Optimal Auction Design for WiFi Procurement *

Information Systems Research, forthcoming

Liangfei Qiu†, Huaxia Rui‡, and Andrew Whinston§

Abstract

The unprecedented growth of cellular traffic driven by the use of smartphone for web surfing, video streaming, and cloud-based services poses bandwidth challenges for cellular service providers. To manage the increasing data traffic, cellular service providers are experimenting the use of third-party WiFi hotspots to augment its cellular capacity. We develop an analytical framework to study the optimal procurement auction for WiFi capacity. Such an auction design is complicated by the fact that WiFi networks have much more limited spatial coverages compared with the cellular network. Neither a global auction that includes all WiFi hotspots nor multiple local auctions that include only hotspots in each local WiFi region is optimal. We find that the optimal mechanism is an integration of one global auction which includes hotspots from an *endogeneously* determined set of WiFi regions and many separate local auctions which are only held in the rest of the WiFi regions. To implement the optimal mechanism, we also provide an efficient algorithm whose computation complexity is of the order of the number of WiFi regions. Our work contributes to the literature by designing the optimal mechanism for a unique type of IT procurement auction problem which is a tight integration of economics and computational technology.

Keywords: WiFi offloading, procurement auction, mechanism design

*We thank Xianjun Geng, Dale Stahl, Maxwell Stinchcombe, Thomas Wiseman for useful comments. We also thank seminar participants at University of Florida, George Washington University, Peking University, the 2013 International Conference on Information Systems, the 48th Hawaii International Conference on System Sciences (HICSS), the 26th Annual POMS Conference, and the 2017 China Summer Workshop on Information Management (CSWIM) for useful discussions. We are grateful to Chee Ching, Yu Jin, Wen-Ling Hsu from AT&T for valuable inputs from the industry perspective. Existing errors are our sole responsibility.

†University of Florida, Warrington College of Business, e-mail: liangfeiqiu@ufl.edu

‡The University of Rochester, Simon Business School, e-mail: huaxia.rui@simon.rochester.edu

§The University of Texas at Austin, McCombs School of Business, e-mail: abw@uts.cc.utexas.edu

1 Introduction

The increasing popularity of smartphones has triggered an explosive growth of mobile data traffic driven by web surfing, video streaming, online gaming, and many other digital goods industries (Tan et al. 2016; Tan and Carrillo 2017). According to Cisco VNI Global Mobile Data Forecast Update (2016-2021), global mobile data traffic grew 63 percent in 2016 and reached 7.2 exabytes per month at the end of 2016, up from 4.4 exabytes per month at the end of 2015 [1]. Moreover, mobile data traffic is expected to grow at a compound annual growth rate of 47 percent from 2016 to 2021, reaching 49.0 exabytes per month by 2021. The huge amount of mobile data traffic poses a challenge to the network infrastructure: Cellular networks are overloaded and congested during peak hours because of insufficient capacity, which leads to poor user experience and churn.

Researchers have proposed several solutions from both technical and economic aspects: (1) increasing the number of cellular towers or deploying the cell-splitting technology; (2) upgrading the network to fourth-generation (4G) networks such as Long Term Evaluation (LTE), High Speed Packet Access (HSPA) and WiMax; (3) expanding capacity by acquiring the spectrum of other networks, such as the attempted purchase of T-Mobile USA by AT&T; (4) adopting smart data pricing mechanisms (e.g. usage- and app-based pricing plans) to constrain the heaviest mobile data users, instead of using flat-rate pricing plans with unlimited data (Sen et al. 2012); and (5) offloading data traffic to WiFi networks (Bulut and Szymanski 2012).

Although all these solutions help alleviate the problem, each has its disadvantages. The first two solutions require large investments, and getting government approval for building new cell towers can take years. From the economic perspective, it is extremely expensive to increase the number of cellular base stations.[2] As a result, all cellular networks augment

[1] http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf.

[2] According to Balachandran et al. (2008), although cell-splitting provides capacity benefits, it could be quite expensive and economically infeasible since in addition to the base station hardware/deployment cost, each of the new bases needs to be provided with backhaul connectivity either via wireline access or microwave

the first two solutions with other approaches to expand capacity. The third solution suffers from regulatory constraints. Cramton et al. (2007) showed that an important market failure arises in spectrum auctions with dominant incumbents. They suggest that the Federal Communications Commission (FCC) should place limits on how much spectrum AT&T and Verizon are allowed to buy.[3] Although the average net benefits realized under congestion-based pricing tend to be higher than the average net benefits realized under flat-rate pricing (Gupta et al. 2011), usage based plans may also backfire by alienating the smartphone users who are likely the customer segment with the highest revenue growth potential.

Because of these technical, economic and regulatory constraints, the fifth solution—using WiFi hotspots for mobile data traffic offloading—seems to be one of the most promising approaches in augmenting solutions (1) and (2). A straightforward approach is for the cellular service providers to build and manage their own hotspots. In fact, we have seen some pilot projects for self-managed hotspots (Aijaz et al. 2013). Even though the option of service providers directly managing hotspots is often available, it is still expensive (Iosifidis et al. 2015) and may not be cost-effective. For example, Paul et al. (2011) found that 28% of subscribers generate traffic only in a single hour during peak hours in a day. Clearly, building and managing hotspots just for that peak hour is not efficient. Offloading traffic to third party hotspots overcomes the obstacle of managing a hotspot and ensures the high availability of WiFi resources. This strategy could potentially be a win-win solution: The cellular service provider saves the cost of building more cellular base stations or hotspots just for the peak traffic demands. The WiFi hotspots profit from sharing their otherwise wasted spare capacity. Indeed, such practice of sharing unused capacity is gaining traction in the industry (e.g., Airbnb, Uber) thanks to the advancement in technology, and the study of such sharing economy is also on the rise (Weber 2014).

We follow this paradigm of sharing economy and focus on offloading mobile traffic to

links.

[3]This concern is also reflected in the action taken by the FCC to block the recent merger between AT&T and T-Mobile USA.

third-party WiFi hotspots owned by entities such as local restaurants, bookstores, and hotels. Cellular service providers have shown great interest in such an approach. In 2012, for example, KDDI Corporation, a principal telecommunication provider in Japan, had already cooperated with about 100,000 commercial WiFi hotspots (Aijaz et al. 2013). However, offloading data traffic to third-party WiFi hotspots is not purely a technology augmenting the existing cellular network. Considering the economic incentives of third-party WiFi hotspots, WiFi offloading is also a practical mechanism design problem. Therefore, effectively leveraging third-party WiFi capacity requires the combination of both information technology and economic theory, which is in the spirit of designing smart markets (Bichler et al. 2010). Because WiFi capacity is a type of product with quite standardized characteristics, competitive bidding should be a better way to select the lowest cost bidder than negotiations.[4]

In the present study, we aim to model and solve the optimal procurement auction of third-party WiFi capacity. Because WiFi networks usually have a more limited range than cellular resources, the range of a cellular tower should be partitioned into several WiFi regions. The cellular capacity can serve data traffic in any WiFi region, whereas WiFi networks can serve only local traffic. However, the procurement auction design is not equivalent to running one local auction in each WiFi region because of the presence of the cellular resource. Buying more resources from one local WiFi hotspot frees up more cellular capacity to serve demand in other WiFi regions, thereby creating an inter-region competition. On the other hand, the procurement auction design is not equivalent to one global auction either where hotspots in all WiFi regions participate in one auction. This is because implementing a global auction may not always be feasible: WiFi capacity in one region cannot be transferred to other regions. The inflexibility of WiFi capacity causes difficulty for implementing a completely global auction when regions with heavy mobile traffic have insufficient WiFi capacity and those with light mobile traffic have excessive WiFi capacity.

[4]See for example, Bajari et al. (2009), who considered several determinants that may influence the choice of auctions versus negotiations. For complex projects, auctions may stifle communication between the buyer and the contractor. Clearly, WiFi capacity satisfies the standard assumption of well-defined products in the auction literature.

We find that the optimal mechanism is equivalent to an integration of one global auction which includes hotspots from an *endogeneously* determined set of WiFi regions and one local auction on each of the rest of the WiFi regions. This integration of global and local auctions is both theoretically interesting and practically important. It is the consequence of two unique features of procuring WiFi capacity for mobile traffic offloading: 1) the coupling of local auction because of the existence of the more flexible cellular capacity; 2) the heterogeneity in terms of both the demand for mobile bandwidth and the supply of WiFi capacity in different WiFi regions. To implement the optimal mechanism, we also provide an efficient algorithm whose computation complexity is of the order of the number of WiFi regions.

The insights from the present paper apply more generally to a class of procurement auction problems. The key issue in the procurement of WiFi capacity is to design the optimal auction mechanism in the presence of product flexibility and information asymmetry between suppliers (i.e., WiFi hotspots) and the downstream firm (i.e., the cellular service provider). This procurement problem in the wireless industry is an example of a general setting where (1) the downstream firm owns product-flexible in-house capacity that can be used for multiple products; (2) the product-flexible capacity is limited, and the firm needs to procure products from multiple upstream suppliers; and (3) each supplier is specialized and can produce only one product. Given the limitation of product-flexible capacity (in-house capacity) and the information asymmetry between the downstream firm and the suppliers, the downstream firm needs to solve the complex problem of designing an optimal procurement auction. This procurement scenario is common when companies are investing in product-flexible capacity that entails the ability to produce multiple products with the same capacity, and the ability to reallocate capacity between products (Goyal and Netessine 2011). Many manufacturing and service companies use flexible capacity to hedge against uncertainty in future demand (Fine and Freund 1990; Van Mieghem 1998).[5]

It should be noted that we have abstracted away the capacity problem of the broadband.

[5] In the automotive industry, the plants for most of the automobile companies are much more flexible than before: Ford's Rouge Plant can manufacture nine different products (Goyal and Netessine 2011).

Our key argument is that broadband technology is fundamentally different from cellular technology in terms of capacity constraints. First, broadband technology has advanced faster than cellular technology in past decades, and currently increasing broadband capacity is much cheaper than increasing cellular capacity. Unlike fiber optics in the case of broadband capacity, cellular capacity is inherently and technically constrained by radio spectrum capacity. Second, even if significant advances in the cellular technologies may take place in a near future, the cellular service providers still need to worry much about the traffic because of various regulatory constraints. We elaborate these differences between broadband technology and cellular technology in Online Appendix C.

2 Literature Review

The technology aspect and implementation of this study are clearly related to the vast literature in computer science on mobile data offloading (Balasubramanian et al. 2010; Iosifidis et al. 2015; Dong et al. 2014). We refer interested readers to Aijaz et al. (2013) for an overview of the technical and business perspectives of mobile data offloading and to Kang et al. (2014) for a discussion on mobile data offloading through third-party WiFi hotspots. The theoretical aspect of the present study is mostly related to two streams of literature, optimal auction design and supply chain management, which we will review in more detail.

In many procurement situations, the buyer cares about other attributes in addition to price when evaluating the submitted bids. Dasgupta and Spulber (1990) extended the standard fixed quantity auction and studied a quantity auction that allows the quantity of goods purchased to be endogenously determined by the submitted bids. In a multi-attribute scoring auction, suppliers submit multidimensional bids, and the contract is awarded to the supplier who submitted the bid with the highest score according to a scoring rule. Che (1993) developed a scoring procurement auction in which suppliers bid on two dimensions of the good. This scoring auction allows only sole sourcing. However, offloading data traffic to multiple WiFi hotspots is naturally done in our procurement setting. Duenyas et al. (2013) showed that a simple version of the open-descending auction can implement the optimal procurement mechanism for a newsvendor problem. The model in the present study differs from such auctions because of the unique challenge in our application setting.

Much of the literature on supply chain management has focused on scenarios where adding product-flexible capacity is beneficial (Goyal and Netessine 2011). Janakiraman et al. (2014) considered a firm that produces multiple products each period, using a shared resource with limited capacity, in a periodically reviewed stochastic inventory model. A natural question is, with limitations on product-flexible capacity, how should a downstream firm design its procurement auction mechanism. In the present study, we introduce an auction design problem with asymmetric information in the presence of product-flexible capacity. The downstream firm procures capacity from the suppliers to optimally combine with its in-house capacity to produce different products. In this process, the downstream firm makes the following decisions: How to allocate its product-flexible capacity to produce different products? How much quantity should be procured from each supplier? What is the corresponding payment scheme for each supplier? Our theoretical analysis provides insights to these questions in the context of the telecommunication industry and complement the existing literature on product line designs when the product-flexible capacity is limited. Netessine et al. (2002) analytically characterized the critical effects of increasing demand correlation between products on the flexible capacity decisions. We also find that the demand correlation as well as the level of in-house product-flexible capacity plays a crucial role in the optimal design of the procurement mechanism. When the demand is highly positively correlated or when the in-house product-flexible capacity is sufficiently large, the optimal procurement mechanism simplifies to a global auction including all upstream suppliers; but in general, it is equivalent to an integration of one global auction and multiple local auctions.

