Subsidization Competition: Vitalizing the Neutral Internet

Richard T. B. Ma
National University of Singapore
and Advanced Digital Science Center
.tbma@comp.nus.edu.sg

ABSTRACT

Unlike telephone operators, which pay termination fees to reach the users of another network, Internet Content Providers (CPs) do not pay the Internet Service Providers (ISPs) of users they reach. While the consequent cross subsidization to CPs has nurtured content innovations at the edge of the Internet, it reduces the investment incentives for the access ISPs to expand capacity. As potential charges for terminating CPs’ traffic are criticized under the net neutrality debate, we propose to allow ISPs to voluntarily subsidize the usage-based fees induced by their content traffic for end-users.

We model the regulated subsidization competition among CPs under a neutral network and show how deregulation of subsidization could increase an access ISP’s utilization and revenue, strengthening its investment incentives. Although the competition might reduce the throughput of certain CPs, we find that the main cause comes from high access prices rather than the existence of subsidization. Our results suggest that subsidization competition will increase the competitiveness and welfare of the Internet content market; however, regulators might need to regulate access prices if the access ISP market is not competitive enough. We envision that subsidization competition could become a viable model for the future Internet.

Keywords

Internet Economics; Subsidization Competition; Net Neutrality

1. INTRODUCTION

Driven by bandwidth-intensive video applications and pervasive Internet access via mobile devices, Internet inter-domain traffic has been growing more than 50% per annum [24] and is expected to grow more than 100% per annum for mobile networks [38]. To sustain such rapid traffic growth, Internet Service Providers (ISPs), especially the mobile access ISPs, need to upgrade their network infrastructures and expand capacities. However, revenues from online services are growing more than twice as fast as those from Internet access [38], whose market capitalizations have stagnated as investors weigh high capital requirements against continued margin pressure. This raises serious challenges regarding the viability of the current Internet model in the future. The fundamental problem lies in the two-sided market structure [8, 31] of the Internet, where the ISPs provide a platform connecting end-users to the Internet Content Providers (CPs). The CP-side transit ISPs are often the backbone ISPs that provide IP transit services. Due to market competition, recent years have witnessed a constant decrease in IP transit prices [56]. The more serious problem lies in the last-mile connections toward the end-users, which are often the bottlenecks of the Internet [13]. For these last-mile access ISPs, including the mobile providers, the main source of revenue is from their end-users. Traditionally, they also need to pay transit ISPs for transit services, although settlement-free peering is common in practice.

Existing proposals to resolve this issue include: 1) allowing ISPs to impose termination fees [9] on CPs’ traffic towards their end-users, and 2) allowing ISPs to prioritize traffic, differentiate services and thus charge users non-uniformly. However, both proposals have triggered a heated debate over network neutrality [49, 14], whose advocates argue that zero-pricing [26] and the neutrality of physical networks are needed to protect and encourage content innovations of the Internet. After all, the limited capacities and rising access costs will dampen user demand, if CPs cannot return the created value to ISPs.

Traditionally, wireline access ISPs charge flat-rate prices and implicitly subsidize CPs for content delivery; however, usage-based pricing is commonly used by wireless providers. The FCC chairman has recently backed usage-based pricing for broadband to penalize heavy Internet users [22, 42] and major U.S. broadband and cable providers, e.g., Verizon [43] and AT&T [47], have adopted tiered schemes where charges are imposed to the metered usage above a predefined data cap. We assume the adoption of usage-based pricing by access ISPs, under which heavy users share the burden of subsidizing CPs. Our goal is to realign the created value and stimulate the needed investment in the Internet value chain and our high-level idea is illustrated in Figure 1 where we propose to allow CPs to voluntarily subsidize the usage-based access costs caused by their content traffic, partially or fully, for the end-users. Our solution modularizes the issue along the tussle boundary such that the physical network is kept neutral while innovative competition among CPs is enabled at higher service layers so
as to incentivize ISP investments. As a special case of full subsidization, AT&T's recent sponsored data [1] plan will likely be available by the end of 2014. Nevertheless, it raised concerns over anti-competitive practices, and the FCC says that they will be monitoring and prepared to intervene if necessary [24].

To understand and demonstrate the potential impact of subsidization, we take an analytical approach as no market data is available for an empirical study yet. We develop a macroscopic Internet model and study the status quo one-sided pricing of the access ISPs in Section 3. Building upon this framework, we model the policy that regulates subsidization and analyze the competition game among CPs in Section 4. We evaluate the policy implications on the ISP's revenue and system welfare in Section 5 and discuss the regulatory and implementation issues, as well as the limitations of our model in Section 6. Our analytical results include:

- We characterize the impact of capacity, user population (Theorem 1) and one-sided ISP pricing (Theorem 2) on the system utilization and the CPs’ throughput.
- We analyze the CPs’ subsidization competition game and characterize the Nash equilibrium (Theorem 3), its uniqueness and its dynamics (Theorem 4 and 6).
- We characterize the impact of subsidization policy on the CPs’ throughput, user population, and the system’s utilization (Theorem 5) and welfare (Corollary 1).

Through our theoretical results and qualitative evaluations, we also identify important policy implications as follows.

- If allowed, profitable CPs tend to subsidize users more (Theorem 1) and achieve higher throughput (Lemma 3), leading to higher system welfare (Corollary 1).
- For fixed ISP pricing, deregulation of subsidization will increase the utilization and revenue of the access ISPs (Corollary 1), strengthening their investment incentives.
- Certain CPs’ throughput might decrease due to the ISP’s high price (Theorem 8), which could be triggered by the deregulation of subsidization. Thus, access prices might need to be regulated if the access ISP market is not competitive enough.

In short, subsidization provides a feedback channel for the highly profitable CPs to repay their users and indirectly transfer value to the access ISPs. It helps the access ISPs attract investment for capacity expansion while keeping their physical networks neutral. Our results suggest that regulators should promote this subsidization competition; however, they might need to regulate the ISP’s pricing if the access market is not competitive enough. We envision that subsidization competition could become a viable model for the future Internet so as to accommodate the growing demand for data traffic.

2. RELATED WORK

A report from A.T. Kearney [38] studied the performance, economic and policy pressures of the Internet and identified four models for the future Internet: 1) modification of retail pricing schemes, 2) traffic-dependent charges, 3) enhanced-quality services over the public Internet, and 4) bilateral service agreements. Subsidization competition is a new regulatory model that preserves the neutrality of the physical networks. We discuss the above existing models and their relationships with our subsidization competition model.

2.1 Modification of Retail Pricing Schemes

Historically, the Internet adopted flat-rate prices [37, 44] for simplicity. Economists [30] and computer scientists [12, 21, 23] advocated usage-based pricing, which was shown to provide congestion control [21], quality of service [23] and economic efficiency [12], and is adopted by mobile providers [32]. Subsidization competition assumes the adoption of usage-based pricing by access ISPs, under which consumers’ usage charges could be subsidized by the CPs.

2.2 Imposition of Termination Fees on CPs

One core question under the net neutrality debate [49] is whether ISPs should be allowed to use two-sided pricing and charge CPs for terminating their content traffic. Prior work [10, 22, 35] studied the investment incentives in Internet-like two-sided markets and drew different conclusions. Njoroge et al. [35] found that through CP-side pricing, ISPs could extract higher surplus and maintain higher investment levels. However, Choi et al. [10] found that expending capacity will decrease the CP-side sale price. Musacchio et al. [33] found that although the returns on investment under one-sided or two-sided pricings are comparable, the size of investment and profit depend on the advertising rates and users’ price sensitivity. From an economic efficiency point of view, [35] concluded that two-sided pricing results in higher social welfare; however, Lee and Wu [26] argued that the zero-pricing at the CP-side could be optimal in theory [8, 21]. Also, [10] found that the short-run welfare is higher under one-sided neutral regulation. Our subsidization mechanism does not impose charges on the CPs by the ISPs, but provides an indirect mechanism for them to voluntarily return the zero-pricing subsidy back to the ISPs via their end-users.

