2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Data Demand Dynamics and Profit Maximization in Communications Markets†

Shaolei Ren and Mihaela van der Schaar

***Abstract*—In this paper, we focus on the users' aggregate data demand dynamics in a wireless communications market served by a monopolistic wireless service provider (WSP). Based on the equilibrium data demand, we optimize the WSP's data plans and long-term network capacity decisions to maximize its profit. For a market where two different data plans are offered, it is shown that the existence of a unique equilibrium data demand depends on the data plans, and the convergence of data demand dynamics is subject to the network congestion cost, which is closely related to the WSP's network capacity. A sufficient condition on the network congestion cost indicates that the WSP needs to provide a sufficiently large network capacity to guarantee the convergence of data demand dynamics. Then, we formalize the problem of optimizing the WSP's data plans and network capacities to maximize its profit, and solve it numerically.**

I. Introduction

We have witnessed over the last decade a successful proliferation of wireless networks, which support a variety of services and applications, and increasingly heated competition among the wireless service providers (WSPs). To sustain their competitive positions in the market and increase revenues, WSPs themselves will need to appropriately price their scarce network resources and expand their network capacities to support the unprecedented amount of wireless traffic. Hence, it becomes of paramount importance for these WSPs to understand how the aggregate data demand of all the subscribers evolves and how the demand is affected by various pricing plans.

We consider a wireless market with a monopolistic WSP serving a sufficiently large number of users. For the sake of analysis, we consider that the WSP can offer only two data plans, while each user can subscribe to one of the available data plans. Due to the resource constraint (e.g., network capacity), congestion effects are observed when multiple users share the same network, degrading the network performance (e.g., increasing delays). Essentially, congestion effects have similar impacts to prices on the users' experiences (i.e., utilities) and are also referred to as *congestion costs* in the literature [8]. Taking into consideration the charged price and congestion cost, each user can dynamically decide whether to subscribe to the WSP's service and which data plan to subscribe to. First, we show that the existence of a unique equilibrium data demand depends on the data plans. Moreover, the convergence of data demand dynamics is subject to the network congestion cost (and hence, the WSP's network capacity, too). We derive a sufficient condition for the convergence of data demand dynamics, indicating that that the WSP needs to provide a sufficiently large network capacity. Then, the problem of optimizing the WSP's data plans and network capacities is formalized and solved by numerical methods to maximize its profit. Finally, numerical results shows that, to maximize its profit, the WSP needs to increase the network capacity for its capped data plan while reducing the network capacity for its unlimited data plan. This coincides with the current trend that some WSPs have discontinued the offering of unlimited data plans [12].

Because of the space limitation, we now only provide an incomplete list of related literature. In our previous work [1], we study the user subscription dynamics and revenue maximization in both monopoly and duopoly communications markets by assuming a general distribution of users' valuation of quality-of-service (QoS) and a general QoS function that captures negative network externalities. By taking into account the congestion cost (i.e., negative network externality), [4] studies the feasibility of Paris Metro pricing (PMP) and shows sufficient conditions on the congestion cost functions, under which PMP leads to a higher revenue or social welfare than flat-rate pricing. Pricing decisions (restricted to unlimited data plans) and network capacity decisions in the presence of network congestion effects are studied in [6], where a missing part is the analysis of users' subscription decisions. In [7], time-dependent pricing is studied from the perspective of its efficiency in terms of revenues.

The rest of this paper is organized as follows. Section II describes the model. In Section III, we study the equilibrium and convergence of data demand dynamics, while in Section IV we formalize the WSP's profit maximization problem. Finally, we conclude this paper in Section V.