3 Model

3.1 Model Setup

A cellular network provides service to its customers who demand bandwidth. We think of the packets requested by the consumers as being serviced in a queuing system. The expected waiting time a typical customer experiences can be written as $W(\mu)$ where μ is the service capacity. Clearly, $W(\mu)$ should be decreasing in μ and be bounded below. We further make the technical assumption that $W(\mu)$ is twice differentiable and is strictly convex in μ:

$$\frac{dW}{d\mu} < 0, \ \frac{d^2W}{d\mu^2} > 0.$$

This is a mild technical assumption that is satisfied, for example, by an $M/M/c$ queue.[6] In the special case of an $M/M/1$ queue, as is assumed in Cheng et al. (2011), $W(\mu)$ is simply $W(\mu) = \frac{1}{\mu - \lambda}$ where λ is the customer demand rate.

Given the expected waiting time, the cellular service provider incurs a cost of $\chi(W)$, which is a strictly increasing function of W. To capture the rapidly rising cost of service degradation due to customers' increased expected waiting times (e.g., dissatisfied customers, or churn), we assume that the function $\chi(\cdot)$ is convex. As a special case, Cachon and Feldman (2011) assumed the waiting cost is a linear function of W. For convenience, we refer to the composition of $\chi(\cdot)$ and $W(\cdot)$ as $\omega(\mu) \equiv \chi(W(\mu))$. It is straightforward to show that $\omega(\mu)$ is strictly decreasing and strictly convex in μ.[7]

Given the unprecedented growth rate of mobile data demand and the high cost associated with congestion, the cellular network is interested in procuring spare resources from third-party WiFi hotspots. Although both can be used to meet the user demand, cellular resources and WiFi resources have different spatial coverages. In suburban areas, a typical cellular base station covers 1-2 miles (2-3 km) and in dense urban areas, it may cover one-fourth

[6] For the proof, please refer to Lemma 1 in Online Appendix A.

[7] $\omega'(\mu) = \chi'(W) \cdot W'(\mu) < 0$, $\omega''(\mu) = \chi''(W)\big(W'(\mu)\big)^2 + \chi'(W)W''(\mu) > 0$.

to one-half mile (400-800 m). A typical WiFi network has a range of 120 feet (32 m) indoors and 300 feet (95 m) outdoors.[8] To model this unique feature of bandwidth supply, we partition a cell sector into several WiFi regions so that WiFi hotspots within the same region are relatively close together. In particular, we assume the cell sector can be divided into M WiFi regions.[9] Cellular resources can serve traffic in any region m, whereas WiFi hotspots in region m can only serve local traffic. A unique challenge in the procurement auction is that the longer range cellular resource introduces coupling between the shorter range WiFi hotspots. The procurement problem in one WiFi region is not independent of the procurement problem in another region, because purchasing more WiFi capacity from a local WiFi hotspot in one region frees up more cellular capacity that can be used to serve the demand in another region.

Serving mobile demand for the cellular network provider incurs cost to a hotspot which differs among hotspots. We assume the cost function for hotspot i to provide capacity Q to the cellular network is

$$C(Q, \theta_i) \equiv \int_0^Q c(q, \theta_i) dq, i = 1, 2, \cdots, N.$$

where $c(q, \theta_i) \geq 0$ is the marginal cost function for hotspot i, and θ_i reflects each hotspot's private information about the cost of bandwidth provision which differs among different hotspots. In reality, the private information θ_i can be interpreted as each hotspot's sensitivity to its WiFi congestion rate. For example, some hotspots, like coffee shops, might be more sensitive to their WiFi congestion because some customers go there primarily for their WiFi services, while other hotspots, like restaurants, might be less sensitive. We assume $c_q(q, \theta_i) \geq 0$ to capture the fact that the marginal cost of providing capacity for each hotspot increases

[8]See http://en.wikipedia.org/wiki/Wifi, and http://en.wikipedia.org/wiki/Cell_site.

[9]For example, one can generate regions by clustering the WiFi hotspots using k-means method. Note that for simplicity, we assume that cellular capacity can be reallocated seamlessly from one WiFi region to another. In practice, some cellular capacity can be redirected (e.g., core processing for the base station), and some capacity cannot be redirected (e.g., radio capacity for directional antennas – these cover only a certain direction and angular range).

as more capacity is provided to the cellular network. Following previous literature (Dasgupta and Spulber 1990), we also assume that the marginal cost is increasing and convex in the cost parameter, $c_\theta \geq 0$, $c_{\theta\theta} \geq 0$, and that $c_{q\theta} \geq 0$. Hotspots' cost parameters are independently and identically distributed with a continuously differentiable cumulative distribution function $F(\cdot)$ defined on $[\underline{\theta}, \bar{\theta}]$ which is common knowledge. We further assume $H(\theta) \equiv F(\theta)/F'(\theta)$ is an increasing function of θ which is satisfied by common distribution functions such as the uniform distribution.

To model potential revenue sharing between a hotspot and its Internet Service Provider (ISP), we assume that a hotspot gets a proportion ($\tilde{\alpha} \in (0,1)$) of the payment from the cellular service provider for providing WiFi capacity. The other proportion $(1-\tilde{\alpha})$ goes to the ISP. We define $\alpha = \tilde{\alpha}^{-1}$ for notational convenience. Finally, let θ^* be the threshold cost parameter chosen by the cellular service provider so that any hotspot with $\theta > \theta^*$ will not participate in the procurement auction.

To determine the optimal auction mechanism, we first need to determine the value of procuring WiFi capacity. Let y_m be the amount of WiFi capacity the cellular service provider purchased from hotspots in region m, and let $y = \sum_{m=1}^{M} y_m$ be the total WiFi capacity purchased in all regions. Because congestion costs in different regions naturally involve different customers at any given time, a cost function that is separable across regions captures such cost structure. This is also consistent with the tradition in economics of using additive utilitarian social welfare function which is also widely used in the information systems literature. In particular, we model the congestion cost of each region, ω_m, as a region-specific function of $y_m + \mu_m$, for $m = 1, \cdots, M$. For example, different regions might have a different customer demand rate and the cellular service provider might also place different weights in different regions. The total congestion cost is then $\sum_{m=1}^{M} \omega_m(\mu_m + y_m)$ where μ_m is the amount of cellular capacity allocated to region m, and the congestion cost minimization problem can be written as

$$J(y_1 \cdots, y_M) \quad = \quad \text{Min}_{\mu_1, \cdots, \mu_M} \sum_{m=1}^{M} \omega_m\big(\mu_m + y_m\big) \tag{1}$$

$$s.t. \qquad \sum_{m=1}^{M} \mu_m \leq \mu, \mu_m \geq 0, \text{for } m = 1, 2, ...M.$$

The cellular service provider purchases WiFi capacity $(y_1, \cdots, y_M)$ for the M regions from hotspots in these regions to supplement its cellular capacity.[10] The objective is to minimize the total cost, including that for congestion $J(y_1, \cdots, y_M)$ and for procurement, which includes both the actual costs of hotspots providing WiFi resource and the information rent due to information asymmetry. The cellular service provider follows a two-step decision procedure: In the first stage, it purchases WiFi capacity from hotspots in different regions. In the second stage, the cellular service provider allocates cellular resources across regions.

3.2 Global Auction

In this section, we assume the non-negativity constraints on μ_m are not binding and call the resulting auction mechanism *global auction*. We will see later that this is an important building block of the actual optimal mechanism.

Proposition 1 *Under global auction, the optimal cellular resource allocation is given by*

$$\mu_m^* = \phi_m\Big(\Psi(y + \mu)\Big) - y_m$$

[10] It is worth noting that our problem can be regarded as a sub-problem of a capacity expansion project for a cellular service provider. Given a cellular sector, the cellular service provider can compare the option of procuring WiFi capacity or building a new cell tower, or not to seek any capacity expansion. By evaluating and comparing the overall benefits (i.e., the difference between the reduced congestion cost and the cost of capacity expansion, whether through WiFi procurement or building cell tower), the service provider can select the best action. We discuss this more formally in Section 4. Intuitively, WiFi procurement will be preferred in situations where (i) the existing tower capacity is not terribly insufficient given the customer demand; or (ii) there are plenty of inexpensive WiFi hotspots in the regions. On the other hand, in regions where the demand or the growth of demand is enormous and the existing cellular capacity is way insufficient, building a new cell tower might be more cost-effective than procuring WiFi capacity from hotspots.

where $\phi_m(\cdot)$ *is the inverse of* $\omega'_m(\cdot)$ *and* $\Psi(\cdot)$ *is the inverse of* $\Phi(\cdot) \equiv \sum_{m=1}^{M} \phi_m\big(\,\cdot\,\big)$. *The optimal congestion cost is*

$$J(y) \equiv \sum_{m=1}^{M} \omega_m\left(\phi_m\Big(\Psi(y+\mu)\Big)\right).$$

Moreover, $J(y)$ *is decreasing and convex.*

Proof. The first-order condition implies that there exists a Lagrange multiplier for the cellular capacity constraint $\nu > 0$ such that

$$\omega'_m(y_m + \mu_m) + \nu = 0, \ \forall m = 1, 2, \cdots, M.$$

Hence,

$$\mu^*_m = \phi_m\big(-\nu\big) - y_m,$$

where ϕ_m is guaranteed to exist due to the strict convexity of $\omega_m(\cdot)$. The cellular capacity constraint is binding at the optimal solution, which implies

$$\mu = \sum_{m=1}^{M} \mu^*_m = \sum_{m=1}^{M} \Big(\phi_m\big(-\nu\big) - y_m\Big) = \sum_{m=1}^{M} \phi_m\big(-\nu\big) - y = \Phi(-\nu) - y.$$

Because $\omega_m(\cdot)$ is strictly convex and $\phi_m(\omega'_m(x)) = x$, we see that ϕ_m is monotone increasing:

$$\phi'_m(\omega'_m(x)) = \frac{1}{\omega''_m(x)} > 0.$$

Hence, $\Phi(x)$ is also monotone increasing:

$$\Phi'(x) = \sum_{m=1}^{M} \phi'_m\big(x\big) > 0,$$

which guarantees the existence of the inverse of $\Phi(\cdot)$, which is denoted by $\Psi(\cdot)$.

Therefore, $\nu = -\Psi(y+\mu)$ and the claim follows. Because $J(y_1, \cdots, y_M)$ is a function of

$y_1, \cdots, y_M$ only through their sum, y, with slight abuse of notation, we write it simply as $J(y)$ when none of the non-negativity constraint is binding.

Now, we show that $J(y)$ is decreasing and convex in y. Denote $z = y + \mu$ so that $\Phi(-\nu) = z$ and $\Psi(z) = -\nu$. Because $\Psi(\Phi(x)) = x$, we have

$$\Psi'(z) = \frac{1}{\Phi'(-\nu)} = \Big(\sum_{m=1}^{M} \phi'_m\big(-\nu\big)\Big)^{-1}.$$

The first and second derivative of $J(y)$ are

$$\begin{aligned} J'(y) &= \sum_{m=1}^{M} \omega'_m\bigg(\phi_m\Big(\Psi(z)\Big)\bigg)\phi'_m\Big(\Psi(z)\Big)\Psi'(z) \\ &= \Psi(z)\frac{1}{\sum_{m=1}^{M}\phi'_m\big(-\nu\big)}\sum_{m=1}^{M}\phi'_m\Big(-\nu\Big) \\ &= \Psi(z) = -\nu < 0, \\ J''(y) &= \Psi'(z) = \Big(\sum_{m=1}^{M}\phi'_m\big(-\nu\big)\Big)^{-1} > 0. \end{aligned}$$

Therefore, $J(y)$ is decreasing and convex in y. ■

For this optimal allocation to be feasible, we need $\mu^*_m \geq 0$, or equivalently,

$$\mu \geq \Phi\Big(\omega'_m(y_m)\Big) - y, \ \forall m = 1, \cdots, M. \tag{2}$$

As we stated at the beginning of this section, we assume the non-negativity constraints of μ_m are non-binding, hence the inequality (2) is always satisfied.

Intuitively, when condition (2) is satisfied, WiFi resource in one region is a perfect substitute of WiFi resource in another region, from the perspective of the cellular service provider. Given our solution for the second-stage problem, the first-stage problem is a direct revelation game in which hotspots truthfully report their types in the Bayesian Nash equilibrium. We adopt the notational convention of writing $\theta_{-i} = (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_N)$. The procurement auction can be implemented via a direct revelation mechanism where

- The cellular service provider announces a payment-bandwidth schedule $P_i = P(\theta_i, \theta_{-i})$, and a bandwidth allocation schedule $q_i = Q(\theta_i, \theta_{-i})$;
- Hotspot i truthfully reports the private cost parameter θ_i, given $P(\theta_i, \theta_{-i})$ and $Q(\theta_i, \theta_{-i})$;
- Hotspot i provides bandwidth $q_i = Q(\theta_i, \theta_{-i})$ to the cellular service provider and its revenue is $\tilde{\alpha} P(\theta_i, \theta_{-i})$.