2.3 Differentiated Services over the Internet

Another core debate under net neutrality is whether ISPs should be allowed to differentiate services and prices. Wu [49] surveyed the discriminatory practices of broadband operators and proposed solutions to manage bandwidth and police ISPs. Sidak [30] focused on consumer utility and argued that differential pricing is essential to maximize utility. Shetty et al. [45] considered a two-class service model and studied capacity planning, regulation, as well as differentiated pricing to consumers. Economides et al. [16] examined various regulations for quality of service and price discrimination, and drew conclusions about desirable regulation regimes. As net neutrality is undergoing heated debate, regulation of desirable differentiation is a rich area for further study. Under our solution, ISPs do not make any physical or economic differentiation and CPs’ subsidies are made voluntarily. Subsidization serves as a policy mechanism that encourages competition and provides investment incentives and economic efficiency for the Internet ecosystem.

2.4 Bilateral Settlements for Valued Services

A fundamental cause of investment shortage is the misalignment between the value created and the revenue obtained in the Internet value chain. Clark et al. [11] discussed various economic tussles among the ISPs, CPs and users. Ma et al. [28] advocated the use of the Shapley profit-sharing mechanism for multi-lateral ISP settlements and studied bilateral implementations [29]. Laskowski [25] identified that the network’s lack of accountability is an obstacle to implement desirable contracting systems for the Internet. Because accounting of traffic towards the end-users is feasible for access ISPs, our subsidization mechanism can be regarded as a contracting system between the CPs and access ISPs, even through they might not be directly connected to each other. Through the feedback channel from CPs to end-users, subsidization provide a new
avenue for the Internet ecosystem to realign the profit distribution among different stakeholders.

3. MACROSCOPIC INTERNET MODEL

To understand the implications of policy on the Internet, we start with a macroscopic model that captures the physical and business dynamics in the status quo Internet among an access ISP, its users and the CPs used by these consumers.

3.1 Basic Physical System Model

We denote \( m \) as the capacity of an access ISP serving the users of a certain geographical region and \( N \) as the set of CPs used by the ISP’s consumers. For each CP \( i \in N \), we denote \( m_i \) as its user population and define \( \theta_i = m_i \lambda_i \), as its total throughput, where \( \lambda_i \) denotes the CP’s average per user throughput. We define \( \Theta(\theta, \mu) \) as the aggregate system throughput. Under any fixed capacity \( \mu \) and aggregate throughput \( \theta \), we define \( \phi = \Phi(\theta, \mu) \) as the system utilization, a function of \( \theta \) and \( \mu \). We assume the average user throughput is a function of the utilization, i.e., \( \lambda_i = \lambda_i(\phi) \).

Assumption 1. \( \Phi(\theta, \mu) \) and any \( \lambda_i(\phi) \) are differentiable. \( \Phi(\theta, \mu) \) is strictly decreasing in \( \mu \), strictly increasing in \( \theta \), and satisfies \( \lim_{\theta \to \infty} \Phi(\theta, \mu) = 0 \) for all \( \mu > 0 \). \( \lambda_i(\phi) \) is strictly decreasing in \( \phi \) and satisfies \( \lim_{\phi \to \infty} \lambda_i(\phi) = 0 \).

Assumption 1 captures the physics of capacity utilization: utilization increases when the system accommodates higher throughput or has less capacity, and vice-versa. Because the system utilization and congestion are two sides of the same coin (congestion happens when the system is over-utilized), it also captures the physics of user throughput: users achieve higher throughput when the system is less congested.

Figure 2: The basic system model \((m, \mu)\).

We denote \( m \) as the vector of user populations of the CPs. Figure 2 illustrates our basic model \((m, \mu)\), where the user populations \( m \) and system capacity \( \mu \) collectively determine the system utilization \( \phi \) and each CP’s throughput \( \theta_i \). Because utilization increases with the accommodated throughput, which decreases with the utilization (congestion), the resulting system utilization should be in an equilibrium.

Definition 1 (Utilization). \( \phi \) is the utilization of a system \((m, \mu)\) if it satisfies the following condition:

\[
\phi = \Phi \left( \sum_{k \in N} m_k \lambda_k(\phi), \mu \right). \tag{1}
\]

Definition 1 states that the utilization \( \phi \) should induce the aggregate system throughput \( \Theta(\phi, \mu) = \sum_{k \in N} m_k \lambda_k(\phi) \) such that it leads to exactly the same level of utilization \( \phi = \Phi(\theta, \mu) \).

We define \( \Theta(\phi, \mu) \equiv \Phi^{-1}(\phi, \mu) \) as the inverse function of \( \Phi(\theta, \mu) \) with respect to \( \phi \). \( \Theta(\phi, \mu) \) can be interpreted as the implied amount of throughput that induces a utilization level \( \phi \) for a system with capacity \( \mu \).

By Assumption 1 \( \Theta(\phi, \mu) \) is strictly increasing in both \( \phi \) and \( \mu \). To characterize the utilization of a system \((m, \mu)\), we define a gap function \( g(\phi) \) between the supply and demand of throughput under a fixed level of utilization \( \phi \) as \( g(\phi) \equiv \Theta(\phi, \mu) - \sum_{k \in N} m_k \lambda_k \).

Lemma 1. Given fixed capacity \( \mu \) and user populations \( m \), \( g(\phi) \) is strictly increasing in \( \phi \). The system operates at a unique level of utilization \( \phi \) which solves \( g(\phi) = 0 \).

Lemma 1 characterizes the uniqueness of the system utilization \( \phi \) under which the throughput supply \( \Theta(\phi, \mu) \) equals the aggregate throughput demand \( \sum_{k \in N} m_k \lambda_k(\phi) \).

We define the marginal change of the throughput gap \( g(\phi) \) due to the marginal change in the system utilization \( \phi \) as

\[
\frac{dg}{d\phi} \equiv \frac{\partial \Theta(\phi, \mu)}{\partial \phi} - \sum_{k \in N} m_k \frac{\partial \lambda_k(\phi)}{d\phi} > 0, \tag{2}
\]

where the first term (second summation) captures the change of throughput in supply (total demand). We characterize the impact of user population \( m_i \) and capacity \( \mu \) on the utilization \( \phi \) and throughput \( \theta_i \), as a function of \( dg/d\phi \) as follows.

Theorem 1 (Capacity and User Effect). The system utilization \( \phi \) decreases if the capacity \( \mu \) increases or any user population \( m_i \) decreases. In particular, we have

\[
\frac{\partial \phi}{\partial \mu} = - \left( \frac{dg}{d\phi} \right)^{-1} \frac{\partial \Theta(\phi, \mu)}{\partial \mu} < 0, \tag{3}
\]

and

\[
\frac{\partial \phi}{\partial m_i} = \left( \frac{dg}{d\phi} \right)^{-1} \lambda_i > 0, \quad \forall i \in N. \tag{4}
\]

Any CP \( i \)'s throughput \( \theta_i \) increases with capacity \( \mu \) and its user population \( m_i \); however, the increase in \( m_i \) reduces the throughput \( \theta_j \) of all other CPs \( j \neq i \). In particular,

\[
\frac{\partial \theta_i}{\partial m_i} = m_i \frac{\lambda_i}{\partial \phi} = m_i \frac{\partial \phi}{\partial m_i} > 0, \quad \frac{\partial \theta_i}{\partial m_k} = \lambda_i + m_i \frac{\lambda_k}{\partial \phi} = 0,
\]

and

\[
\frac{\partial \theta_j}{\partial m_i} = m_j \frac{\lambda_j}{\partial \phi} = m_j \frac{\partial \phi}{\partial m_i} < 0, \quad \forall j \neq i.
\]

Theorem 1 intuitively states that 1) when the system capacity becomes more abundant, the throughput of any CP increases and the system utilization decreases, and 2) when more users of a CP join the system, the system utilization and the throughput of that CP increase, while that of any other CP decreases. Equation 3 explicitly shows that the capacity effect on the system utilization can be expressed as its impact on the feasible system throughput \( \partial \Theta(\phi, \mu)/\partial \mu \), normalized by \( dg/d\phi \), which re-balances the supply and demand of throughput so as to fill the throughput gap \( g(\phi) \) under the new system utilization. Similarly in Equation 4, \( \lambda_i \) can be interpreted as the marginal increase in throughput when a new user of CP \( i \) joins the system. It also implies that the user impact on the utilization is proportional to the per user throughput, i.e., \( \partial \phi/\partial m_i : \partial \phi/\partial \mu = \lambda_i : \lambda_i \).

CPs’ throughput depends on whether their users are sensitive to congestion. This can be characterized by utilization-elasticity of throughput or \( \phi \)-elasticity of \( \lambda_i \), defined as \( \epsilon_i \).