II. Model

Consider a wireless communications market where one monopolistic WSP, denoted by $\mathcal{W}$, offers to N users data communications service, which takes up an overwhelming majority of the wireless traffic. By assuming that N is sufficiently large such that each user is negligible, we use a continuum user population model and normalize the number of users to 1 [1]–[6]. In general, WSP $\mathcal{W}$ may offer multiple data plans, and users can choose any of the plans depending on their own preferences (the user choice shall be detailed later). As in [6], to keep the analysis tractable, we assume that WSP $\mathcal{W}$ offers up to two data plans, represented by $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively. For notational convenience, we also refer to users that subscribe to the plan $\mathcal{P}_i$ as $\mathcal{P}_i-$users (or $\mathcal{P}_i-$subscribers),

†This work is supported in part by National Science Foundation under Grant No. 0830556

The authors are with Electrical Engineering Department, University of California, Los Angeles. E-mail: {rsl,mihaela}@ee.ucla.edu

for $i = 1, 2$. Next, we shall provide the modeling details of the WSP and users.

A. WSP Model

Before entering a market, the WSP needs to first make investment in infrastructure. In this paper, we concentrate on the WSP' capacity deployment which, once determined, is difficult to adjust and hence is an irreversible long-term decision [5]. Denote by $C_i \geq 0$ the network capacity (normalized by the number of users N) that the WSP allocates to its data pricing plan $\mathcal{P}_i$, for $i = 1, 2$. Assuming that the WSP incurs an average cost of τ per unit capacity,[1] we can express the WSP's equilibrium profit per short-term stage (i.e., users' subscription period) as

$$\Pi_{\mathcal{W}} = \sum_{i=1,2} \{R_i - \tau C_i\}, \tag{1}$$

where R_i is the equilibrium revenue per short-term stage derived from $\mathcal{P}_i-$users. Note that in (1), we neglect the recurring cost of serving the users, which can also be absorbed into the revenue R_i [5]. To maximize its profit given the users' rational decisions, the WSP shall strategically determine its capacity $\mathbf{C} = \{C_1, C_2\}$. After building the network, the WSP decides its data plans and may alter them throughout the network's lifespan.

In today's wireless market, the most popular data plans are "unlimited", "capped" and "usage-based", all of which can be represented by a unified pricing model specified by (p, d^*, γ): each subscriber pays a fixed subscription fee p that allows it to transmit and receive up to d^* units of data; for each unit of additional data usage exceeding the capped data limit d^*, the subscriber pays γ. In special cases, a capped data plan characterized by (p, d^*, γ) becomes a usage-based one if $p = 0$ and $d^* = 0$, and an unlimited data plan if $d^* = \infty$ or $\gamma = 0$. For analytical tractability and to gain insights on how the congestion costs affect the data demand dynamics, we assume that the WSP's data plan $\mathcal{P}_1 = (p_1, +\infty, 0)$ is "unlimited" whereas its data plan $\mathcal{P}_2 = (p_2, d_2^*, \gamma_2)$ is "capped".[2] This assumption, which may seem strong, can be justified by noting that some WSPs have (partially) resorted to capped data plans in view of the soaring wireless data service demand that frequently clogs their network infrastructure.[3] Moreover, even if the WSP offers two capped data plans, it is likely that one of the data plans has a very high data limit, which only a negligible fraction of subscribers can exceed in practice, and thus this data plan is almost "unlimited" (see, e.g., [12]).

[1]The cost is averaged over the lifespan of the network infrastructure. For instance, if a network with a lifespan of T short-term stages (i.e., users' subscription period) is built at a cost of $\tilde{\tau}$ per unit capacity, then the average cost per unit capacity is $\tau = \tilde{\tau}/T$.

[2]In the most general case where both data plans are "capped", the approach of analysis in this paper is still applicable, although the analysis becomes more complicated.

[3]Starting from June 7, 2010, AT&T discontinued the offering of unlimited data plans to its new iPhone users and adopts a capped data plan as considered in this paper [12].

