An Iterative Double Auction for Mobile Data Offloading

George Iosifidis[†], Lin Gao*, Jianwei Huang*, and Leandros Tassiulas[†]
*Dept. of Information Engineering, The Chinese University of Hong Kong, Hong Kong
[†]Dept. of Computer and Communication Engineering, University of Thessaly, and CERTH, Greece

Abstract—Mobile data offloading through complementary network technologies such as WiFi and femtocell can significantly alleviate network congestion and enhance users' QoS. In this paper we consider a market where mobile network operators (MNOs) lease third-party deployed WiFi or femtocell access points (APs) to dynamically offload the traffic of their mobile users. We assume that each MNO can employ multiple APs and each AP can concurrently serve traffic from multiple MNOs. We design an iterative double auction mechanism that ensures the efficient operation of the market, where MNOs maximize their offloading benefits and APs minimize their offloading costs. Such a mechanism incorporates the special characteristics of the wireless network, such as the coupling of MNOs' offloading decisions and APs' capacity constraints. The proposed market scheme does not require full information about the MNOs and APs, incurs minimum communication overhead, and creates nonnegative revenue for the market broker.

I. INTRODUCTION

Today we are witnessing an unprecedented worldwide growth of mobile data traffic that is expected to reach 10.8 exabytes per month in 2016, an 18-fold increase compared to 2011 [1]. These developments pose new challenges to mobile network operators (MNO) who have to significantly enhance their infrastructure accordingly. However, traditional network expansion methods such as acquiring new spectrum licences and upgrading technologies (e.g., from WCDMA to LTE/LTE-A) are often costly [2] and time-consuming, and more importantly are expected to be outpaced in less than 4 years by the continuing traffic increase [1]. Clearly, operators must find novel methods to address this problem, and *mobile data offloading* appears as one of the most attractive solutions.

Mobile data offloading refers to the technique of routing the data traffic of mobile users of a macrocellular network using alternative means such as WiFi or femtocell networks. Nowadays, there is consensus that data offloading is a cost-effective [3] and energy-prudent method that benefits both the operators and the mobile users (MUs). Therefore, it is not surprising that many MNOs have already deployed their own WiFi access points (APs) to complement their traditional macrocellular network [4], or initiated collaborations with

The work of L. Gao and J. Huang was supported by the General Research Funds (Project Number CUHK 412710 and CUHK 412511) established under the University Grant Committee of the Hong Kong Special Administrative Region, China. The work of G. Iosifidis and L. Tassiulas was supported by the project SOFON which is implemented under the "ARISTEIA" Action of the Operational Programme "Education and Lifelong Learning" and is cofunded by the European Social Fund (ESF) and National Resources.

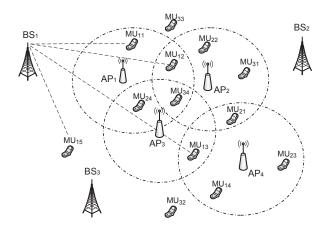


Fig. 1. Each BS is managed by a different MNO and serves a set of MUs. MU_{ij} represents the j^{th} MU associated with BS i. An MU may be covered by multiple APs and each AP may serve more than one MUs.

existing WiFi networks [5]. This approach is facilitated by technological advances such as the HotSpot 2.0 protocol which addresses related security issues [6]. This also leads to interesting new network service models such as seamless hybrid macrocellular - WiFi connection services [7].

Nevertheless, in order to fully reap the benefits of offloading, the MNOs need to ensure that their clients will be able to offload data as frequently as possible. However, a ubiquitous access point deployment by the MNOs is very expensive and even impractical in some cases (e.g., due to site acquisition issues). An ideal method to overcome this obstacle and ensure the high availability of APs is the *employment* (leasing) of third party WiFi and femto APs, which are already installed in homes or other venues (e.g., companies, stores). This strategy will allow operators to handle mobile data traffic with reduced capital and operational expenditures (CAPEX and OPEX), and increase their network capacity on-demand.

However, AP owners are expected to ask for (monetary) compensation, since admitting macrocellular traffic will consume APs' limited wireless resources and broadband connection capacities for their own (internal) traffic demand. Hence, we need to understand: how much traffic each AP should admit for each MNO and how much to charge? or, from the perspective of the MNOs: how much traffic each MNO should offload to each AP and how much to pay? In this paper, we design a mechanism that addresses these issues by taking into account the particular challenges of the wireless

data offloading.

Specifically, we envision an *offloading market* where a set of MNOs (the buyers) compete to lease a set of APs (the sellers) for the offloading service. We assume that the marketplace is managed by a centralized *broker* who can be a state-managed clearing house or a private company, similar as those in the secondary spectrum market [8]. WiFi and femtocell AP owners offer their services (i.e., offloading traffic for the MNOs) with certain reimbursements. As shown in Figure 1, we look at the general case where each AP can serve more than one MNOs, and each MNO may lease multiple APs at different locations to offload the traffic of its users (APs are overlapping). The MNOs declare how much they are willing to pay each AP. The broker collects the MNOs' requests and the APs' offers, and determines how much traffic of each MNO will be offloaded to each AP and at what price.

The challenge here is to design a market mechanism tailored to the wireless offloading problem which, at the same time, satisfies the desirable economic properties. We consider the realistic scenario of a market with incomplete information, i.e., where the broker is not aware of the actual needs of the MNOs and the APs. Therefore, he must employ an incentive compatible mechanism that induces the buyers (MNOs) and the sellers (APs) to reveal truthfully their needs. With this information, the broker tries to maximize the efficiency of the market by properly matching the buyers and the sellers. At the same time, the broker is not willing to lose money. Hence, the mechanism should be (weakly) budget balanced, i.e., the total payments from the buyers should not exceed the total payments to the sellers.