1Due to space limit, we include the proofs of selected results in appendix. Complete proofs of all results can be found in technical report [27].
Definition 2 (ELASTICITY). The elasticity of $y$ with respect to $x$, or $x$-elasticity of $y$, $e^y_x$ is defined as $e^y_x = \frac{\partial y}{\partial x} \cdot \frac{x}{y}$.

Elasticity is often expressed as $e^y_x = \frac{\partial y}{\partial x} / \frac{x}{y}$ and interpreted as the percentage change in $y$ (the numerator) in response to the percentage change in $x$ (the denominator). In particular, $e^y_\phi$ captures the percentage change in $\lambda_i$ in response to the percentage change in the system utilization $\phi$.

Lemma 2. If CP $i$ is replaced by a new CP $j$ which satisfies that $m_i \lambda_i(t_i) = m_j \lambda_j(t_j)$ and $e^\lambda_i(\phi) = e^\lambda_j(\phi)$ for all $\phi > 0$, then the system utilization does not change.

Lemma 2 characterizes CPs’ throughput by $\phi$-elasticity. Mathematically, by keeping the same elasticity $e^\lambda_\phi$, we could scale up the maximum throughput $\lambda_i(0)$ by $\lambda_j(0) = \kappa \lambda_i(0)$ and scale down $m_i$ by $m_j = \kappa^{-1} m_i$ for some constant $\kappa > 0$ without affecting the system utilization $\phi$ and the throughput of all other CPs. Operationally, it implies that we could treat the traffic of a CP $i$ as if it has a single big user, i.e., $m_i = 1$, with the maximum throughput $\lambda_i(0) = \lambda_j(0)$. It also implies that for multiple CPs, e.g., a set $\mathcal{S} \subset \mathcal{N}$, with the same $\phi$-elasticity of throughput, we could conceptually aggregate them as a single CP $k$ with parameters $m_k$ and $\lambda_k(0) = \sum_{\phi \in \mathcal{S}} \lambda_i(0)$. By this result, we will use a single CP to model a group of CPs that have similar traffic characteristics in our later numerical evaluations.

### 3.2 One-Sided ISP Pricing

In the current Internet, most access ISPs only charge their end-users. We denote $t_i$ as the per-unit usage charge for the data traffic of CP $i$. We assume that CP $i$’s user population $m_i$ is a function of its per-unit charge $t_i$, i.e., $m_i = m_i(t_i)$.

Assumption 2. $m_i(t_i)$ is continuously differentiable and decreasing with $\lim_{t_i \to \infty} m_i(t_i) = 0$.

Assumption 2 states that the user demand decreases with the per-unit charge on the content traffic. This model is quite general and coincides with existing models [8, 41] where the heterogeneity of users are modeled by the distribution of their valuations on data traffic and only the users whose valuations are higher than $t_i$ consume data traffic from CP $i$.

Because traffic from different CPs should not be differentiated under net neutrality, ISPs often impose a uniform charge. We denote $p$ as this uniform charge for all types of traffic. When CPs do not provide subsidies, we have $t_i = p$ for any $i \in \mathcal{N}$. We define the ISP’s revenue as $R = \rho m_i \lambda_i$ and each CP $i$’s utility as $U_i = v_i(\theta_i)$, where $v_i$ denotes CP $i$’s average per-unit traffic profit. Because the price $p$ affects the user demand $m_i$, we extend our basic system model to capture the one-sided ISP pricing as follows.

Figure 3 shows that the user populations $m_i$ are influenced by the price $p$. In this case, the utilization $\phi$ affects the ISP’s revenue $R = \rho p \sum_{i \in \mathcal{N}} m_i \lambda_i$. ISPs can use revenue as feedback to guide their pricing and capacity decisions; however, CPs’ utilities are determined by throughput, which have no means to influence in the one-sided pricing model.

Because $t_i = p$ for any CP $i$, by Assumption 2, we express the user populations as a function of price as $m_i = m_i(p)$. For a fixed capacity $\mu$, we define $\phi(p) = \phi(m(p), \mu)$ as the utilization of system $m(p), \mu$ and $\theta_i(p) = \phi(m(p), \lambda_i(p))$ as the corresponding throughput of CP $i$ under price $p$. By Assumption 2 and 3, both $\phi(p)$ and $\theta_i(p)$ are differentiable functions of $p$. We characterize the impact of price $p$ on the system utilization $\phi$ and each CP $i$’s throughput $\theta_i$ in the following theorem. The price’s impact on $U_i$ is simply its impact on $\theta_i$ scaled by $\nu_i$ and we will study its impact on the revenue $R$ under a more general setting in Section 5.

Theorem 2 (PRICE EFFECT). The system utilization $\phi$ decreases with the price $p$. In particular, we have

$$\frac{\partial \phi}{\partial p} = \left( \frac{dg}{d\phi} \right)^{-1} \sum_{i \in \mathcal{N}} \frac{d m_i}{dp} \lambda_i \leq 0. \quad (5)$$

The aggregate throughput $\theta$ decreases with the price $p$, and any CP $i$'s throughput $\theta_i$ increases with $p$ if and only if

$$e^\mu_p / e^\lambda_\phi < -e^\theta_p.$$

Theorem 2 states that when the price increases, due to its direct impact on decreasing the user demand, the system utilization decreases, which implies that the aggregate throughput of all the CPs also decreases. Similar to (5) and (6), the price effect on the system utilization can be expressed as its impact on the aggregate throughput $d\phi/dp$, normalized by $dg/d\phi$ in Equation (5). However, a CP $i$’s throughput might increase or decrease, depending on its utilization elasticity of (per user) throughput $e^{\lambda_i}$ and price elasticity of user demand $e^\mu_p$, i.e., the percentage change in the population $m_i$ in response to the percentage change in price $p$.

In particular, $\theta_i$ increases with $p$ if its users are less price-sensitive and more congestion-sensitive, i.e., $|e^{\lambda_i}|$ is small and $|e^\mu_p|$ is large. Notice that the right hand side of (7) measures the overall impact of price on the system utilization.

To understand the price effect more intuitively, we illustrate a numerical example as follows. We consider a utilization function $\Phi(\theta, \mu) = \theta / \mu$, which uses per capacity throughput as the metric for system utilization. We consider the forms of throughput $\lambda_i(\phi) = e^{-\beta_i \phi}$ and user demand $m_i(t_i) = e^{-\alpha_i t_i}$, that satisfy Assumptions 2 and 3. Under these exponential forms, the $\phi$-elasticity of throughput $e^{\lambda_i}$ equals $-\beta_i$ and the $p$-elasticity of user population $e^{\mu_i}$ equals $-\alpha_i$.

Based on the above setting, we can express a CP $i$’s throughput as $\theta_i = e^{-(\alpha_i + \beta_i) \phi}$, where $\alpha_i$ and $\beta_i$ neatly characterize $\theta_i$’s sensitivity to price and congestion, respectively.

We obtain $dg/d\phi = \mu + \sum_{i \in \mathcal{N}} \beta_i \theta_i$ and

$$\frac{\partial \theta}{\partial p} = \left( \frac{dg}{d\phi} \right)^{-1} \left( \sum_{i \in \mathcal{N}} \frac{d m_i}{dp} \lambda_i \right) = -\frac{\sum_{i \in \mathcal{N}} \alpha_i \theta_i}{\mu + \sum_{i \in \mathcal{N}} \beta_i \theta_i}.$$
By Theorem 2, throughput $\theta_i$ increases at price $p$ if
$$\frac{\alpha_i p}{\beta_i \phi} < \frac{\sum_{j \in N} \alpha_j \theta_j}{\mu + \sum_{k \in N} \beta_k \theta_k}$$.
(8)

We set the system capacity to be $\mu = 1$ and consider a set of 9 types of CPs with values of $(\alpha_i, \beta_i)$ chosen from $\{1, 3, 5\}$.

Figure 4: Aggregate throughput $\theta$ and ISP’s revenue $R$. 

Figure 4 plots the aggregate throughput $\theta$ (left) and the ISP’s revenue $R$ (right) as a function of price $p$ that varies along the x-axis. We observe that the aggregate throughput decreases with the price as indicated by Theorem 2, however, the revenue $R = p\theta$ depends on both the price and aggregate throughput and shows a single-peak pattern.

Figure 5: Throughput $\theta_i$ of different CPs. 

