# Hybrid Data Pricing for Network-Assisted User-Provided Connectivity

Lin Gao, George Iosifidis, Jianwei Huang, and Leandros Tassiulas

***Abstract*—User-provided connectivity (UPC) is a promising paradigm to achieve a low-cost ubiquitous connectivity. In this paper, we study a *network-assisted* UPC service model, where a mobile virtual network operator (MVNO) enables its subscribers to operate as mobile WiFi hotspots (hosts) and provide Internet connectivity for others. A unique aspect of this service model is that the MVNO offers some free data quota to hosts as reimbursements (incentives) for connectivity sharing. This reimbursing scheme, together with a usage-based pricing, constitute a revolutionary hybrid data pricing-reimbursing scheme, which has not been considered before. We characterize the different impacts of data price and reimbursement on the host's forwarding decision systematically. Based on this, we further derive the optimal hybrid pricing-reimbursing policy that maximizes the MVNO's revenue. Our numerical result indicates that by using the proposed hybrid pricing policy, the MVNO can increase its revenue by 20% to 135% under an elastic client demand, and by 20% to 550% under an inelastic client demand, comparing to those achieved under a pricing-only policy.**

## I. Introduction

### A. Motivation

A mobile virtual network operator (MVNO) is a special wireless service provider, who does not own radio spectrum and/or wireless network infrastructure, but *lease* these resources from traditional mobile network operators (MNOs). Usually, an MVNO enters a business agreement with one or multiple MNOs to obtain bulk access to the latter's network resources at wholesale rates, and provides services to its own customers over the leased network resource at retail prices. Such a business model has achieved significant success worldwide [1] due to several factors, including the phenomenal rise of mobile data traffic [2], the cooperative market strategy of MNOs [3], and the increasingly favorable regulatory environment and market maturity [4]. In contrast to traditional MNOs, the MVNOs often employ novel and flexible business models, tailored to the needs of their customers, hence can better reach the niche market that is under-served by the traditional MNOs.

One intriguing business model is the one associated with a recently launched USA MVNO called Karma [5]. Specifically, Karma offers each of its subscribers a 4G Karma device, which allows the subscriber to access the Internet via any mobile

This work is supported by the General Research Funds (Project Number CUHK 412713, CUHK 412710, and CUHK 412511) established under the University Grant Committee of the Hong Kong Special Administrative Region, and the National Natural Science Foundation of China (Project Number 61301118). This work is also supported by the EINS, the Network of Excellence in Internet Science, through FP7-ICT grant 288021 from the European Commission.

L. Gao and J. Huang are with the Department of Information Engineering, The Chinese University of Hong Kong, HK, Email: {lgao, jwhuang}@ie.cuhk.edu.hk; G. Iosifidis and L. Tassiulas are with CERTH, and the ECE Department, University of Thessaly, Greece. Email: {giosifid,leandros}@uth.gr.

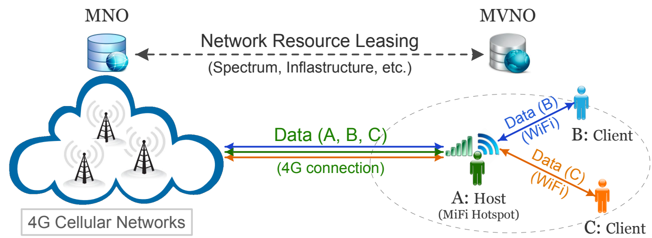



Fig. 1. Illustration of the Karma's User-Provided Connectivity Scheme: Mobile user A (host) is a Karma subscriber and operates as a MiFi hotspot, forwarding data for users B and C (clients) via Wi-Fi.

cellular network that Karma has signed a contract with. Similar as most of today's MNOs and MVNOs, Karma charges a usage-based data pricing ($14 per GB) to its subscribers, with no fixed period contract and expiry date, regardless of the cellular network that actually serves the subscriber. What is unique to Karma is that each 4G Karma device can operate as a mobile Wi-Fi *hotspot* (also known as MiFi[1]), and provide Internet access services for other users who are not subscribers of Karma but are within the vicinity of such a Karma device. Figure 1 illustrates such a scenario, where mobile user A is a Karma subscriber and operates as MiFi hotspot, forwarding data of users B and C from/to the Internet. Obviously, in such a scenario, it is essential to offer necessary incentives to Karma subscriber A such that it is willing to enable this MiFi option and operate as a hotspot (called *host*) in order to forward data for others (called *clients*). Under the current pricing model, Karma reimburses a host 100MB free data for every client that connects through the host for the first time.

Karma's business model differs substantially from that of any existing MVNOs or MNOs, and provides exciting possibilities for the future wireless data market. First, Karma dynamically extends its network coverage and increases its effective customer group, by enabling its subscribers as MiFi hotspots. This constitutes a new model of *User-Provided Connectivity* (UPC) services [6], which have been recently introduced as an alternative and complementary communication method to existing infrastructure-based services [7].[2] Second, the hosts only pay for the data they actually consume, but not for the data they forward for clients. The latter part will be paid by their clients directly. Namely, the hosts only share their 4G connections, but not their data. This simplifies the economic interactions between the hosts and the clients, which greatly facilitate the adoption of this service in practice. Third, Karma's usage-based pricing and free data quota reim-

[1]The term MiFi stands for "My WiFi" and refers to mobile devices with cellular connections (3G or 4G) that can offer Internet access for other devices through WiFi tethering. Such devices are offered by Novatel, Verizon, Sprint and several other companies.

[2]A typical example of UPC services is the internet-based WiFi access model introduced by FON (http://fon.com/). Another instance is cellular traffic offloading through user-owned small-cell base stations [12], [13].

bursement together constitute a revolutionary *hybrid pricing* scheme, which introduces an additional degree of freedom (the reimbursement) in the design of pricing plan.

While only Karma employs such a hybrid pricing scheme at this moment, we believe that such a scheme is leading to a new class of *network-assisted* UPC services. This scheme, which involves the MVNO (network), the hosts, and the clients, enjoys unique technical and economical benefits and challenges. However, previous studies for UPC services (e.g., WiFi sharing [8]–[10]) analyzed only the host-client interaction, and none of them considered the assistance of the network or this hybrid pricing plan.

Although Karma has taken the first step towards such a network-assisted UPC service model, its current hybrid pricing plan has certain restrictions and drawbacks. In particular, the current plan of Karma may *not* provide consistent incentives for the hosts, which in turn may reduce the revenue of Karma, due to the following reasons. First, a host will receive the 100MB free data quota immediately when it starts to serve a new client. That is, the free data quota reimbursed to a host is independent of the amount of data (effort) that it actually forwards for the client. Hence, the host has no incentive to continue serving the client after receiving the quota, which will reduce the revenue of Karma collected from clients. Second, Karma charges the same data price and reimburses the same free data quota to all hosts. That is, the price and free data quota are host-independent. However, a quota that is enough to incentivize a host under certain situation may not be enough under other situations or for other hosts (e.g., those with high traffic needs or low battery energy levels).

Motivated by Karma's business model, we propose a more general network-assisted UPC service model with a hybrid pricing scheme. Our goal is to design a pricing and reimbursing plan that offers consistent incentives to the hosts, and meanwhile brings high revenues for the MVNOs. Such a plan will foster this type of new services, which are seen as an opportunity to improve user-perceived network experience.

### B. Contributions of this paper

Our model generalizes the Karma UPC model in the following aspects. First, the free data quota reimbursed to a host is not fixed but *effort-based*, in the sense that it depends on the amount of data that the host actually forwards for clients. This ensures that the host has a consistent incentive to serve a client. Second, the MVNO can choose different prices and free data quotas for different hosts, that is, the pricing and reimbursing plan is *host-dependent*. Such a differentiation may depend on factors related to the external network environment (such as the user location and the channel condition) as well as the internal user status (such as the traffic need and battery level). For example, the MVNO may reimburse a higher free quota to the hosts in a densely populated area (since these hosts can serve more clients potentially), or the hosts with higher internal traffic needs and/or under lower battery energy levels (since these hosts are less willing to serve clients).

Our goal is to design the optimal pricing and reimbursing strategy for the MVNO that maximizes its total revenue (collected from both hosts and clients), considering the necessary incentives to the hosts. Notice that changing the price or changing the free data quota have different impacts on hosts' forwarding decisions, and hence on the MVNO's revenue. For example, lowing the price to a host will increase the host's willingness to consume data, and thereby reduce its willingness to forward data for clients. This may increase the revenue from the host, but will decrease the revenue from the clients (served through this host). Similar observation can be found when changing the reimbursement. Clearly, such a hybrid pricing structure brings new challenges in the design of the revenue maximizing policy for the MVNO.

We employ a *game-theoretic* analysis, and model the interaction between the MVNO and the hosts as a two-stage leader-follower (Stackelberg) game [11]. In the first stage, the MVNO decides the pricing and free data quota reimbursing plan. Accordingly, in the second stage, every host decides how much data it will consume for its own needs, and how much data it will forward for clients. We analyze the best decisions of both the hosts and the MVNO, and characterize the game equilibrium systematically. Our model and equilibrium analysis capture a variety of system characteristics, including the service (demand) types of hosts or clients, the capacities of hosts' 4G connections, and the energy consumption patterns of hosts. Hence, the derivation of the optimal decisions can be applied to various network scenarios.

