TDS: Time-Dependent Sponsored Data Plan for Wireless Data Traffic Market

Liang Zhang*, Weijie Wu†, Dan Wang*‡
* Department of Computing, The Hong Kong Polytechnic University
†Future Network Theory Lab, Huawei Technologies Co., Ltd.
‡ The Hong Kong Polytechnic University Shenzhen Research Institute
Email: {cslizhang, csdwang}@comp.polyu.edu.hk, wuweijie2@huawei.com

***Abstract*—Mobile data demand is increasing tremendously, and thus new pricing models are in urgent need. One promising new pricing scheme is the "sponsored data plan", i.e., end users may enjoy free access to contents from certain content providers, while these content providers will pay ISPs for corresponding traffic consumed by end users. Proven a number of advantages, the sponsored data plan is still in its infancy. In this paper, we explore some potential of further development of this plan. We extend the design space and propose the idea of *time-dependent sponsoring*, i.e, content providers can decide *when* to sponsor *how much* fractions of traffic. The key intuition is by migrating some traffic consumption from peak to valley times, bandwidth resources can be better utilized. We formulate a game model to study the interactions between the ISP, CPs and users, and derive the optimal sponsoring fractions over various times under this new plan. We show that all parties involved can benefit from this plan, and social welfare increases. We believe our proposal, i.e., time-dependent sponsoring, provides important insights to potential development of the sponsored data plan.**

I. INTRODUCTION

In the past years we have witnessed a tremendous growth of wireless data traffic. This poses huge burden to the Internet service providers (ISPs) since supporting such demand-supply gap requires large investments. The ISPs are only one stakeholder in the Internet. The two-sided Internet market can be captured in Fig. 1, where ISPs are in the middle, end users (EUs) are on one side and content providers (CPs) are on the other side. Facing the surging demands, usage-based plans start prevailing in wireless data markets over flat-rate unlimited plans. For example, Verizon Wireless charges users for \$20 per month for 2 GB amount of data [1]. Such usage-based plans are backed by FCC [2], yet they raise concerns from the CPs because they may intrinsically limit users' willingness to consume data content from the CPs, whose revenue heavily depends on user views. One core problem is the one-sided charge for end users, i.e., ISPs, in particular, the last-mile access ISPs, charge the users as their primary revenue resources. This leads to an unbalanced finance model as neither users want to increase their data consumption thus paying more, nor ISPs want to reduce their price. New pricing models have been proposed to vitalize the Internet market, among which the sponsored data plan (SDP) [3], [4] attracts special interests from both industry and academia.

SDP, or also called tool-free service, means that an ISP and CPs sign some form of contract, such that when end users

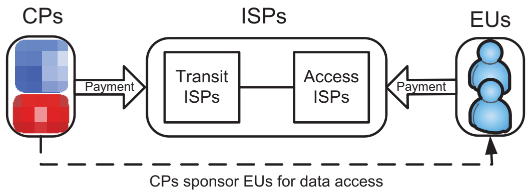


Fig. 1: Two-sided Internet market

access contents from CPs joining SDP, their traffic from/to these CPs will not be charged by the ISP. Instead, CPs will pay for that volume of traffic for end users to the ISP. Since its birth, SDP shows great potential to becoming a major charging pattern over the wireless data network. Intrinsically, SDP balances the finance model of CPs towards ISPs and users as it creates a positive cycle among the three parties: end users are willing to consume more traffic sponsored by CPs; CPs can attract more users and thus more advertisement income; ISPs can obtain more revenue by charging CPs. Thus, they may all benefit from this strategy.

SDP has also been in practice. For example, AT&T announced its sponsored data program in January 2014 [3]. Its sponsored data partner, Syntonic Wireless, launched a toll-free content store six months later [5]. Google has also joined with India's Bharti Airtel to offer free access to certain Google-based services such as Gmail, Google+ and first page of websites via Google search without ringing up data charges [6].

There are emerging studies [7], [8], [9] on SDP, and the research foci are the competition, benefit, equilibrium, fairness, the possible regulations needed, etc., of the Internet market under SDP. These studies have confirmed that SDP can lead to a more balanced finance model for the Internet market. None of the aforementioned works, however, study how to sponsor data. We argue that studies on appropriate sponsoring methods are also important, and even affect the overall success of SDP. In this paper, we propose and study time-dependent sponsoring (TDS). In TDS, CPs can decide the fraction of traffic to sponsor to end users for a given time, and that the fraction may vary over time. The main novelty of TDS is its potential to improve resource utilization.

It is non-trivial to analyze TDS. The key challenges include: 1) It is difficult to model users' behavior as they may differ under TDS; 2) Demands under different times can be correlated due to traffic migration under different subsidizations; and 3) The interactions among users, CPs and ISPs are complex, and we need appropriate models, comprehensive discussions and interpretations to capture them.

We provide a rigid study on TDS. We consider strategic users [10], [11] who may delay data consumption in exchange of a lower price. We establish a Stackelberg game model to capture the interactions among a monopoly ISP, a set of CPs and an arbitrary number of strategic users. We formulate the ISP's and CPs' decision problems as optimizations. We show that a CP's sponsoring decision problem is a non-convex optimization, and we propose a dynamic programming based algorithm in polynomial time. Our main findings include:

- TDS improves CPs' bandwidth utilization and profit, and users' welfare for slightly patient strategic users, but may result in controversial effects for highly patient ones;
- When CPs provide different subsidizations to different groups of users, CPs' bandwidth utilization and profit, as well as users' welfare, can be significantly improved;
- TDS can improve the ISP's capacity utilization, thus increasing the ISP's profit and the social welfare.

This is the outline of this paper. Sec. II states our related work. In Sec. III, we set up a Stackelberg game model to capture the interactions between users, CPs and the ISP. Sec. IV and V analyze CPs' and the ISP's optimal decisions respectively, as well as their impacts to the market. Sec. VI concludes this paper.

II. Related Work

The tremendous growth of wireless data traffic motivates research on better financing models. One category is sponsored data pricing (SDP). Authors in [9], [12] have suggested that SDP benefits all three parties via analyzing the interplay of the CPs, ISPs and end-users. For instance, our previous work [9] built a two-class service model and concluded that ISPs and end users can achieve a win-win trade under SDP if properly regulated. Other works [8], [13] studied the strategy on which content to subsidize, the competition of subsidization, etc. All these studies, in particular [8], have shown that SDP may vitalize the Internet growth. Intrinsically, SDP can establish a more balanced finance model. Nevertheless, interpretations from studies [7], [12] also indicate that SDP may lead to heavier data traffic; since it is the CPs' incentive to deliver more content to end users and it is the users' incentive to consume more when others are paying for them.

Another category is time-dependent pricing (TDP). In TDP, ISPs will set different prices of traffic to charge users at different times. The objective of TDP is to migrate traffic loads from peak times to off-peak times. Studies [14], [15] have shown the potential of TDP in reducing the peak time load, and studies [16], [17] have further discussed various schemes within TDP, e.g., flat-rate, metering, etc. Economists [10], [11] have also studied strategic users in TDP, who can delay their purchase for future prices. Simple TDP schemes have been adopted in practice. For instance, BSNL in India offers unlimited night time (2-8 am) downloads on a monthly data plan of RS 500 (or USD $10) [18]; in US, some ISPs have begun experimenting time dependent pricing plans [15]. Our proposal, i.e., TDS, is also partially inspired by TDP. It marries the advantages of SDP and TDP and further increases the benefits of ISPs and CPs. However, it significantly differs from TDP. TDP is a one-sided charge for end users and thus limits data consumption, while our proposal TDS is a two-sided market model and encourage data consumption. It may also be easier for end users to adopt TDS as compared to TDP, since CPs interact with users directly.