B. User Model

Due to the capacity constraint, the network becomes more congested (i.e., negative network externalities or effect) as more data flow is transmitted [7][8]. Such an effect is quantified by the congestion cost, which has similar impacts to prices on the users' experiences (i.e., utilities) [8]. We denote the congestion cost associated with the data pricing plan $\mathcal{P}_i$ by $g_i(D_i, C_i)$, where $i = 1, 2$ and $D_i \geq 0$ is the aggregate data demand (i.e., the total data demand of all the $\mathcal{P}_i$-users over a certain period) and C_i is the capacity allocated to $\mathcal{P}_i-$users. Without causing ambiguity, we simplify $g_i(D_i, C_i)$ as $g_i(D_i)$ by removing C_i wherever applicable. An implicit assumption in the model is that congestion costs for different data plans are independent of each other, which may be achieved by splitting network capacity among the plans [5].

Users are heterogeneous in the sense that they may have different data service demand and different benefits of utilizing the WSP's communications service. To model the user heterogeneity, each user k is characterized by a two-element tuple (θ_k, d_k), where θ_k indicates user k's benefit from data service and d_k denotes its data demand over a certain period (e.g., a month or a day). The values of θ_k and d_k can be determined by various approaches. For instance, (θ_k, d_k) may be user k's intrinsic characteristic and not influenced by the WSP's pricing schemes. In such scenarios, each individual user has *inelastic* demand [7][8], although the aggregate demand of all the users is still elastic and influenced by the prices. Mathematically speaking, when user k subscribes to the WSP's data plan $\mathcal{P}_i$, its utility is given by

$$u_{k,i} = \theta_k - g_i(D_i) - p_i - \gamma_i [d_k - d_i^*]^+, \tag{2}$$

where $[x]^+ = \max\{0, x\}$, and if its data demand exceeds the granted data limit d_i^*, the term $\gamma_i[d_k - d_i^*]^+$ is positive and represents the additional cost user k incurs. Similar utility functions have been used in [2][4][7] and references therein. The utility function in (2) can be interpreted as follows: θ_k represents the benefit that user k receives from d_k units of data service, $g_i(D_i)$ indicates the congestion cost (i.e., negative network externality), and $p_i + \gamma_i[d_k - d_i^*]^+$ is the payment made to WSP W_i. Users that do not subscribe to any data plans obtain zero utility. Now, we impose some standard assumptions on the users' data demand and their benefits, users' subscription decisions, and the congestion function $g_i(D_i)$.

Assumption 1: The users' benefits and their data demand follow a two-dimensional distribution whose joint density function $f(\theta, d)$ is defined on $\mathcal{U} = \{(\theta, d) \,|\, 0 \leq \theta \leq \theta_{\max}, 0 \leq d \leq d_{\max}\}$. For completeness of definition, we have $f(\theta, d) = 0$ for all $(\theta, d) \notin \mathcal{U}$. The cumulative density function is given by $F(\theta, d) = \int_{-\infty}^{d} \int_{-\infty}^{\theta} f(x, y) dx dy$ for $(\theta, d) \in \mathbb{R}^2$.

Assumption 2: Each user k subscribes to the data plan $\mathcal{P}_i$ if $u_{k,i} > u_{k,j}$ and $u_{k,i} \geq 0$ for $i, j \in \{1, 2\}$ and $i \neq j$. If $u_{k,1} = u_{k,2} \geq 0$, user k subscribes to the unlimited data plan

$\mathcal{P}_1$.[4]

Assumption 3: $g_i(D_i)$ is a non-negative, non-decreasing and differentiable[5] function in $D_i \in [0, D_{\max}]$, where $D_{\max}$ is the maximum possible aggregated data demand, normalized with respect to the total population, and given by

$$D_{\max} \triangleq \int_{y=0}^{d_{\max}} \int_{x=0}^{\theta_{\max}} y f(x,y) dx dy. \quad (3)$$

We briefly explain the above three assumptions. Assumption 1 can be considered as an expression of user diversity in terms of the benefits and their data demand. The lower bound on the interval is set as zero to simplify the analysis, and this will be the case when there is enough diversity in the users so that there are non-subscribers for any positive price [5][6]. Assumption 2 captures the user rationality. A rational user will subscribe to the data plan that provides a higher utility if at least one data plan provides a non-negative utility, and to neither data plan otherwise. Assumption 3 indicates an intuitive fact that the congestion cost that each user experiences when subscribing to the data plan $\mathcal{P}_i$ becomes larger when the aggregate data demand increases.

