class makes more CPs switch to the premium class. As a result, one of the optimal strategies of the monopolistic ISP is to always set $\kappa = 1$. This implies that, if allowed, the selfish ISP will only provide a charged service class $\mathcal{P}$ without contributing any capacity for the ordinary class $\mathcal{O}$. Suppose the ISP is allowed to set $\kappa = 1$, we first study its optimal price c and its impact on the consumer and ISP surplus.

We use the demand function of Equation (3) and the max-min fair mechanism for our numerical simulations. We study a scenario of 1000 CPs, whose α_i, $\hat{\theta}_i$ and v_i are uniformly distributed within $[0, 1]$ and β_i is uniformly distributed within $[0, 10]$. To satisfy all unconstrained throughput for the CPs, the per capita capacity needs to be around $\nu = 250$. Since throughput-sensitive applications, e.g. Skype, bring more utility to consumers in reality, we consider the consumer utility ϕ_i that is uniformly distributed within $[0, \beta_i]$ (the uniform distribution biases utility towards CPs with high throughput sensitivity while introducing some randomness)[5]. Figure 4 plots Ψ and Φ versus the pricing strategy c when ν ranges from 20 to 200. We observe three pricing regimes.

1. When c is small, Ψ increases linearly, i.e. $\Psi = c\nu$. This happens when most of the CPs can afford to join

---

[4]Please refer to [18] for evaluating a competitive equilibrium.

[5]More experiments that confirm our results can be found in [19]. Although the parameters do not come from real CPs, our purpose is to show qualitative trends in general. Our theoretical results do not come from or depend on the particular setting of the experiments.
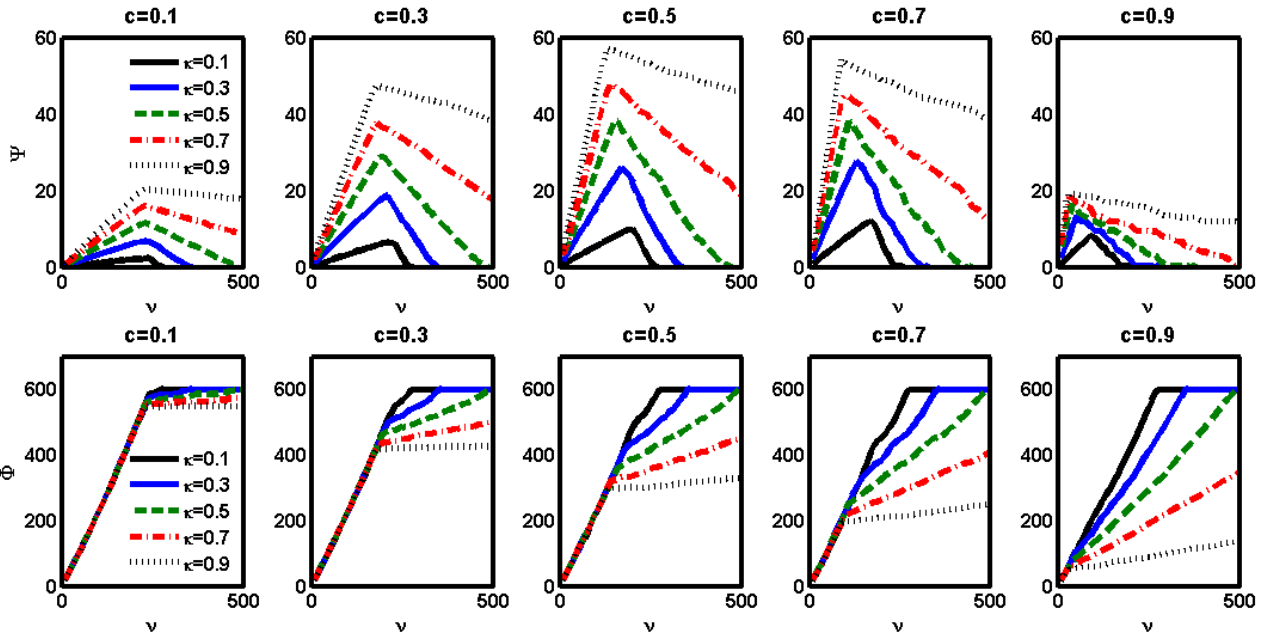
Figure 5: Per capita surplus Ψ and Φ under various strategies $s_I = (\kappa, c)$ versus per capita capacity ν.
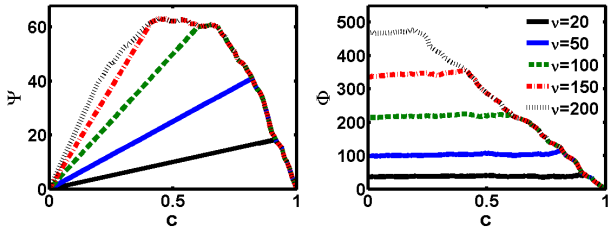

Figure 4: Per capita surplus Ψ and Φ under $\kappa = 1$.

the service and the entire capacity is fully utilized, i.e. $\lambda_\mathcal{P} = \mu$, resulting a high level of consumer surplus Φ.

