A Double Auction Mechanism for Mobile Data Offloading Markets

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

Abstract—The unprecedented growth of mobile data traffic challenges the performance and economic viability of today's cellular networks, and calls for novel network architectures and communication solutions. Mobile data offloading through third-party Wi-Fi or femtocell access points (APs) can significantly alleviate the cellular congestion, and enhance user quality of service (QoS), without requiring costly and time-consuming infrastructure investments. This solution has substantial benefits both for the mobile network operators (MNOs) and the mobile users, but comes with unique technical and economic challenges that must be jointly addressed. In this paper, we consider a market where MNOs lease access points (APs) that are already deployed by residential users for the offloading purpose. 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 by maximizing the differences between the MNOs' offloading benefits and APs' offloading costs. The proposed scheme takes into account the particular characteristics of the wireless network, such as the coupling of MNOs' offloading decisions and APs' capacity constraints. Additionally, it does not require full information about the MNOs and APs, and creates non-negative revenue for the market broker.

Index Terms—Mobile Data Offloading, Network Economics, Auctions, Network Optimization

I. INTRODUCTION

The recent growth of mobile data traffic has been unprecedented. According to several recent industry reports [2], [3], the volume of global mobile data is expected to increase by a 13-fold between year 2012 and 2017. These developments pose new challenges to mobile network operators (MNO), who have to significantly enhance their infrastructure in order to be capable to accommodate all this traffic. 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 [4] and time-consuming. More importantly, they are expected to be outpaced in less than 4 years by the continuing traffic increase [2], [3]. Clearly, operators must find novel methods to address this problem, and *mobile data offloading* appears as one of the most attractive solutions.

A. Motivation

Mobile data offloading refers to delivering mobile data traffic that is originally targeted for macrocellular network, over other networks such as Wi-Fi or femtocell networks

George Iosifidis and Leandros Tassiulas are with University of Thessaly, Electrical and Computer Engineering Department, and CERTH, Greece. Lin Gao and Jianwei Huang are with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. Parts of this work appeared in the proceedings of 11th WiOpt Conference [1].
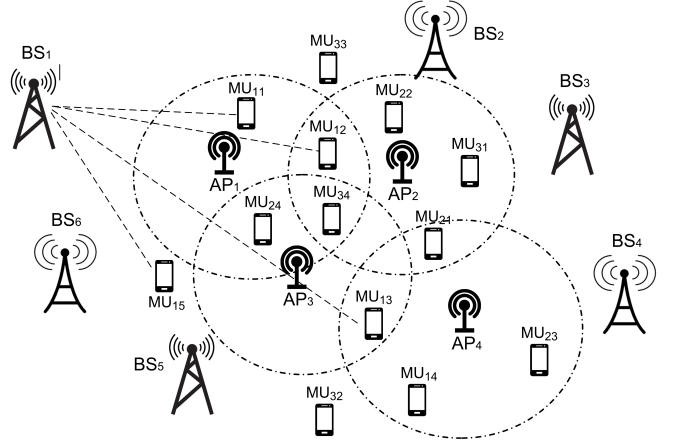

Fig. 1. Different mobile network operators (MNOs) own different sets of base stations (BSs). MU_{kj} denotes the jth mobile user (MU) who is a subscriber of MNO k. An MU may be covered by multiple access points (APs) and each AP may serve more than one MU.

[5]. Nowadays, there is consensus that data offloading is a cost-effective 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 Wi-Fi access points (APs) to complement their traditional macrocellular network [6], or initiated collaborations with existing Wi-Fi networks [7]. This approach is facilitated by recent technological advances such as the HotSpot 2.0 protocol which addresses related security issues [8], and mechanisms that enable seamless Wi-Fi where hand-off from 3G/4G networks is transparent (even for voice calls) [9].

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 AP deployment by the MNOs is very expensive and even impractical in some cases (e.g., due to site acquisition issues [10]). A promising method to overcome this obstacle and ensure the high availability of APs is the *employment* (leasing) of third party Wi-Fi and femto APs which are already installed in homes or other venues (e.g., companies and stores). This strategy will allow operators to handle mobile data traffic with reduced capital and operational expenditures (CAPEX and OPEX). Clearly, this constitutes a method of *outsourcing* part of the network operation to the third-party APs, achieving a flexible and dynamic increase of network capacity on demand and on the spot, i.e., whenever and wherever is required.

However, AP owners are expected to ask for (monetary) compensation, since admitting macrocellular traffic will consume APs' limited wireless resources and broadband connec-

tion capacities for their own (internal) traffic demand. Hence, in order to enable effective third-party offloading, one needs to answer the following questions: *how much traffic should each AP admit for each MNO and how much to charge*? or, from the perspective of the MNOs: *how much traffic should each MNO offload to each AP and how much to pay?* In this paper, we design a mechanism that answers these questions, while addressing the particular challenges of the wireless data offloading.

Specifically, we design 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 [11]. Wi-Fi and femtocell AP owners offer their services (i.e., offloading traffic for the MNOs) in exchange of reimbursements. As shown in Figure 1, we look at the general case where each AP can serve more than one MNO, and each MNO owns several base stations and 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 MNOs and the APs comply with the broker's decisions only if it is of their own interest to do so.

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 asymmetric 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 truthfully reveal their needs. With this information, the broker tries to maximize the efficiency of the market by properly matching the buyers and the sellers. The broker will be rewarded based on the volume of transactions (such as a broker in a stock market), and hence does not have an incentive to deviate from the desirable market goal imposed by the system designer, i.e., maximizing the market efficiency. 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 [12]. They can be inefficient and applicable only to certain simplified settings, e.g., for bidders with single-unit demands (McAfee auction [13]), or they can be budget imbalanced with a high computational complexity (VCG auction [14]). In our case, the double auction design problem is further perplexed due to the following realistic issues that we explicitly take into account:

- (*I*1) *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. On

the other hand, an AP that is located close to the MNO's base station will be less useful.
- (*I*2) *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 heavy the MNO's load is, which in turn depends on its offloading decisions to other APs.
- (*I*3) *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.
- (*I*4) *Interference among APs*. Some APs may interfere with each other, if they are closely located and transmit in the same channel. Therefore, if they are concurrently leased by one or more MNOs, their offloading capacities will be coupled and limited by the interference.
- (*I*5) *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 performance, i.e., consider the issues $(I1) - (I5)$, we choose an alternative market design method that is based on the framework of *Network Utility Maximization* (NUM) [15].

B. Methodology and Contributions

Our starting point is the work of Kelly *et al.* [16] 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) [16] 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 costs[1].

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 [16] 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 [17], and is a reasonable assumption when each bidder is not fully aware of the decisions of other bidders and system parameters. Both aspects are valid for the setting considered in this paper.

To this end, the main technical contributions of this paper are as follows:

1) We study a general market model where multiple operators (MNOs) compete to lease multiple (possibly

---

[1]For the distributed version of [16], it can be argued that there are multiple bandwidth sellers (the links). However, the links only balance the traffic, i.e., they do not have cost functions and do not submit bids.

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 truthfully reveal their needs/demands) under the assumption of price-taking behavior.

3) The proposed scheme has a low computational complexity, generates a small communication overhead, 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 (*asymmetric market information*).

4) The introduced framework considers important realistic issues of the mobile data offloading problem, $(I1) - (I5)$, which, as we explain in details in Section VI, have been overlooked until now by the related prior studies.

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

II. SYSTEM MODEL AND PROBLEM DEFINITION

In this section, we introduce the system model that captures various unique characteristics of the wireless offloading problem. Accordingly, we define the data offloading problem as a market design problem, with the goal to maximize the social welfare for the MNOs and the AP owners.