Figure 5 shows the throughput $\theta_i$ of the 9 individual CPs as a function of the price $p$ in each sub-figure, respectively. In general, the throughput is low with large values of $\alpha_i$ and $\beta_i$ (the lower and right sub-figures), because the user population is more sensitive to price and the per user throughput is more sensitive to congestion. By Theorem 2 when $p$ increases, $\phi$ decreases, and therefore $(\alpha_i/\beta_i)(p/\phi)$ increases. As indicated by condition (8), each $\theta_i$ decreases with $p$ eventually; however, when $p$ is small, we observe that the CPs with a small ratio of $\alpha_i/\beta_i$ (the upper and right sub-figures) demonstrate an increasing trend in throughput. Intuitively, for these CPs, the increase in the per user throughput $\lambda_i$ is much higher than the decrease in the user population $m_i$, so that the aggregate throughput $\theta_i$ could still increase.

Regulatory Implications: Under the existing one-sided pricing, an increase in the access ISP’s price reduces the user demand (by Assumption 2), and consequently reduces the system utilization and the total system throughput (by Theorem 2). Regulators might want to regulate the price of an access ISP if its high price induces low utilization and drives the total system throughput too low (shown in Figure 4). However, as the existing wireless capacities of many carriers are often under-provisioned and highly-loaded, and regulations will further limit the ISPs’ profit margin, we do not see the need for price regulation under the status quo.

4. SUBSIDIZATION COMPETITION

As the access ISPs can neither differentiate services nor charge CPs in the one-sided pricing model in Figure 3, the limited profit margin does not provide enough investment incentive for them to expand capacities. Neither do the CPs have any means to express preferences and improve their utilities. In this section, we propose to create a feedback channel for the CPs to influence the system by allowing them to voluntarily subsidize the usage-based fees for their users.

4.1 The Subsidization Competition Model

We denote $q \geq 0$ as a subsidization policy that limits the maximum subsidy allowed. We denote $s_i \in [0, q]$ as the per-unit usage subsidy provided by CP $i$ for its content traffic for users. We denote $s$ as the vector of subsidies of the CPs. We extend the definition of CP $i$’s utility as $U_i \triangleq (v_i - s_i)\theta_i$.

Figure 6 illustrates the extended model where subsidization from CPs is allowed. In this model, each CP could strategically choose its subsidy $s_i$, to influence its throughput $\theta_i$, via its user population $m_i$ so as to optimize its utility $U_i$. Regulators can also use any welfare metric $W$ to determine the desirable upper-bound $q$ of subsidization for the industry.

Under price $p$ and subsidies $s$, user population $m_i$ satisfies $m_i(t_i) = m_i(p - s_i)$. Given any fixed ISP decision $(p, \mu)$, we denote $\phi(s) \triangleq \phi(m(p, s), \mu)$ as the system utilization and define $\theta_i(s) = m_i(p - s_i)\lambda_i(\phi(s))$ as CP $i$’s throughput.

Lemma 3. For any $s' > s_i$, let $s' = (s'_1, s'_{3-1})$ and $s = (s_i, s_{3-1})$ for some fixed strategy profile $s_{3-1}$. For any price $p$, $\phi(s') \geq \phi(s)$ and $\theta_i(s') \geq \theta_i(s)$. Whenever $s' > s_i$, it unilaterally increases its subsidy $s_i$, its throughput $\theta_i$, and the system utilization $\phi$ might increase; how-
ever, the throughput of any other CP cannot increase. It shows that the subsidization mechanism creates a competitive game among the CPs, where each CP could use subsidy $s_i$ to maximize its utility $U_i$. Notice that besides the policy constraint $q$, each CP’s optimal subsidy strategy depends on all other CPs’ strategies as well as the ISP’s price $p$. In this section, we study the policy implications of subsidization and therefore, we often assume a fixed price $p$ for the access ISP. This setting corresponds to a competitive access market where the ISPs cannot easily manipulate prices, or a case where the ISP’s price is regulated. We will study an ISP’s pricing strategy, its revenue, and the policy implication on the system welfare in the next section.

**Definition 3** (Nash equilibrium). Given any fixed ISP price $p$ and policy $q$, a strategy profile $s$ is a Nash equilibrium if each $s_i$ solves CP $i$’s utility maximization problem:

$$\max_{s_i} U_i(s_i; s_{-i}) = (v_i - s_i)\theta_i(s)$$

subject to $0 \leq s_i \leq q$.

We characterize the Nash equilibrium of the subsidization competition game in the following theorem.

**Theorem 3** (Characterization). For any fixed $p$ and $q$, a strategy profile $s$ is a Nash equilibrium only if

$$s_i = \min\{\tau_i(s), q\}, \quad \forall i \in N,$$

where each $\tau_i$ denotes a threshold for CP $i$, defined by

$$\tau_i(s) \triangleq (v_i - s_i)\epsilon_i^m + (1 + \epsilon_i^\lambda:\phi_i).$$

(9)

In particular, $v_i \leq (\partial \theta_i/\partial s_i)^{-1}\theta_i$ holds if $s_i = 0$. Moreover, if $U_i(s) = U_i(s_i, s_{-i})$ is concave in $s_i$ for all $i \in N$, then the above conditions are also sufficient.

Theorem 3 characterizes the Nash equilibrium in two ways. First, it indicates that any CP does not subsidize if its profitability $v_i$ is lower than $(\partial \theta_i/\partial s_i)^{-1}\theta_i$. Second, if the policy constraint is not tight, i.e., $s_i < q$, the equilibrium subsidy $s_i$ can be characterized by its profit margin $v_i - s_i$ and the elasticity metrics $\epsilon_i^m$, $\epsilon_i^\lambda$, and $\epsilon_i^\phi$. Equation (9) can also be written as $\tau_i(s) = (v_i - s_i)\epsilon_i^m + (v_i - s_i)\epsilon_i^\lambda + (1 + \epsilon_i^\phi)$. This characterization also implies that a CP would subsidize its users more if 1) its traffic profitability $v_i$ (and therefore its profit margin $v_i - s_i$) increases, or 2) its throughput elasticity $\epsilon_i^m$ or user demand elasticity of subsidy $\epsilon_i^\phi$ increases.

By Assumption 1 and 2, each $U_i(s)$ is differentiable in $s_i$. For any strategy profile $s$, we define $u_i(s) = \{u_i(s) : i \in N\}$ as the vector of marginal utilities, where each $u_i(s)$ denotes the marginal utility of CP $i$, defined as $u_i(s) \triangleq \partial U_i(s)/\partial s_i$. Next, we characterize the uniqueness of Nash equilibrium based on a condition on the marginal utilities $u_i(s)$.

**Theorem 4** (Uniqueness). For any fixed price $p$ and policy $q$, if for any distinct pair of feasible strategy profiles $s' \neq s$, there always exist a CP $i \in N$ such that

$$(s_i' - s_i)(u_i(s') - u_i(s)) < 0,$$

then there always exists a unique Nash equilibrium.

Technically, the above sufficient condition for uniqueness requires $u_i$ to be a $P$-function [31]. Notice that this uniqueness property holds for any subset $Q \subset [0, q]^N$ of the strategy space if condition (10) holds in the domain $Q$.

### 4.2 Dynamics of Subsidies in Equilibrium

Since a Nash equilibrium is defined under a fixed price $p$ and policy $q$ by Definition 3, we use $s = s(p, q)$ to indicate a Nash equilibrium under $p$ and $q$. Equation (9) in Theorem 3 hinted that a CP with higher profitability subsidizes users more in an equilibrium. We formally show this as follows.

**Theorem 5** (Profitability Effect on Subsidy). If a CP $i$’s profitability increases from $v_i$ to $\hat{v}_i$ unilaterally, and under the condition (10) of Theorem 2, $s$ and $\hat{s}$ are the corresponding Nash equilibrium, then we must have $\hat{s}_i \geq s_i$.

If a CP’s profitability increases, Theorem 5 tells that it will increase its subsidy in equilibrium, and hints that its throughput will increase by Lemma 1. Through subsidization, CPs can respond to their profitability and influence throughput.

After analyzing the impact of CPs’ profitability, we study how an equilibrium $s(p, q)$ is influenced by the price $p$ and policy $q$. We define $N_-$ and $N_+$ as the set of CPs whose subsidies are 0 and $q$, respectively, and $\hat{N} = N\setminus(N_- \cup N_+)$ as the remaining set of CPs whose subsidies are strictly positive and less than $q$. We define $\hat{s}$ and $\hat{\beta}$ as the subsidies and marginal utilities of the set $\hat{N}$ of CPs.