By selling its spare bandwidth, a hotspot is essentially delivering a certain quantity of requested data for the cellular service provider. Hence, we simply refer to $Q(\theta_i, \theta_{-i})$ as the quantity schedule from now on. To implement the scheme, the cellular service provider can periodically run the procurement auction to obtain WiFi capacity from each region and determines the optimal allocation of cellular capacity across regions. Alternatively, the service provider can re-run the auction whenever the demand rates or the rate distribution across WiFi regions change significantly. As long as the realized demand rates are consistent with the demand rates used to compute the optimal auction rule, the operation is optimal. From individual customers perspective, he or she may not even know how his/her requested data is delivered (cellular tower or WiFi hotspot). It is not technologically difficult to automatically select the "best" (from cellular service provider's financial perspective) source of capacity for the customer data request, which may not necessarily be the geographically closest hotspot. Current technology already can seamlessly switch between cellular towers and WiFi hotspots although the criteria of switching is often purely technological.

In the case of global auction, the expected gain from procuring a total of y Wi-Fi capacity is $J(0) - J(y)$, which is increasing and concave in y.

To describe the optimal global auction mechanism, we first define $\underline{\nu}_i$ as

$$\underline{\nu}_i \equiv \alpha c(0, \theta_i) + \alpha c_\theta(0, \theta_i) H(\theta_i), \forall i = 1, \cdots, N$$

and impose the following technical condition

$$-\Psi(\mu) > \min\{\underline{\nu}_1, \underline{\nu}_2, \cdots, \underline{\nu}_N\}.$$

Note that the function $-\Psi(q)$ can be interpreted as the marginal value of WiFi capacity when the total acquired WiFi capacity is q. Hence, this condition ensures that the marginal benefit of procuring an infinitesimal amount of WiFi capacity is larger than the hotspot with the least (virtual) marginal cost of providing WiFi capacity. Without this condition, WiFi procurement auction is never optimal and should not be considered by the cellular service provider at all.

The following proposition characterizes the optimal global auction mechanism for the cellular service provider and is an application of Dasgupta and Spulber (1990).

Proposition 2 (Dasgupta and Spulber) *In the optimal direct revelation mechanism, all hotspots truthfully announce their cost parameters θ. The optimal procurement quantity schedule $q_i^* = Q^*(\theta_i, \theta_{-i})$, for $i = 1, 2, ...N$ is determined by*

$$-\Psi\left(\mu + \sum_{j=1}^{N} q_j^*\right) = \alpha c(q_i^*, \theta_i) + \alpha c_\theta(q_i^*, \theta_i) H(\theta_i).$$

The optimal payment schedule $P_i = P^*(\theta_i, \theta_{-i})$, for $i = 1, 2, ...N$ is given by:

$$P_i = \alpha\left(C(q_i^*, \theta_i) + \int_{\theta_i}^{\theta^*} C_\theta(Q^*(\theta, \theta_{-i}), \theta) d\theta\right).$$

The cellular service provider's expected gain from the procurement auction is

$$J(0) - \mathbb{E}\Big[J\left(\sum_{i=1}^{N} q_i^*\right) + \alpha \sum_{i=1}^{N} C(q_i^*, \theta_i) + \alpha \sum_{i=1}^{N} C_\theta(q_i^*, \theta_i) H(\theta_i)\Big].$$

The intuition of the above proposition is that the "virtual" marginal costs must be equalized across all hotspots in all regions at the optimal, which determines the optimal quantity functions $Q_i(\vec{\theta})$, $i = 1, \cdots, N$. Because $c_\theta \geq 0$, $c_{\theta\theta} > 0$, and $H(\theta)$ is increasing in θ, it is easy to verify that (1) q_i^* is decreasing in θ_i; that (2) q_j^* is increasing in θ_i for $j \neq i$; and that (3) $\sum_{j=1}^{N} q_j^*$ is decreasing in θ_i.

In the direct revelation game, hotspot i reports its true cost parameter θ_i. The capacity it needs to provide is $q_i = Q^*(\theta_i, \theta_{-i})$, and its payment is $P_i = P^*(\theta_i, \theta_{-i})$. This optimal mechanism is a global auction including all hotspots from different regions. Note that launching separate local auctions within each region is not optimal because running multiple local auctions essentially reduces competition among hotspots which could be exploited by flexibly allocating cellular resources among regions. In equilibrium, the virtual marginal costs $c(q_i, \theta_i) + c_\theta(q_i, \theta_i)H(\theta_i)$ are equalized across hotspots in different regions, and the marginal benefit of procuring WiFi capacity is also equalized across regions. The global auction effectively creates perfect inter-region competition among hotspots which is particularly important when intra-region competition is limited in some regions (e.g., regions with few hotspots).

Based on Proposition 1 and Proposition 2, we design the following procedure to calculate the optimal global auction.

- Invite each of the n hotspots to report its cost parameter θ. Denote the submitted cost parameters as $\{\theta_1, \theta_2, \cdots, \theta_N\}$.
- Define the map $q : \Theta^N \to \mathbb{R}^N$ through the following steps:
  - Given $\nu \geq 0$ and for each $i = 1, 2, \cdots, N$, define $g_i(\nu)$ as

$$g_i(\nu) = \begin{cases} 0 & \text{if } \nu \leq \underline{\nu}_i = \alpha c(0, \theta_i) + \alpha c_\theta(0, \theta_i)H(\theta_i); \\ \infty & \text{if } \nu \geq \bar{\nu}_i \equiv \lim_{x \to \infty} \alpha c(x, \theta_i) + \alpha c_\theta(x, \theta_i)H(\theta_i); \\ x^* & \text{if } \underline{\nu}_i < \nu < \bar{\nu}_i. \end{cases}$$

where x^* is the solution to the equation

$$\alpha c(x,\theta_i) + \alpha c_\theta(x,\theta_i)H(\theta_i) = \nu,$$

in the interval $(0,\infty)$. Because the left-hand-side of the equation is increasing in x and $\nu \in (\underline{\nu}_i, \bar{\nu}_i)$, x^* uniquely exists. Given a value of $\nu \in (\underline{\nu}_i, \bar{\nu}_i)$, we can easily solve for x^* using bisection in an appropriately constructed interval $[0, \bar{q}_i]$. It's also easy to see that the non-negative function $g_i(\nu)$ is (weakly) increasing in ν.

– Let q^* be the unique solution to the following equation:

$$\sum_{i=1}^{N} g_i\left(-\Psi(\mu+q)\right) = q.$$

The uniqueness follows directly from the monotonicity of $\sum_{i=1}^{N} g_i\left(-\Psi(\mu+q)\right) - q$. To see the existence of q^*, note that by our technical assumption, we have $\sum_{i=1}^{N} g_i\left(-\Psi(\mu)\right) > 0$. Let $\bar{q} \equiv \sum_{i=1}^{N} g_i\left(-\Psi(\mu)\right) > 0$. Because $\sum_{i=1}^{N} g_i\left(-\Psi(\mu+q)\right)$ is (weakly) decreasing in q, we have

$$\sum_{i=1}^{N} g_i\left(-\Psi(\mu+\bar{q})\right) \le \sum_{i=1}^{N} g_i\left(-\Psi(\mu)\right) = \bar{q}.$$

By continuity, q^* exists and we can easily solve for q^* using bisection in the interval $(0, \bar{q})$.

– Let $\vec{q} \equiv (q_1, q_2, \cdots, q_N) \equiv (g_1(-\Psi(\mu+q^*)), g_2(-\Psi(\mu+q^*)), \cdots, g_N(-\Psi(\mu+q^*)))$. Clearly, $\sum_i q_i = q^*$. By the definition of $g_i(\nu)$, for any i such that $q_i \equiv g_i(-\Psi(\mu+q^*)) > 0$, we have

$$\alpha c(q_i,\theta_i) + \alpha c_\theta(q_i,\theta_i)H(\theta_i) = -\Psi(\mu+q^*).$$

In other words, the optimality condition of Proposition 2 is satisfied.

- Define payment plan P_i as

$$P_i \equiv P_i(\theta_1, \cdots, \theta_N) \equiv \alpha C(q_i, \theta_i) + \alpha \int_{\theta_i}^{\theta^*} C_\theta(q_i(\theta, \theta_{-i}), \theta) d\theta.$$

- Hotspot i will provide capacity q_i and receive payment P_i.

- The expected gain of the cellular service provider is

$$J(0) - \mathbb{E}\left[J(q^*) + \alpha \sum_{i=1}^{N} C(q_i, \theta_i) + \alpha \sum_{i=1}^{N} C_\theta(q_i, \theta_i) H(\theta_i)\right]$$

3.3 Integrating Global and Local Auctions, $M = 2$

So far, we have assumed that $\mu_m^* \geq 0$ for all m. We call this the feasibility condition and it can be written as

$$\mu \geq \Phi\Big(\omega_m'\big(\sum_{i \in E_m} Q^*(\theta_i, \theta_{-i})\big)\Big) - \sum_{i=1}^{N} Q^*(\theta_i, \theta_{-i}), \ \forall m = 1, \cdots, M.$$

where E_m is the set of participating hotspots in region m. Under the feasibility condition, the cellular capacity μ is sufficiently large for all regions under all realizations of hotspot supply (i.e., $\theta_1, \cdots, \theta_N$). Clearly, this is a very restrictive assumption when the cellular capacity is only moderate. In a realistic environment, the feasibility condition will most likely hold for some realizations of cost parameters (θ_i, θ_{-i}) but not for others.

In this section, we relax the assumption that $\mu_m \geq 0$ is non-binding when $M = 2$. Once we fully solve the optimal procurement auction design problem with $M = 2$, we will solve the most general case with $M > 2$ in Section 3.4.

To further illustrate the feasibility condition, we depict two illustrating examples in Figure 1 based on a simulation.[11] We assume that there are two WiFi regions ($M = 2$), and that each region has four hotspots ($N = 8$). The congestion cost functions for the cellular service

[11]For more details on the simulation, please refer to Online Appendix B.

provider, and WiFi hotspots are $\omega_m(\mu_m + y_m) = \frac{1}{\mu_m + y_m - \lambda_m}$ and $C(q, \theta_i) = \left(\frac{1}{2} + \theta_i\right) q^2$, where the private cost parameters for hotspots, $(\theta_1, \cdots, \theta_8)$, is independently drawn from a uniform distribution $U[0,1]$. In Figure 1, the X-axis corresponds to the demand rate in region 1, and the Y-axis corresponds to the demand rate in region 2.

- A blue $\times$ indicates that the feasibility condition is always satisfied;
- A black star indicates that the feasibility condition is always violated;
- A red dot indicates that the feasibility condition is violated for some realizations of cost parameters (θ_i, θ_{-i}) but not for others.

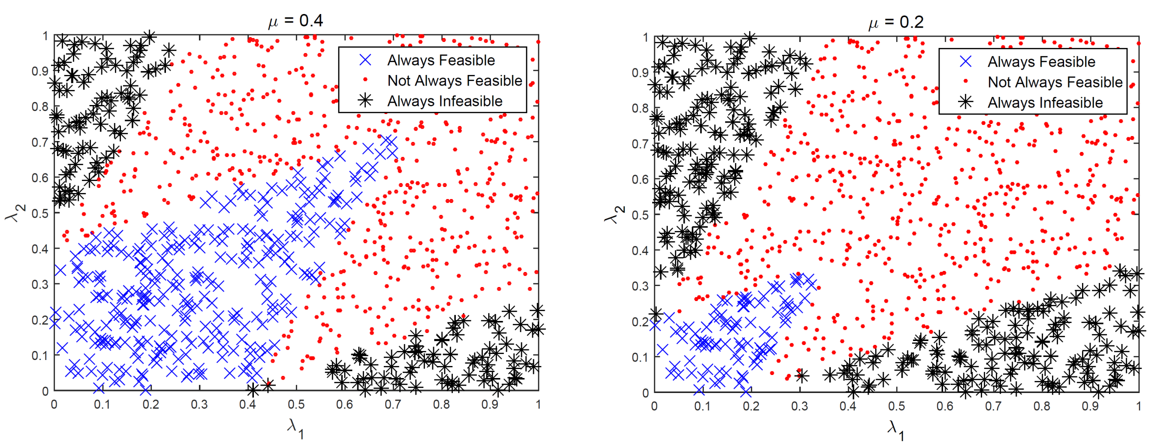

Figure 1: Illustration of the feasibility condition. The cellular capacity, μ, is set to be 0.4 in the left panel and 0.2 in the right panel. The feasibility condition is more likely to be violated when the demands are unbalanced or when μ is small.