In this paper, we migrate the ideas of SDP and TDP, and propose a new charging scheme TDS. This might help the further exploration of design space in wireless network pricing.

III. General Model

In this section, we analyze the market with three parties: a set of CPs $\mathcal{N}$, a monopolistic ISP and a set of end users $\mathcal{M}$. The CPs provide services to end users. We assume that one CP supplies only one service. If a CP provides multiple services, we treat it as multiple virtual CPs. Later, we call a service s and a CP s interchangeably. We consider only one ISP which provides Internet access services to CPs and end users[1]. We define the capacity of the ISP, denoted as μ, as the bottleneck bandwidth of the connection. We use a triple $(\mathcal{N}, \mu, \mathcal{M})$ to represent the whole system.

A. Users' Traffic Demand

We consider a finite time horizon $[0, T]$ with slots $\{1, \cdots, T\}$, each time length being unified as 1. The sponsored price for each slot can be different, so end users can arrive at any slot and consume the traffic at that particular slot, or they can delay the usage to a later slot so as to save money. We denote the *maximal waiting time* as $K \in \{0, \cdots, L-1\}$, and the corresponding *waiting period* of users arriving at t as $S_t^K \triangleq \{t, \cdots, \min\{t+K, L\}\}$. We say that the users are *impatient* if $K = 0$, or *patient* otherwise. We consider heterogeneous users and denote the fraction of those having the maximal waiting time K as $g(K)$, and we have $\sum_{K=0}^{L-1} g(K) = 1$. Given service s, we denote the number of users that arrive at t as m_t^s. Let δ^s be the average per-user traffic volume during one slot for service s. The total (max-possible) traffic at slot t is $\theta_t^s \triangleq m_t^s \delta^s$ if all users are *impatient*.

The prices charged to CPs and end users by the ISP are time-independent. CPs are charged according to the maximal bandwidth they require. For instance, Netflix pays CDNs on a per-megabit-per-second-sustained model so as to guarantee the quality of service [19]. In contrast, end users are charged according to the total traffic volume they consume. Due to net neutrality consideration, we consider a unified price p for per unit traffic for all users, and another unified price q for per unit bandwidth for all CPs. The price charged to end users can be subsided by CPs. Each CP can provide different subsidizations

[1]In reality there might be access ISPs and transit ISPs to connect users and CPs, as illustrated in Fig. 1. However, in this paper we treat them as one entity. Profit sharing among them is out of the scope of this paper.

at different slots. For any CP s, we denote its subsidization of per unit traffic at time t as h_t^s.

If a user arrives at t and targets at service s, she can consume the traffic anytime during the waiting period S_t^K. If the user receives the service, we assume she obtains a valuation v for per-unit traffic. Users may have different valuations, and we assume the valuation v for service s follows a probability density function $f_s(\cdot)$, and the corresponding cumulative function is $F_s(v) \triangleq \int_0^v f_s(x)dx$. Then a particular user's utility for service s at slot k, denoted as $u_t^s(k)$, is:

$$u_t^s(k) = v - (p - h_k^s). \tag{1}$$

A user consumes service s iff she can obtain a non-negative utility, i.e., $\max_{k\in S_t^K}\{u_t^s(k)\} \geq 0$. It indicates that only if a user has $v \geq p - \max_{k\in S_t^K}\{h_k^s\}$, then she will consume service s; further, she will choose the optimal slot that maximizes her utility. Denote this optimal time slot as ϖ_t^s, we have:

$$\varpi_t^s = \arg\max_{k\in S_t^K}\{h_k^s\}. \tag{2}$$

It indicates that users always choose the slot with the maximal subsidization. When two slots have the same utility, we assume a user prefers the earlier slot to break the tie. Then, given the subsidization of service s, i.e., $\mathbf{h}^s = (h_1^s, \cdots, h_L^s)$, and the price charged by the ISP to end users, i.e., p, the maximal possible demand for service s during time slot t, denoted as ρ_t^s, can be expressed as:

$$\rho_t^s(\mathbf{h}^s) = \sum_{K=0}^{L-1} \sum_{l:t\in S_l^K} g(K)\theta_l^s \mathcal{I}\{t = \varpi_l^s\}, \tag{3}$$

where $\mathcal{I}$ is an indicator function. Then, the actual demand, denoted as D_t^s, becomes:

$$D_t^s(\mathbf{h}^s, p) = \big(1 - F_s(p - h_t^s)\big)\rho_t^s(\mathbf{h}^s). \tag{4}$$

B. Utility of CPs

We use r^s to denote the per unit revenue of CP s. CPs may have very different per unit revenues. The revenue can be generated by advertisements (e.g., YouTube), value-added services (e.g., Tencent), e-commerce (e.g., Amazon), etc. The cost of CP s consists of two parts: 1) the cost of sponsored data, i.e., $\mathbf{h}^s$ for per unit traffic, and 2) the price charged by the ISP for bandwidth, i.e., q for per unit bandwidth. We denote the required bandwidth for CP s as λ^s.[2] Thus, the utility of CP s, denoted by Φ^s, is:

$$\Phi^s(\mathbf{h}^s, \lambda^s) = \sum_{t=1}^{L}(r^s - h_t^s)D_t^s(\mathbf{h}^s, p) - q\lambda^s, \tag{5}$$

[2]When the demand of one CP is higher than the required bandwidth, the traffic of this CP will be throttled by the ISP without extra payment.

where p and q are the prices charged by the ISP to end users and CPs, respectively. Thus, the optimal subsidization and required bandwidth for CP s can be determined by:

$$\begin{aligned} \text{OPT-1:} \quad \max_{\{\mathbf{h}^s, \lambda^s\}} \quad & \Phi^s(\mathbf{h}^s, \lambda^s) \\ \text{s.t.} \quad & D_t^s(\mathbf{h}^s, p) \leq \lambda^s, \forall t \in \{1, \cdots, L\}, \quad (6) \\ & \mathbf{0}_{1\times L} \preceq \mathbf{h}^s \preceq r^s \mathbf{1}_{1\times L}. \quad (7) \end{aligned}$$

Let us consider an extreme case where $q = 0$. Any CP $s \in \mathcal{N}$ sets a large enough required bandwidth and a single subsidization h^s to all slots that satisfy $h^s = \arg\max_{h^s\geq 0}(r^s - h^s)\big(1 - F_s(p - h^s)\big)$, making CP s achieve its *maximum possible revenue*, i.e., $\max_{h^s\geq 0}(r^s - h^s)\big(1 - F_s(p - h^s)\big)\sum_{k=1}^{L}\theta_k^s$. We define the monopoly revenue function as $H_s(h^s) = (r^s - h^s)\big(1 - F_s(p - h^s)\big)$. To simplify our analysis, we make the following assumption.