In summary, the main contributions are as follows.

- *Novel UPC Model:* Motivated by Karma's business case, we introduce a new model for the network-assisted UPC services with the hybrid pricing-reimbursing plan. To the best of our knowledge, this is the first work that studies such a hybrid pricing-reimbursing UPC service.
- *Objectives and Challenges:* We aim at designing the optimal pricing-reimbursing plan that maximizes the MVNO's revenue. The problem is challenging, due to the need to provide consistent incentives (to hosts) and to understand thoroughly the impacts of pricing and reimbursing strategies on the hosts' forwarding decisions.
- *Solution Techniques:* We model the interaction of the MVNO and the hosts as a two-stage Stackelberg game, and analyze the game equilibrium (i.e., the MVNO's best pricing-reimbursing decision and the hosts' best forwarding decisions) systematically. Our analysis is generic and captures a variety of system characteristics, and thus is applicable to various network scenarios.
- *Conclusions and Implications:* Numerical study shows that the hybrid pricing policy can significantly increase the MVNO's payoff, for example, by 20% to 135% under an elastic client demand, and by 20% to 550% under an inelastic client demand, comparing to those achieved under a pricing-only policy.

The rest of the paper is organized as follows. In Section II we present the system model. In Sections III we analyze the optimal pricing and reimbursement policies. In Section IV we discuss the complexity and security issues. In Section V we present the simulations. We discuss the related work in Section VI, and finally conclude in Section VII.

## II. System Model

### A. The Model

We consider one MVNO and a set $\mathcal{I} = \{1, 2, \ldots, I\}$ of subscribers (hosts) who can operate as MiFi hotspots. There exists a set $\mathcal{N}_i$ of non-registered users (clients) who is willing (also able) to access the Internet via the MiFi hotspot of host $i$. We consider both *elastic* and *inelastic* demands (services), and the data demand can be *shiftable* in time (delay-tolerant) or *non-shiftable* in time. We assume a time-slotted system, and study the system for one time period which consists of a set $\mathcal{T} = \{1, 2, \ldots, T\}$ of $T$ time slots.[3]

**MVNO.** The MVNO (e.g., Karma) pays the cooperating MNOs (e.g., ClearWire) a wholesale price $w \geq 0$ for every byte of data that its customers (hosts and clients) consume on the MNOs' networks. On the other hand, the MVNO charges the hosts and clients a *usage-based* retail price for the data consumption. Let us denote $p_i \in [0, \bar{p}]$ as the price charged to host $i$ and the clients served through host $i$. Here we allow the MVNO to differentiate the price to hosts (i.e., set different prices for hosts). This differentiation can be based on the host's location (e.g., whether in a populated area or not) or some other criteria such as its QoS requirement.

The MVNO offers some free data quota to the hosts, as reimbursements for extending services to non-registered users (clients). The reimbursement is not fixed, but depends on the amount of data the host forwards for clients. In this work, we consider a *proportional* reimbursing scheme, where the free data quota to a host is proportional to the amount of data it forwards. We denote $\theta_i \in [0, 1]$ as the free data quota *ratio* to host $i$. That is, if host $i$ forwards $X$ bytes of data for clients, it will receive $\theta_i \cdot X$ bytes of free data. Similarly, here we allow the MVNO to differentiate the reimbursement ratio among hosts for similar reasons behind the price differentiation.

The *strategy* of the MVNO includes the pricing vector $\boldsymbol{p} = (p_i)_{i \in \mathcal{I}}$ and the reimbursing vector $\boldsymbol{\theta} = (\theta_i)_{i \in \mathcal{I}}$. The objective of the MVNO is to decide the best strategy to maximize its revenue. Notice that both the price $p_i$ and the free quota ratio $\theta_i$ (to host $i$) remain unchanged within one time period of $T$ slots, but may vary across different periods. In this sense, the MVNO is able to differentiate the pricing and reimbursing strategy across time (periods).

**Hosts.** Each host $i$ has certain communication needs which can be modeled by a utility function $U_i(\cdot)$. Let us denote $x_{it} \geq 0$ as the amount of data host $i$ consumes during slot $t$, and $\boldsymbol{x}_i = (x_{it})_{t \in \mathcal{T}}$ as the data consumption vector over the entire time period. Then,

- A host $i$ with *elastic* services has a concave and increasing utility, e.g., $U_i(\boldsymbol{x}_i) = \log(x_{i1} + ... + x_{it})$, following the principle of diminishing marginal returns [14];
- A host $i$ with *inelastic* services, e.g. downloading $B_i$ bytes of data within a single time period, has a step utility function such as $U_i(\boldsymbol{x}_i) = u_i$ if $\sum_{t=1}^{T} x_{it} \geq B_i$, and $U_i(\boldsymbol{x}_i) = 0$ otherwise.

Both demand types illustrated above are time-shiftable. Note that the model with non-shiftable host demand can be either directly, or after minor modifications, viewed as a special case of our model.

Each host $i$ can access the 4G network in time slot $t$ with a maximum capacity of $R_{it} \geq 0$ bytes (per slot), which allows it to transmit/receive data to/from the Internet. Notice that this capacity may vary over different time slots. For example, when the cellular network is heavily loaded (e.g., in a densely populated area), the capacity for a single host may be small. The host $i$ decides the capacity percentage $\alpha_{it} \in [0, 1]$ for his own data consumption, and the capacity percentage $\beta_{it} \in [0, 1]$ serving other clients. That is, the amount of data that host $i$ consumes (for its own needs) during slot $t$ is

$$x_{it} = \alpha_{it} \cdot R_{it}, \tag{1}$$

and the amount of data that host $i$ forwards for clients is

$$y_{it} = \beta_{it} \cdot R_{it}. \tag{2}$$

The *strategy* of a host $i$ includes the scheduling vectors $\boldsymbol{\alpha}_i = (\alpha_{it})_{t \in \mathcal{T}}$ and $\boldsymbol{\beta}_i = (\beta_{it})_{t \in \mathcal{T}}$ for the entire time period. Obviously, for a feasible strategy $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)$, the following *feasibility constraint* must hold:

$$\alpha_{it} + \beta_{it} \leq 1, \quad \forall t \in \mathcal{T}, \tag{3}$$

which implies that the total scheduled data cannot exceed the capacity. Note that the host may not always fully utilize its capacity when, for example, the battery energy is limited.

**Clients.** As we focus on the interaction of the MVNO and the hosts in this paper, we will model the behavior of the clients in a rather *abstract* way: the clients of host $i$ will demand a total of $D_{it} \geq 0$ bytes of data at each slot $t$. Notice that such demand maybe derived based on a more detailed modeling of clients (as we did for hosts).

Depending on the clients' service types, the client demand can be categorized as *elastic* or *inelastic*. The elastic demand decreases with the data price, while the inelastic demand does not change with the data price. Moreover, depending on whether their services are delay-tolerant or delay-sensitive, the client demand can be further categorized as time *shiftable* or *non-shiftable*. The time shiftable demand can be shifted across the slots within the same period (and thus a maximum $T$ slots delay is allowed), while the time non-shiftable demand cannot be shifted across the slots.

Notice that $D_{it}$ is the maximum amount of data that the clients (of host $i$) demand at slot $t$, beyond which the clients are not interested in consuming and paying. Therefore, we have the following *demand constraints* for a feasible strategy $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)$ of host $i$: (i) for the non-shiftable demand,

$$\beta_{it} \cdot R_{it} \leq D_{it}, \quad \forall t \in \mathcal{T}, \tag{4}$$

and (ii) for the shiftable demand,

$$\textstyle\sum_{t \in \mathcal{T}} \beta_{it} \cdot R_{it} \leq D_i, \tag{5}$$

where $D_i \triangleq \sum_{t \in \mathcal{T}} D_{it}$ is the total client demand in all slots. Note that host $i$ can choose to satisfy part of the client demand if the capacity or the host's battery energy is limited.

**Variables.** For convenience, we list the key variables in Table I. The system is assumed to be quasi-static, as some variables (i.e., those marked with the subscript $t$) may change in different slots $t \in \mathcal{T}$, while others (e.g., the wholesale price $w$, and the MVNO's pricing and reimbursement strategy $p_i$ and $\theta_i$) are fixed across $T$ slots in the same time period. Some of the notations will be further introduced next.

[3]For a typical 4G cellular network (such as LTE), the time length of one slot can be as small as 1ms.