When the feasibility condition is always satisfied (the blue $\times$), even though a hotspot in region 1 cannot directly serve customers in region 2, by serving customers in region 1, it effectively frees up some cellular capacity which can then be used to serve customers in region 2. Thus, a single global auction to obtain bandwidth from all hotspots should outperform multiple local auctions due to increased competition among hotspots. As a result, the optimal procurement mechanism should be the global auction we discussed so far. When the feasibility condition is always violated (the black stars), all cellular resources should be allocated to the region experiencing a surge in demand due to the insufficient amount of cellular capacity. In this case, running two local auctions is optimal.

We focus on the non-trivial scenario (i.e., the red dots) where the feasibility condition is violated for some realizations of cost parameters but not for others. In this more general scenario, the value of procuring $(y_1, \cdots, y_M)$ WiFi capacity from the M regions depends not only on the total procured WiFi capacity y, but also on its distribution across regions. Correspondingly, we denote the value of procuring $(y_1, \cdots, y_M)$ WiFi capacity as $J(0) - J(y_1, \cdots, y_M)$ where $J(y_1, \cdots, y_M)$ is the minimized congestion cost defined in (1). Given $(y_1, \cdots, y_M)$, let $\hat{\mu}_m$ be the optimal amount of cellular capacity allocated to region m ignoring all the non-negativity constraint on μ_i (i.e., in the global auction). From Proposition 1, we know $\hat{\mu}_m = \phi_m\Big(\Psi(y + \mu)\Big) - y_m$.

Intuitively, the optimal auction should be designed so that μ_m coincides with $\hat{\mu}_m$ whenever $\hat{\mu}_m$ is non-negative under some realization of $(\theta_1, \cdots, \theta_N)$. On the other hand, whenever $\hat{\mu}_m$ is negative under other realizations of $(\theta_1, \cdots, \theta_N)$, the optimal quantity schedule should be able to adjust the procurement to take into account the corner solution in the second-stage optimization problem of (1). In other words, the optimal quantity schedule is likely a non-smooth function of $(\theta_1, \cdots, \theta_N)$ with many segments. However, as long as the quantity schedule is non-increasing in hotspot type, we can find a truth-telling mechanism to implement it. Fortunately, as we will show in the proof of our next result, the proposed optimal quantity schedule is continuous everywhere, which essentially upgrades local monotonicity into global monotonicity.

The following proposition gives the optimal quantity schedule and payment function.

Proposition 3 *The optimal quantity $q_i^{**} = Q_i^{**}(\theta_i, \theta_{-i})$ is determined by*

$$-\Psi\left(\mu + \sum_{j=1}^{N} q_j^{**}\right) = \alpha c(q_i^{**}, \theta_i) + \alpha c_\theta(q_i^{**}, \theta_i) H(\theta_i), \forall i = 1, \cdots, N \qquad (3)$$

if the resulting $\hat{\mu}_1$ and $\hat{\mu}_2$ are both non-negative, and is determined by

$$-\omega'_m\left(\mu\mathbf{1}_{\hat{\mu}_m>0}+\sum_{j\in E_m}q_j^{**}\right)=\alpha c\left(q_i^{**},\theta_i\right)+\alpha c_\theta(q_i^{**},\theta_i)H(\theta_i),\forall i\in E_m, m=1,2 \qquad (4)$$

otherwise, where $\mathbf{1}_{\hat{\mu}_m>0}$ is the indicator function for $\hat{\mu}_m>0$.

The optimal payment schedule $P_i^{**}(\theta_i,\theta_{-i})$, for $i=1,2,...n$, is given by:

$$P_i^{**}(\theta_i,\theta_{-i})=\alpha\Big(C\left(q_i^{**},\theta_i\right)+\int_{\theta_i}^{\theta^*}C_\theta\left(Q_i^{**}(\theta,\theta_{-i}),\theta\right)d\theta\Big). \qquad (5)$$

Proof. The revelation principle implies that we only need to focus on direct mechanisms. Hence, we only need to find the quantity schedule $Q(\theta_i,\theta_{-i})$ that maximizes the service provider's gain from the procurement auction, taking into account the information rent required for truth telling.

Our proof has three steps. First, given a quantity schedule, we establish the corresponding payment rule that is necessary for incentive compatibility. Second, we optimally choose the quantity schedule. Third, we verify that the proposed quantity schedule is non-increasing which ensures that the payment rule is not only necessary but also sufficient[12] for incentive compatibility.

Step 1

The expected profit of hotspot provider i with cost parameter θ reporting parameter θ' is

$$\pi(\theta',\theta)=\mathbb{E}_{-i}\big[\tilde{\alpha}P(\theta',\theta_{-i})-C\big(Q(\theta',\theta_{-i}),\theta\big)\big].$$

[12]For proof, please see page 14 of Dasgupta and Spulber (1990).

Incentive compatibility implies that $\pi(\theta,\theta)-\pi(\theta,\theta') \geq \pi(\theta,\theta)-\pi(\theta',\theta') \geq \pi(\theta',\theta)-\pi(\theta',\theta')$, or equivalently,

$$\mathbb{E}_{-i}[C\big(Q(\theta,\theta_{-i}),\theta'\big)-C\big(Q(\theta,\theta_{-i}),\theta\big)] \geq \pi(\theta,\theta)-\pi(\theta',\theta') \geq \mathbb{E}_{-i}[C\big(Q(\theta',\theta_{-i}),\theta'\big)-C\big(Q(\theta',\theta_{-i}),\theta\big)].$$

Let $\pi(\theta) \equiv \pi(\theta,\theta)$. Dividing both sides by $\theta - \theta'$ and taking limits as $\theta' \to \theta$, we have

$$\frac{d\pi(\theta)}{d\theta} = -\mathbb{E}_{-i}\big[C_\theta\big(Q(\theta,\theta_{-i}),\theta\big)\big],$$

Integrating both sides from θ_i to θ^*, and using the fact that $\pi(\theta^*) = 0$, we have

$$\pi(\theta_i) = \int_{\theta_i}^{\theta^*} \mathbb{E}_{-i}\big[C_\theta\big(Q(\theta,\theta_{-i}),\theta\big)\big]d\theta = \mathbb{E}_{-i}\left[\int_{\theta_i}^{\theta^*} C_\theta\big(Q(\theta,\theta_{-i}),\theta\big)d\theta\right]$$

Hence, the expected payment a hotspot provider with cost parameter θ_i will receive must satisfy

$$\mathbb{E}_{-i}[\tilde{\alpha}P(\theta_i,\theta_{-i})] = \mathbb{E}_{-i}\left[C\big(Q(\theta_i,\theta_{-i}),\theta_i\big) + \int_{\theta_i}^{\theta^*} C_\theta\big(Q(\theta,\theta_{-i}),\theta\big)d\theta\right],$$

and the claim on $P_i^{**}(\theta_i,\theta_{-i})$ follows.

Step 2

We now prove the optimal quantity schedule. From the buyer's perspective, the expected payment to any hotspot provider is

$$
\begin{aligned}
\mathbb{E}_i\big[\mathbb{E}_{-i}[P(\theta_i,\theta_{-i})]\big] &= \alpha\mathbb{E}\big[C\big(Q(\theta_i,\theta_{-i}),\theta_i\big)\big] \;+\; \alpha\int_{\underline{\theta}}^{\theta^*}\left(\int_{\theta_i}^{\theta^*}\mathbb{E}_{-i}\big[C_\theta\big(Q(\theta,\theta_{-i}),\theta\big)\big]d\theta\right)dF(\theta_i)\\
&= \alpha\mathbb{E}\big[C\big(Q(\theta_i,\theta_{-i}),\theta_i\big)\big] \;+\; \alpha\left[F(\theta_i)\int_{\theta_i}^{\theta^*}\mathbb{E}_{-i}\big[C_\theta\big(Q(\theta,\theta_{-i}),\theta\big)\big]d\theta\right]\Bigg|_{\underline{\theta}}^{\theta^*}\\
&\qquad +\; \alpha\int_{\underline{\theta}}^{\theta^*}F(\theta_i)\mathbb{E}_{-i}\big[C_\theta\big(Q(\theta_i,\theta_{-i}),\theta_i\big)\big]d\theta_i\\
&= \alpha\mathbb{E}\big[C\big(Q(\theta_i,\theta_{-i}),\theta_i\big)\big] \;+\; \alpha\int_{\underline{\theta}}^{\theta^*}\mathbb{E}_{-i}\big[F(\theta_i)C_\theta\big(Q(\theta_i,\theta_{-i}),\theta_i\big)\big]d\theta_i\\
&= \alpha\mathbb{E}\big[C\big(Q(\theta_i,\theta_{-i}),\theta_i\big)\big] \;+\; \alpha\mathbb{E}_{-i}\left[\int_{\underline{\theta}}^{\theta^*}C_\theta\big(Q(\theta_i,\theta_{-i}),\theta_i\big)H(\theta)dF(\theta_i)\right]\\
&= \alpha\mathbb{E}\big[C\big(Q(\theta_i,\theta_{-i}),\theta_i\big) \;+\; C_\theta\big(Q(\theta_i,\theta_{-i}),\theta_i\big)H(\theta_i)\big].
\end{aligned}
$$

Let $q_i \equiv Q(\theta_i,\theta_{-i})$. The cellular service provider's total expected cost minimization problem can now be written as the following optimization problem:

$$
\begin{aligned}
\min_{\substack{q_i, i=1,\cdots,N\\ \mu_m, m=1,\cdots,M}} \Pi \;&=\; \mathbb{E}\left[\sum_{m=1}^{M}\omega_m\big(\mu_m+y_m\big)+\alpha\sum_{i=1}^{N}C(q_i,\theta_i)+\alpha\sum_{i=1}^{N}C_\theta(q_i,\theta_i)H(\theta_i)\right],\\
s.t. \quad & \sum_{m=1}^{M}\mu_m \le \mu,\\
& \mu_m \ge 0, \forall m=1,2,\cdots,M,\\
& y_m = \sum_{i\in E_m} q_i, \forall m=1,2,\cdots,M,
\end{aligned}
$$

where the expectation $\mathbb{E}[\cdot]$ is taken over $(\theta_1,\theta_2,\cdots,\theta_N)$ and the optimization is taken over $N+M$ functions of $(\theta_1,\theta_2,\cdots,\theta_N)$: $q_i(\vec{\theta})$ and $\mu_m(\vec{\theta})$, $\forall i=1,\cdots,N, m=1,\cdots,M$.

The degenerated structure of this variational calculus problem allows us to solve the problem through pointwise optimization over the space of Θ. Based on this observation, we further simplify the problem by dividing the space of Θ into two areas,

$$\Theta_1 \equiv \{(\theta_1, \theta_2, \cdots, \theta_N) \big| \hat{\mu}_1 \geq 0, \hat{\mu}_2 \geq 0\}, \Theta_2 \equiv \{(\theta_1, \theta_2, \cdots, \theta_N) \big| \hat{\mu}_1 < 0 \text{ or } \hat{\mu}_2 < 0\}.$$

If $\vec{\theta} \in \Theta_1$, the non-negativity conditions of μ_m are not binding. Hence, the optimization problem is equivalent to the one studied in Proposition 1. Therefore, the quantity schedule from Proposition 1 is optimal when $\vec{\theta} \in \Theta_1$.

If $\vec{\theta} \in \Theta_2$, then one of the non-negativity constraints must be binding at the optimal. Without loss of generality, assume $(\mu_1^*, \mu_2^*) = (\mu, 0)$ at the optimal. The pointwise optimization problem can be simplified as

$$\begin{aligned}
\min_{q_1, \cdots, q_N} \Pi &= \omega_1\big(\mu + y_1\big) + \omega_2\big(y_1\big) + \alpha \sum_{i=1}^{N} C(q_i, \theta_i) + \alpha \sum_{i=1}^{N} C_\theta(q_i, \theta_i) H(\theta_i), \\
&= \Big(\omega_1\big(\mu + y_1\big) + \alpha \sum_{i \in E_1} C(q_i, \theta_i) + \alpha \sum_{i \in E_1} C_\theta(q_i, \theta_i) H(\theta_i)\Big) \\
&\quad + \Big(\omega_1\big(y_2\big) + \alpha \sum_{i \in E_2} C(q_i, \theta_i) + \alpha \sum_{i \in E_2} C_\theta(q_i, \theta_i) H(\theta_i)\Big) \\
s.t. &\quad y_m = \sum_{i \in E_m} q_i, \forall m = 1, 2.
\end{aligned}$$

which is the same as that of designing two separate local auctions. Therefore, with $\theta \in \Theta_2$, the optimal mechanism is equivalent to holding two separate local auctions, with all cellular resources allocated to one of the regions.