2. When c is large, Ψ drops sharply. This happens when only a small set of CPs can afford to join the service and the capacity is largely under-utilized, i.e. $\lambda_\mathcal{P} < \mu$, resulting a sharp drop in Φ accordingly.

3. When ν is abundant, e.g. $\nu = 200$, there exists a pricing region where Ψ increases sub-linearly and Φ decreases. Consequently, the ISP's optimal strategy ($c \approx 0.45$) could intentionally keep more CPs away from the (only) service class and under-utilizes the capacity, which hurts the consumer surplus Φ.

Figure 5 illustrates Ψ and Φ under various strategies $s_I = (\kappa, c)$ versus ν ranging up to 500, which doubles the required capacity to serve all unconstrained throughput. For a fixed c in each column, we identify three equilibrium regimes.

1. When ν is small, Ψ increases linearly and Φ increases accordingly. This happens when the premium class capacity is fully utilized, i.e. $\lambda_\mathcal{P} = \kappa\mu$.

2. When ν keeps increasing, Ψ starts to decrease and Φ increases at a much slower rate. This happens when the premium class capacity is not fully utilized, i.e. $\lambda_\mathcal{P} < \kappa\mu$, and more CPs move from $\mathcal{P}$ to $\mathcal{O}$.

3. When ν is large, Ψ drops to zero for small values of κ, where Φ is maximized. This happens when $\mathcal{P} = \emptyset$ and $\mathcal{O}$'s capacity is abundant enough to serve all CPs' unconstrained throughput. However, if κ is big, e.g. $\kappa = 0.9$, it guarantees some revenue for the ISP, but reduces the consumer surplus from its maximum.

Further, under a fixed ν, we observe that higher κ induces higher revenue for the ISP (Theorem 4), even if that results in an under-utilization of the premium class capacity and hurts the consumer surplus. When comparing different prices c, we observe that larger values of c make the premium class becomes under-utilized faster, because fewer CPs can afford to join the premium class when necessary. However, when reaching the turning point where congestion starts to be relieved, κ plays a major role, under which Φ's rate of increase depends on the amount $(1 - \kappa)\mu$ of capacity allocated to $\mathcal{O}$.

Regulatory Implications: In the monopolistic scenario, the ISP would maximize κ for the charged service (Theorem 4).

In the case where the system capacity is abundant, i.e. large values for ν, the ISP would provide more capacity for the premium class than needed, making the premium capacity under-utilized. It also implies that the ordinary service class would be made a *damaged good* [11], where the ISP would have the incentive to degrade service quality, avoid network upgrades or investments for the non-charged service class. Consequently, the consumer surplus is greatly

hurt by the ISP's selfish interest. To remedy this problem, the network neutrality principle should be imposed to some extent to protect consumer surplus. In other words, the non-neutral service differentiation should be limited. The bottom line is that capacity under-utilization should be avoided, which implies that non-work-conserving policies of the ISP should not be allowed. Technically speaking, by imposing a work-conserving policy, we put an upper-bound $\kappa(c)$ for the capacity of the premium class, which can be expressed as a function of the price c. Effectively, the ordinary class would obtain $(1 - \kappa(c))\mu$ amount of capacity.

In the case where the system capacity is scarce, i.e. small values for ν, or under a work-conserving policy, although the system capacity would not be under-utilized, whether the ISP's pricing strategy is beneficial for consumer surplus is still uncertain. In general, an ISP would prefer to set a high price c so as to obtain high surplus $c\lambda_{\mathcal{P}}$ from the premium class. Therefore, the consumer surplus depends on whether the CPs in the premium class would provide higher utility for the consumers, i.e. high ϕ_i values for all $i \in \mathcal{P}$. On the one hand, if the price c is too high, it might limit/reject incubative CPs that are potentially beneficial for the consumers, but not yet profitable (low values of v_i). On the other hand, without enough price differentiation, more useful and probably more profitable CPs cannot provide better services so as to increase the consumer surplus. In the next section, we will show that the problem can be solved by introducing a so-called Public Option ISP for ISP competition.

4. OLIGOPOLISTIC ISP ANALYSIS

In the previous section, we concentrated on a monopolistic ISP I with capacity μ and a strategy $s_I = (\kappa, c)$. In this section, we extend our model to a set $\mathcal{I}$ of oligopolistic ISPs, each $I \in \mathcal{I}$ of which has a capacity μ_I and uses a strategy $s_I = (\kappa_I, c_I)$. We define $\mu = \sum_{I \in \mathcal{I}} \mu_I$ as the total system capacity and $\gamma_I = \mu_I/\mu$ as the capacity share of ISP I. Our oligopolistic models have two major differences from the monopolistic model. First, since consumers connect to the Internet via one of the ISPs, they might make strategic decisions on which ISP to subscribe to. We denote M_I as the consumer size of ISP I, where $\sum_{I \in \mathcal{I}} M_I = M$, and $m_I = M_I/M$ as its market share. Second, besides the ISP surplus, a more important objective of any ISP I is to maximize its market share m_I. This is because the core revenue of the last-mile ISPs relies on the subscription payments of the users and the market share is also what the last-mile ISPs can leverage to generate the CP-side revenue in the first place.