**Theorem 6** (Equilibrium Dynamics). If a Nash equilibrium $s$ satisfies $1) i \in \hat{N}$ if $u_i(s) = 0$, and 2) the condition (10) for a neighborhood of $[0, q]$ centered at $s$ in the strategy space, then in a neighborhood of $(p, q)$, a unique Nash equilibrium is a differentiable function $s(p, q)$, satisfying

$$\partial s_i/\partial q = \begin{cases} \frac{0}{1} \text{ if } i \in N_-; \\ \sum_{k \in \hat{N}} \psi_{ik} \partial u_k/\partial q, \text{ if } i \in \hat{N}, \end{cases}$$

and

$$\partial s_i/\partial p = \begin{cases} 0 \text{ if } i \notin \hat{N}; \\ -\sum_{k \in \hat{N}} \psi_{ik} \partial u_k/\partial p, \text{ if } i \in \hat{N}, \end{cases}$$

where $\Psi = \{\psi_{ij}\} \triangleq (\nabla_s^2 U)^{-1}$, i.e., $\psi_{ij}$ is at the $i$th row and $j$th column of the inverse of the Jacobian matrix $\nabla_s^2 U$.

Theorem 6 states that a marginal change in the ISP’s price or the regulatory policy will not affect the behavior of the set $N_-$. We define $N_-$ and $N_+$ as the CPs who do not subsidize remain the same and the CPs who subsidize the amount $q$ keep subsidizing at the maximum level as $q$ increases. When $p$ changes, the set $\hat{N}$ of CPs readjust subsidies among themselves in a new equilibrium such that their marginal utilities remain zero in (12): when $q$ changes, since the set $N_+$ of CPs increase subsidies accordingly, their impacts have to be counted in the new equilibrium in (11). By applying the KKT condition to Definition 3, we know $u_i(s) = 0$ for all $i \in \hat{N}$, and the first condition of Theorem 6 is a bit stronger which guarantees that the equilibrium is $s$ is regular, i.e., locally differentiable. The second condition only assumes (10) locally, which guarantees the local uniqueness of equilibrium in a neighborhood centered at $s$ by Theorem 6.

**Corollary 1** (Deregulation). Under a fixed ISP price $p$ and the conditions of Theorem 6, in a neighborhood of $q$, we can express $\phi(q)$ and $R(q)$ as the system utilization and the ISP’s revenue under the Nash equilibrium $s(q)$. If $\phi$ is off-diagonally monotone, i.e., $\partial u_i(s)/\partial s_j \geq 0$ for all $i \neq j$,

$$\frac{\partial \phi_i}{\partial q} \geq 0, \quad \frac{\partial R}{\partial q} \geq 0 \text{ and } \frac{\partial s_i}{\partial q} \geq 0, \quad \forall i \in N.$$
5. ISP REVENUE AND SYSTEM WELFARE

Although the policy \( q \), the price \( p \) and the subsidies \( s \) all impact the system, these decisions are not made simultaneously and independently. The system is fundamentally driven by a regulatory policy \( q \), under which the ISP determines its price \( p(q) \) and then the CPs respond with strategies \( s(p, q) \). Theorem 6 characterizes the dynamics of \( s(p, q) \) and Corollary 7 characterizes the impact of policy \( q \) when \( p \) is fixed. In practice, ISPs often use price to optimize revenue, and by Theorem 6 the subsidies \( s \) are also influenced by \( p \). In this section, we study the impact of the ISP’s price on its revenue and the impact of policy on the system welfare when the response of ISP’s pricing is taken into account.

5.1 Impact of ISP’s Pricing on Its Revenue

Under a fixed policy \( q \), by Theorem 6 we can write \( s(p) \) as the induced subsidies as a function of price \( p \). Thus, the induced system utilization is \( \phi(s(p)) \) and the ISP’s induced revenue is \( R(p) = \sum_{i \in N} \theta_i(p) = \sum_{i \in N} m_i(p - s_i(p))\lambda_i(\phi(s(p))) \).

Theorem 7 (Marginal Revenue). If the ISP’s revenue is differentiable at \( p \), then its marginal revenue is

\[
\frac{dR(p)}{dp} = \sum_{i \in N} \theta_i + \Upsilon \sum_{i \in N} \epsilon^{m_i}_p \theta_i,
\]

where \( \Upsilon \) and the \( m_i \)-elasticity of price \( p \) satisfy

\[
\Upsilon = 1 + \sum_{j \in N} \epsilon^{s_j}_i + \epsilon^{m_i}_p = \frac{dm_i}{dt_i} \left( 1 - \frac{\partial s_i}{\partial p} \right).
\]

Theorem 7 characterizes the change in revenue as the price varies and isolates the effect of subsidization into the elasticities \( \epsilon^{m_i}_p \), where \( \partial s_i/\partial p \) plays a role. It also generalizes the case of one-sided pricing, i.e., \( \partial s_i/\partial p = 0 \) for all \( i \in N \). The second term in (13) measures the price’s impact on the aggregate throughput, i.e., \( p(\partial s_i/\partial p) \), which can be factorized by \( \Upsilon \), a parameter determined by the physical model in Figure 2 because \( \epsilon^{s_j}_i \) can be decomposed as \( \epsilon^{s_j}_i = \epsilon^{s_i}_i \epsilon^{s_j}_i \) and

\[
\epsilon^{s_j}_i \epsilon^{s_j}_i = m_j \lambda_j \frac{\partial \phi}{\partial \theta_j} = m_j \lambda_j \left( \frac{dq}{d\phi} \right)^{-1}.
\]

5.2 Impact of Regulatory Policy on Welfare

By Theorem 7 we understand how the ISP’s pricing affects its revenue. We assume that the ISP would adapt its price \( p \) as a function of a given policy \( q \). The next theorem captures the impact of the policy \( q \) on the system states, i.e., \( \phi, m_i \) and \( \theta_i \), where the responses of both the ISP’s price \( p(q) \) and the CPs’ subsidies \( s(p, q) \) are taken into account.

Theorem 8 (Policy Effect). If the ISP’s price is a differentiable function \( p(q) \) of policy \( q \) and \( s \) is a Nash equilibrium of \( p(q) \) and \( q \), under the conditions of Theorem 2 in a neighborhood of \( q \), we can express the Nash equilibrium as \( s(q) \triangleq s(p(q), q) \). The corresponding system utilization \( \phi(q) \), user population \( m_i(q) \) and throughput \( \lambda_i(q) \) satisfy

\[
\frac{dm_i}{dq} = \frac{dm_i}{dt_i} \frac{dt_i}{dq} = \frac{dm_i}{dt_i} \left( 1 - \frac{\partial s_i}{\partial p} \right) \frac{dp}{dq} - \frac{\partial s_i}{dq},
\]

\[
\frac{d\phi}{dq} = \left( \frac{dq}{d\phi} \right)^{-1} \sum_{i \in N} m_i \lambda_i \quad \text{and} \quad \frac{d\lambda_i}{dq} = \frac{d\lambda_i}{d\phi} \frac{d\phi}{dq}.
\]

Any CP \( i \)’s throughput \( \theta_i \) increases with \( q \) if and only if

\[
\epsilon^{m_i}_i \epsilon^{s_i}_i / \epsilon^{s_i}_i < -\epsilon^{\phi}_i.
\]

Theorem 8 shows that the policy effect on utilization \( \frac{d\phi}{dq} \) in (16) and the condition (17) for throughput \( \theta_i \) have similar forms as the price effect \( \frac{dp}{dq} \) in (15) and the condition (16). This is because the policy effect is carried out by its impact on the user populations, i.e., \( \epsilon^{s_i}_i = \epsilon^{m_i}_i \epsilon^{s_i}_i \) via its impact on the price and subsidies. By Assumption 1 and 2, both \( \epsilon^{s_i}_i \) and \( \epsilon^{m_i}_i \) are negative. Inequality (17) tells that a CP’s throughput decreases if only if \( -\epsilon^{m_i}_i \epsilon^{s_i}_i < -\epsilon^{s_i}_i \epsilon^{\phi}_i \), which implies that the subsidy could help increase throughput via \( -\epsilon^{m_i}_i \epsilon^{s_i}_i \), however, \( -\epsilon^{m_i}_i \epsilon^{s_i}_i \) could be decreased due to the ISP’s increasing price, which will reduce the CP’s throughput.