Step 3

Finally, to ensure that the payment schedule is not only necessary but also sufficient for incentive compatibility, we need to verify that $\mathbb{E}_{-i}[Q_i^{**}(\theta_i, \theta_{-i})]$ is non-increasing in θ_i. It suffices to show that $q_i^{**} = Q_i^{**}(\theta_i, \theta_{-i})$ is decreasing in θ_i given any θ_{-i}. Notice that by our assumptions on hotspot cost structure and $H(\theta)$, q_i^{**} is decreasing within the region of θ_i where either the global auction is optimal or the local auction is optimal. The only possible violation of the monotonicity property is when the value of θ_i crosses some threshold below (above) which local (global) is optimal or vice versa. Clearly, if q_i^{**} is continuous at such thresholds, local monotonicity implies global monotonicity.[13]

To establish continuity, we first denote the region to which hotspot i belongs as region 1 and note that global auction is chosen if and only if $\hat{\mu}_1 = \phi_1\big(\Psi(y+\mu)\big) - y_1 > 0$ and $\hat{\mu}_1 < \mu$, or equivalently, $-\omega_1'(y_1) > -\Psi(y+\mu)$ and $-\omega_1'(y_1+\mu) < -\Psi(y+\mu)$. Define the threshold $\tilde{\theta}_i$ as the value of θ_i such that $-\omega_1'(y_1) = -\Psi(y+\mu) \equiv \tilde{m}$ and the threshold $\hat{\theta}_i$ as the value of θ_i such that $-\omega_1'(y_1+\mu) = -\Psi(y+\mu) \equiv \hat{m}$. Clearly, $\tilde{\theta}_i \neq \hat{\theta}_i$ due to the strict convexity of $\omega_1(\cdot)$. Denote the solution to the following equation by $\tilde{q}_i$,

$$\tilde{m} = \alpha c(q_i^{**}, \theta_i) + \alpha c_\theta(q_i^{**}, \theta_i) H(\theta_i),$$

and the solution to the following equation by $\hat{q}_i$,

$$\hat{m} = \alpha c(q_i^{**}, \theta_i) + \alpha c_\theta(q_i^{**}, \theta_i) H(\theta_i).$$

Let $\epsilon > 0$ be small enough. Then, for any $\theta \in (\tilde{\theta}_i - \epsilon, \tilde{\theta}_i + \epsilon)$, q_i^{**} is either the solution to the equation

$$-\Psi\left(\mu + y\right) = \alpha c(q_i^{**}, \theta_i) + \alpha c_\theta(q_i^{**}, \theta_i) H(\theta_i) \tag{6}$$

or the solution to the equation

$$-\omega_1'\big(y_1\big) = \alpha c\left(q_i^{**}, \theta_i\right) + \alpha c_\theta(q_i^{**}, \theta_i) H(\theta_i) \tag{7}$$

Because the left-hand-side of both equation (6) and equation (7) equal $\tilde{m}$ at $\theta_i = \tilde{\theta}_i$, and that the right-hand-side of both equations is the same continuous function of θ_i, we have $\lim_{\theta_i \to \tilde{\theta}_i -} q_i^{**} = \lim_{\theta_i \to \tilde{\theta}_i +} q_i^{**} = \tilde{q}_i$. Hence, q_i^{**} is continuous at $\tilde{\theta}_i$. Similarly, we can show that q_i^{**} is also continuous at $\hat{\theta}_i$. Therefore, q_i^{**} is everywhere continuous and is thus decreasing

[13] In Lemma 2 of Online Appendix A, we prove a stronger result that a continuous and locally decreasing function is globally decreasing, which is sufficient for the case of any $M \geq 2$.

in its range. ■

As we noted, the above auction mechanism has a nice economic interpretation as the integration of global auction and local auctions. Whenever the feasibility condition is satisfied, the mechanism is equivalent to the optimal mechanism described in Proposition 1 which is essentially a *global* auction that includes all hotspots from both regions. Whenever the feasibility condition is violated, the optimal mechanism is to allocate all cellular capacity to one region and to organize one *local* auction for each region. This integration of global and local auctions in the optimal procurement auction is the consequence of two unique features of procuring WiFi capacity for mobile traffic offloading: 1) the coupling of local auction because of the existence of the more flexible cellular capacity; and 2) the heterogeneity of demand for mobile bandwidth and supply of WiFi capacity in different regions.

It should be clarified, however, that there is only one auction, and the choice between running a global auction and running two local auctions is *endogeneously* determined by the auctioneer based on the realization of $(\theta_1, \cdots, \theta_N)$. From the perspective of a hotspot, ex ante, it does not know whether it will participate in a global auction or a local auction. It does not need to know. What matters to a hotspot is only the payment and quantity schedule designed by the auctioneer. Based on these schedules and its expectation of the types of all other hotspots, it is optimal for the hotspot to truthfully report its type by our mechanism design.

3.4 Integrating Global and Local Auctions, $M > 2$

With more than two regions, the basic idea of integrating multiple local auctions and one global auction remains the same, although the optimal grouping of WiFi regions becomes more complicated. We denote by R_g the set of regions where cellular capacity will be allocated (i.e., regions that participate in global auction) and denote by R_l the set of regions where cellular capacity will not be allocated (i.e., regions that participate in local auctions). The optimal auction involves one local auction for each region in R_l where no cellular resource will be allocated and one global auction for all regions in R_g where all cellular resource will be allocated.

The key is to *optimally* divide the set of regions into R_g and R_l. Intuitively, whether a region—say region m—should be in R_l or R_g depends on whether the non-negativity constraint $\mu_m \geq 0$ will be binding or not if it participates in the global auction. But we can evaluate whether $\mu_m \geq 0$ only after we construct R_g and R_l. Because the number of ways of dividing regions into R_g and R_l increases exponentially with the number of regions, a brute-force approach of checking each possible division is practically infeasible. Hence, we must find a division algorithm whose complexity is polynomial in the number of regions. To achieve this goal, we first note in the following proposition that R_g should be as large as possible to achieve optimality.

Proposition 4 *Given $M \geq 2$ and $(\theta_1, \cdots, \theta_N)$, suppose there are two different schemes of dividing the regions into global and local auctions, both of which lead to feasible allocation of cellular capacity: If (R_g, R_l) and $(\tilde{R}_g, \tilde{R}_l)$ where $\tilde{R}_g \subset R_g$, then the optimal gain corresponding to the auction design with (R_g, R_l) is larger than the optimal gain corresponding to the auction design with $(\tilde{R}_g, \tilde{R}_l)$.*

Proof. We write down the optimal auction design problem with (R_g, R_l) as

$$\begin{aligned}
\min_{\substack{q_i, i=1,\cdots,n \\ \mu_m, m=1,\cdots,M}} \Pi \quad &= \quad \sum_{m=1}^{M} \omega_m \big(\mu_m + y_m\big) + \alpha \sum_{i=1}^{N} C(q_i, \theta_i) + \alpha \sum_{i=1}^{N} C_\theta(q_i, \theta_i) H(\theta_i), \\
s.t. \quad & \quad \sum_{m \in R_g} \mu_m \leq \mu, \\
& \quad \mu_m \geq 0, \forall m \in R_g, \\
& \quad \mu_m = 0, \forall m \in R_l, \\
& \quad y_m = \sum_{i \in E_m} q_i, \forall m = 1, 2, \cdots, M.
\end{aligned}$$

Because $\tilde{R}_g \subset R_g$, the optimal auction design problem with $(\tilde{R}_g, \tilde{R}_l)$ is the same as the above problem except with the additional constraints that $\mu_m = 0$, $\forall m \in R_g \backslash \tilde{R}_g$. Clearly, the minimized total cost corresponding to (R_g, R_l) should be smaller than the minimized total cost corresponding to $(\tilde{R}_g, \tilde{R}_l)$. ■

Given that we need to find the largest possible R_g to achieve optimality, we clearly should start with all of the regions—that is, $R_g = \{1, \cdots, M\}$. If doing this leads to infeasible allocation of cellular capacity, we will have to shrink R_g in some way. Intuitively, we should exclude those regions with $\mu_m^* < 0$ from R_g to restore feasibility, which would naturally give rise to a sequential procedure of constructing R_g and R_l. However, the main concern with the sequential procedure is whether "exclusion" should be irreversible. In other words, if a region is excluded from R_g, would it be beneficial to put it back into R_g at some later steps in this sequential procedure? Our next result shows that the answer is no. The key insight is that if a region is in R_l at some stage, then it will be in R_l in later stages had it remained in R_g, which justifies the irreversible shrinking of R_g and guarantees the algorithm complexity of the order of $O(M)$.

To describe the procedure, we first introduce the notations for the k-subproblem. Let $Y_k = \sum_{m \in R_g^k} y_{m,k}$ where $y_{m,k} = \sum_{i \in E_m} q_{i,k}$ and $q_{i,k}$ is determined by the following equation:

$$-\Psi\left(Y_k + \mu\right) = \alpha c(q_{i,k}, \theta_i) + \alpha c_\theta(q_{i,k}, \theta_i) H(\theta_i), \forall i \in \bigcup_{m \in R_g^k} E_m.$$

Let $\mu_{m,k}^* = \phi_m\big(\Psi(Y_k + \mu)\big) - y_{m,k}$. Let $R_+^k \equiv \{m \in R_g^k | \mu_{m,k}^* \geq 0\}$, and $R_-^k \equiv \{m \in R_g^k | \mu_{m,k}^* < 0\}$.

Proposition 5 *Given $M \geq 2$ and $(\theta_1, \cdots, \theta_N)$, the optimal quantity schedule q_i^{**} is given by*

$$\begin{aligned} -\Psi\Big(\mu + \sum_{j \in E_m, m \in R_g} q_j^{**}\Big) &= \alpha c(q_i^{**}, \theta_i) + \alpha c_\theta(q_i^{**}, \theta_i) H(\theta_i), \forall i \in E_m, m \in R_g \\ -\omega_m'\big(\sum_{j \in E_m} q_j^{**}\big) &= \alpha c\left(q_i^{**}, \theta_i\right) + \alpha c_\theta(q_i^{**}, \theta_i) H(\theta_i), \forall i \in E_m, m \in R_l \end{aligned}$$

where R_g and R_l are constructed through the following iterative procedure:

- *Step 0: Let* $k = M$*,* $R_g^M = \{1, 2, \cdots, M\}$*, and* $R_l^M = \emptyset$*.*
- *Step 1: If* $R_-^k = \emptyset$*, let* $R_g = R_g^k$ *and* $R_l = R_l^k$*. Stop the procedure.*
- *Step 2: If* $R_-^k \neq \emptyset$*, let* $R_g^{k-1} = R_+^k$ *and* $R_l^{k-1} = R_l^k \cup R_-^k$*. Decrease* k *by 1 and repeat Step 1.*

The optimal payment schedule P_i^{**}*, for* $i = 1, 2, ...n$ *is given by:*

$$P_i^{**} = \alpha\left(C\left(q_i^{**}, \theta_i\right) + \int_{\theta_i}^{\theta^*} C_\theta\left(q_i^{**}, \theta\right) d\theta\right). \tag{8}$$

Proof. From the proof of Proposition 3, we know that the optimal mechanism is an integration of one global auction and at most $M - 1$ local auctions. With $M > 2$, the key is to determine the optimal division of regions into R_g and R_l.

The first part of this proposition is a straightforward generalization of Proposition 3 and the proof is omitted. The second part offers an efficient algorithm for constructing the optimal R_g and R_l with complexity of $O(M)$.

The key to prove the effectiveness of this algorithm is to show that if a region is in R_l at some stage, then it will always be in R_l in later stages. Therefore, at each stage, we should shrink R_g^k by moving all of those regions in R_-^k to R_l^{k-1} and none of those regions in R_+^k to R_l^{k-1}. In this way, we keep R_g as large as possible (i.e., by not moving those regions in R_+^k) while attempting to restore feasibility (i.e., by moving those regions in R_-^k to R_l).

Mathematically, we need to show that given $t \in R_-^k$ and $s \neq t$, if we let $R_g^{k-1} = R_g^k \backslash \{s\}$, then $t \in R_-^{k-1}$. Note s could be in either R_+^k or R_-^k.

Consider the k-subproblem. The WiFi capacity procurement $q_{i,k} = Q^*\left(\theta_i, \theta_{-i}\right)$, is determined by:

$$-\Psi(\mu + Y_k) = \alpha c(q_{i,k}, \theta_i) + \alpha c_\theta(q_{i,k}, \theta_i) H(\theta_i), \forall i \in \cup_{m \in R_g^k} E_m, \tag{9}$$

where $Y_k = \sum_{i \in \cup_{m \in R_g^k} E_m} q_{i,k}$.