Similar to the monopolistic ISP game $(M, \mu, \mathcal{N}, I)$, we denote $(M, \mu, \mathcal{N}, \mathcal{I})$ as the two-stage oligopolistic ISP game, under which the set of ISPs $\mathcal{I}$ choose their strategies $s_{\mathcal{I}} = \{s_I : I \in \mathcal{I}\}$ simultaneously in the first stage, and then the set of CPs $\mathcal{N}$ and the M consumers make their strategic decisions simultaneously in a second-stage game $(M, \mu, \mathcal{N}, s_{\mathcal{I}})$. In the second-stage game, we denote $s_M = \{M_I : I \in \mathcal{I}\}$

as the consumers' strategy that determines all ISPs' market shares, and $s_{\mathcal{N}} = \{s_{\mathcal{N}}^I = (\mathcal{O}_I, \mathcal{P}_I) : I \in \mathcal{I}\}$ as the CPs' strategy, where each $s_{\mathcal{N}}^I$ denotes the decision made at ISP I.

We denote Φ_I as the per capita consumer surplus achieved at ISP I, defined as $\Phi_I(M_I, \mu_I, \mathcal{N}, s_I) = \Phi_I(\nu_I, \mathcal{N}, s_I) = \Phi((1 - \kappa_I)\nu_I, \mathcal{O}_I) + \Phi(\kappa_I\nu_I, \mathcal{P}_I)$, where $\nu_I = \mu_I/M_I$. We assume that consumers will move towards the ISPs that provide higher per capita surplus as follows.

Assumption 5. *Under any fixed strategy profile $s_{\mathcal{I}}$ and $s_{\mathcal{N}}$, for any pair of ISPs $I, J \in \mathcal{I}$, consumers will move from I to J if $\Phi_I < \Phi_J$. This process stops when $\Phi_I = \Phi_{\mathcal{I}}$ $\forall I \in \mathcal{I}$ for some system-wise per capita consumer surplus $\Phi_{\mathcal{I}}$.*

Although consumers might not be totally elastic or/and accessible to all available ISPs in practice, our assumption takes a macro perspective and assumes that if an ISP provides worse user-experience on average, there must exist some consumers who can and will move to other better ISPs. Based on Assumption 5, we define the equilibrium of the second-stage game $(M, \mu, \mathcal{N}, s_{\mathcal{I}})$ as follows.

Definition 4. *A strategy profile $(s_M, s_{\mathcal{N}})$ is an equilibrium of the multi-ISP game $(M, \mu, \mathcal{N}, s_{\mathcal{I}})$ if 1) for any $I \in \mathcal{I}$, $s_{\mathcal{N}}^I$ is an equilibrium of the single-ISP game $(M_I, \mu_I, \mathcal{N}, s_I)$, and 2) $\Phi_I = \Phi_J$ for any $I, J \in \mathcal{I}$.*

4.1 Duopolistic ISP Game

We first study a two-ISP game with $\mathcal{I} = \{I, J\}$. Before that, we formally define a Public Option ISP as follows.

Definition 5. *A Public Option ISP PO is an ISP that uses a fixed strategy $s_{PO} = (0, 0)$ and does not divide its capacity or charge the CPs.*
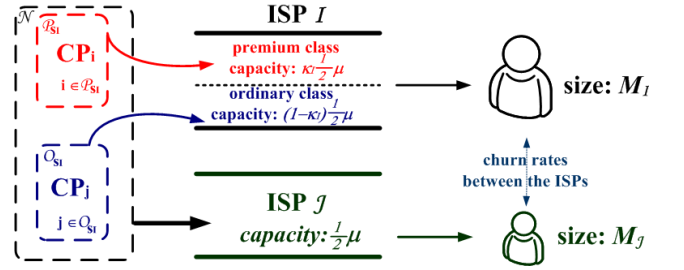

Figure 6: A two-ISPs model.

We assume that ISP J is a Public Option ISP. Figure 6 illustrates an example of the above duopolistic model, where both ISPs have the same amount of capacity, the CPs choose a service class at ISP I and the consumers move between the ISPs. The above setting of the duopolistic game applies for two real scenarios. First, it models the competition between two ISPs, where one of them is actively a Public Option ISP and the other actively manages a non-neutral service differentiation. Second, it also models a situation where a single ISP owns the entire last-mile capacity μ; however, by regulation [3], it is required to lease its capacity to other service providers that do not own the physical line. The leasing

ISP might be technologically limited from providing service differentiation on the leased capacity, but actually have customers in the region. For both scenarios, we will answer 1) whether the non-neutral ISP could obtain substantial advantages over the neutral Public Option ISP (or whether the Public Option could survive under competition), and 2) how the competition is going to impact the consumer surplus.

We study the same set of 1000 CPs as in the previous section. We further assume that $\mu_I = \mu_J = \mu/2$ in our numerical example. We take the same route to numerically evaluate the competitive equilibria of the CPs under $\kappa_I = 1$.
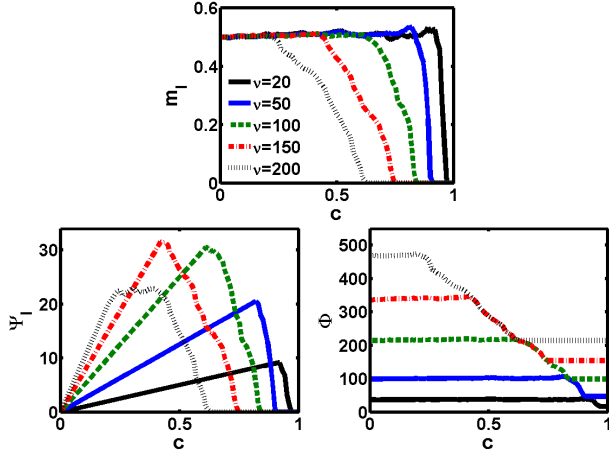

Figure 7: ISP I's market share m_I and per capita surplus Ψ_I and per capita consumer surplus Φ under $\kappa = 1$.