For the considered wireless market, we can describe the timing (i.e., order of moves) as follows.

Stage 1 (long-term): The WSP decides the network capacity C_2 to deploy to maximize its profit.

Stage 2 (medium-term): Given C_2, the WSP chooses its optimal data plan $\mathcal{P}_2 = (p_2, d_2^*, \gamma_2)$ by specifying p_2, d_2^* and γ_2 to maximize its revenue.

Stage 3 (short-term): By jointly considering the congestion cost and offered data plan, users decide whether or not to subscribe to the WSP's service.

From the described timing, it can be seen that the WSP can be regarded as the leader whereas the users are followers. Thus, in order to identify the optimal data plan and network capacity, the WSP needs to first know how the users make their subscription decisions. Therefore, in what follows, we proceed with our analysis using backward induction. Before concluding this section, it is worthwhile to provide the following remarks regarding our model.

Remark 1: As in [7], we assume for the convenience of analysis that each individual user k has an inelastic and fixed demand d_k (and benefit θ_k, too). Alternatively, d_k can be determined by solving a utility maximization problem and θ_k is the maximum benefit that user k receives [9]. Nevertheless, given the WSP's data plans, (θ, d) still follows a certain distribution over all the users and thus, our approach can be viewed as a proxy to determine the users' demand and benefit, provided that the distribution does not change significantly with the data plans.

[4]Online surveys show that users generally prefer an unlimited data plan to a capped one [11]. Moreover, specifying an alternative tie-breaking rule (e.g., random selection between the two data plans) in case of $u_{k,1} = u_{k,2} \geq 0$ will not significantly affect the analysis of this paper.

[5]Since $g_i(\cdot)$ is defined on $[0, D_{\max}]$, we use a one-sided limit to define the derivative of $g(\cdot)$ at 0 and $D_{\max}$, e.g., $g_i'(0) = \lim_{D_i \to 0^+}[g_i(D) - g_i(0)]/(D_i - 0)$.

Remark 2: Compared to the congestion cost function used in the existing literature that disregards the user heterogeneity in terms of data demand and is defined solely in terms of the number of subscribers [1][4][6], $g_i(D_i)$ is more accurate in modeling the congestion effect. Whilst the actual congestion cost also depends on when the users utilize the network, we consider the congestion cost *averaged* over time and ignore the time dependency to keep the analysis tractable [8].

Remark 3: The shape of the congestion cost function $g_i(D_i)$ may be determined by various factors, including the network capacity, resource allocation schemes and/or scheduling algorithms used for the data plan $\mathcal{P}_i$. While our analysis applies to a general function $g_i(D_i)$ satisfying Assumption 3, we shall explicitly focus on the impacts of network capacities on $g_i(D_i)$ when we derive specific results or study the WSP's long-term capacity decision. For instance, a concrete example is given by $g_i(D_i) = D_i/C_i$, which has been widely used (with minor modification, e.g., assuming all the users have the same data demand) in the prior work [4][6].[6]

III. Equilibrium and Convergence of Data Demand Dynamics

The subscription decision stage can be formalized as a non-cooperative game with an infinite number of players, the solution to which is (Nash) equilibrium. At an equilibrium, if any, no users can gain more benefits by deviating from their decisions. In other words, the aggregate data demand of those users subscribing to the WSP's data plans does not change at the equilibrium. Thus, we study the users' equilibrium subscription decisions by specifying the equilibrium data demand (D_1^*, D_2^*). By Assumption 2, we see that the equilibrium data demand (D_1^*, D_2^*) satisfies the following equations