A suitable scheme for this setting is a double auction mechanism. Unfortunately, double auctions are notoriously hard to design and implement [9]. They can be inefficient and applicable only to certain simplified settings, e.g., for bidders with single-unit demands (McAfee auction [10]), or they can be budget imbalanced with a high computational complexity (VCG auction [11]). In our case, the double auction design problem is further perplexed due to the following realistic issues that we explicitly take into account:

- (11) The offloading benefit (utility) of each MNO is AP-specific. For example, an AP that is located at the boundary of an MNO's cell is the most important for that MNO, since it can offload the traffic that otherwise would be very costly for the MNO to serve directly due to poor channel condition between the MU and the base station. However, an AP that is located close to the MNO's base station will be less useful.
- (12) The offloading decisions of the MNOs are coupled.
 The accrued benefit from offloading a given amount of
 traffic to a certain AP depends on how loaded the MNO
 already is, which in turn depends on its decisions of
 offloading traffic to other APs.
- (13) The APs are heterogeneous. That is, different APs may have different costs for serving cellular traffic from the same MNO. Besides, the same AP may also incur different costs for serving traffic from different MNOs

- due to different quality of service (QoS) requirements.
- (14) The offloading decisions of the APs are coupled. An AP's cost for offloading certain amount of data for one MNO also depends on the total traffic that the AP has committed to offload for other MNOs.

In order to overcome the difficulties in double auction design without compromising the system modeling, we choose an alternative method based on the framework of *Network Utility Maximization* (NUM) [12]. Our starting point is the work of Kelly *et al.* [13] which introduced a Walrasian auction for link capacity allocation in networks. In that scheme, multiple buyers (the nodes) bid for bandwidth, and a single seller (the network) determines the unit price for each link so as to balance demand and supply. Here, we generalize this one-side approach (i.e., with one seller, the network) [13] to the case with many sellers (the APs) and many buyers (the MNOs). Moreover, the prices in our scheme not only reflect the APs' capacity constraints but also their offloading costs¹.

Specifically, our proposed mechanism is an iterative algorithm that enables the broker to gradually reach the socially efficient solution, without any prior knowledge for the market. The MNOs and APs submit request and offer bids respectively, in each round, responding to the prices announced by the broker. A basic assumption of [13] is that bidders (i.e., the MNOs and APs in our scheme) are *price-takers*. This means that they do not anticipate (or cannot estimate) the impact of their bids on the prices. Price-taking behavior is often observed in markets with many buyers and sellers [14], or when each bidder is not aware of the decisions of other bidders and/or system parameters.

The main contributions of this paper are as follows:

- We study a general market model where multiple operators (MNOs) compete to lease multiple (possibly overlapping) access points (APs) for data offloading. Each MNO can concurrently lease several APs and each AP can offload traffic for several MNOs at the same time.
- 2) We apply an iterative double auction scheme which is efficient (maximizes the social welfare), weakly budget balanced (the broker does not lose money), individually rational (MNOs and APs are willing to participate), and incentive compatible (MNOs and APs reveal their truthful needs/demands) under the assumption of pricetaking behavior.
- 3) The proposed scheme has low computational complexity, induces small communication overheads, and clears the market for general MNO utility (benefit) functions and AP cost functions (only concavity is required). The broker does not need to know these functions in advance (incomplete market information).
- 4) The introduced framework considers important realistic issues of the mobile data offloading problem, (I1) –

¹For the distributed version of [13], it can be argued that there are multiple bandwidth sellers, i.e. the links. However, the links only balance the traffic, i.e. they do not have cost functions and do not submit bids.

(I4), which, as we explain in details in Section II, have been overlooked until now by the related works.

The rest of this paper is organized as follows. In Section II we review the literature and emphasize how it differs to our work. In Section III we provide the system model and formally introduce the problem. In Section IV we present the mechanism and prove its properties, and in Section V we provide numerical results. We conclude in Section VI.

II. RELATED WORK

The benefits of macrocellular data offloading to WiFi networks have recently been studied and quantified [15], [16]. Approximately 80% of mobile data traffic is generated and consumed indoors [17], and hence can be offloaded to APs. In [18] the authors studied optimal offloading strategies for delay tolerant applications that take into account the delay constraints of users. Clearly, the offloading benefits depend on the availability of APs that are open and have extra capacity to offload cellular traffic. Interestingly though, the problem of incentivizing WiFi open access has received very little attention until today.

Another option for offloading are femtocell access points (FAPs) [19]. This presumes that FAPs operate in the so-called open access mode and admit traffic from non-registered macrocellular users. However, FAP owners are expected to be reluctant to serve other users without proper compensation [20]. This compensation can be either a price discount [21], or a direct payment from the operator. A few related works in this area study monopolistic markets (i.e., with one operator) [22], or do not consider the challenges, (I1) - (I4), that arise for the MNOs [23], [25].

In our previous work [24] we studied the interaction of multiple MNOs and APs in offloading markets assuming that there is complete information for all the market participants. The problem is substantially different when one relaxes this assumption. Markets with many buyers and sellers under incomplete information are usually cleared through double auctions. One option is to use the VCG mechanism which exhibits a very high computational complexity and can yield a budget imbalanced outcome [11]. Another prominent scheme is the McAfee mechanism [10] which has recently been proposed for spectrum allocation in secondary spectrum markets [26] and for traffic relaying [27]. However, this mechanism was originally designed for single-unit demands/offers of homogeneous items, and there are very few extensions for multiple or heterogenous items [26], [27]. In all cases, the outcome is inefficient, which is an inherent characteristic of McAfee auction.

In this paper, we adopt a different approach and use a market mechanism based on the NUM framework [12] and in particular motivated by the scheme in [13] for bandwidth allocation in networks. This is a Walrasian auction which maximizes the market welfare under the assumption that bidders are price-takers. The latter is a valid assumption for large markets where the bid of each player has an infinitesimal impact on the prices, or for the case the bidders have limited

knowledge about the market (e.g., number of players and system resources) and hence cannot estimate their impact [14].

At the expense of this assumption, we propose an iterative double auction mechanism which satisfies all the desirable economic properties. Our work substantially departs from the algorithm in [13]. Namely, the resource sellers (the APs) in our work, unlike the respective sellers (the links) in [13], try to maximize their own net benefit instead of simply balancing the traffic. This renders the proposed scheme appropriate for the many-to-many interactions considered here, and leads to a different behavior compared with [13]. A similar approach was followed in our previous work [28] for bandwidth allocation in peer-to-peer networks where however the objectives and the solution method were significantly different.

III. SYSTEM MODEL AND PROBLEM STATEMENT

In this section, we introduce the system model, and formulate the data offloading problem as a market design problem where the objective is to maximize the social welfare.

A. System Model

We consider a system with a set $\mathcal{M} \triangleq \{1,2,\ldots,M\}$ of base stations (BS) owned by M different MNOs², and a set $\mathcal{I} \triangleq \{1,2,\ldots,I\}$ of APs. Each AP can be a WiFi or a femtocell access point operating in a separate frequency band, and hence does not interfere with the macrocellular network³ Each BS serves a group of mobile users (MUs) which are randomly distributed within the BS's coverage area and have a lot of traffic to send. Each MU is also covered by one or more APs. We assume that time is slotted and we study the market for one time period. The MUs' location and traffic types may change over time but are considered fixed within each period (slot).