A. System Model

We consider a system with a set $\mathcal{K} \triangleq \{1, 2, \ldots, K\}$ of mobile network operators (MNOs), where each operator $k \in \mathcal{K}$ has a set of $\mathcal{M}_k$ base stations (BSs). We denote with $\mathcal{M} \triangleq \{1, 2, \ldots, M\}$ the set of all base stations, i.e., $\mathcal{M} = \bigcup_{k \in \mathcal{K}} \mathcal{M}_k$. Also, we assume that there exists a set $\mathcal{I} \triangleq \{1, 2, \ldots, I\}$ of access points (APs). Each AP can be a Wi-Fi or a femtocell access point which operates in a separate channel, and hence does not interfere with the macrocellular network[2]. Each MNO serves a group of mobile users (MUs) which are randomly distributed within the coverage areas of its base stations, 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 our study focuses on one time period that comprises enough slots for the proposed scheme to converge to the optimal solution (as it will be explained in the sequel).

The MUs' location and traffic types may change over time but are considered fixed within each time period.

Consider the case where each operator $k \in \mathcal{K}$ would like to offload $x_{mi} \geq 0$ bytes of data through AP $i \in \mathcal{I}$ for each BS $m \in \mathcal{M}_k$. We define the *offload request* vector for BS m to all I APs as $\boldsymbol{x}_m \triangleq (x_{mi} : \forall i \in \mathcal{I})$, and total offloaded data of BS m is $X_m = \sum_{i \in \mathcal{I}} x_{mi}$. The offload requests for all BSs of each MNO k are given by the offload request matrix

$$\boldsymbol{x}_k = (\boldsymbol{x}_m : \forall m \in \mathcal{M}_k). \tag{1}$$

The MNOs' requests depend on the locations and traffic of their MUs. We use $J_m(\boldsymbol{x}_m)$, $m \in \mathcal{M}_k$, to denote the utility of MNO k when its BS m offloads traffic $\boldsymbol{x}_m$ to the APs. This utility represents the BS' (and hence the operator's) cost reduction comparing with the case that it serves $\boldsymbol{x}_m$ directly. The servicing cost, for example, can be due to the energy consumption and depends on a number of parameters related to the BS' characteristics and technology [18]. The total offloading benefit for each operator $k \in \mathcal{K}$ is

$$J_k(\boldsymbol{x}_k) = \sum_{m=1}^{M_k} J_m(\boldsymbol{x}_m). \tag{2}$$

We assume that the data service cost for each base station is a strictly convex function on the total traffic, motivated by the energy consumption cost [18], [19]. Hence the utility component functions $J_m(\cdot)$, $m \in \mathcal{M}$, are positive, increasing, and strictly concave functions of vectors $\boldsymbol{x}_m$, $m \in \mathcal{M}$. Notice that the benefit function $J_k(\cdot)$ can capture the users' offloading preferences. Namely, for users who favor offloading, we can include a positive increasing concave function component to the respective base stations' utility functions. This will capture the additional revenue that the operators can have by charging those users when they offload their traffic. On the other hand, for users who suffer when being offloaded (e.g., a fast moving user), we can add a negative convex function to the utility of the corresponding operator. This will account for the dissatisfaction of the users and hence the potential revenue losses of the MNOs.

Moreover, our model captures the following important aspects of the offloading problem. First, for each base station $m \in \mathcal{M}$, the 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 data (the vector $\boldsymbol{x}_m$). For example, an AP located at the boundary of the BS' cell is more important, since it can offload traffic 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 operator $k \in \mathcal{K}$, its offloading decisions to the different APs are coupled. Clearly, the benefit from offloading traffic to an AP depends on the load of each BS $m \in \mathcal{M}_k$, which in turn depends on the operator's offloading decisions for this BS to other APs. In other words, even if two different strategies, $\boldsymbol{x}_m$ and $\hat{\boldsymbol{x}}_m$, suggest equal amount of offloaded data to a certain AP i, $x_{mi} = \hat{x}_{mi}$, the respective utility improvements may differ:

$$J_m(x_{mi}, \boldsymbol{x}_m^{-i}) - J_m(0, \boldsymbol{x}_m^{-i}) \neq J_m(\hat{x}_{mi}, \hat{\boldsymbol{x}}_m^{-i}) - J_m(0, \hat{\boldsymbol{x}}_m^{-i}),$$

---

[2]Wi-Fi 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 [21]. This disjoint subchannel allocation (among APs and BSs) has improved performance especially for dense BSs - APs deployments [22].
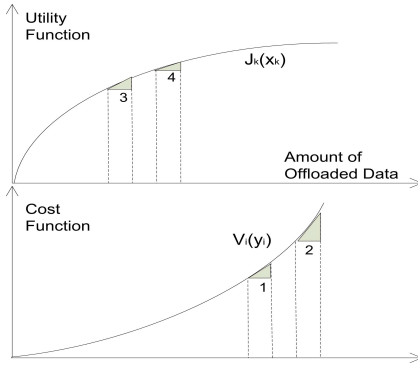
Fig. 2. Concave MNO utility function, $J_k(\cdot)$, and convex AP function, $V_i(\cdot)$. The numbered triangles have the same length (base) - corresponding to equal offloaded data increase - but different height (h): $h_2 > h_1$ and $h_4 < h_3$.

where $\boldsymbol{x}_m^{-i} \triangleq (x_{mj} : \forall j \in \mathcal{I} \setminus \{i\})$, and $\hat{\boldsymbol{x}}_m^{-i} \triangleq (\hat{x}_{mj} : \forall j \in \mathcal{I} \setminus \{i\})$.

Each AP $i \in \mathcal{I}$ responds to offloading requests and admits $y_{im} \geq 0$ bytes in one time period from each BS $m \in \mathcal{M}_k$ of every operator $k \in \mathcal{K}$. We define the *admitted traffic* vector $\boldsymbol{y}_i \triangleq (y_{im} : \forall m \in \mathcal{M})$. Clearly, y_{im} depends on the respective request x_{mi}. We use $V_i(\boldsymbol{y}_i)$ to denote the cost incurred by AP i for serving the mobile operators, which is a positive, increasing, and strictly convex function in vector $\boldsymbol{y}_i$. 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 [18], [19], [20], as illustrated in Figure 2. Notice 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.

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). Similarly, an AP's incurred cost for admitting traffic for a certain BS of an MNO, depends on how loaded the AP already is, i.e., how much traffic offloaded for other BSs. The maximum amount that AP $i \in \mathcal{I}$ can offload is constrained by its capacity C_i bps, and is also affected by the interference that is induced from other closely located APs that transmit in the same channel.

In order to model the inter-AP interference, we use the *interference protocol model* [23], according to which a transmission over a link in a given channel is successful if all the interfering nodes of the receiver do not transmit in this channel. Notice that for Wi-Fi APs that employ the RTS-CTS mechanism, the same condition should also hold for the transmitter, since it needs to receive acknowledgements from the receiver. Here, we do not consider this transmitter side constraint so as to have a unified treatment of femtocell and Wi-Fi APs. The extension, however, is straightforward.

Clearly, to model the interference more accurately, we need to take into account the locations of the specific users that each AP serves. Given that the APs coverage areas are relatively small, and in order to keep our analysis tractable, we employ a more general approach. Namely, for each pair of APs $i, j \in$

$\mathcal{I}$, $i \neq j$, we define the interference parameter $\gamma_{ij} \in [0, 1]$ which captures the average interference that AP i induces to AP j. We assume that $\gamma_{ij} = \gamma_{ji}$ and $\gamma_{ii} = 1$. When two APs are in distant locations, then $\gamma_{ij} = 0$. Such interference parameterization only takes effect when two APs transmit in the same channel.