Assumption 1 (Unimodal Property)**.** *There exist some monopoly price, denoted as h_M^s that $H_s(\cdot)$ is increasing for all $h^s < h_M^s$ and decreasing for all $h^s > h_M^s$.*

Assumption 1 can be satisfied under a wide range of distributions, e.g., uniform and exponential distributions. Under assumption 1 and negligible q, the subsidization for any slot is just the monopoly price if $h_M^s \geq 0$, or 0 otherwise. We also assume that the monopoly price is non-decreasing with respect to p. When the ISP charges a higher price for per unit traffic, CPs will not reduce their subsidization. In practice, ISPs always charge CPs a non-negligible price q. In our technical report [20], we use one example to demonstrate that OPT-1 is a non-convex optimization problem under such case.

C. Utility of the ISP

We use the ISP's revenue to represent its utility, which is from two sources: 1) the unit price charged to CPs for the connection services, i.e., q, and 2) the unit price charged to end users, i.e., p. Under TDS, the price charged to end users can be partially subsidized by CPs but the total revenue per unit traffic keeps the same. Thus, the utility[3] (or revenue) of the ISP, denoted by Π, is:

$$\Pi(p, q) = p\sum_{s=1}^{N}\sum_{t=1}^{L} D_t^s(\mathbf{h}^s, p) + q\sum_{s=1}^{N}\lambda^s. \tag{8}$$

The ISP decides its optimal prices by solving:

$$\begin{aligned} \text{OPT-2:} \quad \max_{\{p,q\}} \quad & \Pi(p, q) \\ \text{s.t.} \quad & \sum_{s=1}^{N} D_t^s(\mathbf{h}^s, p) \leq \mu, \forall t \in \{1, \cdots, L\}, \\ & p \geq 0, \quad q \geq 0. \end{aligned}$$

[3]We only consider fixed cost for the ISP and ignore the marginal cost. We further let the fixed cost be zero because it is a constant and will not impact the result of the optimization.

D. A Two-stage Stackelberg Game

We model the interactions of the ISP and CPs $\mathcal{N}$ as a two-stage Stackelberg game. In particular, we consider:

- *Players*: The ISP and the set of CPs $\mathcal{N}$.
- *Strategies*: The ISP decides the unit prices charged to end users for traffic, and to CPs for bandwidth, i.e., its strategy profile is $S_I \in \{(p, q)\}$. Each CP s decides the price subsidization and the required bandwidth, i.e., its strategy profile is $S_s \in \{(\mathbf{h}^s, \lambda^s)\}$.
- *Rules*: The ISP is the first mover and decides S_I. CPs are the second movers and decide $S_s, \forall s \in \mathcal{N}$. Each CP makes its own decision independently.
- *Outcome*: The outcome is determined by backward induction. In particular, for any given S_I, each CP s decides $S_s, \forall s \in \mathcal{N}$ by solving OPT-1. Based on this knowledge, the ISP decides S_I by solving OPT-2.

Note that by using the Stackelberg game model, we assume the ISP is the first mover and the CPs are the second movers. This is the reality in many countries or regions. ISPs usually know ex ante that CPs would observe their actions, e.g., new pricing strategy, and make optimal decisions based on their actions. When the ISP fixes its prices charged to CPs and end users, CPs decide their optimal price subsidization independently. The decision of one particular CP would not be affected by other CPs' decisions. Thus, we can analyze each CP's price subsidization separately. Based on these, we first analyze one particular CP's subsidization in Section IV and then the ISP's optimal choice in section V.

IV. CPs' Subsidization Strategy

In this section, we analyze the CPs' optimal strategy, i.e., the second stage of the Stackelberg game. We consider the optimal subsidization and required bandwidth of one particular CP, i.e., CP s, under impatient users, i.e., $K = 0$, and patient users, i.e., $K \geq 1$. In both cases, we first analyze the optimal subsidization given the required bandwidth. After that, we study the optimal required bandwidth. Before the analysis of CPs for the two cases, we first consider some general characteristics for OPT-1.

Lemma 1. *Given the required bandwidth λ^s, the optimal price subsidization h_t^* in slot t can only be one of the following cases: 1) $h_t^* = 0$; 2) $h_t^* = h_M^s$; and 3) $h_t^* = h_{\hat{t}}^s$ such that $D_{\hat{t}}^s(\boldsymbol{h}^*, p) = \lambda^s$ and $h_{\hat{t}}^s \in (0, h_M^s)$.*

Proof. Please refer to our technical report [20]. □

Lemma 1 shows three possibilities for the optimal price subsidization. Heavy traffic demand leads to no subsidization, i.e., $h_t^* = 0$, while light traffic demand leads to the highest subsidization, i.e., $h_t^* = h_M^s$. Besides these two cases, partial subsidization is adopted such that the required bandwidth can be fully utilized, i.e., $D_{\hat{t}}^s(\mathbf{h}^*, p) = \lambda^s$. However, it is difficult even to know which case the traffic demand belongs to since it is determined by the relative subsidizations under different times. To simplify the analysis, let us introduce the concept of preference ranking.

Definition 1 (Preference Ranking)**.** *Given any price subsidization $\boldsymbol{h}$, the preference ranking $\mathcal{R} = \{R_1, \cdots, R_L\}$ is a permutation of $\{1, \cdots, L\}$ that satisfies:*

$$R_i \begin{cases} < R_j & \text{if } h_i < h_j, \\ > R_j & \text{if } h_i \geq h_j, \end{cases} \tag{9}$$

for any $i \in \{1, \cdots, L-1\}$ and $j \in \{i+1, \cdots, L\}$.

Definition 1 states that users prefer higher subsidization and early time slot. When $\mathcal{R}$ is given, the traffic demand can also be determined, i.e., $D_t^s(\mathbf{h}, p; \mathcal{R}) = \big(1 - F_s(p - h_t)\big)\rho_t^s(\mathcal{R})$. Using this concept, we can analyze the optimal subsidization further by proposition 1.

Proposition 1. *Given the required bandwidth λ^s and preference ranking $\mathcal{R}$, the optimal price subsidization is:*

$$h_t^* = \begin{cases} h_M^s & \text{if } \rho_t^s < \frac{\lambda^s}{1 - F_s(p - h_M^s)}, \\ \max\left\{0, p - F_s^{-1}(1 - \lambda^s/\rho_t^s)\right\} & \text{otherwise}, \end{cases} \tag{10}$$

where $F_s^{-1}(\cdot)$ is the inverse function of $F_s(\cdot)$.

Proof. Please refer to our technical report [20]. □

Now we will go further into details of the optimal subsidization. Let us discuss when users are *patient* or *impatient*.