TABLE I
KEY VARIABLES

| | | |
|---|---|---|
| **MVNO Parameters and Variables ($*$ denotes a decision variable)** | | |
| | $w$ | The unit wholesale price paid by the MVNO to MNOs; |
| $*$ | $p_i$ | The unit price charged by the MVNO to host $i$; |
| $*$ | $\theta_i$ | The free data quota ratio offered by the MVNO to host $i$; |
| **Host Parameters and Variables ($*$ denotes a decision variable)** | | |
| | $R_{it}$ | The 4G connection capacity of host $i$ at slot $t$; |
| | $\epsilon_{it}$ | The unit energy cost of host $i$ on the 4G connection; |
| | $\xi_{it}$ | The unit energy cost of host $i$ on the Wi-Fi connection; |
| | $x_{it}$ | The capacity scheduled for host $i$ itself at slot $t$; |
| | $y_{it}$ | The capacity scheduled for clients at slot $t$; |
| $*$ | $\alpha_{it}$ | The percentage of capacity scheduled for host $i$ at slot $t$; |
| $*$ | $\beta_{it}$ | The percentage of capacity scheduled for clients at slot $t$; |
| **Client Parameter** | | |
| | $D_{it}$ | The demand of clients (served by host $i$) at time slot $t$. |

### B. Problem Definition

We focus on the interaction of the MVNO and the hosts, and formulate it as a two-stage leader-follower (Stackelberg) game. The game *players* are the MVNO and the hosts.[4] In the first stage, the MVNO (leader) decides the pricing and reimbursing plan $(\boldsymbol{p}, \boldsymbol{\theta})$. In the second stage, every host $i$ (follower) decides its scheduling strategy $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)$, i.e., data allocation between itself and its clients.

**Host Objective.** The objective of each host $i$ is to maximize its payoff, which depends on the perceived utility $U_i$ from its own communication needs, the payment to the MVNO, the free data quota reimbursement from serving clients, and finally, the incurred energy cost. Note that the energy cost is important, as the hosts are mobile users and are expected to be highly sensitive about their battery energy consumptions. Specifically, given the strategy $(\boldsymbol{p}, \boldsymbol{\theta})$ of the MVNO, the *payoff* of host $i$, when choosing a strategy $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)$, is

$$\begin{aligned} J_i(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i; p_i, \theta_i) = U_i(\boldsymbol{x}_i) - \sum_{t\in\mathcal{T}} p_i \cdot (x_{it} - \theta_i \cdot y_{it}) \\ - \sum_{t\in\mathcal{T}} \epsilon_{it} \cdot (x_{it} + y_{it}) - \sum_{t\in\mathcal{T}} \xi_{it} \cdot y_{it}, \end{aligned} \tag{6}$$

where $x_{it} = \alpha_{it} \cdot R_{it}$ and $y_{it} = \beta_{it} \cdot R_{it}$. Notice that $\theta_i \cdot y_{it}$ is the free data quota to host $i$ (for forwarding $y_{it}$ bytes of client data), $\epsilon_{it}$ and $\xi_{it}$ are the unit energy costs incurred by host $i$ for transmitting one byte of data via the 4G link and the WiFi link during slot $t$, respectively.

**MVNO Objective.** The objective of the MVNO is to maximize its payoff, which consists of the revenue from the hosts and clients, and the payment (negative) to the respective MNOs. Formally, the MVNO's *payoff* can be written as

$$\begin{aligned} V(\boldsymbol{p}, \boldsymbol{\theta}; (\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)_{i\in\mathcal{I}}) = \sum_{i\in\mathcal{I}} \sum_{t\in\mathcal{T}} \big(p_i \cdot (x_{it} - \theta_i \cdot y_{it}) \\ + p_i \cdot y_{it} - w \cdot (x_{it} + y_{it})\big). \end{aligned} \tag{7}$$

The first two terms in the sum operation denote the revenue from host $i$ and the clients served through host $i$, and the last term denotes the payment to MNOs.

## III. OPTIMAL PRICING-REIMBURSING STRATEGY

In this section, we study the MVNO-host game under complete information, where both the MVNO and the hosts know all system parameters mentioned above. We solve the game by backward induction. First, we solve the host's best scheduling strategy in the second stage. Then, we study the MVNO's best pricing strategy in the first stage.

[4] Notice that we do not treat clients as game players, as their demands are system parameters and do not actively affect the MVNO' or hosts' decisions.

### A. Host's Best Decision in Stage II

Without loss of generality, we consider the decision problem of a particular host $i$. For the ease of presentation, we assume that the host service is elastic, and the client service is delay-sensitive (either elastic or inelastic). Note that our analysis can be easily extended to other scenarios.

Specifically, given the MVNO's pricing and reimbursing strategy $(p_i, \theta_i)$, host $i$ can derive the optimal scheduling strategy $(\boldsymbol{\alpha}_i^*, \boldsymbol{\beta}_i^*)$ by solving the following problem

$$\begin{aligned} \max_{\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i} \quad & J_i(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i; p_i, \theta_i) \\ \text{s.t.,} \quad & \text{(a) } \alpha_{it} + \beta_{it} \le 1, \quad \forall t \in \mathcal{T}, \\ & \text{(b) } \beta_{it} \cdot R_{it} \le D_{it}, \quad \forall t \in \mathcal{T}, \\ & \text{(c) } \alpha_{it},\ \beta_{it} \ge 0, \quad \forall t \in \mathcal{T}. \end{aligned} \tag{8}$$

It is easy to check that (8) is a convex optimization. Hence, it admits an optimal solution that can be characterized by the KKT conditions. To understand the structure and engineering insights of the optimal solution, in what follows we will first study the optimal scheduling $(\alpha_{it}^*, \beta_{it}^*)$ in a particular slot $t$ (fixed the scheduling decisions in other $T-1$ slots), and then study the optimal scheduling $(\boldsymbol{\alpha}_i^*, \boldsymbol{\beta}_i^*) = (\alpha_{it}^*, \beta_{it}^*)_{t\in\mathcal{T}}$ of all $T$ slots jointly, i.e., the solution of (8).

*1) **Optimal scheduling of slot** $t$:* Now we consider the scheduling of a single slot $t$. We first present a scheduling rule that converges to the optimal single-slot scheduling. Then, we characterize this optimal scheduling systematically.

Let $f_{it}$ and $g_{it}$ denote the first-order derivatives of host $i$'s payoff $J_i(\cdot)$ with respect to $\alpha_{it}$ and $\beta_{it}$, respectively, i.e.,[5]

$$f_{it} \triangleq \frac{\partial J_i}{\partial \alpha_{it}} = (U_i'(\boldsymbol{x}_{i,-t}^*, x_{it}) - p_i - \epsilon_{it}) \cdot R_{it}, \quad \alpha_{it} \in [0,1],$$

$$g_{it} \triangleq \frac{\partial J_i}{\partial \beta_{it}} = \begin{cases} (p_i \cdot \theta_i - \epsilon_{it} - \xi_{it}) \cdot R_{it}, & \beta_{it} \in [0, \frac{D_{it}}{R_{it}}], \\ 0, & \beta_{it} \in [\frac{D_{it}}{R_{it}}, 1], \end{cases}$$

where $x_{it} = \alpha_{it} \cdot R_{it}$ and $\boldsymbol{x}_{i,-t}^* = (\alpha_{ik}^* \cdot R_{ik})_{k\in\mathcal{T}, k\neq t}$. Notice that $\frac{\partial J_i}{\partial \beta_{it}} = 0$ when $\beta_{it} \ge \frac{D_{it}}{R_{it}}$. This is because the host payoff $J_i$ does not change with $\beta_{it}$ when $\beta_{it} \ge \frac{D_{it}}{R_{it}}$, as the clients will not accept data forwarding beyond their demands.

Mathematically, $f_{it}$ and $g_{it}$ denote the *increasing rates* of host $i$'s payoff along two different directions, reflecting the change of its payoff when slightly changing $\alpha_{it}$ or $\beta_{it}$. We can see that (i) $f_{it}$ decreases with $\alpha_{it}$ (by the concavity of $U_i$) but is independent of $\beta_{it}$, hence can be written as a decreasing function $f_{it}(\alpha_{it})$, and (ii) $g_{it}$ is independent of $\alpha_{it}$, and can be written as a step function $g_{it}(\beta_{it})$.

Based on the above, we immediately have the following observations: (i) if $f_{it} > 0$ (or $f_{it} < 0$), then host $i$ can increase its payoff by increasing $\alpha_{it}$ (or decreasing $\alpha_{it}$); similarly, if $g_{it} > 0$ (or $g_{it} < 0$), then host $i$ can increase its payoff by increasing $\beta_{it}$ (or decreasing $\beta_{it}$); and (ii) if $f_{it} > g_{it}$ (or $f_{it} < g_{it}$), then host $i$ can achieve a higher (or lower) payoff by increasing $\alpha_{it}$ than by increasing $\beta_{it}$ the same amount. These observations help us understand how to change $\alpha_{it}$ and $\beta_{it}$ to increase host $i$'s payoff $J_i(\cdot)$.

[5] For notational convenience, we write $U_i'(\boldsymbol{x})$ with $\boldsymbol{x} = (\boldsymbol{x}_{i,-t}^*, x_{it})$ as $U_i'(\boldsymbol{x}_{i,-t}^*, x_{it})$. The same for $J_i(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, p_i, \theta_i)$ in (10).