Note that our proposed model generalizes the protocol interference model, which is a special case of our scheme with the additional restriction of $\gamma_{ij} \in \{0, 1\}$, i.e., two APs either interfere completely or not. Our more general model allows two interfering APs to transmit concurrently in the same channel, but their transmission rates are reduced due to the decrease of their respective SINRs caused by each other's transmission.

The values of the γ_{ij} parameters, $\forall i, j \in \mathcal{I}$, are considered known and available through experiences, e.g., from measurements that have been conducted in the past, and/or can be estimated using location and channel information for the APs. Using this model, we capture the impact of interference on the offloading capacity of every AP $i \in \mathcal{I}$, with the following constraint [23]:

$$\sum_{m=1}^{M} \frac{y_{im}}{C_i} + \sum_{j \in \mathcal{I} \setminus \{i\}} \sum_{m=1}^{M} \frac{\gamma_{ij} y_{jm}}{C_j} \leq T. \qquad (3)$$

The interpretation of (3) is the following: when the APs that may induce interference to AP i (i.e., for all $j \in \mathcal{I} \setminus \{i\}$ with $\gamma_{ij} > 0$) do not transmit, then AP i can offload during a time slot of length T seconds an amount only constrained by its capacity C_i, i.e., $C_i \cdot T$ bytes. However, when an interfering AP transmits, then the effective capacity of AP i is reduced proportionally[3] to the transmissions of the interferers (y_{jm}) and the impact of the transmissions on AP i (γ_{ij}). Without loss of generality, hereafter we normalize the slot duration to be $T = 1$. The key notations of our model are summarized in Table I.

B. Problem Definition

Clearly, the objectives of the MNOs 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. Therefore, there is a need for a market controller (a broker) who will intervene and ensure that the market will operate efficiently[4]. Such an entity will undertake the task of determining the offload request matrix:

$$\boldsymbol{x} \triangleq (\boldsymbol{x}_k : \forall k \in \mathcal{K}) = (x_{mi} : \forall m \in \mathcal{M}, \forall i \in \mathcal{I}), \quad (4)$$

and the admitted traffic matrix:

$$\boldsymbol{y} \triangleq (\boldsymbol{y}_i : \forall i \in \mathcal{I}) = (y_{im} : \forall m \in \mathcal{M}, \forall i \in \mathcal{I}), \quad (5)$$

---

[3]This presumes that a proper scheduling or contention resolution mechanism exists. In the case of Wi-Fi APs, this is implemented by the MAC contention mechanism. For femtocells, we assume this is implemented through radio resource management (RRM) (scheduling) schemes such as the enhanced inter-cell interference coordination techniques (eICIC) proposed in LTE Rel. 10.

[4]Notice that such brokers are usually paid by the volume of the transactions they facilitate [11]. Therefore, they have no incentives to distort the market efficiency.

that ensure the efficient operation of the market. This is achieved when the difference between the total benefit for the MNOs and the aggregate cost for the APs is maximized.

Specifically, the broker can find the optimal $\boldsymbol{x}$ and $\boldsymbol{y}$ by solving the following social welfare maximization (SWM) problem:

$$\text{SWM}: \ \max_{\boldsymbol{x},\boldsymbol{y}} \sum_{k=1}^{K} J_k(\boldsymbol{x}_k) - \sum_{i=1}^{I} V_i(\boldsymbol{y}_i), \quad (6)$$

$s.t.$

$$\sum_{m=1}^{M} \frac{y_{im}}{C_i} + \sum_{j\in\mathcal{I}\backslash\{i\}} \sum_{m=1}^{M} \frac{\gamma_{ij} y_{jm}}{C_j} \leq 1, \ \forall i \in \mathcal{I}, (7)$$

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

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

where constraints (8) indicate that the amount of offloaded data that each AP decides to admit, should satisfy the respective amount requested by the MNOs. It is easy to see that at the equilibrium it will hold $y_{im} = x_{mi}, \ \forall m, i$. It is also important to notice that with this inequality constraint, we ensure that the requested amount of data will not exceed the available capacity of each AP. In other words, due to (7) and (8), we implicitly impose the constraint $x_{mi} \leq C_i, \ \forall m \in \mathcal{M}, \ \forall i \in \mathcal{I}$. 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 characterized using the necessary and sufficient Karush-Kuhn-Tucker (KKT) conditions derived next [24].

Specifically, we relax constraints (7) and (8) and define the Lagrangian of the SWM problem as follows:

$$\begin{aligned} L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{x}, \boldsymbol{y}) &= \sum_{k=1}^{K} \sum_{m=1}^{M_k} J_m(\boldsymbol{x}_m) - \sum_{i=1}^{I} V_i(\boldsymbol{y}_i) \\ &- \sum_{i=1}^{I} \lambda_i \cdot \Big(\sum_{j=1}^{M} \sum_{m=1}^{M} \frac{\gamma_{ij} y_{jm}}{C_j} - 1 \Big) \\ &- \sum_{m=1}^{M} \sum_{i=1}^{I} \mu_{mi} \cdot (x_{mi} - y_{im}), \quad (10) \end{aligned}$$

where $\boldsymbol{\lambda} \triangleq (\lambda_i \geq 0 : \forall i \in \mathcal{I})$ is the vector of Lagrange multipliers corresponding to constraints (7), and $\boldsymbol{\mu} \triangleq (\mu_{mi} \geq 0 : \forall m \in \mathcal{M}, \forall i \in \mathcal{I})$ is the matrix of Lagrange multipliers for constraints (8). Notice also that we have used (2), and have written (7) in a compact form, i.e., included the first term in the double summation term for $j = i$. The KKT conditions that yield the optimal dual variables, $\boldsymbol{\lambda}^\circ$, $\boldsymbol{\mu}^\circ$, and optimal primal variables $\boldsymbol{x}^\circ$, $\boldsymbol{y}^\circ$, for the SWM problem are given by the following set of equations $\forall k \in \mathcal{K}, \ m \in \mathcal{M}_k, \ \forall i \in \mathcal{I}$,

$$(A1): \frac{\partial J_m(\boldsymbol{x}_m^\circ)}{\partial x_{mi}} = \mu_{mi}^\circ, \ (A2): \frac{\partial V_i(\boldsymbol{y}_i^\circ)}{\partial y_{im}} = \mu_{mi}^\circ - \sum_{j=1}^{I} \frac{\gamma_{ji} \lambda_j^\circ}{C_i},$$

$$(A3): \lambda_i^\circ \cdot \Big(\sum_{m=1}^{M} \sum_{j=1}^{I} \frac{\gamma_{ij} y_{jm}^\circ}{C_j} - 1 \Big) = 0, \ (A4): y_{im}^\circ \geq x_{mi}^\circ,$$

$$(A5): \mu_{mi}^\circ \cdot (x_{mi}^\circ - y_{im}^\circ) = 0, \ (A6): x_{mi}^\circ, y_{im}^\circ, \lambda_i^\circ, \mu_{mi}^\circ \geq 0.$$

However, it is infeasible for the broker to derive the optimal SWM solution through solving $(A1)$-$(A6)$ directly, due to the

Symbol	Physical Meaning
x_{mi}	Offload request of BS $m \in \mathcal{M}$ from AP $i \in \mathcal{I}$ (bytes)
y_{im}	Admitted data of BS $m \in \mathcal{M}$ from AP $i \in \mathcal{I}$ (bytes)
$\boldsymbol{x}_k$	Offload request matrix of operator $k \in \mathcal{K}$
$\boldsymbol{y}_i$	Admitted data vector of AP $i \in \mathcal{I}$
$J_k(\cdot)$	Utility function for operator $k \in \mathcal{K}$
$V_i(\cdot)$	Cost function for AP $i \in \mathcal{I}$
γ_{ij}	Interference induced to AP j from AP i
$h_k(\cdot)$	Total Payment from MNO $k \in \mathcal{K}$
$h_m(\cdot)$	Payment from MNO $k \in \mathcal{K}$ for its BS $m \in \mathcal{M}_k$
$l_i(\cdot)$	Payment to AP $i \in \mathcal{I}$

limited information he has for the market. Namely, we assume that the broker is unaware of the utility and cost functions $J_k(\cdot), \ \forall k \in \mathcal{K}$, and $V_i(\cdot), \ \forall i \in \mathcal{I}$. Therefore, he has to use a mechanism (a double auction in this paper) to elicit this hidden information. Such a mechanism should ideally have the following economic properties:

- $(E1)$ *Efficiency*. Maximizes the social welfare, i.e., achieves the optimal solution of the SWM problem.
- $(E2)$ *Incentive compatibility*. The bidders are induced to truthfully reveal (directly or indirectly) their private information.
- $(E3)$ *Individually rationality*. The MNOs and the APs should not get worse and get negative payoffs by participating, since they can always choose not to participate and get zero payoffs.
- $(E4)$ *(Weakly) budget balance*. The aggregate payments from the buyers to the broker should be equal or larger than the total payment of the broker to the sellers. That is, the broker does not need to inject additional money into the market to make the mechanism work.