$$D_1^* = h_{d,1}(D_1^*, D_2^*) = \int_{y=\tilde{d}}^{d_{\max}} \int_{x=g_1(D_1^*)+p_1}^{\theta_{\max}} y f(x,y) dx dy \quad (4)$$

$$D_2^* = h_{d,2}(D_1^*, D_2^*) = \int_{y=0}^{\tilde{d}} \int_{x=g_2(D_2^*)+p_2+\gamma_2[y-d_2^*]^+}^{\theta_{\max}} y f(x,y) dx dy \quad (5)$$

if $p_1 + g_1(D_1^*) > p_2 + g_2(D_2^*)$, and by

$$D_1^* = h_{d,1}(D_1^*, D_2^*) = \int_{y=0}^{d_{\max}} \int_{x=g_1(D_1^*)+p_1}^{\theta_{\max}} y f(x,y) dx dy \quad (6)$$

$$D_2^* = h_{d,2}(D_1^*, D_2^*) = 0 \quad (7)$$

if $p_1 + g_1(D_1^*) \leq p_2 + g_2(D_2^*)$. In (4) and (5), $\tilde{d}$ is given by

$$\tilde{d} = d_2^* + \frac{1}{\gamma_2}[p_1 - p_2 + g_1(D_1^*) - g_2(D_2^*)], \quad (8)$$

which specifies the data demand of marginal users that are "indifferent" between subscribing to the plan $\mathcal{P}_1$ and the plan

[6]Another congestion cost function widely adopted in the literature is $g_i(D_i) = 1/(C_i - D_i)$, which satisfies Assumption 3. Thus, our analysis is also applicable if $g_i(D_i) = 1/(C_i - D_i)$ is considered.

$\mathcal{P}_2$ (see [1][4] for a detailed explanation of "indifferent"). Note that there are two regimes of the equilibrium data demand in the market with two data plans, and which regime governs the equilibrium depends on the relative values of the *effective* full price (not including the additional cost if the data demand exceeds the granted data limit), i.e., $p_1 + g_1(D_1^*)$ and $p_2 + g_2(D_2^*)$. Next, we give the formal definition of the equilibrium point (D_1^*, D_2^*).

Definition 1: When two data plans $\mathcal{P}_1$ and $\mathcal{P}_2$ are offered, (D_1^*, D_2^*) is an *equilibrium* data demand if it satisfies satisfies

$$h_{d,1}(D_1^*, D_2^*) = D_1^* \text{ and } h_{d,2}(D_1^*, D_2^*) = D_2^*, \tag{9}$$

where $h_{d,1}(D_1^*, D_2^*)$ and $h_{d,2}(D_1^*, D_2^*)$ are given in (4)–(7).

We note that there may not exist an equilibrium data demand if the plan $\mathcal{P}_2$ is "capped" (i.e., $d_2^* < d_{\max}$ and $\gamma_2 > 0$). Next, we provide a sufficient condition that establishes the existence and uniqueness of an equilibrium point in Proposition 1, whose proof is available in [10].

Proposition 1. *For any data plans $\mathcal{P}_1 = (p_1, +\infty, 0)$ and $\mathcal{P}_2 = (p_2, d_2^*, \gamma_2)$, there exists a unique equilibrium data demand (D_1^*, D_2^*) satisfying (4)–(7) if*

$$d_2^* = 0 \text{ and } \gamma_2 > 0. \tag{10}$$

Moreover, the equilibrium data demand (D_1^*, D_2^*) satisfies $D_1^* = h_{d,1}(D_1, 0^*)$ and $D_2^* = 0$ if $p_2 + g_2(0) \geq p_1 + g_1(D_1^*)$.

Proposition 1 indicates that, if the two data plans $\mathcal{P}_1$ and $\mathcal{P}_2$ are unlimited and usage-based, respectively, then the data demand admits a unique equilibrium point. It also shows that, if the effective subscription cost of for the data plan $\mathcal{P}_1$ evaluated at D_1^* is always smaller than or equal to that of the data plan $\mathcal{P}_2$, then no users subscribe to the data plan $\mathcal{P}_2$ at the equilibrium point.