In the $(k-1)$-subproblem, $q_{i,k-1}$, is determined by the following equation:

$$-\Psi(\mu + Y_{k-1}) = \alpha c(q_{i,k-1}, \theta_i) + \alpha c_\theta(q_{i,k-1}, \theta_i) H(\theta_i), \forall i \in \cup_{m \in R_g^{k-1}} E_m, \tag{10}$$

where $Y_{k-1} = \sum_{i \in \cup_{m \in R_g^{k-1}} E_m} q_{i,k-1}$.

We first show that $Y_{k-1} \le Y_k$ by contradiction. Suppose $Y_{k-1} > Y_k$, then $-\Psi(\mu + Y_k) > -\Psi(\mu + Y_{k-1})$ because $\Psi(\cdot)' > 0$. Hence, the right-hand-side of equation (9) is greater than the right-hand-side of equation (10) for all $i \in \bigcup_{m \in R_g^{k-1}} E_m$. But $\alpha c(q, \theta_i) + \alpha c_\theta(q, \theta_i) H(\theta_i)$ is increasing in q. Therefore, we must have $q_{i,k} > q_{i,k-1}$, $\forall i \in \cup_{m \in R_g^{k-1}} E_m$, which implies

$$Y_k = \sum_{i \in E_s} q_{i,k} + \sum_{i \in \bigcup_{m \in R_g^{k-1}} E_m} q_{i,k} \ge \sum_{i \in \bigcup_{m \in R_g^{k-1}} E_m} q_{i,k} > \sum_{i \in \bigcup_{m \in R_g^{k-1}} E_m} q_{i,k-1} = Y_{k-1}.$$

Contradiction. Because $Y_{k-1} \le Y_k$, we immediately see that $q_{i,k} \le q_{i,k-1}$, $\forall i \in \cup_{m \in R_g^{k-1}} E_m$ which implies $y_{m,k} \le y_{m,k-1}$, $\forall m \in R_g^{k-1}$.

Second, we show that $\mu^*_{m,k-1} \le \mu^*_{m,k}, \forall m \in R_g^{k-1}$. To see this, notice that

$$\mu^*_{m,k-1} = \phi_m\big(\Psi(Y_{k-1}+\mu)\big) - y_{m,k-1} \le \phi_m\big(\Psi(Y_k+\mu)\big) - y_{m,k-1} \le \phi_m\big(\Psi(Y_k+\mu)\big) - y_{m,k} = \mu^*_{m,k},$$

where the first inequality is because $\phi'_m > 0, \Psi' > 0$ and $Y_{k-1} \le Y_k$, and the second inequality is because $y_{m,k} \le y_{m,k-1}$.

Therefore, $\mu^*_{t,k-1} \le \mu^*_{t,k} < 0$, or equivalently, $t \in R_-^{k-1}$. ■

4 Concluding Remarks

In the present study, we designed an optimal auction mechanism for WiFi procurement so that cellular service providers can offload mobile data. The integration of both cellular and WiFi resources significantly improves mobile bandwidth availability. A unique challenge in this procurement auction is that the longer-range cellular resource introduces coupling among the shorter range WiFi hotspots. We solved for the optimal auction mechanism and show that the optimal auction can be interpreted as an endogeneously determined combination of a global auction and many local auctions.

The actual auctions and offloading to WiFi would need to be integrated with the policy management infrastructure, which supplies some of the key variables in the auction valuation: (1) the currently offered data traffic, (2) the capacity of each cell tower, and (3) the structure of the congestion cost. The proposed procurement auction integrates all relevant information into the supply chain through wireless networks. Our procurement mechanism extends beyond the limits of service providers' cellular resource to interconnect multiple hotspots in different regions. The conventional data offloading is on the basis of the access network discovery and selection function (ANDSF)[14] that processes static WiFi offload policies. Recently, the intelligent mobile solution company, Tekelec, Inc., has developed its Mobile Policy Gateway (MPG)[15] to implement complex WiFi offload policies. The Tekelec MPG enables support for our smart data offloading based on the auction approach.

Our procurement auction design problem can be regarded as a sub-problem of a capacity expansion project for a cellular service provider. Given a cellular sector, the cellular service provider can compare the option of using WiFi capacity procurement or building a new cell tower. By comparing the net gain (i.e., the difference between the reduced congestion cost and the cost of capacity expansion, whether through WiFi procurement or building tower), the service provider can select the one that yields higher net gain. In our model, the benefit of using WiFi capacity procurement is the expected reduction of congestion cost, which is given by $\mathbb{E}\left[J(0,\cdots,0) - J(y_1^*,\cdots,y_M^*)\right]$, where y_m^* is the optimal amount of WiFi capacity procured in region m and $J(y_1^*,\cdots,y_M^*)$ is the corresponding congestion cost defined in equation (1). The total payment to WiFi hotspots and the Internet service provider is $\mathbb{E}\left[\sum_{i=1}^{N} P_i^{**}(\theta_i,\theta_{-i})\right]$. Therefore, the net gain from WiFi procurement is

[14] The purpose of the ANDSF is to assist user equipment to discover and select non-3GPP networks such as WiFi and WiMax.

[15] See http://www.tekelec.com/2012-press-releases/tekelec-and-roke-partner-to-deliver-policyonthemobile-solutions.aspx.

$\mathbb{E}\left[J(0,\cdots,0) - J(y_1^*,\cdots,y_M^*) - \sum_{i=1}^N P_i^{**}(\theta_i,\theta_{-i})\right]$. To determine the optimal amount of cellular capacity, we assume that the cost of increasing the cellular capacity from μ to $\mu+\mu_0$ is $V(\mu_0)$, where $V(\cdot)$ is mostly likely discontinuous at 0 due to the fixed cost of building a new cell tower. The minimized total cost from increasing cellular capacity is the optimized objective of the following problem:

$$K(\mu_0) = \min_{\mu_0 \geq 0} \left(V(\mu_0) + \min_{\substack{\mu_1,\cdots,\mu_m \in \mathbb{R}^+ \\ \sum_i \mu_i \leq \mu+\mu_0}} \sum_{m=1}^{M} \omega_m(\mu_m) \right).$$

Let μ_0^* be the optimal μ_0 to the above problem. Then, the cellular service provider prefers (1) to add μ_0^* cellular capacity if $K(\mu_0^*) < \min\left\{J(0,\cdots,0), \mathbb{E}\big[J(y_1^*,\cdots,y_M^*)+\sum_{i=1}^N P_i^{**}(\theta_i,\theta_{-i})\big]\right\}$; (2) to procure bandwidth from WiFi hotspots if $\mathbb{E}\left[J(y_1^*,\cdots,y_M^*) + \sum_{i=1}^N P_i^{**}(\theta_i,\theta_{-i})\right] < \min\{J(0,\cdots,0), K(\mu_0^*)\}$; and (3) to seek no additional capacity otherwise.

The model in the present study can also be useful for certain supply chain problems. Consider a firm that produces multiple products using a shared resource (in-house capacity) that is common to products 1 and 2. Because of capacity limitations, the firm also needs to procure the products from different suppliers. Supplier 1 only produces product 1; suppliers 2, 3, and 4 only produce product 2. Because the in-house capacity is a shared resource that can be used for all products, it is suboptimal to decompose this supply chain problem into two independent procurement problems. Our theoretical model provides an auction framework for the downstream firm to optimally integrate the upstream capacity with its own product-flexible capacity.

We recognize several limitations in the present research. First, only one cellular service provider is considered in our procurement auction. One direction for future research is to extend our model to a setting with multiple cellular service providers. Second, we assumed the marginal cost function of all hotspots can be approximated using a one-parameter function family. This is a simplifying assumption. It could be interesting to explore how multiple dimensions of hotspot heterogeneity interact with the optimal auction design. Finally, our study focused on the supply-side reaction to congestion by assuming exogenous demand rates to improve analytical tractability. As a future research direction, it is important to study the demand-side reaction by estimating consumer response to congestion using empirical data.

References

Aijaz, A., Aghvami, H., and Amani, M. (2013). A survey on mobile data offloading: technical and business perspectives. *Wireless Communications, IEEE*, 20(2):104–112.

Bajari, P., McMillan, R., and Tadelis, S. (2009). Auctions versus negotiations in procurement: an empirical analysis. *Journal of Law, Economics, and Organization*, 25(2):372–399.

Balachandran, K., Kang, J., Karakayali, K., and Singh, J. (2008). Capacity benefits of relays with in-band backhauling in cellular networks. In *Proceedings of IEEE International Conference on Communications*, pages 3736–3742.

Balasubramanian, A., Mahajan, R., and Venkataramani, A. (2010). Augmenting mobile 3g using wifi. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 209–222.

Bichler, M., Gupta, A., and Ketter, W. (2010). Designing smart markets. *Information Systems Research*, 21(4):688–699.

Bulut, E. and Szymanski, B. (2012). Wifi access point deployment for efficient mobile data offloading. In *Proceedings of the first ACM international workshop on practical issues and applications in next generation wireless networks*, pages 45–50.

Cachon, G. and Feldman, P. (2011). Pricing services subject to congestion: charge per-use fees or sell subscriptions? *Manufacturing & Service Operations Management*, 13(2):244–260.

Che, Y. (1993). Design competition through multidimensional auctions. *RAND Journal of Economics*, 24(4):668–680.

Cheng, H. K., Bandyopadhyay, S., and Guo, H. (2011). The debate on net neutrality: A policy perspective. *Information Systems Research*, 22(1):60–82.

Cramton, P., Skrzypacz, A., and Wilson, R. (2007). The 700 mhz spectrum auction: An opportunity to protect competition in a consolidating industry. *Working Paper*.

Dasgupta, S. and Spulber, D. (1990). Managing procurement auctions. *Information Economics and Policy*, 4(1):5–29.

Dong, W., Rallapalli, S., Jana, R., Qiu, L., Ramakrishnan, K. K., Razoumov, L., Y., Z., and Cho, T. W. (2014). ideal: Incentivized dynamic cellular offloading via auctions. *IEEE/ACM Transactions on Networking*, 22(4):1271–1284.

Duenyas, I., Hu, B., and Beil, D. R. (2013). Simple auctions for supply contracts. *Management Science*, 59(10):2332–2342.

Fine, C. H. and Freund, R. M. (1990). Optimal investment in product-flexible manufacturing capacity. *Management Science*, 36(4):449–466.

Goyal, M. and Netessine, S. (2011). Volume flexibility, product flexibility, or both: The role of demand correlation and product substitution. *Manufacturing & Service Operations Management*, 13(2):180–193.

Gupta, A., Jukic, B., Stahl, D., and Whinston, A. (2011). An analysis of incentives for network infrastructure investment under different pricing strategies. *Information Systems Research*, 22(2):215–232.

Iosifidis, G., Gao, L., Huang, J., and Tassiulas, L. (2015). A double auction mechanism for mobile data-offloading markets. *IEEE/ACM Transactions on Networking*, 23(5):1634–1647.

Janakiraman, G., Nagarajan, M., and Veeraraghavan, S. (2014). Simple policies for managing flexible capacity. *Working Paper*.

Kang, X., Chia, Y. K., Sun, S., and H., C. (2014). Mobile data offloading through a third-party wifi access point: An operator's perspective. *IEEE Transactions on wireless communications*, 13(10):5340–5351.

Netessine, S., Dobson, G., and Shumsky, R. (2002). Flexible service capacity: Optimal investment and the impact of demand correlation. *Operations Research*, 50(2):375–388.

Paul, U., Subramanian, A., Buddhikot, M., and Das, S. (2011). Understanding traffic dynamics in cellular data networks. In *INFOCOM, 2011 Proceedings IEEE*, pages 882–890.

Sen, S., Joe-Wong, C., Ha, S., and Chiang, M. (2012). Incentivizing time-shifting of data: a survey of time-dependent pricing for internet access. *IEEE Communications Magazine*, 50(11):91–99.

Tan, Y. and Carrillo, J. E. (2017). Strategic analysis of the agency model for digital goods. *Production and Operations Management*, 26(4):724–741.

Tan, Y., Carrillo, J. E., and Cheng, H. K. (2016). The agency model for digital goods. *Decision Sciences*, 47(4):628–660.

Van Mieghem, J. (1998). Investment strategies for flexible resources. *Management Science*, 44(8):1071–1078.

Weber, T. (2014). Intermediation in a sharing economy: Insurance, moral hazard, and rent extraction. *Journal of Management Information Systems*, 31(3):35–71.