A. Impatient Users ($K = 0$)

When $K = 0$, all users are impatient, i.e., $\rho_t^s = \theta_t^s$. This usually happens for real time services like live telecast videos. Given the required bandwidth λ^s, the optimal subsidization becomes:

$$h_t^* = \begin{cases} h_M^s & \text{if } \theta_t^s < \frac{\lambda^s}{1 - F_s(p - h_M^s)}, \\ \max\left\{0, p - F_s^{-1}(1 - \lambda^s/\theta_t^s)\right\} & \text{if } \theta_t^s \geq \frac{\lambda^s}{1 - F_s(p - h_M^s)}. \end{cases} \tag{11}$$

Given the optimal subsidization, we then analyze the optimal required bandwidth. We define $\tilde{\theta}_i = \big(1 - F_s(p - h_M^s)\big)\theta_i^s$ and rearrange the time slots such that $\tilde{\theta}_i < \tilde{\theta}_j$ if $i < j$. We also add one dummy slot $t = 0$ with traffic demand $\tilde{\theta}_0 = \big(1 - F_s(p)\big)\theta_1^s$. Note that when $\lambda^s < \tilde{\theta}_0$, $\mathbf{h} = \mathbf{0}$; when $\lambda^s > \tilde{\theta}_L$, $\mathbf{h} = h_M\mathbf{1}$. Then, the optimal required bandwidth, denoted as λ^*, should appear within interval $[\tilde{\theta}_0, \tilde{\theta}_L]$. We divide this interval into several subintervals $[\tilde{\theta}_{l-1}, \tilde{\theta}_l] (l \in \{1, \cdots, L\})$. If $\lambda^s \in [\tilde{\theta}_{l-1}, \tilde{\theta}_l]$, then the bandwidth is under-utilized for any $t \in \{1, \cdots, l-1\}$ and fully-utilized for any $t \in \{l, \cdots, L\}$. Let us define the concept traffic revenue as all CPs' total revenue minus the fees paid to the ISP for traffic subsidization, and denote its value during relabeled slots $\{i, \cdots, j\}$ as $\phi_{i,j}$. Then, we can divide the traffic revenue of CP s into two parts: 1) the traffic revenue from all under utilized slots, i.e.,

$$\phi_{1,l-1} = (r^s - h_M^s) \sum_{i=1}^{l-1} \tilde{\theta}_i, \tag{12}$$

and 2) the traffic revenue from all fully utilized slots, i.e.,

$$\phi_{l,L}(\lambda^s) = \lambda^s \sum_{i=l}^{L} \left[r^s - \max\{0, p - F_s^{-1}(1 - \lambda^s/\theta_i^s)\}\right]. \tag{13}$$

Thus, we can simplify OPT-1 as:

$$\text{OPT-3:} \max_{\{\lambda^s, l\}} \quad \phi_{1,l-1} + \phi_{l,L}(\lambda^s) - q\lambda^s$$
$$\text{s.t.} \quad \tilde{\theta}_{l-1} \le \lambda^s \le \tilde{\theta}_l, l \in \{1, \cdots, L\}. \quad (14)$$

To determine the optimal required bandwidth, we introduce a new variable l and separate the domain region $[\tilde{\theta}_0, \tilde{\theta}_L]$ into several ones, i.e., $[\tilde{\theta}_{l-1}, \tilde{\theta}_l](l \in \{1, \cdots, L\})$. The above optimization can be solved by first fixing l and obtaining local optimal required bandwidth λ_l^* and then choosing the global optimal λ^* from $\{\lambda_1^*, \cdots, \lambda_L^*\}$. Note that even when we fix l, the above optimization may still be non-convex. The following lemma gives the conditions to have a convex problem.

Lemma 2. *If the monopoly function $H_s(\cdot)$ is concave in $[0, h_M^s]$ and $F_s(\cdot)$ is a concave function, then the above optimization is convex given a fixed value of l.*

Proof. Please refer to our technical report [20]. □

The conditions in Lemma 2 guarantee that the revenue $\phi_{l,L}(\lambda)$ from fully-utilized slots is concave. This makes OPT-3 convex for any fixed l. Besides, the conditions also guarantee the CP's utility $\Phi^s(\lambda)$ is concave in $[\tilde{\theta}_0, \tilde{\theta}_L]$. Under these conditions, we can obtain the optimal required bandwidth by the following theorem.

Theorem 1. *Assume the conditions in Lemma 2 satisfy. If there exists an l such that $\frac{\partial \phi_{l+1,L}}{\partial \lambda^s}|_{\lambda^s = \tilde{\theta}_l + \epsilon} \le q \le \frac{\partial \phi_{l,L}}{\partial \lambda^s}|_{\lambda^s = \tilde{\theta}_l - \epsilon}$ for any small ϵ, then $\lambda^* = \tilde{\theta}_l$; otherwise, there exists an l and $\lambda^* \in [\tilde{\theta}_{l-1}, \tilde{\theta}_l]$ such that $\frac{\partial \phi_{l,L}}{\partial \lambda^s}|_{\lambda^s = \lambda^*} = q$.*

Proof. Please refer to our technical report [20]. □

Theorem 1 shows the optimal required bandwidth is either the end point of some subinterval, e.g., $\tilde{\theta}_l$, that makes $\frac{\partial \Phi^s}{\partial \lambda^s}|_{\lambda^s = \tilde{\theta}_l - \epsilon} \ge 0$ and $\frac{\partial \Phi^s}{\partial \lambda^s}|_{\lambda^s = \tilde{\theta}_l + \epsilon} \le 0$ for any small ϵ, or the point within some subinterval $[\tilde{\theta}_{l-1}, \tilde{\theta}_l]$ that makes $\frac{\partial \Phi^s}{\partial \lambda^s}|_{\lambda^s = \lambda^*} = 0$, i.e., $\frac{\partial \phi_{l,L}}{\partial \lambda^s}|_{\lambda^s = \lambda^*} = q$.

B. Patient Users ($K \ge 1$)

When users are patient, i.e., $K \ge 1$, the price subsidization may not be determined easily even if the required bandwidth is given. This problem is non-convex and we will prove it in our technical report [20]. In this section, we design a dynamic algorithm to obtain the optimal price subsidization with polynomial time complexity. Let us first introduce the definition of peak slot.

Definition 2 (Peak slot)**.** *A slot t is a peak slot in the period $\{i, \cdots, j\}$ if $R_t > R_s$ for any $s \in \{i, \cdots, j\} \setminus \{t\}$.*

A user prefers consuming the content in a peak slot, compared to other slots in period $\{i, \cdots, j\}$. The concept of peak slot helps us design dynamic algorithms that separate the original problem into subproblems.

Consider the optimal subsidization $\mathbf{h}^*$ with preference ranking $\mathcal{R}^*$. Suppose slot t is the peak slot in period $\{1, \cdots, L\}$ under $\mathcal{R}^*$. Thus, we have $R_t^* = L$. Therefore, users with the maximal waiting time K arriving at any slot during $\{\max\{t-K, 1\}, \cdots, t-1\}$ will delay their consumption to slot t; thus, the potential demand for slot t is $\rho_t^s(\mathcal{R}^*) = \sum_{K=0}^{L-1} \sum_{l: t \in S_l^K} g(K)\theta_l^s$. If we know the peak slot t and h_t^*, then we can separate OPT-1 into two subproblems: maximizing the traffic revenue during $\{1, \cdots, t-1\}$ and that during $\{t+1, \cdots, T\}$. Note that the population of users in the former subproblem with the maximal waiting time K arriving at any slot $l \in \{\max\{t-K, 1\}, \cdots, t-1\}$ is not $g(K)\theta_l^s$ any more. Since all such users delay their traffic to the peak slot t, the population becomes zero instead. Therefore, we need to redefine the subproblems.