In practice, the users do not have complete information regarding each other and hence, they may not make directly the subscription decisions that strikes an equilibrium. Instead, an adjustment process where the users update their subscription decisions based on limited information is required. A natural and well-studied approach to modeling the adjustment process is the best-response dynamics, in which each decision maker chooses the best action in response to the decisions made by the others. As in [1][3], we consider the best-response dynamics based on naive (or static) expectation, and assume that the users can only change their subscription decisions (e.g., opt out of the plan $\mathcal{P}_2$) at discrete time periods indexed by $t = 1, 2, \cdots$. The users expect that the congestion cost incurred when subscribing to a data plan in the time period t is equal to that in the previous period $t-1$ and make their subscription decisions to myopically maximize their utility in the time period t [1][2][3]. We assume that, other than the subscription price, there is no cost involved (e.g., initiation fees, termination fees, device prices) when users switch between the data plans $\mathcal{P}_1$ and $\mathcal{P}_2$[2]. By Assumption 2, at period $t = 1, 2\cdots$, user k subscribes to the data plan $\mathcal{P}_1$ if and only if

$$\begin{aligned} &\theta_k - g_1(D_1^{t-1}) - p_1 \geq \theta_k - g_2(D_2^{t-1}) - p_2 - \gamma_2[d_k - d_2^*]^+ \\ &\text{and } \theta_k - g_1(D_1^{t-1}) - p_1 \geq 0, \end{aligned} \tag{11}$$

to the data plan $\mathcal{P}_2$ if and only if

$$\begin{aligned} &\theta_k - g_2(D_2^{t-1}) - p_2 - \gamma_2[d_k - d_2^*]^+ > \theta_k - g_1(D_1^{t-1}) - p_1 \\ &\text{and } \theta_k - g_2(D_2^{t-1}) - p_2 - \gamma_2[d_k - d_2^*]^+ \geq 0, \end{aligned} \tag{12}$$

and to neither data plan if and only if

$$\begin{aligned} &\theta_k - g_1(D_1^{t-1}) - p_1 < 0 \\ &\text{and } \theta_k - g_2(D_2^{t-1}) - p_2 - \gamma_2[d_k - d_2^*]^+ < 0. \end{aligned} \tag{13}$$

Therefore, given the data plans $\mathcal{P}_1 = (p_1, +\infty, 0)$ and $\mathcal{P}_2 = (p_2, \gamma_2, d_2^*)$, the data demand dynamics is described by a sequence $\{(D_1^t, D_2^t)\}_{t=0}^{\infty}$ in $\mathcal{D} = \{(D_1, D_2) \in \mathbb{R}_+^2 \mid D_1 + D_2 \leq D_{\max}\}$ generated by $D_1^t = h_{d_1}(D_1^{t-1}, D_2^{t-1})$ and $D_2^t = h_{d_2}(D_1^{t-1}, D_2^{t-1})$, where $h_{d_1}(D_1^{t-1}, D_2^{t-1})$ and $h_{d_2}(D_1^{t-1}, D_2^{t-1})$ are obtained by substituting (D_1^{t-1}, D_2^{t-1}) into (4)–(7).

Since an equilibrium point may not exist if the data plan $\mathcal{P}_2$ is unlimited or capped, we restrict the analysis in the remainder of this paper to the case that the plan $\mathcal{P}_2$ is usage-based (although an initial subscription fee p_2 may be charged) such that a unique equilibrium point is guaranteed to exist. Next, we provide a sufficient condition for the data demand dynamics $\{(D_1^t, D_2^t)\}_{t=0}^{\infty}$ to converge.