**Algorithm 1:** Single-Slot Scheduling Rule.

**Initialization:** $\alpha_{it} = 0$ and $\beta_{it} = 0$;
**while** $(\alpha_{it} + \beta_{it} < 1)$ *and* $(f_{it}(\alpha_{it}) > 0$ *or* $g_{it}(\beta_{it}) > 0)$ **do**
  **if** $f_{it}(\alpha_{it}) > g_{it}(\beta_{it})$ *and* $f_{it}(\alpha_{it}) > 0$ **then**
    $\alpha_{it} = \alpha_{it} + \Delta$ (Schedule capacity to host $i$);
  **else if** $g_{it}(\beta_{it}) > f_{it}(\alpha_{it})$ *and* $g_{it}(\beta_{it}) > 0$ **then**
    $\beta_{it} = \beta_{it} + \Delta$ (Schedule capacity to clients);

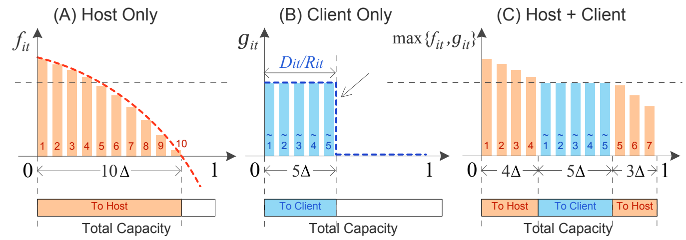



Fig. 2. Illustration of the optimal scheduling under different scenarios.

By the above observations, we have the following **scheduling rule**: (i) schedule a small capacity $\Delta$ to host $i$ (i.e., increase $\alpha_{it}$ with $\Delta$), if $f_{it}$ is positive and larger than $g_{it}$, and (ii) schedule a small capacity $\Delta$ to clients (i.e., increase $\beta_{it}$ with $\Delta$), if $g_{it}$ is positive and larger than $\alpha_{it}$. The above process starts from an initial empty schedule ($\alpha_{it} = \beta_{it} = 0$), and ends until one of the following situations occurs: (i) the entire 4G capacity is scheduled, i.e., $\alpha_{it} + \beta_{it} = 1$, or (ii) the host neither benefits from consuming data ($f_{it}(\alpha_{it}) \le 0$), nor from forwarding data for clients ($g_{it}(\beta_{it}) \le 0$). Formally, we present the complete scheduling rule in Algorithm 1, and show its optimality in the following Lemma 1.

**Lemma 1.** *The scheduling rule in Algorithm 1 converges to the optimal single-slot scheduling* $(\alpha^*_{it}, \beta^*_{it})$ *when* $\Delta \to 0$, *where* $\Delta$ *is the capacity increment in each iteration.*

Figure 2 illustrates the optimal scheduling achieved by the Algorithm 1 under different scenarios. In scenario A (with the host service only), the host continues scheduling capacity to itself until $f_{it} \le 0$, and the optimal scheduling is $\alpha^*_{it} = 10\Delta$. In scenario B (with the client service only), the host continues scheduling capacity to clients until $g_{it} = 0$, and the optimal scheduling is $\beta^*_{it} = \frac{D_{it}}{R_{it}} = 5\Delta$. In scenario C (with both host and client services), the host first schedules $4\Delta$ units of capacity to itself (as $f_{it} > g_{it}$ when $\alpha_{it} \le 4\Delta$), then schedules $5\Delta$ units to clients (as $f_{it} < g_{it}$ when $\alpha_{it} > 4\Delta$ and $\beta_{it} \le 5\Delta$), and finally schedules the last $3\Delta$ units to itself again (as $g_{it} = 0$ when $\beta_{it} > 5\Delta$ but $f_{it} > 0$). The corresponding optimal scheduling is $\alpha^*_{it} = 7\Delta$ and $\beta^*_{it} = 5\Delta$.

Next we characterize the optimal scheduling solution $(\alpha^*_{it}, \beta^*_{it})$ analytically. Notice that in the trivial case of $g_{it} \le 0$ (which implies that the free quota received by the host can not compensate the energy cost for forwarding data for clients), the host will not schedule any capacity to clients according to Algorithm 1, and the optimal scheduling is simply $\alpha^*_{it} = [\arg_\alpha f_{it}(\alpha) = 0]_0^1$ and $\beta^*_{it} = 0$, where $[x]_0^y = \max\{0, \min\{x, y\}\}$. Thus, in what follows, we will focus on the non-trivial case of $g_{it} > 0$.

To better characterize the optimal scheduling solution in this non-trivial case, we define a new equivalent problem with nice structures and properties as follows. First, we divide the total capacity $R_{it}$ into two parts in the following way:[6]

$$R^{\mathrm{A}}_{it} = \min\{R_{it},\ D_{it}\}, \quad R^{\mathrm{B}}_{it} = R_{it} - R^{\mathrm{A}}_{it}, \tag{9}$$

where $R^{\mathrm{A}}_{it}$ is the *shared* capacity which can be scheduled to either the host or the clients, and $R^{\mathrm{B}}_{it}$ is the *dedicated* capacity which can only be scheduled to the host. Let us denote $\alpha^{\mathrm{A}}_{it}$ and $\beta_{it}$ as the percentages (comparing with the total capacity) in $R^{\mathrm{A}}_{it}$ scheduled to host $i$ and to its served clients, respectively, and denote $\alpha^{\mathrm{B}}_{it}$ as the percentage (comparing with the total capacity) in $R^{\mathrm{B}}_{it}$ scheduled to host $i$. Then, we define the following new scheduling problem: *Finding the optimal scheduling* $(\alpha^{\mathrm{A}}_{it}, \beta_{it})$ *of the shared capacity* $R^{\mathrm{A}}_{it}$ *(to host* $i$ *and the clients, respectively), and the optimal scheduling* $\alpha^{\mathrm{B}}_{it}$ *of the dedicated capacity* $R^{\mathrm{B}}_{it}$ *(to host* $i$*).* Formally,

$$\begin{aligned} \max_{\alpha^{\mathrm{A}}_{it}, \alpha^{\mathrm{B}}_{it}, \beta_{it}} \quad & J_i(\boldsymbol{\alpha}^*_{i,-t}, \alpha_{it}, \boldsymbol{\beta}^*_{i,-t}, \beta_{it}; p_i, \theta_i) \\ \text{s.t.,} \quad & \text{(a) } \alpha^{\mathrm{A}}_{it} + \beta_{it} \le \tfrac{R^{\mathrm{A}}_{it}}{R_{it}}, \\ & \text{(b) } \alpha^{\mathrm{B}}_{it} \le \tfrac{R^{\mathrm{B}}_{it}}{R_{it}}, \\ & \text{(c) } \alpha^{\mathrm{A}}_{it},\ \alpha^{\mathrm{B}}_{it},\ \beta_{it} \ge 0, \end{aligned} \tag{10}$$

where $\alpha_{it} = \alpha^{\mathrm{A}}_{it} + \alpha^{\mathrm{B}}_{it}$ is the total capacity scheduled to host $i$, and $\beta_{it}$ is the total capacity scheduled to clients. Note that the client demand constraint (i.e., $\beta_{it} \le \frac{R^{\mathrm{A}}_{it}}{R_{it}}$) is implied in (a). Intuitively, in this new problem we restrict the client service within the shared capacity $R^{\mathrm{A}}_{it}$, and thus reconcile (but not fully decouple) the scheduling among the host and the clients.

In the following lemmas, we show that the new problem (10) is equivalent to the original single-slot scheduling problem, and further characterize the optimal solution of (10).

**Lemma 2.** *If* $(\alpha^{\mathrm{A}*}_{it}, \alpha^{\mathrm{B}*}_{it}, \beta^*_{it})$ *is the optimal solution of (10), then* $(\alpha^*_{it} = \alpha^{\mathrm{A}*}_{it} + \alpha^{\mathrm{B}*}_{it},\ \beta^*_{it})$ *is the optimal scheduling of slot* $t$ *in the original problem.*

**Lemma 3.** *The optimal solution* $(\alpha^{\mathrm{A}*}_{it}, \alpha^{\mathrm{B}*}_{it}, \beta^*_{it})$ *of (10) is*

$$\begin{cases} \alpha^{\mathrm{B}*}_{it} = \left[\arg_\alpha f^{\mathrm{B}}_{it}(\alpha) = 0\right]_0^{R^{\mathrm{B}}_{it}/R_{it}} \\ \alpha^{\mathrm{A}*}_{it} = \left[\arg_\alpha f^{\mathrm{A}}_{it}(\alpha) = g_{it}\right]_0^{R^{\mathrm{A}}_{it}/R_{it}} \\ \beta^*_{it} = \frac{R^{\mathrm{A}}_{it}}{R_{it}} - \alpha^{\mathrm{A}*}_{it} \end{cases} \tag{11}$$

*where* $f^{\mathrm{A}}_{it} = \frac{\partial J_i}{\partial \alpha^{\mathrm{A}}_{it}}$, $f^{\mathrm{B}}_{it} = \frac{\partial J_i}{\partial \alpha^{\mathrm{B}}_{it}}$, *and* $g_{it} = \frac{\partial J_i}{\partial \beta_{it}}$.