Consider the case where each BS $m \in \mathcal{M}$ would like to offload $x_{mi} \geq 0$ bytes of data through AP $i \in \mathcal{I}$. We define BS m's offload request vector to all I APs as $x_m \triangleq (x_{mi})_{i \in \mathcal{I}}$, and total offloaded data of BS m is $X_m = \sum_{i \in \mathcal{I}} x_{mi}$. A BS's request depends on the locations and traffic of its MUs. We use $J_m(x_m)$ to denote BS m's utility when offloading traffic x_m to the APs, which equals to BS's cost reduction comparing with the case that it serves x_m directly⁴. We assume that $J_m(\cdot)$ is a positive, increasing, and strictly concave function of vector x_m , satisfying the principle of diminishing marginal returns [29].

Our model captures the following important aspects of the offloading problem. First, the BS's offloading benefit is in general AP-specific. It not only depends on the total offloading traffic (X_m) , but also depends on which AP offloads how much (the vector \boldsymbol{x}_m). For example, an AP located at the boundary of the BS's cell is more important since it can offload traffic

²For the rest of the paper we will use "BS" and "MNO" interchangeably. ³WiFi operates in the unlicensed ISM band, which is naturally separate from the licensed cellular band. The femtocell may use a different band from the macrocell base stations under the "separate carrier" scheme [30].

⁴If an MNO manages more than one BSs, then the respective utility function represents the joint offloading benefit.

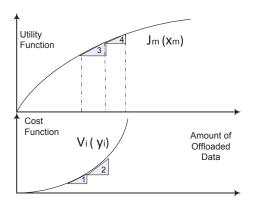


Fig. 2. Concave BS utility function, $J_m(\cdot)$, and convex AP function, $V_i(\cdot)$. The numbered triangles have the same length - corresponding to equal offloaded data increase - but different height (h): $h_2 > h_1$ and $h_4 < h_3$.

from MUs that are costly for the BS to serve directly (e.g., due to poor channel condition between the MU and the BS).

Second, for each BS $m \in \mathcal{M}$, its offloading decisions to the different APs are coupled. Clearly, the benefit from offloading traffic to an AP depends on the load of the BS, which in turn depends on its offloading decisions to other APs. In other words, even if two different strategies, x_m and \hat{x}_m , suggest equal amount of offloaded data to a certain AP i, $x_{mi} = \hat{x}_{mi}$, the respective utility improvements may differ:

$$\begin{split} J_m(x_{mi}, \boldsymbol{x}_m^{-i}) - J_m(0, \boldsymbol{x}_m^{-i}) &\neq J_m(\hat{x}_{mi}, \hat{\boldsymbol{x}}_m^{-i}) - J_m(0, \hat{\boldsymbol{x}}_m^{-i}) \\ \text{where } \boldsymbol{x}_m^{-i} &\triangleq (x_{mj})_{j \in \mathcal{I} \setminus \{i\}}, \text{ and } \hat{\boldsymbol{x}}_m^{-i} &\triangleq (\hat{x}_{mj})_{j \in \mathcal{I} \setminus \{i\}}. \\ \text{Each AP } i &\in \mathcal{I} \text{ responds to offloading requests and admits} \end{split}$$

Each AP $i \in \mathcal{I}$ responds to offloading requests and admits $y_{im} \geq 0$ bytes from each BS m in one time period. We define the admitted traffic vector $\mathbf{y}_i \triangleq (y_{im})_{m \in \mathcal{M}}$. Clearly, a viable market solution exists only if the BSs and APs finally agree on the market outcome, i.e. if $x_{mi} = y_{im}$, $\forall m \in \mathcal{M}, \forall i \in \mathcal{I}$. We use $V_i(\mathbf{y}_i)$ to denote the cost incurred by AP i for serving the BSs, which is a positive, increasing and strictly convex function in vector \mathbf{y}_i . This property captures the fact that as the admitted traffic by each AP increases, its operation cost for admitting one more unit of traffic increases due to the congestion effect and because less of its resources are available for serving its own traffic [29], as illustrated in Figure 2.

The AP's offloading cost depends on its own traffic demand as well as the MUs' traffic characteristics (e.g., the average distance from the AP and the requested QoS). Similarly, an AP's incurred cost for admitting traffic for a certain BS, depends on how loaded the AP already is, i.e. how much traffic offloaded for other BSs. Finally, AP $i \in \mathcal{I}$ has a capacity of C_i bps and hence the maximum amount of data that can be admitted within a certain time period is $C_i \cdot T$ bytes:

$$\sum_{m=1}^{M} y_{im} \le C_i \cdot T. \tag{1}$$

Without loss of generality, we normalize the slot duration to be T=1. A key difference of our model from previous works is that we use general utility and cost functions which allow us to capture a wide range of wireless offloading systems. For example, if the MUs of a BS are not covered by any AP, then the respective offloading utility component is zero.

B. Problem Statement

Clearly, the objectives of the BSs and APs are conflicting with each other. If they decide independently how much data to offload or to admit, it is very difficult to reach an agreement (i.e., $x_{im} = y_{mi}$, $\forall m \in \mathcal{M}, \forall i \in \mathcal{I}$). Therefore, there is a need for a market controller, a broker, with the task to find the offload request matrix $\mathbf{x} \triangleq (\mathbf{x}_m)_{m \in \mathcal{M}} = (x_{mi})_{m \in \mathcal{M}, i \in \mathcal{I}}$ and the admitted traffic matrix $\mathbf{y} \triangleq (\mathbf{y}_i)_{i \in \mathcal{I}} = (y_{mi})_{m \in \mathcal{M}, i \in \mathcal{I}}$ that ensure the efficient operation of the market. This is achieved when the total benefit for the MNOs is maximized and the aggregate cost for the APs is minimized.