In our algorithm, we consider the subproblem as maximizing the traffic revenue during $\{i, \cdots, j\}$, with traffic from users with the maximal waiting time K arriving at slot l being:

$$\bar{\theta}_l^K = \begin{cases} 0 & \text{if } j \ne T, \max\{i, j-K+1\} \le l \le j, \\ g(K)\theta_l^s & \text{otherwise.} \end{cases} \quad (15)$$

We denote the optimal traffic revenue during $\{i, \cdots, j\}$ as $W(i, j, \bar{h})$, where $\bar{h}$ is the subdiziation during peak slot for its original problem. Note that $\bar{h}$ is also the upper bound for all subsidizations during $\{i, \cdots, j\}$. We extend the definition of waiting period to sub-period $\{i, \cdots, j\}$ denoted as $S_t^K(i,j) \triangleq \{t, \cdots, \min\{t+K, j\}\}$. Then, we have:

$$W(i, j, \bar{h}) \triangleq \max_{\{\mathbf{h}\}} \sum_{t=i}^{j} (r^s - h_t)\bar{D}_t^s(\mathbf{h}, p)$$
$$\text{s.t.} \quad \bar{D}_t^s(\mathbf{h}, p) \le \lambda^s, \forall t \in \{1, \cdots, L\}, \quad (16)$$
$$\mathbf{0}_{1\times L} \preceq \mathbf{h} \preceq v^s \mathbf{1}_{1\times L}, \quad (17)$$
$$h_t \le \bar{h}, \forall t \in \{i, \cdots, j\}, \quad (18)$$

where $\bar{D}_t^s = \big(1 - F_s(p - h_t)\big) \sum_{K=0}^{L-1} \sum_{l: t \in S_l^K(i,j)} \bar{\theta}_l^K \mathcal{I}\{t = \varpi_l^s\}$. Note that given the required bandwidth λ^s, the optimal subsidization of OPT-1 is equivalent to that of $W(1, T, h_M^s)$. In addition, $W(i, i, \bar{h}) = g(0)\theta_i^s \max_{h \in [0, \bar{h}]} H_s(h)$. Then, we have:

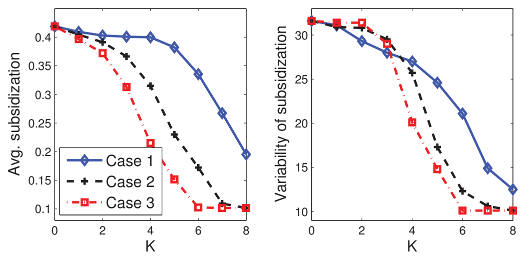
$$W(i, j, \bar{h}) = \max_{k \in \{i, \cdots, j\}} \Big\{ \max_{h \in [0, \bar{h}]} \big\{ W(i, k-1, h) + \gamma_{i,j}^k(h) + W(k+1, j, h) \big\} \Big\}, \quad (19)$$

where

$$\gamma_{i,j}^k(h) = \begin{cases} \min\{\sum_{K=0}^{L-1} \sum_{l: t \in S_l^K(i,k)} \bar{\theta}_l^K H_s(h), \lambda^s\} \\ \text{if } h \le \max\{0, p - F^{-1}(1 - \lambda^s / \bar{\rho}(i,j,k))\}, \\ -\infty \quad \text{otherwise,} \end{cases} \quad (20)$$

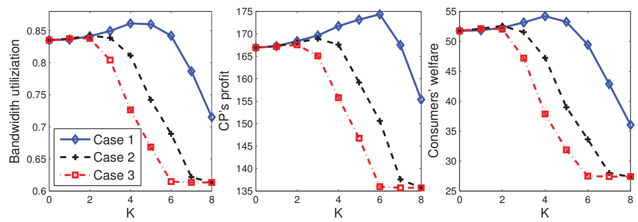
and $\bar{\rho}(i,j,k) = \sum_{K=0}^{L-1} \sum_{l: t \in S_l^K(i,k)} \bar{\theta}_l^K$.

To understand why the recursion in Eq. 19 holds, we consider the optimal subsidization of CP s and assume that $k \in \{i, \cdots, j\}$ is the peak slot with corresponding subsidization $h_k \le \bar{h}$. Then, the users with the maximal waiting time K arriving at any slot during $\{\max\{i, k-K\}, \cdots, k\}$ will delay their consumption to slot k, and thus the demand in slot k is $\sum_{K=0}^{L-1} \sum_{l: t \in S_l^K(i,k)} \bar{\theta}_l^K \big(1 - F_s(p - h_k)\big)$ constrained by the required bandwidth. Thus we can obtain the traffic revenue at slot k as $\gamma_{i,j}^k(h_k)$. Since users delay consumption to slot k, the



(a) (b)

Fig. 2: Subsidizations under various K



(a) (b) (c)

Fig. 3: Effects of the maximal waiting time K

traffic from users with the maximal waiting time K arriving at any slot during $\{i, \cdots, k-1\}$ is:

$$\bar{\theta}_l^K = \begin{cases} 0 & \text{if } \max\{i, k-K\} \le l \le k-1, \\ g(K)\theta_l^s & \text{otherwise.} \end{cases}$$

By definition, the traffic revenue obtained during $\{i, \cdots, k-1\}$ is exactly $W(i, k-1, h_k)$. We then consider the period $\{k+1, \cdots, j\}$. Recall Eq. 15 and that the traffic demand during $\{k+1, \cdots, j\}$ holds unchanged after separation, so the traffic revenue obtained during $\{k+1, \cdots, j\}$ is also $W(k+1, j, h_k)$. Thus, we have $W(i, j, \bar{h}) = W(i, k, h_k) + \gamma_{i,j}^k(h_k) + W(k, j, h_k)$.

Note that $W(i, j, h)$ is non-decreasing in h. This is also true for $\gamma_{i,j}^k(h)$ if $h \le \max\left\{0, p - F^{-1}\left(1 - \lambda^s/\bar{\rho}(i, j, k)\right)\right\}$. Then, the optimal subsidization during $[0, \bar{h}]$ in Eq. 19 is:

$$\bar{h}_{i,j}^k = \min\left\{\bar{h}, \max\left\{0, p - F^{-1}\left(1 - \lambda^s/\bar{\rho}(i, j, k)\right)\right\}\right\}. \quad (21)$$

Thus, we can simplify Eq. 19 as:

$$W(i, j, \bar{h}) = \max_{k \in \{i, \cdots, j\}} \left\{W(i, k, \bar{h}_{i,j}^k) + \gamma_{i,j}^k(\bar{h}_{i,j}^k) + W(k, j, \bar{h}_{i,j}^k)\right\}.$$

Let us state the complexity to solve OPT-1.

Theorem 2. *Given λ^s, the time complexity of solving OPT-1 is $O(L^6)$.*

Proof. Please refer to our technical report [20]. □

This shows we can obtain the optimal subsidization for the non-convex optimization OPT-1 in polynomial time $O(L^6)$. Using this dynamic algorithm, we search the optimal required bandwidth by a linear search algorithm $LSearchAlg()$ (please refer to our technical report [20] for details).