Nevertheless, a fundamental result in mechanism design is that there does not exist a double auction that possesses all theses properties [12].

To overcome this difficulty, we take here 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 an infinitesimal impact on the market prices.

III. THE IDA MECHANISM

In this section, we introduce an Iterative Double Auction (IDA) mechanism for price-taking bidders. We prove that it solves the mobile data offloading problem, taking into account the technical issues $(I1)-(I5)$, and at the same time possesses the desirable economic properties $(E1)$-$(E4)$ under the assumption of price-taking behaviors.

A. IDA Resource Allocation and Pricing Rules

The basic idea of this mechanism is that the broker solves an optimization problem (different from the SWM problem) to determine $\boldsymbol{x}$ and $\boldsymbol{y}$, which can be combined with the proper pricing (for the MNOs) and reimbursement (for the APs) rules

in order to ensure the optimal solution of SWM. This scheme corresponds to a double auction where many buyers (MNOs) and many sellers (APs) interact, facilitated by the broker, in an iterative fashion and adjust their bids until the market reaches an efficient point, i.e., the market clearing solution.

The Broker. The mechanism comprises two stages in each iteration. In the first stage, each MNO $k \in \mathcal{K}$ submits a bid $p_{mi} \geq 0$ for each one of its BSs $m \in \mathcal{M}_k$, and for each[5] AP $i \in \mathcal{I}$. Similarly, each AP $i \in \mathcal{I}$ submits a bid $\alpha_{im} \geq 0$ for every[6] BS $m \in \mathcal{M}_k$ of each MNO $k \in \mathcal{K}$. We define the bid matrix and the bid vector for each MNO $k \in \mathcal{K}$ and each AP $i \in \mathcal{I}$, respectively, as follows

$$\boldsymbol{p}_k \triangleq (p_{mi} : \forall m \in \mathcal{M}_k, \forall i \in \mathcal{I}),$$
$$\boldsymbol{\alpha}_i \triangleq (\alpha_{im} : \forall m \in \mathcal{M}). \quad (11)$$

These bids signal the offloading needs and serving costs for the MNOs 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 MNOs' 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):

$$\text{BAP}: \max_{\boldsymbol{x},\boldsymbol{y}} \sum_{m=1}^{M} \sum_{i=1}^{I} \left(p_{mi} \log x_{mi} - \frac{\alpha_{im}}{2} y_{im}^2 \right) \quad (12)$$
$$s.t.$$
$$\sum_{m=1}^{M} \frac{y_{im}}{C_i} + \sum_{j \in \mathcal{I} \setminus \{i\}} \sum_{m=1}^{M} \frac{\gamma_{ij} y_{jm}}{C_j} \leq 1, \forall i \in \mathcal{I}, \quad (13)$$
$$y_{im} \geq x_{mi}, \forall m \in \mathcal{M}, \forall i \in \mathcal{I}, \quad (14)$$
$$x_{mi} \geq 0, y_{im} \geq 0, \forall m \in \mathcal{M}, \forall i \in \mathcal{I}. \quad (15)$$

Notice that the objective function is motivated by the allocation rule of [16], with the additional convex component capturing the (convex) increasing cost functions of the APs.

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

$$\tilde{L}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{x}, \boldsymbol{y}) = \sum_{m=1}^{M} \sum_{i=1}^{I} \left(p_{mi} \log x_{mi} - \frac{\alpha_{im}}{2} y_{im}^2 \right)$$
$$- \sum_{i=1}^{I} \lambda_i \left(\sum_{j=1}^{I} \sum_{m=1}^{M} \frac{\gamma_{ij} y_{jm}}{C_j} - 1 \right)$$
$$- \sum_{m=1}^{M} \sum_{i=1}^{I} \mu_{mi} (x_{mi} - y_{im}), \quad (16)$$

and denote the optimal solution of the BAP problem as $\boldsymbol{x}^*$, $\boldsymbol{y}^*$, $\boldsymbol{\lambda}^*$, and $\boldsymbol{\mu}^*$. The respective KKT conditions yield a set of equations $(B1) - (B6)$, where $(B3) - (B6)$ are identical to $(A3) - (A6)$ of the SWM problem, but $(B1) - (B2)$ differ

---

[5]From a BS m's point of view, if an AP i is not within BS m's coverage, then the utility of BS m will not increase with the component x_{mi}. In this case, BS m will never submit a positive bid p_{mi}.

[6]In many cases, an AP may lie in the coverage area of more than one base stations of the same operator. This is highly probable in heterogeneous networks where base stations have often overlapping coverages [25].

from $(A1) - (A2)$. Namely, $\forall k \in \mathcal{K}$, $m \in \mathcal{M}_k$, $\forall i \in \mathcal{I}$, we have

$$(B1) : x_{mi}^* = \frac{p_{mi}}{\mu_{mi}^*}, \quad (B2) : y_{im}^* = \frac{\mu_{mi}^* - \sum_{j=1}^{I} \frac{\gamma_{ji} \lambda_j^*}{C_i}}{\alpha_{im}},$$

$$(B3) : \lambda_i^* \cdot \left(\sum_{m=1}^{M} \sum_{j=1}^{I} \frac{\gamma_{ij} y_{jm}^*}{C_j} - 1 \right) = 0, \quad (B4) : y_{im}^* \geq x_{mi}^*,$$

$$(B5) : \mu_{mi}^* \cdot (x_{mi}^* - y_{im}^*) = 0, \quad (B6) : x_{mi}^*, y_{im}^*, \lambda_i^*, \mu_{mi}^* \geq 0.$$

It is important to note that $(B1)$ and $(B2)$ define the **allocation rule** of our proposed mechanism.

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

$$p_{mi} = x_{mi}^* \cdot \frac{\partial J_m(\boldsymbol{x}_m^*)}{\partial x_{mi}}, \quad \alpha_{im} = \frac{1}{y_{im}^*} \cdot \frac{\partial V_i(\boldsymbol{y}_i^*)}{\partial y_{im}}, \quad (17)$$

then the two-stage scheme defined in this section yields a solution identical to the unique optimal solution of the SWM problem , i.e., $\boldsymbol{x}^\circ \triangleq \boldsymbol{x}^*$ and $\boldsymbol{y}^\circ \triangleq \boldsymbol{y}^*$. Clearly, our goal is to *derive the proper payment and reimbursement rules that will induce the MNOs and APs to bid according to (17)*.

The Bidders. We now look at the bidders' behavior in the first stage. Let $h_k(\boldsymbol{x}_k)$ denote the MNO k's payment to the broker for the service it receives from the APs (matrix $\boldsymbol{x}_k$). This can be further decomposed in the payments for each base station of the operator, i.e., $h_m(\boldsymbol{x}_m)$, $\forall m \in \mathcal{M}_k$. 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 payoffs.