The optimal solution in Lemma 3 implies that the host will schedule the dedicated capacity $R^{\mathrm{B}}_{it}$ to itself when $f^{\mathrm{A}}_{it}(\alpha^{\mathrm{A}}_{it}) > 0$, but will schedule the shared capacity $R^{\mathrm{A}}_{it}$ to itself only when $f^{\mathrm{A}}_{it}(\alpha^{\mathrm{A}}_{it}) > g_{it}$ (due to the competition of the client demand). Consider the example in Figure 2.C, we have: $R^{\mathrm{A}}_{it} = 5\Delta$ and $R^{\mathrm{B}}_{it} = 7\Delta$. By the first equation of (11), we can obtain: $\alpha^{\mathrm{B}*}_{it} = 7\Delta$. By the second equation of (11), we can obtain: $\alpha^{\mathrm{A}*}_{it} = 0$ (as $f^{\mathrm{A}}_{it}(\alpha) < g_{it}$ since the host is already assigned with $\alpha^{\mathrm{B}*}_{it} = 7\Delta$ units of dedicated capacity). Finally, by the last equation of (11), we can obtain: $\beta^*_{it} = 5\Delta$. Obviously, this solution is exactly same as the one shown in Figure 2.C.

We further notice that $\frac{\partial J_i}{\partial \alpha^{\mathrm{A}}_{it}} = \frac{\partial J_i}{\partial \alpha^{\mathrm{B}}_{it}} = (U'_i(\boldsymbol{x}^*_{i,-t}, x_{it}) - p_i - \epsilon_{it}) \cdot R_{it}$, where $x_{it} = (\alpha^{\mathrm{A}}_{it} + \alpha^{\mathrm{B}}_{it}) \cdot R_{it}$. Then the optimal

[6] We would like to point out that the following analysis also applies to the trivial case of $g_{it} < 0$, by simply choosing $R^{\mathrm{A}}_{it} = 0$ and $R^{\mathrm{B}}_{it} = R_{it}$.

scheduling $\alpha_{it}^{\text{B}*}$ and $\alpha_{it}^{\text{A}*}$ in (11) can be rewritten as

$$\alpha_{it}^{\text{B}*} = \big[\arg_{\alpha_{it}^{\text{B}}} U_i'(\boldsymbol{x}_{i,-t}^*, x_{it}) = p_i + \epsilon_{it}\big]_0^{R_{it}^{\text{B}}/R_{it}},$$
$$\alpha_{it}^{\text{A}*} = \big[\arg_{\alpha_{it}^{\text{A}}} U_i'(\boldsymbol{x}_{i,-t}^*, x_{it}) = p_i \cdot (1+\theta_i) - \xi_{it}\big]_0^{R_{it}^{\text{A}}/R_{it}}.$$

Next we discuss some insights from the optimal scheduling solution in Lemma 3. In economics, (i) $U_i'(\cdot)$ is the *marginal utility*, which reflects the utility gain when slightly increasing the capacity (data) consumed by the host; (ii) $p_i + \epsilon_{it}$ is the *marginal cost*, which reflects the loss of payment and energy when slightly increasing the *dedicated* capacity consumed by the host; (iii) $p_i \cdot (1+\theta_i) - \xi_{it}$ is the *opportunity cost*, which reflects the loss of payment, energy, and all potential gain (from other choices such as scheduling the capacity to clients) when slightly increasing the *shared* capacity consumed by the host.[7] Then, the optimal scheduling solution in Lemma 3 implies that the optimal scheduling would either equalize the marginal utility and the marginal cost (for the dedicated capacity), or equalize the marginal utility and the opportunity cost (for the shared capacity).

*2)* ***Optimal scheduling of all slots****:* Now we study the optimal scheduling $(\boldsymbol{\alpha}_i^*, \boldsymbol{\beta}_i^*) = (\alpha_{it}^*, \beta_{it}^*)_{t\in\mathcal{T}}$ of the entire period of $T$ slots. Similarly, we will first present an optimal multi-slot scheduling rule, and then characterize the optimal multi-slot scheduling solution systematically.

With a similar analysis in the single-slot case, we have the following **multi-slot scheduling rule**: schedule a small capacity $\Delta$ in each slot $t \in \mathcal{T}$ *with un-scheduled capacity* to (i) host $i$ (i.e., increasing $\alpha_{it}$ with $\Delta$) if $f_{it}$ is positive and larger than $\{f_{ik}\}_{k\neq t, k\in\mathcal{T}}$ and $\{g_{ik}\}_{k\in\mathcal{T}}$, or (ii) clients (i.e., increasing $\beta_{it}$ with $\Delta$) if $g_{it}$ is positive and larger than $\{f_{ik}\}_{k\in\mathcal{T}}$ and $\{g_{ik}\}_{k\neq t,k\in\mathcal{T}}$. Note this scheduling rule is a natural multi-slot extension of that in the single-slot case (Algorithm 1). Namely, in the single-slot case, the host only has two choices in each iteration depending on the value of $f_{it}$ and $g_{it}$, i.e., scheduling to itself (when $f_{it} \geq \max\{0, g_{it}\}$) or to clients (when $g_{it} \geq \max\{0, f_{it}\}$). In this multi-slot case, the host has a total of $2T$ choices in each iteration depending on the values of $\{f_{it}\}_{t\in\mathcal{T}}$ and $\{g_{it}\}_{t\in\mathcal{T}}$, i.e., scheduling the capacity in each of these $T$ slots to itself (when $f_{it} \geq \max\{0, f_{ik}, g_{ik}\}_{k\in\mathcal{T}}$) or to clients (when $g_{it} \geq \max\{0, f_{ik}, g_{ik}\}_{k\in\mathcal{T}}$).

Next we characterize the optimal multi-slot scheduling decision $(\boldsymbol{\alpha}_i^*, \boldsymbol{\beta}_i^*)$. Similarly, we divide the total capacity $R_{it}$ in each slot $t$ into two parts $R_{it}^{\text{A}}$ and $R_{it}^{\text{B}}$ in the following way: (i) if $g_{it} \geq 0$, then $R_{it}^{\text{A}} = \min\{R_{it}, D_{it}\}$ and $R_{it}^{\text{B}} = R_{it} - R_{it}^{\text{A}}$, and (ii) if $g_{it} < 0$, then $R_{it}^{\text{A}} = 0$ and $R_{it}^{\text{B}} = R_{it}$. Then, we define the new equivalent scheduling problem:

$$\begin{aligned} \max_{\{\alpha_{it}^{\text{A}}, \alpha_{it}^{\text{B}}, \beta_{it}\}_{t\in\mathcal{T}}} \quad & J_i(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i; p_i, \theta_i) \\ \text{s.t.,} \quad & \text{(a) } \alpha_{it}^{\text{A}} + \beta_{it} \leq \tfrac{R_{it}^{\text{A}}}{R_{it}}, \quad \forall t\in\mathcal{T}, \\ & \text{(b) } \alpha_{it}^{\text{B}} \leq \tfrac{R_{it}^{\text{B}}}{R_{it}}, \quad \forall t\in\mathcal{T}, \\ & \text{(c) } \alpha_{it}^{\text{A}},\ \alpha_{it}^{\text{B}},\ \beta_{it} \geq 0, \quad \forall t\in\mathcal{T}, \end{aligned} \tag{12}$$

where $\alpha_{it} = \alpha_{it}^{\text{A}} + \alpha_{it}^{\text{B}}$ and $\beta_{it} = \beta_{it}$.

[7]An opportunity cost is the loss of potential gain from other choices when one choice is chosen. In our model, when scheduling a unit of shared capacity to host $i$, the host will suffer a cost $p_i + \epsilon_{it}$. When scheduling this capacity to clients, however, the host gains a payoff of $p_i \cdot \theta_i - \epsilon_{it} - \xi_{it}$. Thus, the loss of potential gain is $p_i \cdot \theta_i - \epsilon_{it} - \xi_{it} + p_i + \epsilon_{it} = p_i \cdot (1+\theta_i) - \xi_{it}$.

**Algorithm 2:** Iterative Water-filling Procedure.

```
Initialization: W = 0, and W_VIR = −1;
while W ≠ W_VIR do
    for t = 1 to T do                                  /* scheduling */
        if W > p_i + ε_it then
            α^B_it = R^B_it / R_it  (Schedule capacity R^B_it to host i);
        if W > p_i · (1 + θ_i) − ξ_it then
            α^A_it = R^A_it / R_it  (Schedule capacity R^A_it to host i);
    Compute the virtual water-level: W_VIR = U'_i(x_i)
    if W_VIR > W then                                  /* WL updating */
        Increase the water-level: W = W + δ
    else if W_VIR < W then
        Decrease the water-level: W = W − δ
```

We can similarly show that the new problem (12) is equivalent to the original problem (8), in the sense that if $(\alpha_{it}^{\text{A}*}, \alpha_{it}^{\text{B}*}, \beta_{it}^*)_{t\in\mathcal{T}}$ is the solution of (12), then $(\alpha_{it}^* = \alpha_{it}^{\text{A}*} + \alpha_{it}^{\text{B}*}, \beta_{it}^*)_{t\in\mathcal{T}}$ is the solution of (8). Moreover, we give the optimal solution of (12) in the following lemma.