Optimal Auction Design for WiFi Procurement

Online Appendix

Liangfei Qiu*, Huaxia Rui†, and Andrew Whinston‡

*University of Florida, Warrington College of Business, e-mail: liangfeiqiu@ufl.edu
†The University of Rochester, Simon Business School, e-mail: huaxia.rui@simon.rochester.edu
‡The University of Texas at Austin, McCombs School of Business, e-mail: abw@uts.cc.utexas.edu

A Lemmas

Lemma 1 *In an $M/M/c$ queue, the expected waiting time is convex in service rate.*

Proof. The expected waiting time of an $M/M/c$ queue can be written as

$$W = \frac{1}{\mu} + \frac{A}{\mu^{c+1} - b\mu^c}$$

where both A and b are constant with $b < \mu$. Clearly, $W(\mu)$ is decreasing in μ because

$$\frac{d}{d\mu}\left(\frac{1}{\mu^{c+1} - b\mu^c}\right) = -\frac{(c+1)\mu^c - bc\mu^{c-1}}{\left(\mu^{c+1} - b\mu^c\right)^2} < 0.$$

For the convexity of $W(\mu)$, it suffices to show

$$\frac{d^2}{d\mu^2}\left(\frac{1}{\mu^{c+1} - b\mu^c}\right) > 0,$$

which is equivalent to

$$\begin{aligned}
\left(c(c+1)\mu^{c-1} - bc(c-1)\mu^{c-2}\right)\left(\mu^{c+1} - b\mu^c\right)^2 &< 2\left(\mu^{c+1} - b\mu^c\right)\left((c+1)\mu^c - bc\mu^{c-1}\right)^2 \\
\Leftrightarrow \left(c(c+1)\mu - bc(c-1)\right)(\mu - b) &< 2\left((c+1)\mu - bc\right)^2 \\
\Leftrightarrow (c+1)(c+2)\mu^2 - 2bc(c+2)\mu + b^2(c^2+c) &> 0. \\
\Leftrightarrow (c+1)(c+2)\left(\mu - \frac{bc}{c+1}\right)^2 + \frac{b^2c}{c+1} &> 0.
\end{aligned}$$

■

Lemma 2 *Let $\Theta \in \mathbb{R}$ be a closed interval and $f : \Theta \to \mathbb{R}$ be a function that is locally decreasing, that is, $\forall \theta \in \Theta$, there exists $\epsilon > 0$ such that $f(\theta)$ is monotone decreasing on $[\theta, \theta + \epsilon]$. If f is continuous, then f is monotone decreasign on Θ.*

Proof. Pick any $\theta_1, \theta_2 \in \Theta$ and assume $\theta_1 < \theta_2$. Let $m = \inf_{\theta \in \Theta} f(\theta)$ and $K = \{\theta \in [\theta_1, \theta_2] \big| f(\theta) = m\}$. By the Weierstrass theorem, K is non-empty. Let $\tilde{\theta} \equiv \sup(K)$, then

$\tilde{\theta} \in K$ because K is closed which is due to the continuity of f. Suppose $\tilde{\theta} \neq \theta_2$, then $\exists \epsilon > 0$ such that $\tilde{\theta} + \epsilon < \theta_2$ and

$$m = f(\tilde{\theta}) \geq f(\tilde{\theta} + \epsilon) \geq m.$$

Hence, $\tilde{\theta} + \epsilon \in K$. But $\tilde{\theta} \equiv \sup(K)$. Contradiction. Therefore, $\sup(K) = \theta_2$ and $f(\theta_2) = m \leq f(\theta_1)$. ■

B An Application with Simulation

Applying our proposed auction mechanism to the network data from one of the largest U.S. service providers, we address the following question in this section: Compared with the standard Vickrey-Clarke-Groves (VCG) auction for mobile data offloading suggested in the computer science literature (Dong et al. 2014), how much can our optimal procurement auction improve the cellular network's expected payoff? Since the VCG-type auction is a welfare maximizing mechanism, it is not surprising that our mechanism can outperform the standard VCG auction. However, our Monte Carlo simulation results demonstrate that the improvement is considerable: As compared with the standard VCG auction, our procurement auction significantly improves the cellular network's expected payoff and reduces procurement cost by more than 50%. We also evaluate the impact of the cellular capacity and the relative cost of deploying cellular resources on the performance difference between these two mechanisms.

B.1 Derivations with Specific Functional Forms

To compute numerical examples, we first assume the following parameterization of $C(q, \theta_i)$: $C(q, \theta_i) = (0.5 + \theta_i)q^2$, which implies $c(q, \theta_i) = (1 + 2\theta_i)q$, $c_\theta(q, \theta_i) = 2q$.

In this case, we can explicitly solve for $g_i(\nu)$ as

$$g_i(\nu) = \frac{\nu}{\alpha\,(1 + 2\theta_i + 2H(\theta_i))}.$$

We also assumed that

$$\omega_m(x) = \frac{\kappa_m}{x - \lambda_m},$$

where $\kappa_m > 0$ is the weight placed on region m. Hence, $\omega'_m(x) = -\kappa_m(x - \lambda_m)^{-2}$, and

$$\phi_m(x) = \lambda_m + \sqrt{-\frac{\kappa_m}{x}}.$$

So we have

$$\begin{aligned}\Phi(x) &= \sum_{m=1}^{M} \lambda_m + \sqrt{-\frac{1}{x}} \sum_{m=1}^{M} \sqrt{\kappa_m} \\ \Psi(x) &= -\frac{\left(\sum_{m=1}^{M} \sqrt{\kappa_m}\right)^2}{\left(x - \sum_{m=1}^{M} \lambda_m\right)^2}\end{aligned}$$

Therefore, q^* is determined by the following equation

$$\sum_{i=1}^{N} g_i \left(\frac{\left(\sum_{m=1}^{M} \sqrt{\kappa_m}\right)^2}{\left(\mu + q - \sum_{m=1}^{M} \lambda_m\right)^2} \right) = q.$$

Substituting the functional form of $g_i(\nu)$, we have q^* as the solution to the following cubic equation:

$$q \left(\mu + q - \sum_{m=1}^{M} \lambda_m \right)^2 = \left(\sum_{m=1}^{M} \sqrt{\kappa_m} \right)^2 \sum_{i=1}^{N} \frac{1}{\alpha \left(1 + 2\theta_i + 2H(\theta_i)\right)}.$$

q^* can be solved either explicitly or by using bisection to search in the interval $(0, \bar{q})$ where $\bar{q}$ is defined as

$$\bar{q} \equiv \sum_{i=1}^{N} g_i \left(-\Psi(\mu)\right) = \left(\sum_{m=1}^{M} \sqrt{\kappa_m} \right)^2 \frac{1}{(\mu - \sum_{m=1}^{M} \lambda_m)^2} \sum_{i=1}^{N} \frac{1}{\alpha \left(1 + 2\theta_i + 2H(\theta_i)\right)}.$$

B.2 VCG Auction

Before we do the comparison, we review the multi-unit VCG auction for procurement in our context. The VCG auction involves the following steps:

- Invite each hotspot to report its cost parameter θ. Denote the submitted cost parameters as $(\theta_1, \theta_2, \cdots, \theta_N)$.

- Under the VCG mechanism, the socially efficient allocation minimizes the sum of the expected congestion cost of the cellular service provider and the cost of hotspots. Hence, the minimization problem can be formalized as follows:

$$\begin{aligned} \min_{q_1, q_2, \dots, q_N} \quad & \mathbb{E}\Big[J\left(\sum_{i=1}^{N} q_i^*\right) + \sum_{i=1}^{N} C(q_i^*, \theta_i)\Big] \\ \text{s.t.} \quad & q_i \geq 0, \forall i = 1, 2, \cdots, N. \end{aligned}$$

- Let $\pi(\theta_1, \theta_2, \cdots, \theta_k)$ be the optimal value of the objective function, and let $(q_1^*, q_2^*, \cdots, q_n^*)$ be an optimal solution to the cost minimization problem. Let $\pi_{-i}(\theta_{-i})$ be the optimal value of the objective function with the additional constraint $q_i = 0$ (i.e., hotspot i does not participate in the auction).

- The cellular service provider will pay hotspot i according to the following:

$$P_i = \pi_{-i}(\theta_{-i}) - \pi(\theta_1, \theta_2, \cdots, \theta_N) + C(q_i^*, \theta_i)$$

  where $\pi_{-i}(\theta_{-i}) - \pi(\theta_1, \theta_2, \cdots, \theta_N)$ is the bonus payment to hotspot i, representing the positive externality that hotspot i is imposing on the cost minimization problem. The cellular service provider pays hotspot i its cost $C(q_i^*, \theta_i)$, plus its contribution to the cost minimization problem. This payment internalizes the externality.

- Hotspot i provides capacity q_i^* and receives payment P_i.

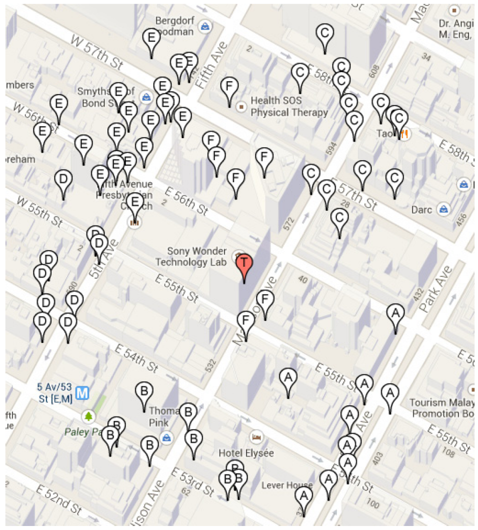

Figure 1: Area Map of A Typical Cell Sector

Note that the VCG auction is both truth-telling and socially efficient by standard arguments. All hotspots bid their cost parameters truthfully, irrespective of other hotspots' bids. The VCG mechanism guarantees the minimum total cost. However, it leads to an overpayment to hotspots that is shown in the simulation.

B.3 Simulation

In our simulations, we consider a typical urban neighborhood in New York City, as shown in Figure 1. We define a cell sector as the range of the cell tower. Our dataset consists of the location information of 14,576 cell towers from a large cellular provider in the U.S. In our simulation study, we pick a cell tower in New York City from the full list of cell towers and simulate the mobile data demand in this sector. In Figure 1, the cell tower is represented by the marker labelled with the letter "T", and the 69 WiFi hotspots in the given cell sector are represented by other markers.[1] We set the communication range for a cell tower as 250m, and set the communication range for Wi-Fi as 100m. The following steps describe the procedure of simulations:

[1] Locations of commercial WiFi hotspots are from http://wigle.net.

- Generate traffic demands in the given cell sector: To gain a sense of the population density in the coverage area of the cell tower, we use 2010 census data, which contains the land area coverage and population density of each zip code. Combining the market share of this service provider for the first quarter 2013,[2] we estimate the number of users in the given cell sector. On average, smartphone users consume about 1GB data per month, but the usage patterns of mobile data is highly uneven.[3] Paul et al. (2011) and Jin et al. (2012) found that a small number of heavy users contribute to a majority of data usage in the network. To consider the heterogeneity of data usage and the effects of peak hours, we simulate individual data usage from the byte distribution in Jin et al. (2012).[4]

- Generate WiFi regions in the cell sector: Dong et al. (2014) showed that the appropriate number of WiFi regions in a cell sector is six. Following their approach, we generate six WiFi regions by clustering the WiFi hotspots using k-means. In Figure 1, Region A, Region B, ... , and Region F indicate which region the WiFi hotspots belong to.

- Generate traffic demands in each WiFi region: We use two different methods to place users in the cell sector and assign them to the corresponding WiFi regions according to their locations. (1) All users are randomly placed in the cell sector. (2) All users are placed according to the densities of the hotspots.[5] After placing all the users, a nearest hotspot is calculated for each user location. If the distance between the nearest hotspot found and the user location is less than the hotspot range (100m), the user is counted

[2]See http://www.talkandroid.com/159929-t-mobile-loses-market-share-while-verizon-and-att-continue-to-dominate.

[3]See http://www.fiercewireless.com/special-report/average-android-ios-smartphone-data-use-across-tier-1-wireless-carriers-through.

[4]We obtain the quantiles of the byte distribution from Jin et al. (2012) and generate inidvidual usage using the Johnson System. We also adjust the usage by considering the effect of peak hours, see http://chitika.com/browsing-activity-by-hour.

[5]To calculate the densities of the hotspots for different locations, we divide the square circumscribing the cell sector into a 20 by 20 array of grids. By default, each grid has a weight of 1, except the grids whose centers are not in the range of the tower. The grid's weight is increased by the number of hotspots whose locations are inside the grid. Then, a list of grid indices is created according to the weight of each grid. Finally, for each user, a grid index is first uniformly chosen from the list, and then the location of the user is uniformly chosen from the range of the grid with the grid index just picked.