Specifically, given the allocation rule defined in $(B1)$ and $(B2)$, MNOs and APs solve their own optimization problems in order to find their optimal bids. Each MNO $k \in \mathcal{K}$ finds the optimal bid matrix $\boldsymbol{p}_k^*$ by solving the following payoff maximization problem:

$$\text{MNOP-k}: \quad \max_{\boldsymbol{p}_k} \left(J_k(\boldsymbol{x}_k) - h_k(\boldsymbol{x}_k) \right), \quad (18)$$
$$s.t.$$
$$p_{mi} \geq 0, \forall i \in \mathcal{I}, m \in \mathcal{M}_k. \quad (19)$$

The unique optimal solution of the MNOP-k problem satisfies the following optimality conditions:

$$\frac{\partial J_m(\boldsymbol{x}_m)}{\partial x_{mi}} = \mu_{mi} \frac{\partial h_m(\boldsymbol{x}_m)}{\partial p_{mi}}, \quad \forall i \in \mathcal{I}, m \in \mathcal{M}_k, \quad (20)$$

where we have used from $(B1)$ the equality $\partial x_{mi} / \partial p_{mi} = 1/\mu_{mi}$. Similarly, each AP $i \in \mathcal{I}$ finds the optimal bid vector $\boldsymbol{\alpha}_i^*$ by solving the following payoff maximization problem:

$$\text{APP-i}: \quad \max_{\boldsymbol{\alpha}_i} \left(-V_i(\boldsymbol{y}_i) + l_i(\boldsymbol{y}_i) \right), \quad (21)$$
$$s.t.$$
$$\alpha_{im} \geq 0, \forall m \in \mathcal{M}. \quad (22)$$

The unique optimal solution of the APP-i problem satisfies the following optimality conditions:

$$\frac{\partial V_i(\boldsymbol{y}_i)}{\partial y_{im}} = \frac{\alpha_{im}^2}{\frac{\sum_{j=1}^{I} \lambda_j \gamma_{ji}}{C_i} - \mu_{mi}} \frac{\partial l_i(\boldsymbol{y}_i)}{\partial \alpha_{im}}, \quad \forall m \in \mathcal{M}, \quad (23)$$

where we have used the derivative $\partial y_{im}/\partial \alpha_{im} = \mu_{mi} - \left(\sum_{j=1}^{I} \lambda_j \gamma_{ji}/C_i \right)$ that stems from $(B2)$.

Consider now the socially optimal solution $(A1)$ - $(A6)$, and equations (20) and (23) which yield the best responses of the MNOs and the APs respectively under the BAP allocation rule $(B1)$ - $(B6)$. In order to induce the MNOs and APs to submit the desirable bids described by (17) (i.e., the ones that maximize the social welfare), the **pricing rules** need to satisfy:

$$h_k(\boldsymbol{p}_k) = \sum_{i=1}^{I} \sum_{m=1}^{M_k} p_{mi}, \quad (24)$$

and

$$l_i(\boldsymbol{\alpha}_i) = \sum_{m=1}^{M} \frac{\left(\sum_{j=1}^{I} \frac{\gamma_{ji}\lambda_j}{C_i} - \mu_{mi} \right)^2}{\alpha_{im}}, \quad (25)$$

where we have written the payment $h_k(\boldsymbol{p}_k)$ by each MNO k, and the compensation $l_i(\boldsymbol{\alpha}_i)$ to each AP i, as functions of the respective bids.

The rules in (24) and (25) are intuitive: *each MNO pays exactly its bid* as in (24), i.e., the amount it declared that is willing to pay. Using the relationship $(B2)$ the reimbursement (25) can be written as

$$l_i(\boldsymbol{\alpha}_i) = \sum_{m=1}^{M} y_{im} \left(\frac{\sum_{j=1}^{I} \lambda_j \gamma_{ji}}{C_i} - \mu_{mi} \right), \quad (26)$$

which indicates that *each AP is reimbursed proportionally to the amount of data it offloads for each BS* weighted by a factor indicating (i) the interference and congestion on its link, and (ii) the difference between the requested and admitted traffic (μ_{mi}) by the specific BS m.

The above allocation and pricing rules of the IDA mechanism ensure that the market will maximize the social welfare. In the sequel, we explain how the proposed mechanism achieves this desirable outcome.

B. The IDA Implementation Algorithm

With the proper allocation rule $(B1)$ and $(B2)$, and payment rule (24) - (25), the MNOs 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 MNOs' utility functions and APs' cost functions). However, as the MNOs 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.

A high level description of the market operation is as follows. The broker announces the newly calculated dual variables, and the MNOs and APs solve their own problems to determine their bids, i.e., according to (18) - (19) and (21) - (22) respectively[7]. Then, they submit the bids to the

[7]We assume that the interference parameters γ_{ij}, $\forall i, j \in \mathcal{I}$ are known to APs i, j and to the base stations (MNOs) who compete for leasing them.

broker who finds the current optimal allocation and the new prices by solving the BAP problem (12) - (15). Notice that the latter is parameterized by the bids of the buyers and sellers. Solving the BAP problem by the broker yields the matrices $\boldsymbol{x}$ and $\boldsymbol{y}$ and the Lagrange multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. Similarly, the APP and MNOP problems are parameterized by the Lagrange multipliers of the BAP problem. That is why these problems (BAP, APP, and MNOP problems) need to be solved iteratively based on the feedback from the market.

For the BAP problem in particular, we will exploit its decomposable structure and solve it through a primal-dual Lagrange decomposition method [24]. This enables the parallel and hence fast execution of the algorithm. Moreover, this decomposition enables the decentralized operation of the market. That is, we can apply this mechanism to a setting where there are multiple brokers, each one responsible for a certain geographic subarea with a subset of access points. In this case, the different brokers will jointly solve the BAP problem, defined for the entire market, through minimum signaling that is necessary for their coordination. In the sequel, we focus first on the parallel execution of the scheme (with one broker), and accordingly we explain how the auction can be implemented by a set of different brokers.

The entire IDA scheme is described in details in Algorithm 1. The broker initializes the primal variables (i.e., $\boldsymbol{x}$ and $\boldsymbol{y}$) and dual variables (i.e., $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$) (line 2). For such an initialization, we can choose any set of values that satisfy the complementary slackness constraints $(B3)$ - $(B6)$. For example, we can choose $\lambda_i^{(0)} = 0$ and $x_{mi}^{(0)} = y_{im}^{(0)}$, with any positive values of $x_{mi}^{(0)}$ and $\mu_{mi}^{(0)}$.

Then the IDA algorithm will iteratively compute the primal and dual BAP variables (by the broker) and bids (by MNOs and APs) until convergence. First, the broker announces the dual variables of the BAP problem, to the MNOs and APs (line 5). Each MNO and each AP calculate their optimal bids by solving their respective optimization problems (lines $7-8$). All the MNOs and APs submit their new bids to the broker (line 9). Then, the broker collects the new bids and computes the updated allocations $\boldsymbol{x}^{(t)}$ and $\boldsymbol{y}^{(t)}$, using the allocation rule of problem BAP $(B1)$ - $(B2)$ (line 10). In the sequel, the broker finds the updated dual variables $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ (line 11), by using a gradient descent method:

$$\lambda_i^{(t+1)} = \left(\lambda_i^{(t)} - s^{(t)} \frac{\theta L(\cdot)}{\theta \lambda_i} \right)^+, \forall i \in \mathcal{I},$$

$$\mu_{mi}^{(t+1)} = \left(\mu_{mi}^{(t)} - s^{(t)} \frac{\theta L(\cdot)}{\theta \mu_{mi}} \right)^+, \forall i \in \mathcal{I}, m \in \mathcal{M},$$

where $(\cdot)^+$ denotes the projection onto the non-negative orthant and ensures that feasibility constraints $\lambda_i^{(t+1)} \geq 0$ and $\mu_{mi}^{(t+1)} \geq 0$ are satisfied.