Specifically, the broker can find the optimal x and y by solving the social welfare maximization (SWM) problem:

SWM:
$$\max_{\boldsymbol{x},\boldsymbol{y}} \sum_{m=1}^{M} J_m(\boldsymbol{x}_m) - \sum_{i=1}^{I} V_i(\boldsymbol{y}_i),$$
 (2)

s.t

$$\sum_{m=1}^{M} y_{im} \le C_i, \, \forall \, i \in \mathcal{I}, \tag{3}$$

$$x_{mi} = y_{im}, \, \forall \, m \in \mathcal{M}, \, \forall \, i \in \mathcal{I},$$
 (4)

$$x_{mi} \ge 0, y_{im} \ge 0, \forall m \in \mathcal{M}, \forall i \in \mathcal{I}.$$
 (5)

Notice that, technically, we could remove from SWM the variables x by using (4). However, we keep this formulation in order to facilitate the analysis and make clear the relation of the SWM problem with the respective problems that each BS and AP solves.⁵ The objective function of SWM is strictly concave and the constraint set is compact and convex. Hence, SWM admits a unique optimal solution that can be described using the Karush-Kuhn-Tucker (KKT) conditions [31].

Specifically, we relax the constraints and define the Lagrangian:

$$L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{x}, \boldsymbol{y}) = \sum_{m=1}^{M} J_m(\boldsymbol{x}_m) - \sum_{i=1}^{I} V_i(\boldsymbol{y}_i)$$
$$-\sum_{i=1}^{I} \lambda_i \cdot \left(\sum_{m=1}^{M} y_{im} - C_i\right) + \sum_{m=1}^{M} \sum_{i=1}^{I} \mu_{mi} \cdot (y_{im} - x_{mi}),$$

where $\lambda \triangleq (\lambda_i \geq 0)_{i \in \mathcal{I}}$ is the vector of Lagrange multipliers corresponding to constraints (3), and $\mu \triangleq (\mu_{mi} \in \mathcal{R})_{m \in \mathcal{M}, i \in \mathcal{I}}$ is the matrix of Lagrange multipliers for constraints (4). The KKT conditions that yield the optimal solution, λ° , μ° , x° , y° , for the SWM problem are given by the following set of equations: $\forall m \in \mathcal{M}, \forall i \in \mathcal{I}$,

$$(A1): \quad \frac{\partial J_m(\boldsymbol{x}_m^\circ)}{\partial x_{mi}} = \mu_{mi}^\circ, \ (A2): \frac{\partial V_i(\boldsymbol{y}_i^\circ)}{\partial y_{im}} = \mu_{mi}^\circ - \lambda_i^\circ,$$

(A3):
$$\lambda_i^{\circ} \cdot \left(\sum_{m=1}^M y_{im}^{\circ} - C_i\right) = 0, (A4): x_{mi}^{\circ} = y_{im}^{\circ},$$

$$(A5): \quad \mu_{mi}^{\circ} \cdot (y_{im}^{\circ} - x_{mi}^{\circ}) = 0, \ (A6): x_{mi}^{\circ}, y_{im}^{\circ}, \lambda_{i}^{\circ} \ge 0.$$

However, the direct SWM solution from the broker is not possible due to the limited information the broker has for the

⁵This formulation is similar to the *consistency pricing* in [12] where different nodes must agree on a common - system wide - solution

market. Namely, we assume that the broker is unaware of the utility and cost functions $J_m(\cdot)$, $\forall m \in \mathcal{M}$, and $V_i(\cdot)$, $\forall i \in \mathcal{I}$. Therefore, he has to use a mechanism (double auction in this paper) to elicit this hidden information. Such a mechanism should ideally be (i) *efficient*: maximize the social welfare, (ii) *individually rational*: bidders do not get worse by participating, (iii) *incentive compatible*: bidders truthfully reveal (directly or indirectly) their private information, (iv) (weakly) budget balanced: total payments to the broker are nonnegative. Nevertheless, a fundamental result in mechanism design is that there does not exist a double auction that possesses all theses properties [9].

Here, we take a different approach and assume that bidders are *price takers*. This assumption, which corresponds to a perfect competition market, is reasonable for bidders with bounded computational capabilities and/or limited information, or for large markets with many participants where each one has infinitesimal impact on the market prices.

IV. THE IDA MECHANISM

In this section, we present an Iterative Double Auction (IDA) mechanism for price taking bidders. We prove that it solves the mobile data offloading problem, taking into account (I1) - (I4), and satisfies the desirable economic properties.

A. IDA Resource Allocation and Pricing Rules

The basic idea of this mechanism is that the broker solves a different optimization problem to determine \boldsymbol{x} and \boldsymbol{y} which, if it is combined with the proper pricing (for the BSs) and reimbursement (for the APs) rules, ensures the optimal solution of SWP. This scheme corresponds to a double auction where many buyers (BSs) and many sellers (APs) interact in an iterative fashion until the market reaches an efficient, i.e. market clearing point.

The mechanism consists of two stages during each iteration. In the first stage, each BS $m \in \mathcal{M}$ submits a bid $p_{mi} \geq 0$, for each AP $i \in \mathcal{I}$, and each AP $i \in \mathcal{I}$ submits a bid $\alpha_{im} \geq 0$ for every BS $m \in \mathcal{M}$. These bids signal the offloading needs and serving costs for the BSs and the APs respectively, and are used as inputs in the allocation rule. Later on we will explain the precise relationship between these bids and the actual BSs' payments and APs' reimbursements. In the second stage, the broker determines the allocation (how much traffic each AP will admit from each BS) based on the bids from two sides by solving the broker allocation problem (BAP):

BAP:
$$\max_{x,y} \sum_{m=1}^{M} \sum_{i=1}^{I} \left(p_{mi} \log x_{mi} - \frac{\alpha_{im}}{2} y_{im}^2 \right),$$
 (6)

s.t.

$$\sum_{m=1}^{M} y_{im} \le C_i, \, \forall \, i \in \mathcal{I}, \tag{7}$$

$$x_{mi} = y_{im}, m \in \forall M, \forall i \in \mathcal{I},$$
 (8)

$$x_{mi} \ge 0, y_{im} \ge 0, \forall m \in \mathcal{M}, \forall i \in \mathcal{I}.$$
 (9)

Notice that the objective function is motivated by the allocation rule of [13], with the additional convex component capturing

the (convex) increasing cost functions of the APs. We also define the bid vectors $\boldsymbol{p}_m \triangleq (p_{mi})_{i \in \mathcal{I}}$ and $\boldsymbol{\alpha}_i \triangleq (\alpha_{im})_{m \in \mathcal{M}}$ for every BS $m \in \mathcal{M}$ and AP $i \in \mathcal{I}$ respectively.