Proposition 2. *For data plans $\mathcal{P}_1 = (p_1, +\infty, 0)$ and $\mathcal{P}_2 = (p_2, 0, \gamma_2)$ where $\gamma_2 > 0$, the data demand dynamics converges to the unique equilibrium point starting from any initial point $(D_1^0, D_2^0) \in \mathcal{D} = \{(D_1, D_2) \in \mathbb{R}_+^2 \mid D_1 + D_2 \leq D_{\max}\}$ if the condition in (14) is satisfied, where d_{max} is the maximum individual demand, $\theta_{\max}$ is the maximum benefit derived from subscribing to the WSP's service and $K = \max_{(\theta,d)\in\mathcal{U}} f(\theta, d)$*

Proof: See [10]. □

We can obtain more specific condition regarding the network capacities for the convergence of data demand dynamics by plugging $g_1(D_1) = D_1/C_1$ and $g_2(D_2) = D_2/C_2$ into (14). The condition (14) provides us with an insight that, if congest costs increase too rapidly, the data demand dynamics may exhibit oscillation or divergence. Another important observation from (14) is that the two data plans also affect the convergence. Specifically, given higher prices, it is easier for the congestion costs to satisfy the convergence condition. Intuitively, higher prices result in lower aggregate data demand. Therefore, there is less fluctuation in the data demand dynamics and the requirement on the congestion costs becomes less stringent.

We finally note that interested readers may refer to [10] for remarks regarding the applicability of the considered user subscription model and the cost in updating the subscription decisions (e.g., time spent in calling the customer service, early termination fees).

$$\max_{(D_1,D_2)\in[0,d_{\max}]^2} \{g_1'(D_1), g_2'(D_2)\} < \frac{1}{K \cdot \left(\frac{d_{\max}^2}{2} + \frac{d_{\max}}{\gamma_2}[\theta_{\max} - p_1]^+ + \frac{d_{\max}}{\gamma_2}[\theta_{\max} - p_2]^+ \right)} \quad (14)$$

IV. WSP's Profit Maximization

Over the entire lifespan of the network infrastructure, the WSP can change its data plans to maximize its revenue, although the change of data plans is sufficiently slow compared to the users' subscription decisions. We write the the WSP's equilibrium revenues for the data plans $\mathcal{P}_1$ and $\mathcal{P}_2$ as

$$R_1 = \int_{y=\tilde{d}}^{d_{\max}} \int_{x=g_1(D_1^*)+p_1}^{\theta_{\max}} p_1 f(x,y) dx dy$$
$$\text{and } R_2 = \int_{y=0}^{\tilde{d}} \int_{x=g_2(D_2^*)+p_2+\gamma_2 y}^{\theta_{\max}} (p_2 + \gamma_2 y) f(x,y) dx dy \quad (15)$$

if $p_1 + g_1(D_1^*) > p_2 + g_2(D_2^*)$, and as

$$R_1 = \int_{y=0}^{d_{\max}} \int_{x=g_1(D_1^*)+p_1}^{\theta_{\max}} p_1 f(x,y) dx dy \text{ and } R_2 = 0 \quad (16)$$

if $p_1 + g_1(D_1^*) \leq p_2 + g_2(D_2^*)$, where $\tilde{d}$ is given by $\tilde{d} = \frac{1}{\gamma_2}[p_1 - p_2 + g_1(D_1^*) - g_2(D_2^*)]$. The expressions of equilibrium revenues in (15) and (16) are quite complicated and hence, lose analytical tractability. As a consequence, we resort to numerical search to identify the optimal $\mathcal{P}_1 = (p_1, +\infty, 0)$ and $\mathcal{P}_2 = (p_2, 0, \gamma_2)$ maximizing $R_1 + R_2$. Moreover, we shall find the WSP's optimal capacities through exhaustive search.

We see from Fig. 1 that, to maximize its profit, the WSP needs to increase the network capacity for its capped data plan while reducing the network capacity for its unlimited data plan. This can be explained as follows: when an unlimited data plan is offered, subscribers with high data demand will cause excessive congestion costs for the other subscribers, reducing the profitability of the unlimited data plan. This also coincides with the current trend that some WSPs have began to discontinue the offering of unlimited data plans [12].