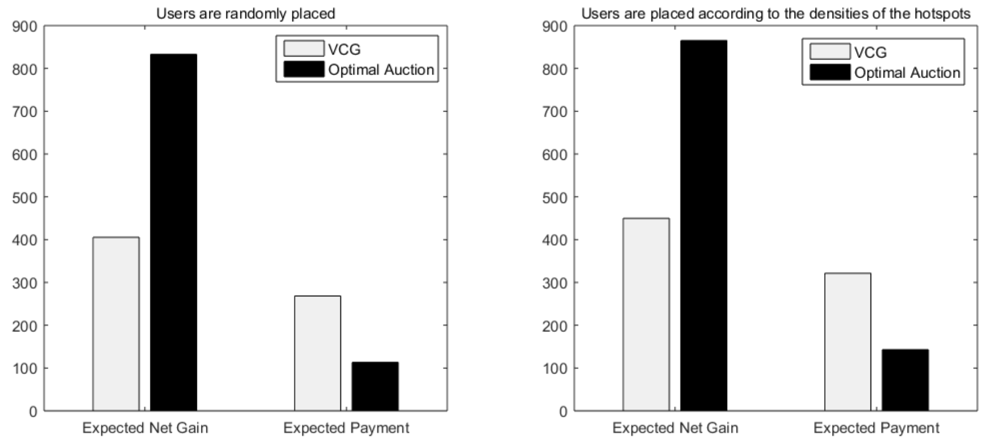

Figure 2: Performance Comparison of the Procurement Mechanisms for the Service Provider

as one of the regional population according to the WiFi region; otherwise, the user is considered as in the region with no hotspots (region 0). We run 1,000 simulations to generate traffic demands in each WiFi region.

- Generate cell tower capacity: The cell tower capacity is set to three times 3.84 MHz (Dong et al. 2014). Data spectral efficiency varies across towers from 0.5 to 2 bps/Hz.[6] We set spectral efficiency to be 1 by default and then vary the spectral efficiency to evaluate its impact. Note that when the user demand for mobile data is below 80% of the cell tower capacity, the cellular service provider faces no congestion cost.

We conduct a variety of simulations to compute the corresponding allocation under the VCG mechanism and under our optimal mechanism. The relative cost of deploying cellular resources as compared with WiFi resources affects the bandwidth allocation result. Joseph et al. (2004) assumed that the relative cost of deploying cellular resources as compared with WiFi resources is 4:1. We follow their assumptions and set the parameter values: $\omega_m\left(\mu_m + y_m\right) = 0.5a\left(\frac{1}{\mu_m + y_m - \lambda_m}\right)$ where a is set to 4 and $C\left(Q, \theta_i\right) = \left(0.5 + \theta_i\right) Q^2$. In the simulation, we vary the relative cost parameter a and find that the results are robust. A hotspot's private cost parameters θ_i is drawn from a uniform distribution $U[0, 1]$ for 1,000

[6]See http://www.rysavy.com/Articles/2011_05_Rysavy_Efficient_Use_Spectrum.pdf.

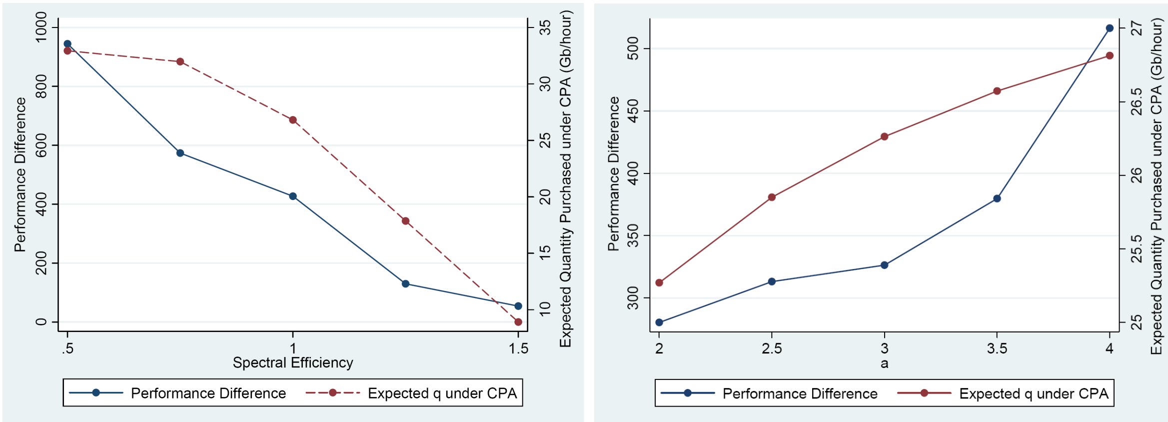

Figure 3: Performance Difference and Cell Tower Capacity (Left); Performance Difference and Relative Cost of Deploying Cellular Resources (Right)

times.

The simulation result of the performance comparison is shown in Figure 2. In the left panel, the users are randomly placed in the cell sector. In the right panel, the users are placed according to the densities of the hotspots. The two panels show similar results: our proposed procurement auction significantly outperforms the VCG mechanism in terms of the expected net gain of the cellular service provider (the expected net gain = the reduction of congestion cost - the payment to hotspots).

Data spectral efficiency varies across cell towers using different wireless technologies. An increase in spectral efficiency significantly contributes to tower capacity. The left panel of Figure 3 evaluates the impact of spectral efficiency (cell tower capacity) on the performance difference, which is defined as the difference between the service provider's expected net gain under the proposed mechanism and the gain under the VCG mechanism.[7] Note that the unit of the performance difference is normalized, and we are only interested in the trend. As the cellular capacity increases, the advantage of our proposed mechanism, in comparison with the VCG mechanism, decreases. This is because the bandwidth purchased from the WiFi hotspots also decreases as cellular capacity increases, as is indicated by the dashed line

[7]The simulation results are similar when the users are randomly placed or are placed according to the densities of the hotspots, so here we only present the result when the users are randomly placed.

in the left panel of Figure 3. The service provider is less willing to purchase WiFi resources when it owns a relatively large cellular capacity, and the overpayment problem in the VCG mechanism is thus less severe.

The right panel of Figure 4 shows that as a increases, the advantage of our mechanism as compared with the VCG mechanism increases, which is expected because, with congestion being more costly, the service provider is more willing to procure from the WiFi hotspots, thereby exacerbating the overpayment problem in the VCG mechanism.

C Cellular Technology and Broadband Technology

We elaborate the differences between broadband technology and cellular technology from the two perspectives.

(1) Broadband Technology

With the rapid deployment of fiber optics, broadband capacity constraints are gradually becoming less of an issue, even as the consumption of online content continues to grow at a rapid pace. More specifically, our assumption was guided by the fact that over the past years, broadband providers have increased capacity, and thanks to rapid advances in fiber technology (whose rate of growth is even faster than Moore's Law in semiconductors; at the same time, the networking equipment have been getting cheaper by around 25-40% every year following Moore's Law), broadband providers have been able to increase capacity at a very low cost, even as consumers have increased their consumption for online content. The cost of provisioning the marginal customer at large broadband providers today is less than \$1/month: about half of that cost is till the point of peering (the "backhaul" cost, in industry terminology), and the other half is incurred while carrying the data from the point of peering to the local exchange. Thus, broadband capacity has not been a bottleneck even as consumption for data has increased. Choi et al. (2014) highlighted the difference between fixed and mobile networks: Mobile networks encounter technical and physical constraints in expanding capacity due to the limited availability of spectrum. The Federal Communications Commission (FCC) also stated the difference: "Mobile broadband is an earlier-stage platform than fixed broadband, ..., Mobile broadband speeds, capacity, and penetration are typically much lower than for fixed broadband. ... In addition, existing mobile networks present operational constraints that fixed broadband networks do not typically encounter." (FCC Order, page 94-95).

An independent verification that broadband capacity is not a bottleneck comes from empirical observations. FCC comes out with an annual state of the broadband report every year (called "Measuring Broadband America"), and the latest report that is available currently is for the year 2014. In this report, one of the performance metrics that the Commission measures is the "24 Hour versus Peak Performance Variation by Technology." The data shows that there is hardly any dip in performance during peak periods (for example, for fiber, during peak periods, the performance drops from an average 115% of advertised speeds to 112% of the advertised speed during peak hours; for cable, the drop is from 105% of advertised speeds to 101%; and for DSL, the drop is from 95% to 91%), and these numbers have arguably become better since (currently, the FCC has the raw data available for the 2015 report on its website). Therefore, broadband providers are gradually becoming more able to handle their peak load without any degradation in speed of delivery.

(2) Cellular Technology

The cellular capacity is determined by amount of spectrum, number of cell towers, and spectral efficiency of technology, as is illustrated in the following figure which is from Rysavy-Research (2014).

Spectrum is a limited and finite resource for mobile networks (Rysavy-Research 2014). In the U.S., cellular systems use roughly 500 MHz, although an individual operator's access to spectrum is much smaller and is subject to spectrum aggregation rules. On the other hand, wired network can access far more frequencies in the mediums (e.g., coax cable, fiber-optic cable, etc.) they use, and they can carry their spectrum within the physical medium with near-complete control. Once the capacity of one cable is exhausted, another one can be placed alongside. This is in stark contrast with wireless networks which rely on the propagation of signals through the air and the same frequencies cannot be used without interference until some distance away.

Because of the finiteness of the spectrum resource, obtaining radio spectrum is very costly: In the United States, the Federal Communications Commission (FCC) conducts competitive auctions of licenses for electromagnetic spectrum. Since July 1994, the FCC has conducted 87 spectrum auctions, which raised over $60 billion for the U.S. Treasury (Cramton et al. 2002). Therefore, additional radio spectrum is not always available and obtaining it from spectrum auctions is expensive. Additionally, due to antitrust concerns in the wireless industry, several influential economists suggested that FCC should place limits on how much spectrum AT&T and Verizon are allowed to buy (Cramton et al. 2007). Such concern is reflected in the action taken by the FCC to block the merger between AT&T and T-Mobile in 2011. Due to these regulatory constraints, it is very difficult for cellular service providers to acquire additional spectrum resources.

Given limited spectrum, the cellular industry has been using sophisticated modulation and encoding methods to extract as much capacity as possible from available spectrum to meet the growing demand from mobile users. However, today's networks already operate at close to maximum theoretical spectral efficiency constrained by the laws of physics. It is also far more challenging to increase efficiency in radio technology than to increase efficiency in wire or fiber cables because radio connections in open environment have more noise than shielded wires.

Although building more cell towers can also increase wireless capacity, building a new cell tower is very expensive and time consuming. In the United States, the number of cell towers increased from 12,824 in 1993 to 304,360 in 2013. However, the increased number of cell towers has not allowed capacity to come even close to matching the capacity of wired network. Some industry expert estimates that it will cost at least $150,000 to construct a tower.

Moreover, there are health concerns on the radiation from cell towers. Many governmental bodies require that cellular service providers share cell towers so as to decrease environmental and cosmetic impact. This issue is an influential factor of rejection of installation of new cell towers in communities. For example, in February 2009, the French telecom company Bouygues Telecom was ordered to take down a cell tower due to uncertainty about its effect on health. Residents in the commune Charbonnieres in the Rhone department had sued the company claiming adverse health effects from the radiation emitted by the 19-meter-tall cell tower.

Because of the unique characteristics of the information and communications technology (ICT) industries, broadband capacity has not been a constraining factor in the past several years. In contrast, cellular capacity is limited by the finite amount of radio spectrum and the inherent limitations of radio as a medium. Even with a breakthrough in cellular technologies in the near future, cellular capacity will still be limited by various regulatory constraints, which is less of an issue for broadband because it does not rely on radio spectrum and cell towers.

References

Choi, J., Jeon, D., and Kim, B. (2014). Asymmetric neutrality regulation and innovation at the edges: fixed vs. mobile networks. *Working Paper*.

Cramton, P., Skrzypacz, A., and Wilson, R. (2002). Spectrum auctions. In Cave, M., Majumdar, S., and Vogelsang, I., editors, *Handbook of Telecommunications Economics*.

Cramton, P., Skrzypacz, A., and Wilson, R. (2007). The 700 mhz spectrum auction: An opportunity to protect competition in a consolidating industry. *Working Paper*.

Dong, W., Rallapalli, S., Jana, R., Qiu, L., Ramakrishnan, K. K., Razoumov, L., Y., Z., and Cho, T. W. (2014). ideal: Incentivized dynamic cellular offloading via auctions. *IEEE/ACM Transactions on Networking*, 22(4):1271–1284.

Jin, Y.and Duffield, N., Gerber, A., Haffner, P., Hsu, W. L., Jacobson, G., Sen, S., Venkataraman, S., and Zhang, Z. L. (2012). Characterizing data usage patterns in a large cellular network. In *Proceedings of the ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design*, pages 7–12.

Paul, U., Subramanian, A., Buddhikot, M., and Das, S. (2011). Understanding traffic dynamics in cellular data networks. In *INFOCOM, 2011 Proceedings IEEE*, pages 882–890.

Rysavy-Research (2014). How wireless is different: Considerations for the open internet rulemaking. Technical report, Rysavy Research.