Corollary 2 (Policy Impact on Welfare). Under the conditions of Theorem 8 let \( W(q) \triangleq \sum_{i \in N} \theta_i(q)v_i \) define the system welfare. Suppose \( \frac{dW}{dq} \) in (16) is positive, then the marginal welfare \( \frac{dW}{dq} \) is positive if and only if

\[
\sum_{i \in N} \sum_{k \in N} w_{ik} v_i > \sum_{i \in N} -\epsilon^{m_i}_i v_i, \quad \text{where} \quad w_i \triangleq \lambda_i \frac{dm_i}{dq}.
\]

We measure the welfare by \( W = \sum_{i \in N} \theta_i v_i \) for two reasons. First, it internalizes the subsidy transfer from CPs to ISP. Second, because CP profits are often positively correlated to their values to users, it also serves an estimate for user welfare. In Corollary 8, \( w_i v_i \) can be interpreted as the increase in welfare due to the policy’s impact on the user population \( m_i \), and therefore, the left side of the inequality represents the normalized increase in welfare due to the changes in the user populations. The right side of the inequality represents the normalized decrease in welfare due to the policy’s impact on each average throughput \( \lambda_i \) via \( m_i \). Corollary 8 states that the welfare increase if and only if the increasing component is larger than the decreasing component. Notice that the decreasing component only depends on the physical characteristics of the system.
CPs, while the weight $w_i / \sum_{j \in C} w_j$ for each $v_i$ tends to be large if $v_i$ is large, because profitable CPs will have stronger tendencies to subsidize their users so as to attract a larger population $m_i$.

To understand the policy effect more intuitively, we perform a numerical evaluation as follows. We use the same setting $m_i = e^{-\alpha_i t_1}, \lambda(t) = e^{-\beta t}$, and $\phi(\theta, \mu) = \theta/\mu$ as in Section 3. We model 8 types of CPs with $\alpha_i, \beta_i \in (2.5, 3)$ and $v_i \in (0.5, 1)$. By Lemma 2, each CP represents the aggregation of a group of CPs with similar characteristics of traffic and user demand. In each of the following figures, we vary the policy $q$ at 5 levels from 0 to 2.0 and vary the ISP’s price $p$ from 0 to 2 along the x-axis. When $q = 0$, the figures show the baseline case where subsidization is not allowed.

**Figure 7:** ISP’s Revenue $R$ and System Welfare $W$.

Figure 7 plots the ISP’s revenue $R$ (left) and the system welfare $W$ (right) when the price $p$ varies along the x-axis. We observe that under any fixed price $p$, both the ISP’s revenue and the system welfare are higher when the CPs’ subsidization is less regulated, i.e., $q$ is large. However, if the deregulation of subsidization will trigger a higher ISP price, it also shows that the system welfare will decrease with the price $p$ under any fixed policy $q$.

Figures 8 and 7 plot the individual subsidy $s_i$ and throughput $\theta_i$ of the 8 types of CPs under equilibrium, respectively. In Figure 8, we observe that the CPs that have a higher profitability, i.e., $v_i = 1$ in the lower 4 sub-figures, or a higher demand elasticity, i.e., $\alpha_i = 5$ in the right 4 sub-figures, provide much higher subsidies compared to their counterparts. We also observe that when the price $p$ is small, except for the two CPs with $\alpha_i = 2$ and $v_i = 0.5$, most CPs want to subsidize at the maximum level $q$ constrained by the policy; however, when $p$ increases, subsidies may stay flat and then decrease due to the decrease in profit margin. By comparing Figure 8 with Figure 7, we observe that when $q = 2$, the ISP maximizes its revenue by setting $p$ a bit less than 1, where the subsidies from the CPs are kept at a high level. In Figure 7, we observe that the CPs with higher profitability, i.e., $v_i = 1$, or lower congestion elasticity, i.e., $\beta_i = 2$, achieve higher throughput compared to their counterparts. When comparing the throughput with that of the baseline, i.e., $q = 0$, we observe that the CPs with high profitability achieve higher throughput, with the only exception for the case $(\alpha_i, \beta_i, v_i) = (2, 5, 1)$ when $p$ is small. As $p$ is small, the system attracts large user demand and is relatively congested. Because this CP is more congestion sensitive, i.e., $\beta_i$ is large, and by Corollary 2, the ISP maximizes its revenue by setting $p$ a bit less than 1, where the CPs with higher profitability have the higher throughput for the CPs of higher profitability, the system welfare increases when the policy $q$ is relaxed.

**Regulatory Implications:** In a deregulated market where CPs are allowed to subsidize their users, highly profitable CPs will provide greater subsidies (by Theorem 7) and obtain more users (by Assumption 2), achieving higher throughput (by Theorem 7) and leading to a higher system welfare (by Corollary 2). However, deregulation of subsidization might trigger an increase in the ISP’s price, leading to a decrease in the system welfare (shown in Figure 7). High ISP prices also discourage user demand (by Assumption 2) and consequently decrease their throughput (shown in Figure 6). Our results suggest that the subsidization competition will increase the ISP’s revenue and system welfare (by Corollary 7); however, price regulations might be needed if the access market is not competitive and the ISP’s price is set too high.

6. DISCUSSIONS

In this section, we discuss related issues of subsidization, its connections to other recent developments in the Internet industry and comment on some limitations of our proposed model.

6.1 Implementation and Regulatory Issues

Subsidization works with the usage-based pricing schemes of the access ISPs. For mobile providers, implementing data caps and accounting for data traffic consumptions are common in practice. Broadband providers in countries like India and Australia have been using usage-based pricing, and major U.S. ISPs like Verizon [27] and AT&T [27] have also adopted usage-based pricing after gaining support from the FCC [27]. To differentiate content and enable subsidization from different CPs, platforms such as FreeBand [3] have been developed to provide data traffic statistics to mobile carriers such as to lower the bills of consumers, based on the type of applications. AT&T’s recent sponsored data [1] plan allows CPs to fully subsidize their users, and Syntonic Wireless [7] provides the corresponding Sponsored Content Store for end-users to use sponsored applications. Our model allows partial subsidization and the implementation would be easy if the CP has a direct financial relationship with its end-users, e.g., a shopping site. Otherwise, ISPs could implement accounting mechanisms to realize the subsidization as an intermediary, however, they would be preferred not to be involved in the payment transfer if net neutrality is concerned.

Deregulation of subsidization might reduce the throughput of certain CPs, e.g., the upper-left ones in Figure 2. These CPs either do not have incentives to subsidize, because their users are not price-sensitive, i.e., the elasticity $e_{\text{m}}$ is small, or they cannot afford to subsidize, because their profitability $v_i$ is low. In the former case, the decrease in throughput is mainly due to a higher utilization of the system, which will be relieved in the long term when the ISPs expand capacities. Startup companies with low profits are of the latter case. The decrease in throughput is mainly due to the ISP’s high price that limits the user demand, and therefore, regulators might want to regulate ISP’s monopolistic pricing or introduce competition, e.g., municipal infrastructures. However, if they are promising, we also expect that venture capital would provide the funding source for them to subsidize their users and achieve their potential high profitability. After all, we believe in a transparent and competitive market where users can choose CPs based on quality and prices, and the businesses drive the evolution of the Internet ecosystem.

Although our analyses imply that a deregulation of subsidization will provide CPs more freedom to subsidize and increase the competitiveness of the content market, regulators might still want to impose policies that prevent discriminatory subsidization of the access ISPs. For example, Comcast’s Xbox XFinity [28] service enables its users to access on-demand content on their Xbox devices, which does not count against their 250GB data cap. By implicitly subsidizing their users, this service favors the ISP’s vertically integrated business over other content competitors, e.g., Netflix. Some...
CPs work with access ISPs to provide better services and cheaper prices for their users, e.g., Google Apps for Verizon [4]; however, this might also inversely affect the competitors of those favored CPs. We believe that the subsidization option should be identical, and should be given to all CPs equally. Through this, the access ISPs would treat CPs more neutrally and the subsidization would create a transparent and uniform platform for the CPs to compete.