Figure 7 plots ISP I's market share m_I, per capita surplus Ψ_I, defined as $\Psi_I = \frac{c}{M}\lambda_{\mathcal{P}_I}$, and Φ versus ISP I's charge c_I. By the same reasons as before, the revenue of I increases linearly when its capacity is fully utilized, i.e. $\lambda_{\mathcal{P}_I} = \kappa_I \mu_I$. However, we observe three differences: 1) after $\lambda_{\mathcal{P}_I}$ drops below $\kappa_I \mu_I$, Ψ_I drops to zero much steeper than before, 2) Φ_I does not drop down to zero when c_I increases to 1, and 3) the maximum Ψ_I is lower in the case of $\nu = 200$ than in the case of $\nu = 150$, which means that under $\kappa_I = 1$, capacity expansion could reduce ISP I's revenue from the CPs. All these observations can be explained by checking the market share of ISP I in the upper sub-figure. The market share m_I starts to increase with c_I until ISP I's capacity becomes under-utilized, i.e. $\lambda_{\mathcal{P}_I} < \kappa_I \mu_I$. Afterwards, the market share drops dramatically. This explains that under congestion, i.e. $\lambda_{\mathcal{P}_I} = \kappa_I \mu_I$, by increasing c_I, ISP I restricts the number of CPs in its service class and maintains less congestion, which could result higher consumer surplus, and therefore, attract more consumers from ISP J. After $\lambda_{\mathcal{P}_I}$ drops below $\kappa_I \mu_I$, further increase of c_I reduces the number of CPs in the service as well as the total throughput. This reduces consumer surplus, and therefore, consumers start to depart from ISP I to J. When c_I reaches 1, no CP survives in I's service class and all consumers move to ISP J, which guarantees a non-zero consumer surplus in equilibrium.

Parallel to Figure 5, Figure 8 illustrates the per capita surplus Ψ_I, Φ and ISP I's market share m_I under various strategies s_I versus ν ranging up to 500. Compared to the monopolistic case, we observe two differences in Ψ_I and Φ: 1) under any strategy s_I, ISP I's revenue drops sharply to zero after reaching a maximum point where $\lambda_{\mathcal{P}_I}$ drops below $\kappa_I \mu_I$, and 2) the increase of consumer surplus does not get affected by ISP I's strategy too much. By observing the market share of ISP I, we identify two capacity regimes. First, when ν is extremely scarce, the differential pricing slightly benefits the consumer; and therefore, ISP I can obtain a slightly larger percentage of the market [6]. Second, when the per capita capacity ν is abundant, ISP I obtains at most an equal share of the market if it uses a small value of κ. Under this case, the capacity under $\mathcal{O}$ can support half of the population's unconstrained throughput and in fact, the premium class is empty, i.e. $\mathcal{P} = \emptyset$. As a result, ISP I follows the Public Option ISP by using some kind of neutral policy (small κ) and maximizes the consumer surplus.

Theorem 5. *In the duopolistic game $(M, \mu, \mathcal{N}, \mathcal{I})$, where an ISP J is a Public Option, i.e. $s_J = (0,0)$, if s_I maximizes M_I under an equilibrium $(s_M, s_{\mathcal{N}})$, then the per capita consumer surplus $\Phi_{\mathcal{I}}$ is also maximized under that equilibrium.*

Theorem 5 implies that the existence of a Public Option ISP is superior to a network neutral situation, where $s_I = (0,0)$. This is because given the freedom of choosing an optimal s_I to maximize market share, ISP I's strategy will induce a maximum consumer surplus under $s_J = (0,0)$.

Based on our results, we answer the previously raised two questions: 1) The non-neutral ISP cannot win substantially over the Public Option ISP, which can still be profitable under the competition, confirming the independent findings from [12]. 2) Regardless of the capacity size, the competition induces higher consumer surplus in equilibrium than under network neutral regulations. The strategic ISP could obtain slightly over 50% of the market; however, if it differentiates services in the way that hurts consumer surplus, its market share will drop sharply.

Regulatory Implications: In the duopolistic scenario with one of the ISPs being a Public Option, contrary to the monopolistic case, the non-neutral strategy s_I is always aligned with the consumer surplus (Theorem 5). This result shows an interesting alternative to remedy the network neutrality issue under a monopolistic market. Instead of enforcing the ISP to follow network neutrality, the government (or a private organization, if it can be profitable [12]; otherwise, the government would bear a social cost so as to achieve the maximization of consumer surplus) can provide the consumers with a Public Option ISP that is neutral to all CPs. Given such a neutral entity in the market, consumers will move to their public option if it provides higher consumer

---

[6]By limiting the number of CPs in $\mathcal{P}$, the proportion of throughput-sensitive traffic is larger, which yields higher consumer surplus.