Finally, the broker checks the termination criterion (line 12). Termination happens when relative changes of the bids during consecutive two iterations are sufficiently small, which is determined by the positive constants $\epsilon_1 > 0$ and $\epsilon_2 > 0$. If this condition is not satisfied, then the algorithm will enter a new iteration and repeat lines 5 to 12. After the algorithm has converged, the broker calculates the final payments $h_k(\boldsymbol{x}_k^{(t)})$ and reimbursements $l_i(\boldsymbol{y}_i^{(t)})$, for each BS and AP respec-

---

Algorithm 1: Iterative Double Auction (IDA)

output: $\boldsymbol{x}^*$, $\boldsymbol{y}^*$, $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$

1 $t \leftarrow 0$;

2 Initialize $x_{mi}^{(0)}$, $y_{im}^{(0)}$, $\mu_{mi}^{(0)}$, $\lambda_i^{(0)}$, ϵ_1, ϵ_2, $\forall m \in \mathcal{M}, i \in \mathcal{I}$;

3 conv_flag $\leftarrow 0$;

4 while conv_flag $= 0$ **do**

5 $\quad$ The broker announces $\mu_{mi}^{(t)}$, $\lambda_i^{(t)}$, $\forall m \in \mathcal{M}, i \in \mathcal{I}$;

6 $\quad$ $t \leftarrow t + 1$;

7 $\quad$ Each MNO k computes the optimal bids $\boldsymbol{p}_k^{(t)}$ by (18) - (19) ;

8 $\quad$ Each AP i computes the optimal bids $\boldsymbol{\alpha}_i^{(t)}$ by (21) - (22);

9 $\quad$ Each AP i and MNO k submit their bids $\boldsymbol{\alpha}_i^{(t)}$ and $\boldsymbol{p}_k^{(t)}$ to the broker;

10 $\quad$ The broker computes the new $\boldsymbol{x}^{(t)}$, $\boldsymbol{y}^{(t)}$ by $(B1)$ - $(B2)$;

11 $\quad$ The broker uses gradient updates for dual variables:

$$\lambda_i^{(t)} = \left(\lambda_i^{(t-1)} + s^{(t)} \cdot \left(\sum_{m=1}^{M} \sum_{j=1}^{I} \frac{\gamma_{ij} y_{jm}^{(t-1)}}{C_j} - 1\right)\right)^+,$$

$$\mu_{mi}^{(t)} = \left(\mu_{mi}^{(t-1)} + s^{(t)} \cdot \left(x_{mi}^{(t-1)} - y_{im}^{(t-1)}\right)\right)^+,$$

$\quad$ $\forall m \in \mathcal{M}, i \in \mathcal{I}$, where $s^{(t)} = 0.05$ is the step size;

12 $\quad$ The broker checks the termination criterion:

$\quad$ **if** $\left|\frac{p_{mi}^{(t)} - p_{mi}^{(t-1)}}{p_{mi}^{(t-1)}}\right| < \epsilon_1$ and $\left|\frac{\alpha_{im}^{(t)} - \alpha_{im}^{(t-1)}}{\alpha_{im}^{(t-1)}}\right| < \epsilon_2$,

$\quad$ $\forall m \in \mathcal{M}, i \in \mathcal{I}$ **then**

$\quad\quad$ | conv_flag $\leftarrow 1$;

$\quad$ **end**

end

13 The broker computes $h_k(\boldsymbol{x}_k^{(t)})$, $l_i(\boldsymbol{y}_i^{(t)})$, $\forall k \in \mathcal{K}, \forall i \in \mathcal{I}$, using (24) and (25);

---

tively, by using (24) and (25) (line 13). Notice that, in such tatonnement algorithms, the payments are only valid after the convergence of the mechanism [20].

The above algorithm assumes the parallel execution of the necessary IDA computations by a single broker. Interestingly, the same algorithm is applicable, with very slight modifications, for the case that we have multiple different brokers that are cooperating to jointly clear the market in a decentralized fashion. For such a setting, the bids from each AP and for each BS are submitted only to the respective local broker. The latter collects the local bids and calculates the primal variables for these APs and BSs, using eq. $(B1)$ - $(B2)$. Also, he updates the respective dual variables through the gradient update formula. Notice that this decentralized market operation, which significantly improves the scalability of the mechanism, requires the coordination of different brokers. This can be achieved through the exchange of the y_{im}, y_{jm} variables for each pair of APs $i, j \in \mathcal{I}$ that lie in different areas (i.e., managed by different brokers) but still interfere with each other, i.e., $\gamma_{ij} > 0$.

Algorithm 1 is executed in a synchronous fashion, which requires a common clock of all users and a small delay for message passing. This time synchronization is a reasonable assumption for a local market where the information transmission delay is small. Finally, Algorithm 1 has a relatively small communication overhead since there is only a polynomial number of messages that need to be circulated in the market. Namely, in each round there are $2 \cdot I \cdot M$ bids that must be communicated from the MNOs, for all their base stations, and the access points to the broker. Similarly, the broker announces the $I \cdot M + I$ dual variables to the $I \cdot M$ bidders. Therefore, the message passing overhead is $O\left((I \cdot M)^2\right)$ per round.

IV. CONVERGENCE ANALYSIS AND ECONOMIC PROPERTIES OF IDA MECHANISM

In this section we present the convergence analysis for the IDA mechanism, and prove that it has the desirable economic properties $(E1)$-$(E4)$.

A. Convergence Analysis

Algorithm 1 converges to the optimal solution of problem SWM under some mild conditions, since it has a strictly concave objective function. Namely, we assume that the time slot of the dual variables update is very small (or equivalently, the step size is very small), hence we approximate the algorithm with its continuous-time counter-part. Specifically, the following theorem holds:

Theorem 1. *Algorithm* 1 *converges to the optimal solution of the SWM problem globally, i.e., from any initial point that satisfies the complementary slackness conditions* $(A3) - (A6)$.

Proof. We consider a very small time slot and hence assume that the Lagrange multipliers are updated according to the differential equations:

$$\frac{\mathrm{d}\lambda_i}{\mathrm{d}t} = \left(\sum_{m=1}^{M} \sum_{j=1}^{I} \frac{\gamma_{ij} y_{jm}^*}{C_j} - 1\right)_{\lambda_i}^+, \qquad (27)$$

$$\frac{\mathrm{d}\mu_{mi}}{\mathrm{d}t} = \left(x_{mi} - y_{im}\right)_{\mu_{mi}}^+ \qquad (28)$$

which are derived by the gradient update rules (for very small step sizes). Note that we have used from [26] the notation $\left(g(x)\right)_y^+$:

$$\left(g(x)\right)_y^+ = \begin{cases} g(x) & \text{if } y > 0, \\ \max\left(g(x), 0\right) & \text{if } y = 0. \end{cases} \qquad (29)$$

We prove the convergence of our two-sided algorithm following the rationale of the proof in [47, Ch. 22], which was used to prove the one-side version of our algorithm in [16]. Specifically, we define the Lyapunov function:

$$Z(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \sum_{i=1}^{I} \frac{(\lambda_i - \lambda_i^*)^2}{2} + \sum_{m=1}^{M} \sum_{i=1}^{I} \frac{(\mu_{mi} - \mu_{mi}^*)^2}{2}, \quad (30)$$

and we prove that $\mathrm{d}Z(\boldsymbol{\lambda}, \boldsymbol{\mu})/\mathrm{d}t \leq 0$.