6.2 Connections to Industrial Developments

Our proposed subsidization framework is mostly close to AT&T's sponsored data [47], under which CPs could pay the access ISP so as to fully subsidize their users. However, there are two major differences. First, our framework is more flexible such that partial subsidization is allowed and determined by CPs. Second, our framework is more fair and transparent, i.e., the usage-based pricing for users are the same for all CPs and the subsidizations paid by CPs are equally treated. Much efforts have been made by CPs to provide better services for their users, e.g., Netflix’s OpenConnect [6] and Google Global Cache (GGC) [5]. These efforts make the Internet transit more efficient and competitive. Nevertheless, to support quality of service for content delivery, more severe issues are related to the monopolistic position of the last-mile broadband providers and the scarce capacity of the wireless providers. This is why Netflix has agreed to pay access ISPs like Comcast [40] and Verizon [19] via paid-peering agreements for better content delivery. Although paid-peering provides higher revenue for the access ISPs and some CPs do have incentives to pay for it, access ISPs need to actively manage their network capacity and differentiate CPs’ traffic, in other words, maintaining a non-neutral physical network. Our proposed subsidization framework, on the other hand, is a pricing mechanism, which preserves the neutrality of the physical network.

6.3 Model Limitations

Our analytical framework is based on a macroscopic equilibrium model that captures the interactions among an access ISP, a set of CPs and end-users via the CPs’ subsidization and the ISP’s capacity and pricing decisions. Our theoretical results are mostly qualitative, which provides understanding and predictions of the changes in the system states when some of the driving factors change. The limitation is that it might not be able to capture short-term off-equilibrium types of system dynamics, where players’ decisions are not ratio-
ational or optimal. Our numerical evaluations are limited to a styled model of CPs to capture qualitative trends; however, more detailed experiments and validations could be challenging, because market data are needed so as to obtain the characteristics of the CPs, e.g., their profitability and elasticities, and an ISP needs to execute such a plan in the market. With the emerging sponsored data plan from AT&T, we expect this type of market data could be available for regulatory authorities as well as research communities. This study focuses on a single access ISP; however, we believe that competition between ISPs will also incentivize them to adopt subsidization schemes, through which users can obtain subsidized services. Finally, our result shows that under the subsidization mechanism, the access ISP could increase utilization and revenue. Because ISPs care about their profit, which is a function of both utilization and revenue, the increase in profit will further provide investment incentives for them to expand capacity. We do not have sufficient space to study the ISP’s capacity planning decision in detail, which is a direction of our future work.

7. CONCLUSIONS

The Internet is facing a dilemma of increasing traffic demand and decreasing incentives for the access ISPs to expand capacity, because service prioritization and two-sided pricing are still under debate in terms of their implications for net neutrality. We propose to allow CPs to voluntarily subsidize the data usage costs incurred at the access ISPs for their end-users. Our solution modularizes the issue along the tussle boundary such that the physical network is kept neutral. Through modeling and analyses, we show that a deregulated subsidization policy will incentivize profitable CPs to subsidize users and increase the system welfare and the revenue of the access ISPs. With the improved profit margins, the access ISPs would obtain more investment incentives under subsidization. However, the deregulation of subsidization might trigger an ISP to increase price, which would decrease the system welfare as well as the throughput of certain CPs. We suggest that price regulation might be needed if the access market is not competitive enough and the price is too high. Because subsidization creates a feedback channel for CPs to return value to the access ISPs and realign the created value and investment in the value chain, we believe that subsidization competition could vitalize neutral Internet for the future.

8. ACKNOWLEDGMENTS

The author would like to thank the shepherd of this paper, Matthew Andrews, for useful advice on revising this work. The author would also like to thank Vishal Misra for discussions and suggestions, and the anonymous reviewers for their insightful comments.

This study is supported by Ministry of Education of Singapore AcRF grant R-252-000-526-112 and the research grant for the Human Sixth Sense Programme at the Advanced Digital Sciences Center from Singapore’s Agency for Science, Technology and Research (A*STAR).

APPENDIX

A. PROOFS OF SELECTED RESULTS

Proof of Lemma 1. By Assumption 1 we know \( \Theta(\phi, \mu) \) is strictly increasing in \( \phi \) and \( \lambda_i(\phi) \) is strictly decreasing in \( \phi \); therefore, \( g(\phi) = \Theta(\phi, \mu) - \sum_{i \in \mathcal{N}} m_i \lambda_i(\phi) \) is strictly increasing in \( \phi \). Because \( \lim_{\phi \to 0} \Theta(\phi, \mu) = 0 \) and \( \lambda_i(0) > 0 \) for all \( i \in \mathcal{N} \), \( \lim_{\phi \to 0} g(\phi) < 0 \); because \( \lim_{\phi \to \infty} \lambda_i(\phi) = 0 \) for all \( i \in \mathcal{N} \), \( \lim_{\phi \to \infty} g(\phi) = \lim_{\phi \to \infty} \Theta(\phi, \mu) > 0 \). As a result, \( g(\phi) = 0 \) must have a unique solution.

Finally, when \( g(\phi) = 0 \), \( \Theta(\phi, \mu) = \sum_{i \in \mathcal{N}} m_i \lambda_i(\phi) \), which is equivalent to Equation 1 of Definition 1.

Proof of Theorem 1. Because \( g(\phi) \) is strictly increasing in \( \phi \) by Lemma 1 and by Equation 2, we have

\[
\frac{dg}{d\phi} = \frac{\partial \Theta}{\partial \phi} - \sum_{i \in \mathcal{N}} m_i \frac{d\lambda_i}{d\phi} > 0.
\]

When considering the capacity effect, we write the system utilization \( \phi(\mu) \) as a function of the system capacity. By Lemma 1

\[
g(\phi(\mu)) = \Theta(\phi(\mu), \mu) - \sum_{i \in \mathcal{N}} m_i \lambda_i(\phi(\mu)) = 0.
\]

By taking derivative of \( \mu \) on both sides, we obtain

\[
\frac{\partial \Theta}{\partial \mu} + \frac{\partial \Theta}{\partial \phi} \frac{d\phi}{d\mu} - \sum_{i \in \mathcal{N}} m_i \lambda_i \frac{d\phi}{d\mu} = 0,
\]

which is equivalent to

\[
\left( \frac{\partial \Theta}{\partial \phi} - \sum_{i \in \mathcal{N}} m_i \frac{d\lambda_i}{d\phi} \right) \frac{d\phi}{d\mu} = \frac{\partial \Theta}{\partial \mu} \quad \text{or} \quad \frac{dg}{d\phi} \frac{d\phi}{d\mu} = \frac{\partial \Theta}{\partial \mu}.
\]

When considering the user effect, we write the system utilization \( \phi(\mu) \) as a function of the user populations. By Lemma 1

\[
g(\phi(\mu)) = \Theta(\phi(\mu), \mu) - \sum_{i \in \mathcal{N}} m_i \lambda_i(\phi(\mu)) = 0.
\]

By taking derivative of \( \mu \) on both sides, we obtain

\[
\frac{\partial \Theta}{\partial \mu} - \sum_{i \in \mathcal{N}} m_i \lambda_i \frac{d\phi}{d\mu} - \lambda_i = 0,
\]

which is equivalent to

\[
\left( \frac{\partial \Theta}{\partial \phi} - \sum_{i \in \mathcal{N}} m_i \frac{d\lambda_i}{d\phi} \right) \frac{d\phi}{d\mu} = \lambda_i \quad \text{or} \quad \frac{dg}{d\phi} \frac{d\phi}{d\mu} = \lambda_i.
\]

Since \( \frac{d\phi}{d\mu} > 0 \), the above implies \( \frac{d\phi}{d\mu} = \left( \frac{d\phi}{d\mu} \right)^{-1} \lambda_i > 0 \). Finally, the results for \( \partial \Theta_i/\partial \mu_i, \partial \Theta_j/\partial \mu_i \) and \( \partial \Theta_j/\partial \mu_i \) can be derived by using \( \theta_i(\phi) = m_i \lambda_i(\phi) \) and chain rules, and applying the above results for \( \partial \Theta/\partial \mu \) and \( \partial \Theta/\partial \mu_i \).