The BAP problem has the same constraint set as the SWM problem, and a different yet strictly concave objective function. Hence, it admits a unique optimal solution. We define the corresponding Lagrange function as

$$L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{x}, \boldsymbol{y}) = \sum_{m=1}^{M} \sum_{i=1}^{I} \left(p_{mi} \log x_{mi} - \frac{\alpha_{im}}{2} y_{im}^{2} \right)$$

$$-\sum_{i=1}^{I} \lambda_i \cdot \left(\sum_{m=1}^{M} y_{im} - C_i\right) + \sum_{m=1}^{M} \sum_{i=1}^{I} \mu_{mi} \cdot (y_{im} - x_{mi}),$$

and denote the optimal solution of the BAP problem as x^* , y^* , λ^* , and μ^* . The corresponding KKT conditions yield a set of equations (B1)-(B6), where (B3)-(B6) are identical to (A3)-(A6) of SWM but the first two sets differ:

$$(B1): x_{mi}^* = \frac{p_{mi}}{\mu_{mi}^*}, \quad (B2): y_{im}^* = \frac{\mu_{mi}^* - \lambda_i^*}{\alpha_{im}}$$
 (10)

for $\forall m \in \mathcal{M}, \forall i \in \mathcal{I}$. It is important to note that (10) defines the **allocation rule** of our proposed mechanism.

Comparing equations (A1) - (A6) and (B1) - (B6), we observe that if the BSs and the APs submit the following bids:

$$p_{mi} = x_{mi}^* \cdot \frac{\partial J_m(\boldsymbol{x}_m^*)}{\partial x_{mi}}, \quad \alpha_{im} = \frac{1}{y_{im}^*} \cdot \frac{\partial V_i(\boldsymbol{y}_i^*)}{\partial y_{im}}, \quad (11)$$

then the previously described two-stage scheme yields a solution identical to the one of problem SWM in a single iteration, i.e., $x^{\circ} \triangleq x^*$ and $y^{\circ} \triangleq y^*$. Clearly, the task of the market designer here is to derive the proper payment and reimbursement rules that will induce the players, i.e., the BSs and APs, to bid according to (11).

We now look at the bidders' behavior in the first stage. Let $h_m(\boldsymbol{x}_m)$ denote the BS m's payment to the broker for the service it receives from the APs (vector \boldsymbol{x}_m). Similarly, let $l_i(\boldsymbol{y}_i)$ denote the AP i's reimbursement from the broker for the data it offloads (vector \boldsymbol{y}_i). Clearly, the payments and reimbursements depend on the respective bids through the auction allocation rule. The bidders are rational, price-taking entities and optimize their bids by maximizing their utility and cost functions.

Specifically, given the allocation rule defined in (B1) and (B2), BSs and APs solve their own optimization problems in order to find their optimal bids. Each BS $m \in \mathcal{M}$ finds the optimal bid vector p_m^* by solving the following problem:⁶

BSP:
$$\max_{\boldsymbol{p}_m \succ \mathbf{0}} \left(J_m(\boldsymbol{x}_m) - h_m(\boldsymbol{x}_m) \right),$$
 (12)

which yields the following optimality conditions:

$$\frac{\partial J_m(\boldsymbol{x}_m)}{\partial x_{mi}} = \mu_{mi} \frac{\partial h_m(\boldsymbol{x}_m)}{\partial p_{mi}}, \quad \forall i \in \mathcal{I}.$$
 (13)

Similarly, each AP $i \in \mathcal{I}$ finds the optimal bid vector α_i^* by solving the following problem:

APP:
$$\max_{\boldsymbol{\alpha}_i \succ \mathbf{0}} \left(-V_i(\boldsymbol{y}_i) + l_i(\boldsymbol{y}_i) \right),$$
 (14)

⁶Here $p_m \succeq \mathbf{0}$ means p_m is a non-negative vector, i.e., $p_{mi} \geq 0, \forall i \in \mathcal{I}$.

which yields the following optimality conditions:

$$\frac{\partial V_i(\boldsymbol{y}_i)}{\partial y_{im}} = \frac{\alpha_{im}^2}{\lambda_i - \mu_{mi}} \frac{\partial l_i(\boldsymbol{y}_i)}{\partial \alpha_{im}}, \quad \forall m \in \mathcal{M}.$$
 (15)

Comparing the best responses of BSs, (13), and of APs, (15), with those required by the socially optimal solution, i.e. (11), we can immediately obtain the **pricing rules** that will lead to the solution of SWM (i.e., the socially optimal solution):

$$h_m(\mathbf{p}_m) = \sum_{i=1}^{I} p_{mi}, \quad l_i(\mathbf{\alpha}_i) = \sum_{m=1}^{M} \frac{(\lambda_i - \mu_{mi})^2}{a_{im}}, \quad (16)$$

where we have written the payment $h_m(p_m)$ by each BS m and the compensation $l_i(\alpha_i)$ to each AP i, as functions of the respective bids. These rules are intuitive: each BS pays exactly its bid, i.e., the amount it declared that is willing to pay. On the other hand, the reimbursement (16) can be written using (B1) as $l_i(\alpha_i) = \sum_{m=1}^M y_{im}(\lambda_i - \mu_{mi})$: each AP is reimbursed proportionally to the amount of data that it offloaded weighted by a factor indicating the congestion on its link (λ_i) and the difference between the requested and admitted traffic $(\mu_{mi},$ which, by (B1) and (B2), is equal to $\sqrt{p_{mi} \cdot \alpha_{im}}$ whenever there is no link congestion).

B. IDA Algorithm

With the proper allocation rule (10) and payment rule (16), the BSs and APs can compute the optimal bids in one round and achieve an efficient market equilibrium, if they know the complete network information (including the BSs' utility functions and APs' cost functions). However, as the BSs and APs do not have this information, there is a need for an iterative algorithm that gradually adjusts the market operation point to reach the desirable one.

The proposed scheme consists of the following consecutive steps in each iteration. The broker solves the BAP problem to determine the prices, and the BSs and APs solve their own problems to determine their bids. Notice that the BAP problem is parameterized by the bids of the buyers and sellers. Solving BAP yields the matrices x and y and the Lagrange multipliers λ and μ , which further determine the prices $h_m(\cdot)$ and $l_i(\cdot)$. Similarly, APP and BSP problems are parameterized by the Lagrange multipliers of the BAP problem. We exploit the decomposable structure of BAP, and solve it by employing a primal-dual Lagrange decomposition method [31]. This enables parallel and fast execution of the algorithm.