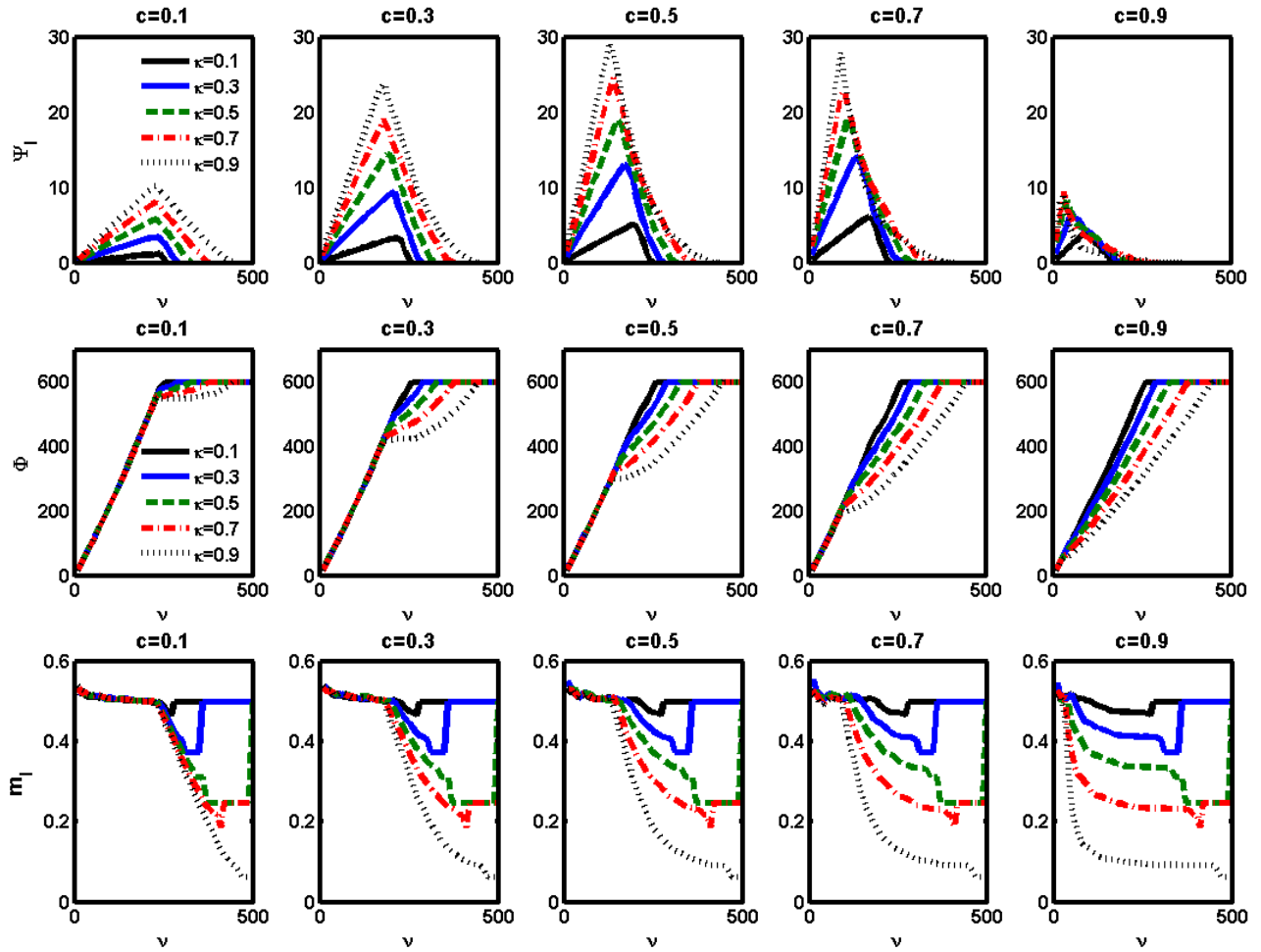**Figure 8: Per capita surplus Ψ, Φ and market share m_I under various strategies $s_I = (\kappa, c)$ vs. per capita capacity ν.**

surplus than the non-neutral ISP that uses differential pricing to the CPs. Meanwhile, in order to maximize its market share, the non-neutral ISP will adapt its strategy to maximize consumer surplus. In conclusion, the introduction of a Public Option ISP is superior to network neutral regulations under a monopolistic market, since its existence aligns the non-neutral ISP's selfish interest with the consumer surplus.

4.2 Oligopolistic ISP Competition Game

After analyzing the duopolistic game between a non-neutral and a Public Option ISP, we further consider a deregulated market under which all ISPs make non-neutral strategies. We consider a multi-ISP game under which each ISP I chooses a strategy s_I to maximize its market share m_I.

We first consider a homogenous strategy $s = (\kappa, c)$, which can be a preferred or regulated strategy, used by all ISPs.