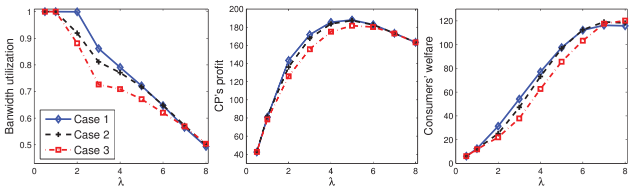
C. Numerical Illustrations

To intuitively demonstrate CPs' subsidization, we provide a numerical example. We divide users into impatient and patient groups with a population ratio $m_1 : m_2$. We consider three cases: 1) $m_1 : m_2 = 3 : 1$, and 2) $m_1 : m_2 = 1 : 1$, and 3) $m_1 : m_2 = 1 : 3$. We consider homogeneous K for patient users and set $K = 4$ by default. Users' per unit valuation of traffic follows a uniform distribution $U([0, 2])$. We analyze two days and define a slot as one hour. The potential traffic pattern of each day is set as: $\theta_t = t + \delta$ if $t \le 12$ and $\theta_t = 25 - t + \delta$ if $13 \le t \le 24$, where δ follows a Gaussian distribution $G(0, 1)$. The revenue of per unit traffic is $r = 2$. The prices charged to users and the CP are $p = 1.5$ and $q = 10$, respectively. The required bandwidth is $\lambda = 3$ unless otherwise specified.

1) Effects of Strategic Behaviors: Users' decision of delaying the consumption for higher subsidization is referred to as *strategic behaviors*, and we will investigate their effects on subsidization. In Fig. 2(a), we show the effect of the maximal waiting time on the average prices of subsidization. When K increases, the average subsidization prices decreases. When the ratio of patient users becomes higher, the average subsidization decreases more rapidly.

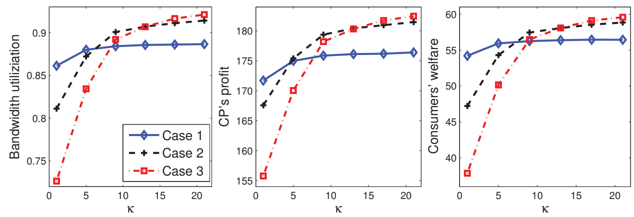
Another important measure of subsidization is its *variability*, which we define as the total number of distinct values that are taken for subsidization prices over the whole period. In extreme cases, the variability of subsidization is 1 (or T) when subsidization prices during all slots take the same (or all different) value(s). The variability, to some extent, reflects the profit gain of subsidization. Intuitively, if its value is small, then the CP does not have many choices for subsidization, and the profit gain is not hopefully high. In Fig. 2(b), we show that the variability is decreasing in K. The intuition is that when a CP provides a slightly higher subsidization in some slot, many patient users will delay their traffic to that slot, resulting in a huge increase of traffic demand and thus high incentives for the CP to reduce such subsidization. When K is larger, the effect becomes more obvious, resulting in lower variability, and this is what CPs are unwilling to see.

In Fig. 3, we show how strategic behaviors affect the bandwidth utilization, CP's profit and users' welfare. Here, we define the users' welfare (aka consumers' welfare) as the difference between their valuation towards the service and their payment. Intuitively, one may guess that when users are patient, the CP provides high subsidizations in slots with low potential demands to attract patient users, who should have otherwise consumed traffic in busy slots, and thus the bandwidth should be better utilized. Indeed, this conjecture can be verified to some extend when the maximal waiting time is not high, e.g., $K \le 4$ for case 1 and $K \le 2$ for case 2 and 3, as shown in Fig. 3(a). However, often times we observe an opposite effect, i.e., low bandwidth utilization due to strategic behaviors. This is because when users become more patient (i.e., larger K and/or smaller $m_1 : m_2$), the CP is forced to provide lower subsidization prices and variability, shown previously in Fig. 2, thus resulting in inefficient bandwidth usage. This hurts both the CP and users (Fig. 3(b) and



(a) (b) (c)

Fig. 4: Effects of the required bandwidth λ



(a) (b) (c)

Fig. 5: Effects of the number of groups κ

Fig. 3(c)). For instance, when $m_1 : m_2 = 1 : 3$ and $K = 6$, the CP's profit and users' welfare reduce by 19% and 47%, respectively, compared to $K = 0$. In later simulations, we demonstrate how the strategy of staggered subsidizations can avoid such phenomenon.

2) Effects of Required Bandwidth: In Fig. 4 we show the effects of required bandwidth on bandwidth utilization, CP's profit and users' welfare. Figure 4(a) shows that when the required bandwidth is very small, e.g., $\lambda = 1$, the bandwidth can be fully utilized. As the required bandwidth increases, the bandwidth utilization decreases. Note that the traffic demand always increases with respect to the required bandwidth since more traffic usage indicates more traffic revenue. Figure 4(b) shows that the CP's profit first increases (due to increase of demand) and then decreases (due to high bandwidth cost) with respect to the required bandwidth. The optimal required bandwidth for this CP is around $\lambda = 4.7$. Figure 4(c) shows that the users' welfare always increases with respect to the required bandwidth.

3) Effects of Staggered Subsidizations: As we discuss previously, homogeneous strategy behavior leads to inefficient bandwidth utilization when users are patient. To avoid the inefficiency, we adopt staggered subsidizations, i.e., to differentiate users into multiple groups with different subsidization strategies. Specifically, we divide users into κ groups and decide the subsidization sequently. The subsidization of one group is determined by the algorithm $DynamicAlg()$, but the bandwidth limitation (or the required bandwidth as defined previously) is the remaining amount after being occupied by previous groups.

In Fig. 5, we show the effects of the number of groups on bandwidth utilization, CP's profit and users' welfare . Figure 5(a) shows that the bandwidth utilization increases significantly with the number of groups. In addition, a larger population ratio of patient users means higher bandwidth utilization. For instance, under the case 3, i.e., $m_1 : m_2 = 1 : 3$, the bandwidth utilization increases from 73% to 92% when the numbr of groups increases from $\kappa = 1$ to $\kappa = 21$. Note that this increase is not obvious when the number of groups is large, e.g, $\kappa = 5$ for case 1. Thus, we only need to decide subsidizations for small number of groups so as to avoid the homogeneous strategy behavior for patient users. Moreover, the efficient usage of bandwidth due to staggered subsidizations benefits both users and the CP, shown in Fig. 5(b) and 5(c).

Summary: When we analyze the CPs with strategic users, the maximal waiting time effects CPs' optimal subsidization significantly. A higher maximal waiting time indicates a lower average subsidization price and a lower variability. In addition, a higher maximal waiting time does not always improve the bandwidth utilization and increase CPs' profit. This only happens when users are slightly patient, e.g., $K = 4$ and $m_1 : m_2 = 3 : 1$. Highly patient users lead to serious homogeneous strategic behavior and thus reduce bandwidth utilization. This homogeneous strategic behavior can be avoided when CPs stagger the subsidizations for different groups. A larger number of groups indicates higher bandwidth utilization and thus higher profit of CPs and welfare of users.

V. Monopolistic ISP's Strategy

In the previous section, we have analyzed the second stage of the Stackelberg game, i.e., CPs' choice on optimal subsidization and required bandwidth. In this section, we consider the first stage of the Stackelberg game, i.e., the ISP's strategy. We analyze the effects of ISP's strategy under two cases: homogeneous and heterogeneous CPs.

A. Homogeneous CPs

Let us start by considering homogeneous CPs, i.e., the per unit revenue ($r^s = r$) and the traffic pattern ($\theta_t^s = \theta_t$) are the same for all CPs. These CPs provide the same subsidization and required bandwidth, and thus simplify our analysis. In this case, we can simplify the ISP's optimization OPT-2 as:

$$\text{OPT-4:} \quad \max_{\{p,q\}} \quad p\sum_{t=1}^{L} D_t(\mathbf{h}, p) + q\lambda(p,q)$$

$$\text{s.t.} \quad \lambda(p,q) \le \mu/N, \tag{22}$$

$$p \ge 0, \quad q \ge 0. \tag{23}$$

Now we can show the effect of expending the capacity in the following theorem.