**Lemma 4.** *The optimal solution of (12) is given by*

$$\begin{cases} \alpha_{it}^{\text{B}*} = \big[\arg_{\alpha_{it}^{\text{B}}} U_i'(\boldsymbol{x}_i) = p_i + \epsilon_{it}\big]_0^{R_{it}^{\text{B}}/R_{it}} \\ \alpha_{it}^{\text{A}*} = \big[\arg_{\alpha_{it}^{\text{A}}} U_i'(\boldsymbol{x}_i) = p_i \cdot (1+\theta_i) - \xi_{it}\big]_0^{R_{it}^{\text{A}}/R_{it}} \\ \beta_{it}^* = \frac{R_{it}^{\text{A}}}{R_{it}} - \alpha_{it}^{\text{A}*} \end{cases} \tag{13}$$

*for all slots $t \in \mathcal{T}$, where $\boldsymbol{x}_i = \{(\alpha_{it}^{\text{A}} + \alpha_{it}^{\text{B}}) \cdot R_{it}\}_{t\in\mathcal{T}}$.*

Note that in (13), the scheduling $\alpha_{it}^{\text{A}*}$ and $\alpha_{it}^{\text{B}*}$ (i.e., the dedicated and shared capacities in slot $t$ scheduled to host $i$ itself) are coupled with each other by the common term – marginal utility $U_i'(\boldsymbol{x}_i)$. To decouple and solve them explicitly, we propose an iterative *water-filling* procedure as follows. First, we interpret (i) the marginal utility $U_i'(\boldsymbol{x}_i)$ as the *water-level*, (ii) the marginal cost $p_i+\epsilon_{it}$ as the *floor* of the dedicated capacity $R_{it}^{\text{B}}$, and (iii) the opportunity cost $p_i \cdot (1+\theta_i) - \xi_{it}$ as the *floor* of the shared capacity $R_{it}^{\text{A}}$. Then we define the following **water-filling procedure**: gradually schedule the capacity (both dedicated and shared) to the host whenever a floor is lower than the water-level, and schedule all of the remaining shared capacity to the clients.

Specifically, in each round, the water-filling procedure consists of two sequential processes: *scheduling* and *water-level updating*. In the scheduling process, the capacity is scheduled in the following way: (i) any dedicated capacity $R_{it}^{\text{B}}$ with a floor lower than the water-level is scheduled to the host (to $\alpha_{it}^{\text{B}}$), (ii) any shared capacity $R_{it}^{\text{A}}$ with a floor lower than the water-level is scheduled to the host (to $\alpha_{it}^{\text{A}}$), and (iii) any remaining shared capacity is scheduled to clients (to $\beta_{it}$). In the water-level updating process, it computes a virtual water-level (the $U_i'(\cdot)$ under the new $\alpha_{it}^{\text{A}}$ and $\alpha_{it}^{\text{B}}$), and then update the (real) water-level in the following way: (i) increase the water-level (with a small amount $\delta$) if the virtual water-level is larger than the real one, and (ii) decrease the water-level otherwise. The whole procedure starts with an initial arbitrary water-level, and ends until the water-level does not change. Formally, we show the complete water-filling process in Algorithm 2.

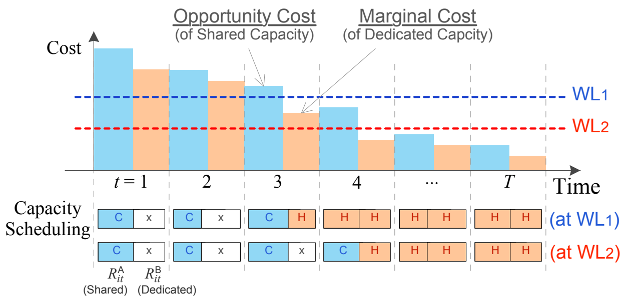



Fig. 3. Illustration of the scheduling under two different water-levels $\mathsf{WL}_1$ and $\mathsf{WL}_2$: C – the capacity scheduled to clients, H – the capacity scheduled to the host, and x – the capacity not scheduled to any user.

Figure 3 illustrates the scheduling solution under different water-levels $\mathsf{WL}_1$ and $\mathsf{WL}_2$. The capacity denoted by "H" is scheduled to the host, the capacity denoted by "C" is scheduled to the clients, and the capacity denoted by "x" is not scheduled to any user. Obviously, with the decrease of the water-level (e.g, from $\mathsf{WL}_1$ to $\mathsf{WL}_2$), the capacity scheduled to the host decreases as well.

*3) **Property of the optimal scheduling**:* Now we provide some useful properties for the derived optimal scheduling. Denote $X_i$ as the total capacity scheduled to host $i$ and $Y_i$ as the total capacity to clients under the optimal scheduling:

$$X_i \triangleq \sum_{t\in\mathcal{T}} \alpha_{it}^* \cdot R_{it}, \quad Y_i \triangleq \sum_{t\in\mathcal{T}} \beta_{it}^* \cdot R_{it}. \tag{14}$$

Note that both $X_i$ and $Y_i$ are functions of $p_i$ and $\theta_i$.

**Proposition 1.** *(i) $X_i$ decreases with $\theta_i$, (ii) $Y_i$ increases with $\theta_i$, and (iii) $X_i + Y_i$ increases with $\theta_i$.*

**Proposition 2.** *If $D_{it}$ is independent of $p_i$ (inelastic client demand), then (i) $X_i$ decreases with $p_i$, (ii) $Y_i$ increases with $p_i$, and (iii) $X_i + Y_i$ may decrease or increase with $p_i$.*

### *B. MVNO's Best Decision in Stage I*

Based on the hosts' best response in Stage II, the MVNO determines the best pricing and reimbursing strategy $(\boldsymbol{p}^*, \boldsymbol{\theta}^*)$ that maximizes its payoff defined in (7). Specifically, the MVNO's optimization problem is

$$\begin{aligned} &\max_{\boldsymbol{p},\boldsymbol{\theta}} \quad V(\boldsymbol{p}, \boldsymbol{\theta}; (\boldsymbol{\alpha}_i^*, \boldsymbol{\beta}_i^*)_{i\in\mathcal{I}}) \\ &\text{s.t.,} \quad (\boldsymbol{\alpha}_i^*, \boldsymbol{\beta}_i^*) \text{ is solved by (8)}, \quad \forall i \in \mathcal{I}, \end{aligned} \tag{15}$$

where $(\boldsymbol{\alpha}_i^*, \boldsymbol{\beta}_i^*)$ is the host $i$'s best scheduling decision under $p_i$ and $\theta_i$, and both $\boldsymbol{\alpha}_i^*$ and $\boldsymbol{\beta}_i^*$ are functions of $p_i$ and $\theta_i$. Substituting (14), we can rewrite the MVNO's payoff as

$$V(\boldsymbol{p}, \boldsymbol{\theta}; (\boldsymbol{\alpha}_i^*, \boldsymbol{\beta}_i^*)_{i\in\mathcal{I}})) = \sum_{i\in\mathcal{I}} V_i(p_i, \theta_i; \boldsymbol{\alpha}_i^*, \boldsymbol{\beta}_i^*), \tag{16}$$

where

$$\begin{aligned} V_i(p_i, \theta_i; \boldsymbol{\alpha}_i^*, \boldsymbol{\beta}_i^*) &= V_i^{\mathrm{H}}(p_i, \theta_i) + V_i^{\mathrm{C}}(p_i, \theta_i) \\ &\triangleq (p_i - w) \cdot X_i(p_i, \theta_i) + (p_i(1-\theta_i) - w) \cdot Y_i(p_i, \theta_i) \end{aligned}$$

denotes the MVNO's payoff achieved through host $i$, where $V_i^{\mathrm{H}}(p_i, \theta_i) \triangleq (p_i - w) \cdot X_i(p_i, \theta_i)$ is the payoff directly from host $i$, and $V_i^{\mathrm{C}}(p_i, \theta_i) \triangleq (p_i(1-\theta_i) - w) \cdot Y_i(p_i, \theta_i)$ is the payoff from the clients served through host $i$.

By (16), we can decouple the MVNO's optimization problem (15) into a set of independent sub-problems, each associated with one host. Without loss of generality, we focus on the $i$-th sub-problem (associated with host $i$):

$$\max_{p_i, \theta_i \geq 0} \quad V_i^{\mathrm{H}}(p_i, \theta_i) + V_i^{\mathrm{C}}(p_i, \theta_i). \tag{17}$$

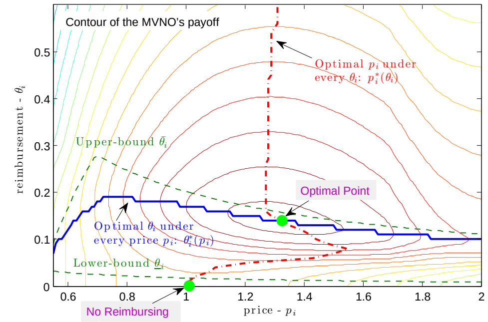



Fig. 4. Illustration of the optimal reimbursing strategy $\theta_i^*(p_i)$ given $p_i$, and the optimal pricing strategy $p_i^*(\theta_i)$ given $\theta_i$.