Lemma 4. *If $s_{\mathcal{I}} = \{s_I = s : I \in \mathcal{I}\}$ for some strategy $s = (\kappa, c)$, then $\{m_I = \gamma_I, s_{\mathcal{N}}^I = s_{\mathcal{N}}(M, \mu, \mathcal{N}, s) : I \in \mathcal{I}\}$ is an equilibrium of the game $(M, \nu, \mathcal{N}, s)$.*

Lemma 4 shows a symmetric equilibrium where market share m_I is proportional to capacity μ_I. It implies that ISPs

will have incentives to invest and expend capacity so as to obtain a larger market share. This equilibrium could be reached when ISPs simply mimic one another's strategy.

A further question is whether the competition of market share among the ISPs would induce equilibria where consumer surplus is high. To address this issue, we first define

$$\epsilon_{s_I} = \sup\{\Phi(\nu_1, \mathcal{N}, s_I) - \Phi(\nu_2, \mathcal{N}, s_I) : \nu_1 < \nu_2\}. \quad (8)$$

We denote s_{-I} as the strategy profile of the ISPs other than ISP I, and define $\delta_{s_I} = \sup\{m_1 - m_2 : \Phi(\nu_1, \mathcal{N}, s_I) \leq \Phi(\nu_2, \mathcal{N}, s_I)\}$ and $\epsilon_{s_{-I}} = \max\{\epsilon_{s_J} : J \in \mathcal{I} \backslash \{I\}\}$.

Theorem 6. *Under any fixed strategy profile s_{-I}, if I's strategy s_I is a best-response to s_{-I} that maximizes its market share m_I in the game $(M, \mu, \mathcal{N}, s_{\mathcal{I}})$, then s_I is a $\epsilon_{s_{-I}}$-best-response for the per capita consumer surplus $\Phi_{\mathcal{I}}$, i.e.*

$$\Phi_{\mathcal{I}} \geq \Phi'_{\mathcal{I}} - \epsilon_{s_{-I}}, \quad \forall s'_I \neq s_I.$$

Moreover, if s_I is a best-response that maximizes consumer surplus $\Phi_{\mathcal{I}}$ in the game $(M, \mu, \mathcal{N}, s_{\mathcal{I}})$, then s_I is a δ_{s_I}-best-response for the market share m_I, i.e.

$$m_I \geq m'_I - \delta_{s_I}, \quad \forall s'_I \neq s_I.$$

Theorem 6 states that, given the fixed strategies of all other ISPs, an ISP's best-responses to maximize 1) its market share and 2) the consumer surplus are closely aligned. Parallel to Theorem 5, it shows that an ISP's selfish interest is, although not perfectly, aligned with the consumer surplus under competition. Technically, the $\epsilon_{s_{-I}}$ imperfection is due to the discontinuity of $\Phi(\nu, \mathcal{N}, s_I)$ in ν when CPs move between service classes. When $\epsilon_{s_{-I}}$ approaches zero, Φ becomes non-decreasing and the objectives of market share maximization and consumer surplus maximization converges.

Definition 6. *A strategy profile $s_{\mathcal{I}} = \{s_I : I \in \mathcal{I}\}$ is a market share Nash equilibrium of the game $(M, \mu, s_{\mathcal{N}}, \mathcal{I})$ if for any $I \in \mathcal{I}$ and any strategy $s'_I \neq s_I$, the market share m_I satisfies $m_I(s'_I, s_{-I}) \leq m_I(s_I, s_{-I})$. Similarly, $s_{\mathcal{I}}$ is a consumer surplus Nash equilibrium of the game $(M, \mu, s_{\mathcal{N}}, \mathcal{I})$ if for any $I \in \mathcal{I}$ and any strategy $s'_I \neq s_I$, the consumer surplus $\Phi_{\mathcal{I}}$ satisfies $\Phi_{\mathcal{I}}(s'_I, s_{-I}) \leq \Phi_{\mathcal{I}}(s_I, s_{-I})$.*

Corollary 1. *If $s_{\mathcal{I}}$ is a market shares Nash equilibrium of the oligopolistic game $(M, \mu, \mathcal{N}, \mathcal{I})$, then it is also a consumer surplus $\epsilon_{s_{\mathcal{I}}}$-Nash equilibrium, where $\epsilon_{s_{\mathcal{I}}} = \max\{\epsilon_{s_I} : I \in \mathcal{I}\}$. Conversely, if $s_{\mathcal{I}}$ is a consumer surplus Nash equilibrium, then it is also a market share $\delta_{s_{\mathcal{I}}}$-Nash equilibrium, where $\delta_{s_{\mathcal{I}}} = \max\{\delta_{s_I} : I \in \mathcal{I}\}$.*

As a direct consequence of Theorem 6, Corollary 1 addresses that the objectives of maximizing market share and maximizing consumer surplus are also closely aligned under Nash equilibria of the oligopolistic game $(M, \mu, \mathcal{N}, \mathcal{I})$.