The scheme is described in details in Algorithm 1. First, the broker initializes the primal variables (i.e., \boldsymbol{x} and \boldsymbol{y}) and dual variables (i.e., $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$), and announces the latter to the BSs and APs (lines 1-3). Any set of initial values that satisfy the complementary slackness constraints (A3, A5) is suitable, for example $y_{im}^{(0)} = C_i/M$ and $x_{mi}^{(0)} = y_{im}^{(0)}$ for any $\lambda_i^{(0)}$, $\mu_{mi}^{(0)}$. Accordingly, the BSs and APs calculate their optimal bids by solving their respective optimization problems (lines 7-8). The broker collects the new bids and checks the termination condition (line 9). Termination happens when the updated bids are equal to the bids that were submitted in the previous

round⁷. Accordingly, the broker finds the new values for \boldsymbol{x} and \boldsymbol{y} using the allocation rule of problem BAP (10) (line 10), and calculates the payments. Next, the new values of the primal variables are used for the update, through a gradient descent method, of the dual variables $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, (line 12). The new dual variables are announced to the bidders which will use them to update their bids in the next step, in case the algorithm has not converged.

```
Algorithm 1: Iterative Double Auction (IDA)
     output: x^*, y^*, \lambda^*, \mu^*
 1 t \leftarrow 0;
2 Initialize x_{mi}^{(0)}, y_{im}^{(0)}, \mu_{mi}^{(0)}, \lambda_i^{(0)}, \forall m \in \mathcal{M}, i \in \mathcal{I};
3 Announce \mu_{mi}^{(0)} and \lambda_i^{(0)}, \forall m \in \mathcal{M}, i \in \mathcal{I};
4 conv_flag \leftarrow 0;
 5 while conv_flag = 0 do
             t \leftarrow t + 1;
             Each BS m computes the optimal bids p_m^{(t)} by (12);
             Each AP i computes the optimal bids \alpha_i^{(t)} by (14);
 8
             The broker collects all bids and checks termination:
              \begin{array}{l} \textbf{if} \ |p_{mi}^{(t)} - p_{mi}^{(t-1)}| < \epsilon \ \text{and} \ |\alpha_{im}^{(t)} - \alpha_{im}^{(t-1)}| < \epsilon, \\ \forall m \in \mathcal{M}, i \in \mathcal{I} \ \textbf{then} \end{array} 
              The broker computes the new x^{(t)}, y^{(t)} by (10);
10
             The broker computes h_m(\boldsymbol{x}_m^{(t)}) and l_i(\boldsymbol{y}_i^{(t)}), \forall m, i; The broker uses gradient update for dual vars:
11
               \lambda_i^{(t+1)} = \left(\lambda_i^{(t)} + s_t \cdot \left(\sum_{m=1}^M y_{im}^{(t)} - C_i\right)\right)^+, \forall i \in \mathcal{I},
                \mu_{mi}^{(t+1)} = \mu_{mi}^{(t)} - s_t \cdot (y_{im}^{(t)} - x_{mi}^{(t)}), \forall m \in \mathcal{M}, i \in \mathcal{I},
```

C. Convergence Analysis of IDA Mechanism

end

Algorithm 1 converges to the optimal solution of problem SWM under some mild conditions, since the SWP problem has a strictly concave objective fucntion. We assume that the time slot of the update is very small (or equivalently, the step size is very small), hence we approximate the algorithm with its continuous-time counter-part. Specifically, we consider that the Lagrange multipliers are updated according to the following differential equations (based on the gradient updates):

where s_t is a properly selected step size [31];

The broker announces $\mu^{(t+1)}$ and $\lambda^{(t+1)}$;

$$\frac{\mathrm{d}\lambda_i}{\mathrm{d}t} = \left(\sum_{m=1}^M y_{im} - C_i\right)^+, \quad \frac{\mathrm{d}\mu_{mi}}{\mathrm{d}t} = x_{mi} - y_{im},$$

where $(\cdot)^+$ is the projection onto the nonnegative orthant.

We prove the convergence of our two-sided algorithm following the rationale of the proof in [32, Ch. 22], which

 $^{^7 \}rm{In}$ order to facilitate the numerical analysis, instead of strict equality we check whether their difference is smaller than $\epsilon>0$ which is a very small number.

was used to prove the one-side version of our algorithm in [13]. Specifically, we define the following Lyapunov function:

$$Z(\lambda, \mu) = \sum_{i=1}^{I} \frac{(\lambda_i - \lambda_i^*)^2}{2} + \sum_{m=1}^{M} \sum_{i=1}^{I} \frac{(\mu_{mi} - \mu_{mi}^*)^2}{2}.$$
 (17)

and we prove that it is $dZ(\lambda, \mu)/dt \leq 0$.

By applying the chain rule, we obtain

$$\frac{\mathrm{d}Z(\boldsymbol{\lambda},\boldsymbol{\mu})}{\mathrm{d}t} = \sum_{i=1}^{I} (\lambda_i - \lambda_i^*) \frac{\mathrm{d}\lambda_i}{\mathrm{d}t} + \sum_{m=1}^{M} \sum_{i=1}^{I} (\mu_{mi} - \mu_{mi}^*) \frac{\mathrm{d}\mu_{mi}}{\mathrm{d}t}$$

which can be written as

$$\frac{dZ(\lambda, \mu)}{dt} = \sum_{i=1}^{I} (\lambda_i - \lambda_i^*) \cdot \left(\sum_{m=1}^{M} y_{im} - C_i\right)^+ + \sum_{m=1}^{M} \sum_{i=1}^{I} (\mu_{mi} - \mu_{mi}^*) \cdot (x_{mi} - y_{im}),$$

or

$$\frac{dZ(\lambda, \mu)}{dt} \le \sum_{i=1}^{I} (\lambda_i - \lambda_i^*) \cdot \left(\sum_{m=1}^{M} y_{im} - C_i \right) + \sum_{m=1}^{M} \sum_{i=1}^{I} (\mu_{mi} - \mu_{mi}^*) \cdot (x_{mi} - y_{im}).$$