Notice that changing the price $p_i$ or changing the free data quota ratio $\theta_i$ have different impacts on $V_i^{\mathrm{H}}$ and $V_i^{\mathrm{C}}$. For example, increasing $p_i$ will improve the payoff $V_i^{\mathrm{C}}$ from clients (since both the unit payoff $p_i(1-\theta_i) - w$ and the capacity consumption $Y_i$ increases), but may hurt the payoff $V_i^{\mathrm{H}}$ from host $i$ (due to the reduction of its capacity consumption $X_i$). Similarly, increasing $\theta_i$ will hurt the payoff $V_i^{\mathrm{H}}$ from host $i$ (due to the reduction of its capacity consumption $X_i$), but may increase or decrease the payoff $V_i^{\mathrm{C}}$ from clients (as the unit payoff $p_i(1-\theta_i) - w$ decreases but the capacity consumption $Y_i$ increases). Thus, the MVNO faces the tradeoff of balancing the payoff from host $i$ and the payoff from clients.

We can easily check that the problem (17) is non-convex (due to the irregular changes of $V_i^{\mathrm{H}}$ and $V_i^{\mathrm{C}}$ with respect to $p_i$ and $\theta_i$), and thus KKT conditions no longer guarantee the optimality of the solution. Fortunately, we can characterize some useful properties of (17), which will help us to find out the optimal solution effectively.

*1) **Optimal Reimbursing Strategy**:* We first study the optimal reimbursing strategy $\theta_i^*$ (also denoted by $\theta_i^*(p_i)$) given the pricing strategy $p_i$. Due to the non-convexity of (17), it is hard to derive the optimal $\theta_i^*$ analytically. To this end, we will characterize the lower-bound and upper-bound of $\theta_i^*$ under $p_i$. Based on these bounds, the optimal $\theta_i^*$ can be effectively found using numerical methods such as the exhaustive search.

**Lemma 5.** *Given any price $p_i$, the optimal $\theta_i^*(p_i)$ is bounded within $[\underline{\theta}_i, \bar{\theta}_i]$, where $\underline{\theta}_i = \min\left\{\frac{\epsilon_{it}+\xi_{it}}{p_i}, \forall t \in \mathcal{T}\right\}$, and $\bar{\theta}_i = \max\left\{\frac{\epsilon_{it}+\xi_{it}}{p_i}, \frac{U_i'(X_i)+\xi_{it}}{p_i} - 1, \forall t \in \mathcal{T}\right\}$.*[8]

We can understand the two bounds as follows. If $\theta_i < \underline{\theta}_i$, then $p_i \cdot \theta_i - \epsilon_{it} - \xi_{it} < 0$, $\forall t \in \mathcal{T}$. In this case, host $i$ will never schedule any capacity to clients, that is, the system is equivalent to a non-reimbursement system. This implies that any $\theta_i$ lower than $\underline{\theta}_i$ is meaningless. If $\theta_i > \bar{\theta}_i$, then $p_i \cdot \theta_i - \epsilon_{it} - \xi_{it} > 0$ and $U_i'(X_i) < p_i \cdot (1+\theta_i) - \xi_{it}$, $\forall t \in \mathcal{T}$. In this case, host $i$ will schedule all of the shared capacity to clients. Thus, further increasing $\theta_i$ will not affect $X_i$ and $Y_i$, but will reduce the MVNO's unit profit achieved from $Y_i$ (since the MVNO needs to reimburse host $i$ more free data quota). This implies that any $\theta_i$ higher than $\bar{\theta}_i$ is dominated by $\bar{\theta}_i$.

Figure 4 illustrates how the lower-bound $\underline{\theta}_i$, upper-bound $\bar{\theta}_i$, and the optimal $\theta_i^*(p_i)$ (the bold solid curve) change with price $p_i$. We can see that $\theta_i^*(p_i)$ first increases with $p_i$ when

[8] With a little abuse of notation, we use $U_i'(X_i)$ to denote the marginal utility of host $i$ when consuming $X_i$ bytes of data.

$p_i$ is small (e.g., when $p_i \leq 0.8$), and then decreases with $p_i$ when $p_i$ is large (e.g., when $p_i \geq 0.8$).

**Observation 1.** *The optimal $\theta_i^*(p_i)$ is a unimodal function. That is, there exists a critical price $\widehat{p}_i$, such that $\theta_i^*(p_i)$ increases with $p_i$ if $p_i \leq \widehat{p}_i$, and decreases with $p_i$ if $p_i \geq \widehat{p}_i$.*

*2) **Optimal Pricing Strategy**:* Now we study the optimal pricing strategy $p_i^*$ or $p_i^*(\theta_i)$ given the reimbursing $\theta_i$.

We first notice that when $\theta_i = 0$ (i.e., no free data quota), host $i$ has no incentive to serve clients (i.e., $\beta_{it}^* = 0$ and $Y_i = 0$) (since serving clients will simply increase the host's energy cost). In this case, the model degenerates to a traditional pricing system without reimbursement (i.e., with pricing only). Thus, we can immediately find that with the optimal reimbursing strategy, the MVNO can always achieve a payoff no worse than that without reimbursement.

Figure 4 illustrates how the optimal price $p_i^*(\theta_i)$ (the bold dash-dot curve) changes with $\theta_i$. We can see that $p_i^*(\theta_i^*)$ first increases with $\theta_i$ when $\theta_i$ is small (e.g., when $\theta_i \leq 10\%$), and then decreases with $\theta_i$ when $\theta_i$ becomes a little bit larger (e.g., when $10\% \leq \theta_i \leq 15\%$), and finally slightly increases with $\theta_i$ when $\theta_i$ is large (e.g., when $\theta_i \geq 15\%$).

**Observation 2.** *There exists two critical prices $\widehat{\theta}_i$ and $\widetilde{\theta}_i$, such that the optimal $p_i^*(\theta_i)$ increases with $\theta_i$ when $\theta_i \leq \widehat{\theta}_i$, and decreases with $\theta_i$ when $\widehat{\theta}_i \leq \theta_i \leq \widetilde{\theta}_i$, and finally slightly increases with $\theta_i$ when $\theta_i \geq \widetilde{\theta}_i$.*

*3) **Optimal Pricing-Reimbursing Strategy**:* Combining the best reimbursing strategy $\theta_i^*(p_i)$ given the price $p_i$ and the best pricing strategy $p_i^*(\theta_i)$ given the reimbursing $\theta_i$, we immediately have the following optimal solution.

**Lemma 6.** *The optimal pricing reimbursing policy $(p_i^*, \theta_i^*)$ occurs at the intersection point of $\theta_i^*(p_i)$ and $p_i^*(\theta_i)$.*

Figure 4 illustrates the optimal pricing and reimbursing point (denoted by the upper dot). In this example, the optimal price is $p_i^* \approx 1.3$ and the optimal reimbursement ratio is $\theta_i^* \approx 14\%$. We also illustrate the optimal pricing point in a pricing-only system (without reimbursing), denoted by lower dot with $\theta_i = 0$ and $p_i^*(0) \approx 1$. Obviously, with the reimbursement, the MVNO can charge a higher (optimal) price to the host, since the clients actually behave as competitors of the host, which will hurt the revenue of the host. On the other hand, the host benefits from the reimbursement.

## IV. Complexity and Security

Now we briefly discuss the complexity and security issues arising in the proposed hybrid pricing UPC model. There are several security issues in a UPC model. For example, how to guarantee the information security when sharing the connectivity among strangers, and how to deal with the "cheating" problem (e.g., a host may report a larger amount of forwarding data than what he actually forwards so as to gain a higher free quota). Furthermore, the deployment complexity is also a big concern for a UPC model. For example, a host needs to report how much data he actually forwards for which clients (such that the MVNO can successfully reimburse the host and charge the clients), which will create a large operational overhead among the host, the clients, and the MVNO. All of the above problems need to be satisfactorily addressed in order for such business model to be widely adopted in practice.

Despite of the different focuses of complexity and security issues, we will show that they can be (at least partly) addressed by *using a dedicated MVNO-issued 4G device as the MiFi hotspot*, instead of using the user's personal 4G device. Specifically, the MVNO will offer every subscriber (who want to act as a host) a specific 4G device, with a small price or even free. This device allows the subscriber itself to access the Internet via any mobile cellular network that the MVNO has signed a contract with. More importantly, it can also operate as a MiFi hotspot, and provide Internet access services for other users. Such devices are already developed by Novatel, Verizon, Sprint and several other companies. Notice that a host will gain reimbursements from the MVNO only when sharing its connectivity via this specific 4G device issued by the MVNO. This is exactly same as what Karma did today [5]. We can see that by using such a specific MiFi hotspot device, the above security and complexity issues can be largely addressed, as all of the data encrypting/decrypting, streaming, classifying, counting, and reporting processes will be automatically performed by the functionalities solidified in the device (which are not programmable to the user). Our proposed hybrid pricing scheme only requires an additional functionality for this 4G device, which allows the host to specify the amount of data forwarded at each time slot. This can be easily achieved by today's hardware capability, or even by software.