Proof of Lemma 2. Because \( s_i' > s_i, t_i' = p - s_i' < p - s_i = t_i \) and by Assumption 2

\[
m_i'(t_i') = m_i(t_i) \geq m_i(t_i).
\]

By Theorem 1 we know that \( \partial \Theta_i/\partial \mu_i > 0 \) and therefore, \( \partial \Theta_i/\partial \mu_i = \partial \Theta_i/\partial \mu_i > 0 \) for all \( j \neq i \). As a result, \( \theta_i(\phi) \geq \theta_i(\phi) \) and \( \theta_j(\phi) \leq \theta_j(\phi) \) for all \( j \neq i \), which implies \( U_i(\phi) = v_i \theta_i(\phi) \) and \( U_j(\phi) = v_j \theta_j(\phi) \) for all \( j \neq i \).

Proof of Theorem 3. If \( s \) is a Nash equilibrium, each \( s_i \in [0, q] \) maximizes \( U_i(s_i; s_{-i}) = (v_i - s_i) m_i(p - s_i) \lambda_i(\phi(s)) \). By Karush-Kuhn-Tucker condition, we can derive that

\[
\frac{\partial U_i}{\partial s_i} \begin{cases} 
0 & \text{if } s_i = 0; \\
\geq 0 & \text{if } s_i = q; \\
= 0 & \text{if } 0 < s_i < q.
\end{cases} \quad \forall i \in \mathcal{N}. \quad (18)
\]

For CP \( i \) with \( s_i = 0 \), the above condition tells that

\[
\frac{\partial U_i}{\partial s_i} = (v_i - s_i) \frac{\partial \theta_i}{\partial s_i} - \theta_i = v_i \frac{\partial \theta_i}{\partial s_i} - \theta_i \leq 0.
\]
By Lemma\textsuperscript{[3]} we know \( \partial \theta_i / \partial s_i > 0 \) and the above implies \( u_i \leq (\partial \theta_i / \partial s_i)^{-1} b_i \). By taking partial derivatives of \( s_i \) on \( U_i = (v_i - s_i) m_i \lambda_i \), we deduce that
\[
\frac{\partial U_i}{\partial s_i} = (v_i - s_i) \left( \frac{\partial \lambda_i}{\partial s_i} + \lambda_i \frac{\partial m_i}{\partial s_i} \right) = m_i \lambda_i.
\]
Dividing \( m_i \lambda_i \) on both sides of (15), we have
\[
(u_i - s_i) \left( \frac{1}{\lambda_i} \frac{\partial \lambda_i}{\partial s_i} + \frac{1}{m_i} \frac{\partial m_i}{\partial s_i} \right) \leq 1 \quad \text{if} \quad s_i = 0; \quad \geq 1 \quad \text{if} \quad s_i = q_i; \quad 0 < s_i < q_i.
\]
When we multiply \( s_i \) on both sides of the above, the left-hand side becomes
\[
(u_i - s_i) \left( \epsilon_i^m + \epsilon_i^\lambda \right) = (v_i - s_i) \left( \epsilon_i^m + \epsilon_i^\lambda \epsilon_i^m \right) = \tau_i(s).
\]
When \( s_i = 0 \), \( \tau_i(s) = 0 = s_i \), and therefore, we have
\[
\tau_i(s) \begin{cases} 
\geq s_i & \text{if} \quad s_i = q_i; \\
es_i & \text{if} \quad 0 < s_i < q_i, \\
\text{or} \quad s_i = \min(\tau_i(s), q_i).
\end{cases}
\]
Finally, if \( U_i(s) \) is concave for all \( s \in N \), local optimality also guarantees the global optimality and therefore, the above first order necessary conditions become sufficient conditions for \( s \) to be a Nash equilibrium.

\textbf{Proof of Theorem 4} The condition (16) implies that \( U_i(s_i; s_{-i}) \) is concave for any \( s_{-i} \). Because the strategy space \([0, q]\) is compact, it guarantees the existence of Nash equilibrium. Suppose there exists two distinct Nash equilibria \( \hat{s} \) and \( \hat{s} \). By concavity of \( U_i(s_i; s_{-i}) \) in the maximum and for any \( s \in [0, q] \) and any \( x_i \in [0, q] \), we have
\[
\begin{align*}
x_i & = 0, \quad \tau_i(s) \leq 0, \quad \text{and} \quad \lambda_i \in \mathbb{R} \quad \text{such that} \quad \hat{s}_i = \frac{\partial \lambda_i}{\partial s_i} + \frac{\partial m_i}{\partial s_i} = \lambda_i.
\end{align*}
\]
By substituting \( x = \hat{s} \) in the first inequality and \( x = \hat{s} \) in the second inequality, we have for any \( s \in N \), \( (\hat{s}_i - s_i) u_i(s) \leq 0 \) and \( (\hat{s}_i - s_i) u_i(s) \leq 0 \). Because the above inequalities, and therefore, the above becomes \( (\hat{s}_i - s_i) u_i(s) \leq 0 \). Finally, we obtain \( (\hat{s}_i - s_i) u_i(s) \leq 0 \). Similarly, we know that the Jacobian \( V_i(\hat{s}, -\hat{u}) \) is a P-matrix, which is always non-singular. Therefore, we deduce
\[
\begin{align*}
\nabla v \hat{s} = -(\nabla v \hat{s})^{-1} \nabla v \hat{u} \hat{E}_q = -\nabla v \hat{s} \hat{E}_q, \quad \text{the same as the third case of (14)}.
\end{align*}
\]
When taking \( p \) as the sensitivity parameter and applying Theorem 3.1 of [15], we can obtain \( Q(\hat{s}_i; \hat{u}(p)) = -M \hat{E}_p \) and \( (I - Q)(\nabla v \hat{s}) (\hat{p}(p) + \nabla v \hat{F}) = 0 \). Because the constraints in \( G \) do not depend on \( p \), \( E_p \) is a zero vector, therefore, the first equation implies that \( \partial s_i / \partial q = 0 \) for all \( i \notin N \). By substituting \( F \triangleq -u \) into the second equation, we have \( (I - Q)(\nabla v \hat{u}) = 0 \). Similarly, by substituting \( s_i / \partial q = 0 \) for all \( i \notin N \), into the above, we obtain \( \nabla v \hat{s} \hat{u} \hat{E}_q = 0 \). Because \( v(\hat{u} - u) \) is non-singular, it is equivalent to the second case of (12) in a matrix form as
\[
\begin{align*}
\nabla v \hat{s} = -(\nabla v \hat{s})^{-1} \nabla v \hat{u} \hat{u} = -\nabla v \hat{s} \hat{u} \hat{u}.
\end{align*}
\]

\textbf{Proof of Corollary 1} By (11) of Theorem 6, \( \partial s_i / \partial q \geq 0 \) is immediate for \( i \notin N \). For the CPs in \( N \), by Theorem 6 we have
\[
\nabla v \hat{s} = -(\nabla v \hat{s})^{-1} \nabla v \hat{u} \hat{E}_q = -\nabla v \hat{s} \hat{u} \hat{E}_q = 0.
\]
Because \( u \) is off-diagonally monotone, we know that \( V_i(\hat{s}, -\hat{u}) \) is a P-matrix with non-positive off-diagonal entries or an M-matrix [19], which implies that all the entries of \( -V_i \) are non-negative. Therefore, by Theorem 6, we know that \( V_i(\hat{s}, \hat{u}) \) is an M-matrix. Because \( u \) is off-diagonally monotone, the components of \( V_i(\hat{s}, \hat{u}) \) are all non-negative, and therefore, \( \nabla v \hat{s} = -\nabla v \hat{s} \hat{u} \hat{E}_q \geq 0 \).

To show that \( \partial s_i / \partial q \geq 0 \) for all \( i \in N \). Since the subsidies are monotonic, by Lemma 3, we deduce that \( \partial s_i / \partial q \geq 0 \). Finally, because the ISP’s revenue can be written as \( R = \gamma(t, \mu) \).
and $\partial \Theta(\phi, \mu)/\partial \phi \geq 0$, we have

$$\frac{\partial R}{\partial q} = \frac{\partial R}{\partial \phi} \frac{\partial \phi}{\partial q} = p \frac{\partial \Theta(\phi, \mu)}{\partial \phi} \frac{\partial \phi}{\partial q} \geq 0.$$  

B. REFERENCES

http://www.att.com/att/sponsoreddata/

http://www.xbox.comcast.net/

http://www.freebandtechnologies.com

http://www.verizon.com/smallbusiness/products/business-apps/google-apps/

http://www.google.com

http://www.netflix.com

http://www.syntonic.com/


http://www.syntonic.com/