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, using (27) and (28), can be written as:

$$\frac{\mathrm{d}Z(\boldsymbol{\lambda}, \boldsymbol{\mu})}{\mathrm{d}t} = \sum_{i=1}^{I}(\lambda_i - \lambda_i^*) \cdot \Big(\sum_{m=1}^{M}\sum_{j=1}^{I} \frac{\gamma_{ij}y_{jm}^*}{C_j} - 1 \Big)_{\lambda_i}^{+}$$
$$+ \sum_{m=1}^{M}\sum_{i=1}^{I}(\mu_{mi} - \mu_{mi}^*) \cdot (x_{mi} - y_{im})_{\mu_{mi}}^{+}.$$

Observe now that we can write the following inequalities for each AP $i \in \mathcal{I}$ and each BS $m \in \mathcal{M}$, respectively:

$$\Big(\sum_{m=1}^{M}\sum_{j=1}^{I} \frac{\gamma_{ij}y_{jm}}{C_j} - 1 \Big)_{\lambda_i}^{+} \le \Big(\sum_{m=1}^{M}\sum_{j=1}^{I} \frac{\gamma_{ij}y_{jm}}{C_j} - 1 \Big), \quad (31)$$

$$(x_{mi} - y_{im})_{\mu_{mi}}^{+} \le (x_{mi} - y_{im}). \quad (32)$$

Let us first focus on inequality (31). If the projection $(\cdot)_{\lambda_i}^{+}$ is not active, then it holds as an equality. On the other hand, if the projection is active, then the right-hand side is positive while the left-hand side is zero, hence it is again satisfied. With a similar reasoning, we can see that (32) holds as well in every case.

Therefore, the derivative of the Lyapunov function is upper bounded as follows:

$$\frac{\mathrm{d}Z(\boldsymbol{\lambda}, \boldsymbol{\mu})}{\mathrm{d}t} \le \sum_{i=1}^{I}(\lambda_i - \lambda_i^*) \cdot \Big(\sum_{m=1}^{M}\sum_{j=1}^{I} \frac{\gamma_{ij}y_{jm}}{C_j} - 1 \Big)$$
$$+ \sum_{m=1}^{M}\sum_{i=1}^{I}(\mu_{mi} - \mu_{mi}^*) \cdot (x_{mi} - y_{im}).$$

Next, we add and subtract in the RHS of the above inequality the terms $\sum_j \sum_m \gamma_{ij}y_{jm}^*/C_j$ and $(x_{mi}^* - y_{im}^*)$ for each AP $i \in \mathcal{I}$ and BS $m \in \mathcal{M}$. Hence, the derivative $\dot{Z} = \mathrm{d}Z(\boldsymbol{\lambda}, \boldsymbol{\mu})/\mathrm{d}t$ can be written:

$$\dot{Z} \le \sum_{i=1}^{I}(\lambda_i - \lambda_i^*) \cdot \Big(\sum_{m=1}^{M}\sum_{j=1}^{I} \frac{\gamma_{ij}y_{jm}}{C_j} - \sum_{m=1}^{M}\sum_{j=1}^{I} \frac{\gamma_{ij}y_{jm}^*}{C_j} \Big)$$
$$+ \sum_{m=1}^{M}\sum_{i=1}^{I}(\mu_{mi} - \mu_{mi}^*)(x_{mi} - y_{im} - x_{mi}^* + y_{im}^*)$$
$$+ \sum_{i=1}^{I}(\lambda_i - \lambda_i^*)\Big(\sum_{m=1}^{M}\sum_{j=1}^{I} \frac{\gamma_{ij}y_{jm}^*}{C_j} - 1 \Big)$$
$$+ \sum_{m=1}^{M}\sum_{i=1}^{I}(\mu_{mi} - \mu_{mi}^*)(x_{mi}^* - y_{im}^*).$$

After some algebraic manipulations, and using $(A1, A2)$, and the complementary slackness conditions that are satisfied at the equilibrium, we get:

$$\dot{Z} \le \sum_{m=1}^{M}\sum_{i=1}^{I}\Big[(y_{im} - y_{im}^*)\Big(\frac{\partial V_i(\boldsymbol{y}_i^*)}{\partial y_{im}} - \frac{\partial V_i(\boldsymbol{y}_i)}{\partial y_{im}}\Big)$$
$$+ (x_{mi} - x_{mi}^*)\Big(\frac{\partial J_m(\boldsymbol{x}_m)}{\partial x_{mi}} - \frac{\partial J_m(\boldsymbol{x}_m^*)}{\partial x_{mi}}\Big)\Big].$$

The RHS of the above inequality is non-positive, because of the following property that holds for each concave function $f(\cdot)$ [24]:

$$f(u) \le f(v) + \nabla f(v)^T(u - v). \quad (33)$$

Therefore, we have proved $\dot{Z} \le 0$, which concludes our convergence proof. $\square$

B. Properties of IDA Mechanism

Next we will show that the IDA algorithm possesses the desirable economic properties $(E1)$-$(E4)$ of efficiency, incentive compatibility, individual rationality, and (weakly) budget balance. This is summarized in Propositions 1 - 3 below.

Proposition 1. *The IDA mechanism is efficient $(E1)$, and incentive compatible $(E2)$ for price-taking bidders.*

Proof. The algorithm, according to Theorem 1, will converge to a point that satisfies $(B1) - (B6)$. Hence, given the pricing and reimbursement functions (24)-(25), and the strategy of the bidders (18) and (21), the algorithm will reach the social optimal solution $(A1) - (A6)$, ensuring the efficient operation of the market.

Moreover, the proposed mechanism is incentive compatible since it induces the bidders to update their bids so as to gradually reveal their utility and cost functions. To see this, observe that in each iteration, the bidders submit their currently optimal bids by solving the respective APP and MNOP problems. Hence, although the bidders do not communicate their actual functions, they eventually submit the socially optimal bids (17), which allows the broker to elicit this hidden information and maximize the social welfare of the market. Hence, the algorithm is incentive compatible. $\square$

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.

Proposition 2. *The IDA mechanism is individually rational $(E3)$.*

Proof. First, notice that not participating in the market leads to a zero payoff for any AP or MNO, as they will not offer or receive any offloading service respectively, nor they will pay any kind of fee. Moreover, for each BS $m \in \mathcal{M}_k$ of each operator $k \in \mathcal{K}$, the IR condition can be translated to the constraint:

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

which, using (20), can be written as

$$J_m(\boldsymbol{x}_m^*) \ge \sum_{i=1}^{I}x_{mi}^*\frac{\partial J_m(\boldsymbol{x}_m^*)}{\partial x_{mi}}. \quad (34)$$

Since $J_m(\cdot)$ is strictly concave and $J_m(\boldsymbol{0}) = 0$, the inequality (34) is always satisfied due to property (33).

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

$$-V_i(\boldsymbol{y}_i^*) + l_i(\boldsymbol{\alpha}_i^*) \ge 0, \quad (35)$$

or

$$-V_i(\boldsymbol{y}_i^*) + \sum_{m=1}^{M} \frac{\big(\sum_{j=1}^{I} \frac{\gamma_{ji}\lambda_j}{C_i} - \mu_{mi}\big)}{\alpha_{im}} \ge 0. \quad (36)$$

Also, we have from (25):

$$\frac{\partial l_i(\boldsymbol{\alpha}_i)}{\partial \alpha_{im}} = \sum_{m=1}^{M} \frac{-1}{\alpha_{im}^2} \Big(\sum_{j=1}^{I} \frac{\gamma_{ji}\lambda_j}{C_i} - \mu_{mi} \Big)^2 \qquad (37)$$

Hence, if we substitute (37) in (23) we get

$$\frac{\partial V_i(\boldsymbol{y}_i)}{\partial y_{im}} = -\Big(\sum_{j=1}^{I} \frac{\gamma_{ji}\lambda_j}{C_i} - \mu_{mi} \Big). \qquad (38)$$

By substituting (38) in (36), we have

$$V_i(\boldsymbol{y}_i^*) \le \sum_{m=1}^{M} y_{im}^* \frac{\partial V_i(\boldsymbol{y}_i)}{\partial y_{im}} \qquad (39)$$

which is always satisfied due to (33) since $-V_i(\cdot)$ is a concave function. $\qquad\square$

Finally, the algorithm is weakly budget balanced, and therefore the broker does not have to inject money in the market. That is, at the equilibrium, the payments from the broker to all the access points, $\sum_{i=1}^{I} l_i(\boldsymbol{\alpha}_i)$, do not exceed the payments from all the operators to the broker, $\sum_{k=1}^{K} h_k(\boldsymbol{p}_k)$.