Theorem 3 (Effects of capacity)**.** *Denote the ISP's optimal strategy in the system $(\mathcal{N}, \mu, \mathcal{M})$ and $(\mathcal{N}, \mu', \mathcal{M})$ as (p, q) and (p', q'), respectively. If $\mu' \ge \mu$, then $\Pi' \ge \Pi$. Moreover, if $\mathcal{R}' = \mathcal{R}$, then at least one of the following conditions hold: (a) $q' \le q$, or (b) $p' \le p$.*

Proof. Please refer to our technical report [20]. □

Theorem 3 indicates that when we increase the capacity, the ISP reduces the price of per unit bandwidth charged to CPs,

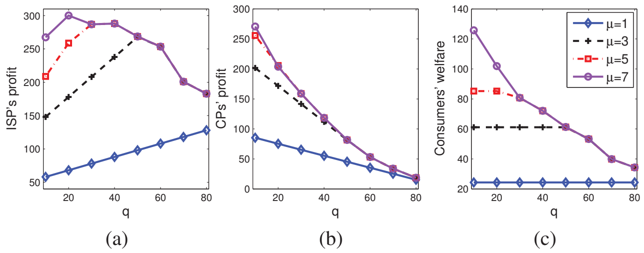


Fig. 6: Effects of q with homogeneous CPs

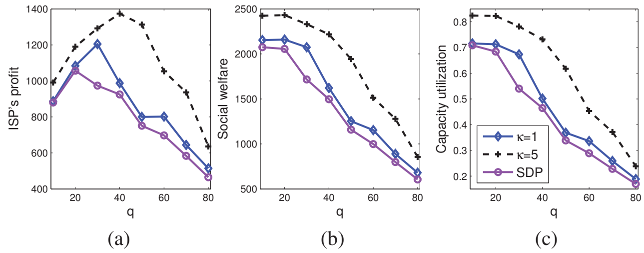


Fig. 8: Effects of q with heterogeneous CPs

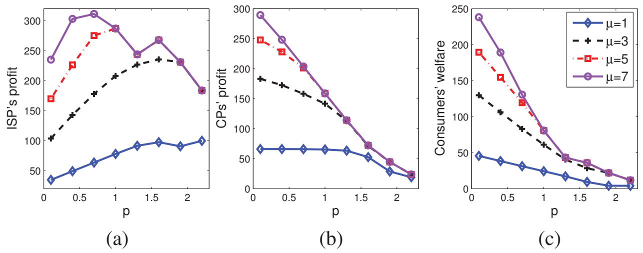


Fig. 7: Effects of p with homogeneous CPs

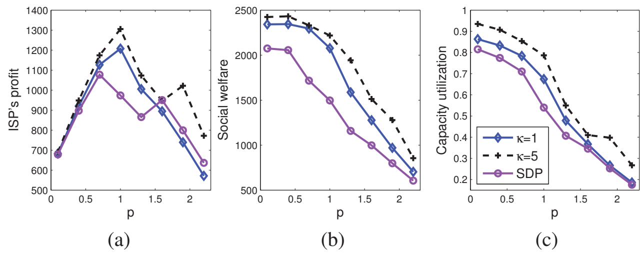


Fig. 9: Effects of p with heterogeneous CPs

or the price of per unit traffic charged to users, or both, so as to obtain a higher profit. In addition, the ISP always benefits from the capacity extension. Then, we consider the effects of ISP's strategy on CPs.

Theorem 4 (Effects of ISP's strategy)**.** *Given any two strategies that satisfy $(p',q') \succeq (p,q)$, we have $\Phi' \leq \Phi$. In addition, if $\mathcal{R}' = \mathcal{R}$, then $\lambda' \leq \lambda$, $h'_t \leq h_t, \forall t \in \{1,\cdots,L\}$.*

Proof. Please refer to our technical report [20]. □

Theorem 4 shows that when the ISP increases its price charged to CPs and/or users, CPs' profit reduces. Moreover, when the rankings of subsidization are the same, CPs reduce their subsidization and their required bandwidth.

To intuitively understand the effects of ISP's strategy, let us consider the following example. Assume the traffic pattern, users' valuation and CPs' revenue are the same as in the Sec. IV-C. We set $m_1 : m_2 = 1 : 3$ and $K = 4$. The prices charged to users and CPs are $p = 1$ and $q = 30$ respectively unless otherwise specified.

In Fig. 6, we investigate the effects of various prices of per unit bandwidth on the ISP's and CPs' profits, and users' welfare . Figure 6(a) indicates that the ISP's profit first increases and then decreases with respect to q. Given the value of μ, there exists such a value q (e.g., $q = 20$ under $\mu = 7$) that maximises the ISP's profit. In addition, a larger capacity indicates a higher profit of the ISP. A small capacity, e.g., $\mu = 3$, may limit the decision for the optimal price of per unit bandwidth. Figure 6(b) shows that CPs' profit decreases significantly with respect to q. For instance, when q increases from 10 to 40, CPs' profit decreases from 271 to 118, or by nearly 56%. A larger capacity also indicates a higher profit of CPs. Moreover, when the capacity is larger, CPs' profit also decreases more rapidly with respect to q. Figure 6(c) shows that users' welfare is non-increasing with respect to q. When q is small, e.g., $q = 20$ under $\mu = 3$, users' welfare keeps a constant with respect to q since the demand is constrained by the capacity. Moreover, a larger capacity indicates a higher welfare of users.

In Fig. 7, we investigate the effects of various prices of per unit traffic on the ISP's and CPs' profit, and users' welfare. Figure 7(a) shows that there exists a price of per unit traffic (e.g., $p = 0.7$ under $\mu = 7$) that maximizes the ISP's profit. A larger capacity indicates a higher profit of the ISP. Moreover, the ISP's profit increases with respect to p when the capacity becomes the constraint, e.g., $\mu = 1$. Figure 7(b) shows CPs' profit decreases concavely with respect to p. For instance, when $\mu = 1$, the CPs' profit keeps almost unchanged when p increases from 0.1 to 1.3, but decreases by 55% when p increases from 1.3 to 1.9. Figure 7(c) shows that users' welfare decreases significantly with respect to p. When the capacity is larger, the decreasing trend becomes more rapidly.

B. Heterogeneous CPs

In this subsection, we analyze the effects of ISP's strategy on the capacity utilization[4], ISP's profit, social welfare under heterogeneous CPs via simulations. We define the social welfare as the sum of all CPs' profit, ISP's profit and users' welfare. We consider five heterogeneous CPs in the simulations. Each CP s has a random phase displacement ξ^s from $G(0, 2^2)$ and random amplitude of traffic pattern γ^s from $U[0.5, 1.5]$, i.e., $\theta^s_t = \gamma^s(t + \xi^s) + \delta$. The revenue of per unit traffic for service $s \in \{1,\cdots,5\}$ is set to be $r^s = 0.5 \times s$. The rest settings are the same as the previous subsection.

In Fig. 8, we show the effects of prices of per unit bandwidth on the ISP's profit, social welfare and capacity utilization.

[4] Capacity utilization here refers to the utliziation of ISP's total capacity, different from bandwidth utilization which refers to the utilization of CPs' required bandwidth.