Regulatory Implications: In the oligopolistic scenario, all ISPs' optimal strategies are closely aligned with the consumer surplus. Even if some ISPs use sub-optimal decisions, any remaining ISPs' optimal strategy would still nearly maximize the system consumer surplus (Theorem 6). This alignment with consumer surplus also sustains under Nash equilibria of the multi-ISP competition game (Corollary 1). Under this case, the existence of a Public Option ISP would be sub-optimal compared to the efficient Nash equilibria; however, its damage is very limited because the Public Option ISP would be the only one that uses a sub-optimal strategy, where all other ISPs can adapt to optimal strategies and more consumers will move from the Public Option to better and non-neutral ISPs. Of course, there is no reason why the Public Option cannot perform the price discrimination that aligns with the consumer surplus, which induces an efficient Nash equilibrium in theory. However, implementing a neutral Public Option will avoid mistakes or accidental "collusion" with the existing ISPs in the market. In contrast, if network neutral regulations are enforced, all ISPs will have to perform a neutral but inefficient strategy, which could reduce the consumer surplus substantially. In conclusion, network neutral regulations are not needed and should not be imposed under a competitive market; however, regulations should enforce the ISPs to be transparent in the sense that ISPs' capacity and strategies should be common knowledge to all ISPs, which would help the market converge to an efficient equilibrium in an easier manner.

5. RELATED WORK

Despite of its short history, a lot of of work on network neutrality can be found in computer science [10, 22, 6, 12, 24, 16], economics [8, 14], and law [26, 25] literature.

From an economics perspective, Sidak [25] looked at the network neutrality regulation from consumer welfare's point of view and argued that differential pricing is essential to the maximization of welfare. We also focus on the consumer welfare and seek the conditions under which ISPs' strategy would be aligned with consumer welfare. Choi et al. [8] analyzed the effect of neutral regulations on ISPs' investment incentive and found that capacity expansion decreases the sale price of the premium service. This coincides with our finding under the monopolistic scenario; however, under oligopolistic competitions, we find that ISPs do have incentives to increase capacity so as to maximize market share.

From an engineering perspective, Dhamdhere et al. [12] took a profitability perspective and concluded that the ISPs can still survive without violating network neutrality. This supports our proposal of a Public Option ISP that can be implemented and sustained by either a government or a private organization. Crowcroft [10] reviewed various technical aspects and concluded that "perfect" network neutrality has never been and should not be engineered. We share the same view that under competition, network neutrality regulation is not necessary; while, under a monopolistic market, a non-regulatory alternative can be a Public Option ISP that incentivizes the existing ISP to maximize consumer surplus.

From a modeling point of view, one departure in our approach from previous analyses is the way we model traffic and congestion in the network. Traditionally, the $M/M/1$ formula for delay has been used to abstract out traffic and congestion [8] in economic analyses. Our view is that a more appropriate approach is to more faithfully model closed loop protocols like TCP that carry most of the traffic on the Internet. Musacchio et al. [22] considered advertising CPs and also used a two-stage model under which ISPs move first. Their focus was primarily on a monopolistic ISP. Caron et al. [6] modeled differentiated pricing for only two application types. Shetty et al. [24] used a similar PMP-like two-class service differentiations and considered capacity planning, regulation as well as differentiated pricing to consumers. Our differentiated pricing focuses on the CP-side, where the CPs choose service classes and consumers choose ISPs. Yuksel et al. [27] also used a two-class service model, but focused on transit ISPs and quantified the equivalent overprovisioning cost when best-effort is used. Our work focuses on the last-mile eyeball ISPs and consumer surplus.

From a regulatory aspect, Wu [26] surveyed the discriminatory practices, e.g. selectively dropping packets, of broadband and cable operators and proposed solutions to manage bandwidth and police ISPs so as to avoid discrimination. Shetty et al. [24] proposed a simple regulatory tool to restrict the percentage of capacity the ISPs dedicate to a premium service class. Economides et al. [13] compared var-

ious regulations for quality of service, price discrimination and exclusive contracts, and drew conclusions on desirable regulation regimes. Ma et al. [16, 17] considered the ISP settlement aspect and advocated the use of Shapley value as profit-sharing mechanism to encourage ISPs to maximize social welfare. Our proposal of a Public Option ISP, on the other hand, is an non-regulatory alternative to the network neutral regulations.

6. DISCUSSION AND CONCLUSIONS

In a monopolistic market, the ISP's selfish non-neutral strategy hurts consumer surplus. Although network neutral regulation might improve consumer surplus, we find a better non-regulatory alternative which is to introduce a Public Option ISP. The existence of a Public Option ISP incentivizes the existing ISP's strategy to be aligned with consumer surplus, and achieve higher consumer surplus than that under network neutral regulations. In an oligopolistic competition, market forces influence ISPs' non-neutral strategies to be aligned with consumer surplus and ISPs will get market shares proportional to their capacities. Although network neutral regulations are not needed and should not be imposed under oligopolistic scenarios, we envision that the Public Option could be implemented as the safety net, or the last/back-up choice, for the consumers if the existing commercial ISPs' strategy hurt consumer surplus.