After some simple algebraic manipulations, we get:

$$\frac{\mathrm{d}Z(\lambda,\mu)}{\mathrm{d}t} \leq \sum_{i=1}^{I} (\lambda_i - \lambda_i^*) \left(\sum_{m=1}^{M} y_{im} - \sum_{m=1}^{M} y_{im}^* \right)
+ \sum_{m=1}^{M} \sum_{i=1}^{I} (\mu_{mi} - \mu_{mi}^*) (x_{mi} - y_{im} - x_{mi}^* + y_{im}^*)
+ \sum_{i=1}^{I} (\lambda_i - \lambda_i^*) \left(\sum_{m=1}^{M} y_{im}^* - C_i \right)
+ \sum_{m=1}^{M} \sum_{i=1}^{I} (\mu_{mi} - \mu_{mi}^*) (x_{mi}^* - y_{im}^*).$$

Using the complementary slackness and (A1), (A2), we have:

$$\frac{\mathrm{d}Z(\boldsymbol{\lambda},\boldsymbol{\mu})}{\mathrm{d}t} \leq \sum_{m=1}^{M} \sum_{i=1}^{I} \left[(y_{im} - y_{im}^*) \left(\frac{\partial V_i(\boldsymbol{y}_i^*)}{\partial y_{im}} - \frac{\partial V_i(\boldsymbol{y}_i)}{\partial y_{im}} \right) + (x_{mi} - x_{mi}^*) \left(\frac{\partial J_m(\boldsymbol{x}_m)}{\partial x_{mi}} - \frac{\partial J_m(\boldsymbol{x}_m^*)}{\partial x_{mi}} \right) \right].$$

The above inequality is satisfied due to the following property that holds for each concave function $f(\cdot)$ [31]:

$$f(y) \le f(x) + \nabla f(x)^T (y - x). \tag{18}$$

D. Properties of IDA Mechanism

From the analysis above, it is clear that the IDA algorithm induces the BSs and the APs to bid truthfully, (12) and (14), and according to the socially optimal bids, (11). In the sequel, we prove that the algorithm is also weakly budget balanced and individually rational.

Lemma 1. The IDA mechanism is weakly budget balanced.

Proof: The budget balance $\Lambda(\cdot)$ is defined as:

$$\Lambda(\boldsymbol{p}, \boldsymbol{\alpha}) = \sum_{m=1}^{M} h_m(\boldsymbol{p}_m) - \sum_{i=1}^{I} l_i(\boldsymbol{\alpha}_i),$$

or, if we use (16), we get

$$\Lambda(\boldsymbol{p}, \boldsymbol{\alpha}) = \sum_{m=1}^{M} \sum_{i=1}^{I} p_{mi} - \sum_{m=1}^{M} \sum_{i=1}^{I} \frac{(\lambda_i - \mu_{mi})^2}{a_{im}}.$$

Notice that (10) is satisfied at the equilibrium. Hence,

$$\Lambda(\boldsymbol{p}, \boldsymbol{\alpha}) = \sum_{m=1}^{M} \sum_{i=1}^{I} x_{mi}^* \cdot \lambda_i^* \ge 0.$$

Additionally, the IDA mechanism ensures the voluntarily participation of the bidders since they are guaranteed to have at least zero net utility for all the possible market outcomes.

Lemma 2. The IDA mechanism is individually rational (IR).

Proof: For each BS $m \in \mathcal{M}$, the IR condition can be translated to the following constraint:

$$J_m(\boldsymbol{x}_m^*) - \sum_{i=1}^{I} p_{mi} \ge 0$$
, or $J_m(\boldsymbol{x}_m^*) - \sum_{i=1}^{I} x_{mi}^* \mu_{mi}^* \ge 0$,

which can be written, using (13), as

$$J_m(\boldsymbol{x}_m^*) \ge \sum_{i=1}^{I} x_{mi}^* \cdot \frac{\partial J_m(\boldsymbol{x}_m^*)}{\partial x_{mi}}.$$
 (19)

Since $J_m(\cdot)$ is strictly concave and $J_m(\mathbf{0}) = 0$, the inequality (19) is always satisfied according to the condition (18).

Similarly, for each AP $i \in \mathcal{I}$, the IR condition

$$-V_i(\boldsymbol{y}_i^*) + l_i(\boldsymbol{\alpha}_i) \ge 0 \tag{20}$$

is satisfied according to the property (18).

V. SIMULATIONS

In this section we provide numerical results to validate our theoretical analysis. We consider a small market with M=5 BSs and I=5 APs. The study can be easily extended to a larger market. The BSs' utility functions $J_m(\cdot)$, $\forall m \in \mathcal{M}$, and APs' cost functions $V_i(\cdot)$, $\forall i \in \mathcal{I}$, are:

$$J_m(\boldsymbol{x}_m) = 10 \cdot \sum_{i=1}^{5} \log(\theta_{mi} x_{mi}), \ V_i(\boldsymbol{y}_i) = 0.1 \cdot \sum_{m=1}^{5} e^{\rho_{im} y_{im}}$$

where $\theta_{mi} \geq 0$ represents the offloading efficiency of AP i for BS m (i.e., the offloading benefit is AP-specific) and parameter $\rho_{im} \geq 0$ captures the fact that each AP may incur different cost by serving a different BS. The capacity of each AP is C=15 Mbps.

Parameters ρ_{im} and θ_{mi} , $\forall\,m\in\mathcal{M},\,\forall\,i\in\mathcal{I}$, were chosen with uniform independent probabilities from interval [0.5,1]. For example, we used $\theta_{11}=0.81$ and $\theta_{21}=0.59$, which means that BS 1 benefits more from service of AP 1 than BS 2 does. Indeed, the mechanism for these system parameters yields $y_{11}=3.33$ and $y_{12}=2.87$. Notice also that, although $\theta_{51}=0.54<\theta_{11}$, BS 5 offloads more data to AP 1, i.e. $y_{15}=3.59$. This is due to the smaller cost that BS 5 induces to AP 1 since the respective cost parameter is $\rho_{15}=0.61$ while

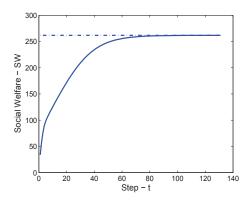


Fig. 3. Evolution of social welfare produced by the IDA.

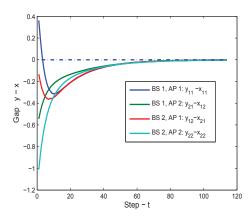


Fig. 4. Evolution of the gap between requested (x) and admitted data (y). $\rho_{11} = 0.74$. Finally, the payments of the BSs 1, 2 and 5 are

 $p_{11} = 7.3$, $p_{21} = 6.29$, and $p_{51} = 6.63$, respectively. Notice that BS 5 pays less than BS 1 although it offloads more data than the latter.