Figure 8(a) shows that the ISP's profit first increases and then decreases with respect to q. In addition, the ISP's profit under TDS is always higher than that under SDP. A larger number of groups indicates a much higher profit for the ISP. For instance, when $q = 40$, the ISP's profit for TDS with $\kappa = 1$ is by 7% larger than that for SDP, while the value increases to 48% when $\kappa = 5$. Figure 8(b) shows that the social welfare decreases with respect to q. Figure 8(c) shows that the capacity utilization is of the similar trend as social welfare: it decreases with respect to q. The intuition is that a higher q shrinks the required bandwidth for each CP, resulting lower utilization.

In Fig. 9, we show the effects of prices of per unit traffic on the ISP's profit, social welfare and capacity utilization. Figure 9(a) shows the ISP's profit is maximized when p is neither too small nor too large. In general, the ISP's profit under TDS is larger than TDP, e.g., by 3% on average when $\kappa = 1$ and by 15% on average when $\kappa = 5$. Figure 9(b) shows that the social welfare decreases with respect to p, and the decreasing trend becomes more rapidly when p is large. In addition, the value of social welfare under TDS is always higher than the value under SDP. Figure 9(c) shows that the capacity utilization has the similar trends as social welfare: it decreases with respect to p, and its value under TDS is also higher than that under SDP.

Summary: The capacity has great impacts on the ISP's and CPs' profit, and users' welfare. A larger capacity indicates higher profits for the ISP and CPs as well as users' welfare. In addition, the optimal prices charged by the ISP to CPs and end users all decrease with respect to the capacity. Well designed TDS always outperforms SDP in terms of ISP's profit, social welfare and capacity utilization. A larger number of groups indicates better performance of TDS. Moreover, the social welfare and capacity utilization are controversial to the ISP's strategy. Higher prices charged to CPs and end users indicate lower social welfare and capacity utilization.

VI. Conclusion

In this paper, we propose time-dependent sponsoring, i.e., each CP can subsidize its users depending on the traffic demand at different time slots. We analyze TDS with strategic users, who can delay their consumption for higher subsidization in future. We formulate a Stackelberg game to capture the interactions among strategic users, CPs and an ISP. Our main conclusions include: 1) highly patient strategic users may reduce the average subsidization and the number of subsidizations, thus reducing bandwidth utilization, CPs' profit and users' welfare, and 2) when CPs provide different subsidizations to different groups, the bandwidth utilization can be improved significantly and so are CPs' profit and users' welfare, and 3) each CP's optimal subsidization under TDS increases capacity utilization, users' welfare and social welfare, and 4) the ISP obtains a higher profit under TDS than SDP. We would also like to mention some limitations in our paper: 1) we consider one monopoly ISP only but does not consider multiple ISPs' competitive market; 2) in our model, users make the decision of consumption time simply according to the lowest price to pay, while in reality, users may prefer to consume the content earlier if the price is not too high. We would like to address these issues in our future work.

Acknowledgement: Dan Wang is supported in part by National Natural Science Foundation of China No. 61272464 and RGC/GRF PolyU 5264/13E. Weijie Wu is supported in part by National Natural Science Foundation of China No. 61402287 and Shanghai Yangfan Project No. 14YF1401900.

References

[1] V. Wireless, "The more everything plan." http://www.verizonwireless.com/landingpages/more-everything/#how-it-works.

[2] A. Odlyzko, B. S. Arnaud, E. Stallman, and M. Weinberg, "Know your limits: Considering the role of data caps and usage based billing in internet access service." Public Knowledge White Paper, May 2012, avaliable at http://www.publicknowledge.org/know-your-limits-considering-role-data-caps-and-us.

[3] AT&T, "Sponsored data from at&t." Available at: http://developer.att.com/apis/sponsored-data.

[4] M. Andrews, U. Ozen, M. I. Reiman, and Q. Wang, "Economic models of sponsored content in wireless networks with uncertain demand," in *Proc. Smart Data Pricing Workshop, IEEE*, 2013.

[5] P. Goldstein, "At&t sponsored data partner syntonic wireless to launch 'toll-free' content store." http://www.fiercewireless.com/story/att-sponsored-data-partner-syntonic-wireless-launch-toll-free-content-store/2014-07-10.

[6] P. Goldstein, "Google joins with india's bharti airtel for toll-free wireless internet service." http://www.fiercemobileit.com/story/att-ceo-content-providers-asking-toll-free-data-plans/2012-07-18.

[7] L. Zhang and D. Wang, "Sponsoring content: Motivation and pitfalls for content service providers," in *Proc. Smart Data Pricing Workshop, IEEE*, 2014.

[8] R. T. B. Ma, "Subsidization competition: Vitalizing the neutral internet," in *Proc. ACM CoNEXT*, 2014.

[9] L. Zhang, W. Wu, and D. Wang, "Sponsored data plan: A two-class service model in wireless data networks," in *Proc. ACM SIGMETRICS*, 2015.

[10] H. soo Ahn, M. Gumus, and P. Kaminsky, "Pricing and manufacturing decisions when demand is a function of prices in multiple periods," *Operations Research*, vol. vol. 55, no. 6, pp. 1039 – 1057, 2007.

[11] C. Borgs, O. Candogan, J. Chayes, I. Lobel, and H. Nazerzadeh, "Optimal multiperiod pricing with service guarantees," *Management Science*, vol. vol. 60, no. 7, pp. 1792 – 1811, 2014.

[12] C. Joe-Wong, S. Ha, and M. Chiang, "Sponsoring mobile data: An economic analysis of the impact on users and content providers," in *Proc. IEEE INFOCOM*, 2015.

[13] M. Andrews, G. Bruns, and H. Lee, "Calculating the benefits of sponsored data for an individual content provider," in *Proc. IEEE CISS*, 2014.

[14] S. Ha, S. Sen, C. J. Wang, Y. Im, and M. Chiang, "Tube: Time-dependent pricing for mobile data," in *Proc. ACM SIGCOMM*, 2012.

[15] C. Wong, S. Ha, and M. Chiang, "Time-dependent broadband pricing: Feasibility and benefits," in *Proc. IEEE ICDCS*, 2011.

[16] L. Zhang, W. Wu, and D. Wang, "Time dependent pricing in wireless data networks: Flat-rate vs. usage-based schemes," in *Proc. IEEE INFOCOM*, 2014.

[17] P. Loiseau, G. Schwartzy, J. Musacchio, and S. Amin, "Incentive schemes for internet congestion management: Raffles versus time-of-day pricing," in *Annual Allerton Conf.*, 2011.

[18] BSNL, "End of a golden era: Bsnl to do away with night unlimited broadband plans?." http://telecomtalk.info/end-of-a-golden-era-bsnl-to-do-away-with-night-unlimited-broadband-plan/118528/.

[19] D. Rayburn, "Will bandwidth caps strangle netflix? not likely." http://www.streamingmedia.com/Articles/Editorial/Featured-Articles/Will-Bandwidth-Caps-Strangle-Netflix-Not-Likely-76134.aspx.

[20] L. Zhang, W. Wu, and D. Wang, "Tds: Time-dependent sponsored data plan for wireless data traffic market," tech. rep. Avialible at https://www.dropbox.com/s/9k5qoxwf5ajrpis/TR.pdf?dl=0.