Theoretically speaking, the existence of a Public Option ISP will be effective if $\mu_{PO} > 0$, regardless of how large its capacity is. This is because, in the idealized game model, we assume that an ISP's sole objective is to maximize its market share. In practice, ISPs will trade off their market share with potential revenue from the CPs, which depends on the characteristics of the CPs, e.g. their profit margin and throughput sensitivity, and the condition of the system, e.g. the available capacity and congestion level. Moveover, ISPs might be able to use the CP-side revenue to subsidize the service fees for consumers so as to increase market share. In general, the more ISPs compete freely in a market, the less the market needs a public option and the less capacity we need to deploy for the Public Option ISP to be effective. In the most hostile case where only one monopolistic ISP exists in the market, a Public Option ISP could be effective as long as it has a capacity that is larger than the percentage of consumers that the monopoly cannot afford to lose. For example, if 10% of the market share is critical for the monopoly, implementing 10% of its capacity would be able to at least "steal" 10% of consumers from the monopoly if it follows a network neutral strategy. If the monopoly applies a worse than neutral strategy for consumer surplus, it will lose even more. In that sense, although 10% of the capacity will not be operating optimally, its existence incentivizes the remaining 90% maximizing for consumer surplus, which could result in much better consumer surplus than requiring the monopoly to follow network neutral regulations.

In summary, we believe our paper sheds new light on and informs the continuing debate on the role of regulation on the Internet and our introduction of the Public Option ISP is an important contribution.

7. REFERENCES

[1] FCC Acts to Preserve Internet Freedom and Openness. News Release(12/21/10). http://www.fcc.gov/Daily_Releases/Daily_Business/2010/db1221/DOC-303745A1.pdf.
[2] Google corporate website. http://www.google.com/corporate/tech.html.
[3] Local Loop Unbundling. http://en.wikipedia.org/wiki/Local-loop_unbundling.
[4] Netflix technology blog. http://techblog.netflix.com/2011/01/netflix-performance-on-top-isp-networks.html.
[5] M. Campbell and J. Browning. Apple, Google asked to pay up as mobile operators face data flood. *Bloomberg News*, December 7 2010.
[6] S. Caron, G. Kesidis, and E. Altman. Application neutrality and a paradox of side payments. *Proceedings of the ACM ReARCH '10*, November 2010.
[7] D. M. Chiu and R. Jain. Analysis of the increase and decrease algorithms for congestion avoidance in computer networks. *Computer Networks and ISDN Systems*, 17(1).
[8] J. P. Choi and B.-C. Kim. Net neutrality and investment incentives. *The Rand Journal of Economics*, 41(3):446–471, Autumn 2010.
[9] C. Courcoubetis and R. Weber. *Pricing Communication Networks: Economics, Technology and Modelling*. John Wiley & Sons Ltd., 2003.
[10] J. Crowcroft. Net neutrality: the technical side of the debate: a white paper. *ACM SIGCOMM Computer Communication Review*, 37(1), January 2007.
[11] R. J. Deneckere and R. P. McAfee. Damaged goods. *Journal of Economics & Management Strategy*, 5(2):149–174, June 1996.
[12] A. Dhamdhere and C. Dovrolis. Can ISPs be profitable without violating network neutrality? *Proceedings of ACM NetEcon*, August 2008.
[13] N. Economides and J. Tag. Network neutrality and network management regulation: Quality of service, price discrimination, and exclusive contracts. *Research Handbook on Governance of the Internet. London: Edward Elgar*, 2012.
[14] B. Hermalin and M. L. Katz. The economics of product-line restrictions with an application to the network neutrality debate. *Information Economics & Policy*, 19(2):215–248, 2007.
[15] C. Labovitz, D. McPherson, S. Iekel-Johnson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. In *Proceedings of the ACM SigComm*, New Delhi, India, 2010.
[16] R. T. B. Ma, D. Chiu, J. C. Lui, V. Misra, and D. Rubenstein. Internet Economics: The use of Shapley value for ISP settlement. *IEEE/ACM Transactions on Networking*, 18(3), June 2010.
[17] R. T. B. Ma, D. Chiu, J. C. Lui, V. Misra, and D. Rubenstein. On cooperative settlement between content, transit and eyeball internet service providers. *IEEE/ACM Transactions on Networking*, 19(3), June 2011.
[18] R. T. B. Ma and V. Misra. Congestion equilibrium for differentiated service classes. *Allerton Conference on Communication, Control & Computing 2011*.
[19] R. T. B. Ma and V. Misra. The public option: a nonregulatory alternative to network neutrality. *CoRR, June 2011. http://arxiv.org/abs/1106.3242*.
[20] A. Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomic theory*. Oxford University Press, 1995.
[21] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5), October 2000.
[22] J. Musacchio, G. Schwartz, and J. Walrand. Network neutrality and provider investment incentives. *Asilomar Conference*, pages 1437–1444, November 2007.
[23] A. Odlyzko. Paris metro pricing for the Internet. *Proceedings of ACM EC'99*, pages 140–147, 1999.
[24] N. Shetty, G. Schwartz, and J. Walrand. Internet QoS and regulations. *IEEE/ACM Transactions on Networking*, 18(6), December 2010.
[25] J. G. Sidak. A consumer-welfare approach to network neutrality regulation of the Internet. *Journal of Competition Law and Economics*, 2(3), 2006.
[26] T. Wu. Network neutrality, broadband discrimination. *Journal of Telecommunications and High Technology Law*, 141, 2005.
[27] M. Yuksel, K. K. Ramakrishnan, S. Kalyanaraman, J. D. Houle, and R. Sadhvani. Quantifying overprovisioning vs. class-of-service: Informing the net neutrality debate. *Proceedings of 19th International Conference on Computer Communications and Networks (ICCCN)*, pages 1–8, 2010.