In Fig. 3, we plot the social welfare achieved by the algorithm in each iteration and we observe that it gradually converges to the optimal one, i.e. to the solution of SWM (dotted line). In Fig. 4, we present the convergence of x and y. Specifically, we plot the gaps between x and y for 4 BS-AP pairs, and observe that the gaps gradually converge to zero. These two figures imply that our IDA algorithm elicits the true hidden information (i.e., the utility and cost functions), and converges to the socially optimal solution.

VI. DISCUSSION AND CONCLUSIONS

In this paper, we considered a market where MNOs lease third-party owned WiFi or femtocell APs to offload their mobile data traffic. This is a promising solution for increasing the user perceived network capacity in a dynamic and scalable fashion, with low CAPEX and OPEX costs. Today, the technologies to implement such solutions are already in place (e.g., secure offloading methods). Data offloading can alleviate congestion of 2G/3G cellular networks, and also serve as a low-cost auxiliary technology for the emerging 4G networks.

We proposed an iterative double auction mechanism, which satisfies the desirable economic properties, and maximizes the welfare of the market, under the assumption of price-

taking bidders. There are very interesting directions for future work. First, one can study what is the impact of strategic, price-anticipating behavior in the market outcome. Similarly, it is challenging to study how colluding behaviors will affect the algorithm. Also, it is important to consider practical implementation issues such as how to hardwire this algorithm to APs and BSs so as to communicate with the broker and execute IDA algorithm in a real-time fashion.

REFERENCES

- [1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update", Feb. 2012.
- Bloomberg: "AT&T Pays 2 Billions to Qualcom for Spectrum", 2009.
- FemtoForum, "Femtocell Business Case Whitepaper", 2009
- AT&T Press Release, "AT&T Expands Wi-Fi Hotzone Pilot Project to Additional Cities", July, 2010.
- [5] BT Wifi Press Release, "O2 Brings 3,000 New Wi-Fi Hotspots to iPhone Customers", July, 2008.
- Cisco, "Making Wi-Fi as Secure and Easy to Use as Cellular", 2012.
- Republic Wireless, http://www.republicwireless.com/ Nov. 2012.
- Spectrum bridge, http://www.spectrumbridge.com/.
- R. B. Myerson, and M. A. Satterthwaite, "Efficient Mechanisms for Bilateral Trading", Journal of Econ. Theory, vol. 29, no. 2, 1983.
- [10] R. P. McAfee, "A Dominant Strategy Double Auction", Journal of Economic Theory, vol. 56, no. 2, pp. 434-450, 1992
- [11] P. Maille, B. Tuffin, "Why VCG Auctions can Hardly be Applied to the Pricing of Interdomain and Ad hoc Networks", in Proc. of NGI, 2007.
- [12] D. P. Palomar, M. Chiang, "A Tutorial on Decomposition Methods for Network Utility Maximization ", IEEE JSAC, vol. 24, no. 8, pp. 1439-1451, 2006.
- [13] F. P. Kelly, A. Maulloo, and D. Tan, "Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability", Journal of Oper. Res. Society, vol. 29, no. 3, pp. 237-252, 1998. [14] L. Johansen, "Price-Taking Behavior", Econometrica, vol. 4, no. 7, 1977.
- [15] K. Lee, I. Rhee, J. lee, S. Chong, and Y. Yi, "Mobile Data Offloading: How Much Can WiFi Deliver?", In Proc. of ACM CoNEXT, 2010.
- [16] N. Ristanovic, J. Y. Le Boudec, A. Chaintreau, and V. Erramilli, "Energy Efficient Offloading of 3G Networks", In Proc. of IEEE MASS, 2011.
- [17] Informa Telecoms & Media, "Femtocell Market Status", Feb. 2011.
- [18] M. H. Cheung, and J. Huang, "Optimal Delayed Wi-Fi Offloading", in Proc. of WiOpt, 2013.
- [19] J. G. Andrews, et al., "Femtocells: Past, Present, and Future", IEEE JSAC, vol. 30, no. 3, pp. 497-508, 2012.
- [20] H. S. Jo, P. Xia, and J. G. Andrews, "Downlink Femtocell Networks: Open or Closed?", In Proc. of IEEE ICC, 2011.
- [21] S. Y. Yun, Y. Yi, D. H. Cho, and J. Mo, "The Economic Effects of Sharing Femtocells", IEEE JSAC, vol. 30, no. 3, 2012.
- [22] Y. Chen, J. Zhang, Q. Zhang, and J. Jia, "A Reverse Auction Framework for Access Permission Transaction to Promote Hybrid Access in Femtocell Network", in Proc. of IEEE Infocom, 2012.
- [23] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Spectrum Leasing as an Incentive towards Uplink Macrocell and Femtocell Cooperation", IEEE JSAC, vol. 30, no.3, pp. 617-630, 2012.
- [24] L. Gao, G. Iosifidis, J. Huang, and L. Tassiulas, "Economics of Mobile Data Offloading", in Proc. of IEEE SDP Workshop, 2013.
- [25] S. Hua, X. Zhuo, and S. Panwar, "A Truthful Auction Based Incentive Framework for Femtocell Access", Arxiv, http://arxiv.org/abs/1210.5012.
- [26] H. Xu, J. Jin, and B. Li, "A Secondary Market for Spectrum", in Proc. of IEEE Infocom, 2010.
- [27] D. Yang, X. Fang, and G. Xue, "Truthful Auction for Cooperative Communications", in Proc. of ACM Mobihoc, 2011.
- [28] G. Iosifidis, and I. Koutsopoulos, "Double Auction Mechanisms for Resource Allocation in Autonomous Networks", IEEE JSAC, vol.28, no.1, pp.95-102, Jan. 2010.
- C. Courcoubetis, and R. Weber, "Pricing Communication Networks: Economics, Technology and Modelling", Willey Press, 2003.
- [30] L. Duan, J. Huang, and B. Shou, "Economics of Femtocell Service Provision", IEEE TMC, forthcoming, 2013.
- [31] D. Bertsekas, "Nonlinear Programming", Athena Scientific, 1999.
- N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, "Algorithmic Game Theory", Cambridge University Press, 2007.