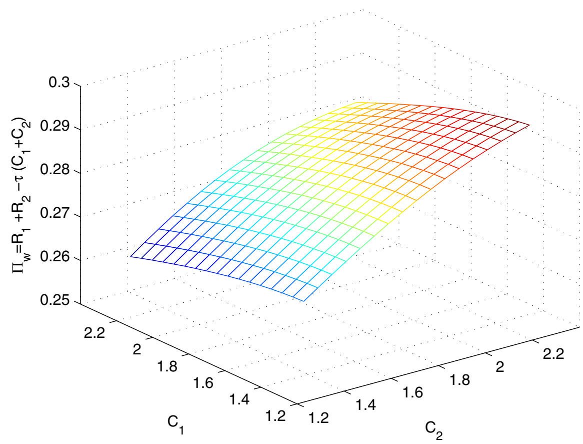

Fig. 1. Two data plans: optimal profit versus network capacities. $\tau = 0.02$.

V. Conclusion

In this paper, we considered a wireless communications market where one monopolistic WSP serves a large number of users. The users' data demand dynamics, the WSP's data plan decision and network capacity decision were studied. In our analysis, the users' heterogeneity in terms of their benefits and data demand, as well as the network congestion costs, were explicitly taken into consideration. For the user's data demand dynamics, we showed that: (1) for certain data plans, there may not exist any equilibrium data demand; (2) in order to guarantee the convergence of data demand dynamics, the congestion costs should not increase too rapidly when the aggregate data demand increases, implying that the WSP needs to deploy a large network capacity to support the users' demand. The WSP's data plan decision and network capacity decision were formalized and solved numerically to maximize the WSP's profit. Finally, numerical results indicate that to maximize its profit, the WSP should increase the network capacity for its capped data plan while decreasing the network capacity for its unlimited data plan.

References

[1] S. Ren, J. Park, and M. van der Schaar "User subscription dynamics and revenue maximization in communication markets" *IEEE Infocom*, Apr. 2011.

[2] Y. Jin, S. Sen, R. Guerin, K. Hosanagar, and Z.-L. Zhang, "Dynamics of competition between incumbent and emerging network technologies," *NetEcon*, Seattle, WA, USA, August 22, 2008.

[3] M. Manshaei, J. Freudiger, M. Felegyhazi, P. Marbach, and J. P. Hubaux, "On wireless social community networks," *IEEE Infocom*, Apr. 2008.

[4] C. K. Chau, Q. Wang, and D. M. Chiu, "On the viability of Paris metro pricing for communication and service networks," *IEEE Infocom*, Mar. 2010.

[5] N. Shetty, S. Parekh, and J. Walrand, "Economics of femtocells," *IEEE Globecom*, Dec. 2009.

[6] N. Shetty, G. Schwartz, and J. Walrand, "Internet QoS and regulations," *IEEE/ACM Trans. Networking*, vol. 18, no. 6, pp. 1725-1737, Dec. 2010.

[7] L. Jiang, S. Parekh, and J. Walrand, "Time-dependent network pricing and bandwidth trading," *IEEE Intl. Workshop Bandwidth on Demand (BoD)*, 2008.

[8] D. Acemoglu and A. Ozdaglar, "Competition and efficiency in congested markets," *Mathematics of Operations Research*, vol. 32, no. 1, pp. 1-31, 2007.

[9] P. Hande, M. Chiang, A. R. Calderbank, and S. Rangan, "Network pricing and rate allocation with content provider participation," *IEEE Infocom*, Apr. 2009.

[10] S. Ren and M. van der Schaar "Data Demand Dynamics in Communications Markets" *UCLA Tech. Report*.

[11] *Mobile users wary of usage-based pricing: Survey*, Reuters, Nov. 18, 2010. http://www.reuters.com/article/idUSN1827905920101118.

[12] *AT&T Wireless Data Plan*, available at http://www.wireless.att.com.