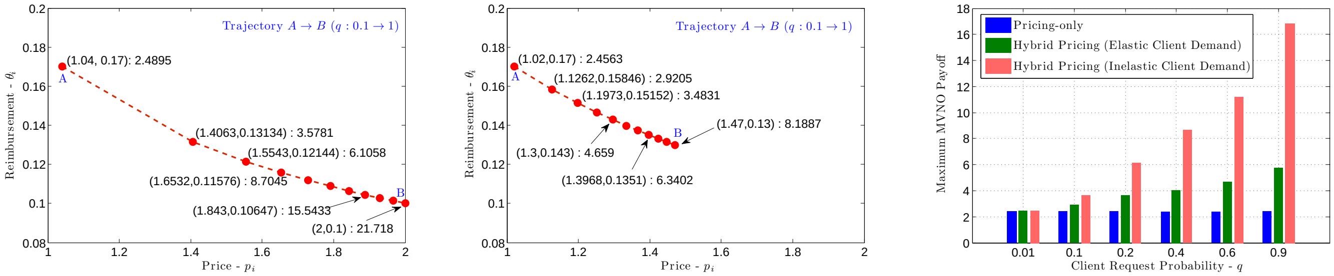
## V. Simulation

We perform numerical studies in a network of $I = 50$ hosts, each serving $|\mathcal{N}_i| = 8$ clients. Each host's 4G capacity is drawn from $[0, 10]$Mbps randomly and uniformly. Each client randomly requests capacity from the corresponding host, with a demand drawn from $[0, 1]$Mbps randomly and uniformly. The host utility is $U_i = 10 \cdot \log(1 + 0.2X_i)$. The length of each period is 1 hour, consisting of $T = 60$ one-minute slots.

**MVNO's Pricing-Reimbursing.** We first look at the MVNO's optimal pricing and reimbursing decisions. Figure 5 (A) and (B) show the trajectories of the MVNO's best pricing-reimbursing strategy under different request probability of clients (denoted by $q$), with inelastic and elastic client demands. Each tuple along the trajectory denotes the MVNO's best pricing and reimbursing decisions, and its achieved profit.[9] We can see that when the client request probability $q$ (or equivalently, the client demand) increases, the MVNO will increase the price, and meanwhile decrease the reimbursement. This is because a larger client demand gives the MVNO more room to elicit the profit. Moreover, the trajectory is steeper under the inelastic client demand. This is because the inelastic client demand does not depend on the price, and thus the MVNO is able to charge a higher price with less reimbursement to maximize its profit.

**MVNO's Payoff.** We now look at the MVNO's maximum payoff under the proposed hybrid pricing-reimbursing scheme.

[9] For example, $(1.04, 0.17) : 2.4895$ denotes that the MVNO's best pricing decision is 1.04 and best reimbursing decision is 0.17, under which it will achieve an expected profit 2.4895.




(A) Trajectory of Optimal Pricing-Reimbursing (Inelastic Client Demand)

(B) Trajectory of Optimal Pricing-Reimbursing (Elastic Client Demand)

(C) MVNO's Maximum Payoff (Inelastic/Elastic Client Demand)

Fig. 5. MVNO's Optimal Pricing-Reimbursing Strategy and Maximum Payoff.

Figure 5 (C) shows the MVNO's maximum payoffs achieved by different pricing schemes under different client request probability $q$. We can see that using the proposed hybrid pricing scheme, the MVNO's maximum payoff increase with $q$ under both elastic client demand (the 2nd bar in each group) and inelastic client demand (the 3rd bar in each group); while using the traditional pricing-only scheme (the 1st bar in each group), it does not depend on $q$ since the MVNO never serves the clients. We can further see that compared to the pricing-only scheme, our proposed hybrid pricing scheme can increase the MVNO's payoff by $20\%$ to $135\%$ under the elastic client demand, and by $50\%$ to $550\%$ under the inelastic client demand. Notice that when $q$ is very small (e.g., $0.01$), the MVNO achieves approximately the same payoff under different pricing schemes. This is because the client demand is so small that can be ignored.

## VI. Related Work

UPC models and services have become increasingly popular in recent years. In 2005, FON (www.fon.com) introduced commercially WiFi sharing, which is a prominent example of UPC services. In these models, users offer low-cost Internet accesses to each other (usually home users serve roaming users) based on reciprocation schemes [6]. In [8], the authors studied optimal pricing policies that enable adoption of UPC services in monopoly markets (e.g., how much FON should charge the service), while [15] analyzed the competition between FON-like operators with conventional cellular operators. On the other hand, [9] and [10] studied pricing schemes for decentralized UPC models, where users negotiate and charge each other directly, or can even form groups (clubs) [16].

This paper considers a network-controlled UPC model that differs from the above scheme both in technical and economic terms. First, the MVNO plays a very crucial role (through the reimbursement) on the decisions of the hosts. Second, unlike other related models (e.g., FON), the pricing is usage-based and hence the hosts' decisions directly affect the revenue of the MVNO. Finally, our hotspots are provided by mobile users with energy limitations. This constraint affects their servicing strategy significantly.

## VII. Conclusions

We believe that the Karma model is the first instance of a next generation FON-like communities, where Internet access is offered through autonomous mobile hotspots that are reimbursed for the services they offer. Clearly, this is a revolutionizing business model that *outsources* part of the network operation to the users. In this work, we provided a first approach for its modeling, and we characterized the optimal pricing-reimbursing strategies for different system parameters and users' demands. Our theoretical and numerical analysis indicate that the hybrid pricing scheme can significantly increase the MVNO's payoff. Interestingly, we show that the optimal price increases with the client demand, and the optimal reimbursement decreases with the client demand. There are many fascinating directions for future work. Our next step is to design a marketplace, where Internet access, in terms of mobile data usage, will be freely traded among MVNOs, mobile hotspots and users, under complete or incomplete network information. Another interesting direction is to study the hybrid pricing problem in a multi-client multi-host UPC model, where each client may have different choices of hosts to connect, and each host can choose different sets of clients to serve as well.

## References


[1] "List of MVNO Worldwide", *Internet Open Resource*, online available at http://www.mobileisgood.com/mvno.php.

[2] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2012–2017," White Paper, February 2013.

[3] GIGaom, "Why are MVNOs so hot right now? Thank the carriers", *www.gigaom.com*, 25 June, 2012.

[4] C. Camaran, and D. D. Miguel, "Mobile Virtual Network Operator (MVNO) Basics: What is behind this mobile business trend", *Valoris Telecom. Practice*, Tech. Rep., 2008.

[5] https://yourkarma.com/

[6] R. Sofia, and P. Mendes, "User-Provided Networks: Consumer as Provider", *IEEE Comm. Mag.*, vol. 46, no. 12, 2008.

[7] 3GPP Standards Specifications http://www.3gpp.org/specifications.

[8] M. H. Afrasiabi, and R. Guerin "Pricing Strategies for User-Provided Connectivity Services", *Proc. IEEE Infocom,* 2012.

[9] J. Musacchio, and J. C. Walrand, "WiFi Access Point Pricing as a Dynamic Game", *IEEE/ACM Trans. Netw.*, vol. 14, no. 2, 2006.

[10] R. K. Lam, D. M. Chiu, and J. C. S. Lui, "On the Access Pricing and Network Scaling Issues of Wireless Mesh Networks", *IEEE Transactions on Computers*, vol. 56, no. 1, 2007.

[11] R. Gibbons, "Game Theory for Applied Economists", *Princeton University Press,* 1992.

[12] L. Gao, G. Iosifidis, J. Huang, and L. Tassiulas, "Economics in Mobile Data Offloading", *IEEE Infocom Workshop on SDP*, 2013.

[13] G. Iosifidis, L. Gao, J. Huang, and L. Tassiulas, "An Iterative Double Auction for Mobile Data Offloading", *Proc. IEEE WiOpt*, 2013.

[14] C. Courcoubetis, and R. Weber, "Pricing Communication Networks: Economics, Technology and Modelling", *Willey Press*, 2003.

[15] M. Manshaei, J. Freudiger, M. Felegyhazi, P. Marbach, and J. P. Hubaux, "On Wireless Social Community Networks", *Proc. IEEE Infocom*, 2008.

[16] E. C. Efstathiou, P. A. Frangoudis, and G. C. Polyzos, "Controlled Wi-Fi Sharing in Cities: A Decentralized Approach Relying on Indirect Reciprocity", *IEEE Trans. on Mobile Comp.*, vol. 9, no. 8, 2010.

[17] L. Gao et al., "Hybrid Data Pricing for Network-Assisted UPC", http://jianwei.ie.cuhk.edu.hk/publication/AppendixKarma.pdf, *Online Appendix*, July 2013.