Proposition 3. *The IDA mechanism is weakly budget balanced* $(E4)$.

Proof. The broker's budget balance $\Lambda(\cdot)$ is defined as:

$$\Lambda(\boldsymbol{p}, \boldsymbol{\alpha}) = \sum_{k=1}^{K} h_k(\boldsymbol{p}_k) - \sum_{i=1}^{I} l_i(\boldsymbol{\alpha}_i),$$

or, if we use (24)-(25),

$$\Lambda(\boldsymbol{p}, \boldsymbol{\alpha}) = \sum_{m=1}^{M} \sum_{i=1}^{I} p_{mi} - \sum_{m=1}^{M} \sum_{i=1}^{I} \frac{\big(\sum_{j=1}^{I} \frac{\lambda_j \gamma_{ji}}{C_i} - \mu_{mi} \big)^2}{\alpha_{im}}.$$

Notice that $(B1)$ and $(B2)$ are satisfied at the equilibrium. Hence,

$$\Lambda(\boldsymbol{p}, \boldsymbol{\alpha}) = \sum_{m=1}^{M} \sum_{i=1}^{I} x_{mi}^* \mu_{mi}^* + \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im}^* \Big(\sum_{j=1}^{I} \frac{\lambda_j^* \gamma_{ji}}{C_i} - \mu_{mi} \Big),$$

or

$$\Lambda(\boldsymbol{p}, \boldsymbol{\alpha}) = \sum_{m=1}^{M} \sum_{i=1}^{I} \mu_{mi}^* (x_{mi}^* - y_{im}^*) + \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im}^* \sum_{j=1}^{I} \frac{\lambda_j^* \gamma_{ji}}{C_i},$$

which is always non-negative because of $(B5)$ and $(B6)$. $\qquad\square$

V. Performance Evaluation

In this section we provide numerical results to validate our theoretical analysis. First, we present an example with a small representative setting of $K = 2$ operators, each one with 1 base station, and $I = 3$ access points, in order to shed light on the basic idea of the IDA algorithm. Then, we consider a small market with $K = 2$ operators, $M = 5$ BSs and $I = 5$ APs and study the impact of various system parameters on the offloading market.
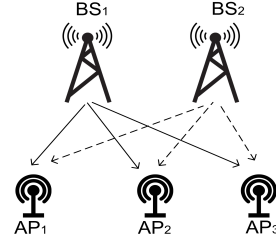

Fig. 3. A toy example with 2 base stations, each one managed by a different MNO, and 3 Access Points. The offloading efficiency and servicing cost parameters are $\rho_{11} = 0.746$, $\rho_{21} = 0.943$, $\rho_{31} = 0.61$, $\rho_{12} = 0.826$, $\rho_{22} = 0.936$, $\rho_{32} = 0.614$, $\theta_{11} = 0.568$, $\theta_{12} = 0.611$, $\theta_{13} = 0.954$, $\theta_{21} = 0.79$, $\theta_{22} = 0.923$, and $\theta_{23} = 0.606$.
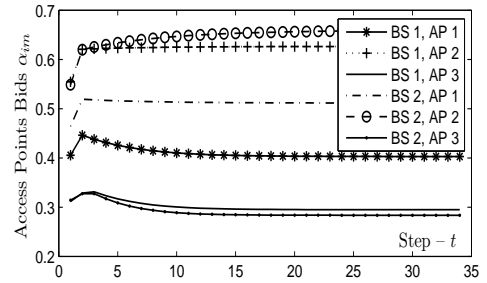

Fig. 4. Evolution off bids of the APs for a small market with $K = 2$ operators with $M = 2$ base stations in total, and $I = 3$ APs.

A. Numerical Example

Consider the setting in Figure 3 where there are 2 operators, and each manages a single base station. The operators compete in order to lease the bandwidth capacity of 3 APs. The utility functions of the operators and the cost functions of the APs are respectively:

$$J_m(\boldsymbol{x}_m) = 10 \cdot \sum_{i=1}^{3} \log(\theta_{mi} x_{mi}), \ m \in \{1, 2\} \qquad (40)$$

and

$$V_i(\boldsymbol{y}_i) = 0.1 \cdot \sum_{m=1}^{2} e^{\rho_{im} y_{im}}, \ i \in \{1, 2, 3\} \qquad (41)$$

where $\theta_{mi} \ge 0$ represents the offloading efficiency of AP i for BS m (i.e., the offloading benefit is AP-specific), and parameter $\rho_{im} \ge 0$ captures the fact that each AP may incur different cost by serving a different BS. Notice that these functions satisfy the desirable properties of concavity/convexity. Unless otherwise specified, the capacity of each AP is assumed to be equal to $C = 15$ Mbps. Finally, for this example we assume that the APs do not interfere with each other, i.e., $\gamma_{ij} = 0$, $\forall i \ne j$, and $\gamma_{ii} = 1$.

In Figure 4 we plot the bids of the APs for each iteration of the IDA Algorithm, which converge to the optimal (social welfare maximizing) values after approximately 10 iterations. The initial values are randomly selected. From the numerical results of Table II (which hold at the equilibrium), we can verify some intuitive properties of the mechanism. For example, AP 1 has smaller cost for serving BS 1 than BS 2, i.e., $\rho_{11} < \rho_{12}$, and therefore admits more traffic from BS 1, i.e.,

TABLE II
PARAMETERS ρ_{im}, θ_{mi}, ADMITTED DATA y_{im}^*, AND PAYMENTS p_{mi}^* FOR
THE APs AND BSs IN FIGURE 3 AT THE EQUILIBRIUM.

		AP 1	AP 2	AP 3
$\{\rho_{im}\}_{i=1,2,\,m=1,2,3}$	BS 1	0.75	0.94	0.61
	BS 2	0.82	0.93	0.61
$\{\theta_{mi}\}_{i=1,2,\,m=1,2,3}$	BS 1	0.57	0.61	0.95
	BS 2	0.79	0.92	0.60
$\{y_{im}\}_{i=1,2,\,m=1,2,3}$	BS 1	4.17	3.26	5.32
	BS 2	3.83	3.39	5.17
$\{p_{mi}\}_{i=1,2,\,m=1,2,3}$	BS 1	7.03	6.66	8.35
	BS 2	7.51	7.58	7.58

$y_{11}^* > y_{12}^*$. Also, BS 1 gets a lower benefit from offloading traffic to AP 1 than to AP 3, hence $y_{11}^* < y_{31}^*$. Finally, AP 1 has lower cost for serving BS 1 than AP 2 does, i.e., $\rho_{11} < \rho_{21}$, hence $y_{11}^* > y_{21}^*$.

Regarding the payments which can be calculated by eq. (24) - (25), operator 1 (BS 1) pays $h_1 = 22$, and operator 2 (BS 2) $h_2 = 22.7$ monetary units. An AP i is paid $l_i = \sum_{m=1}^2 (\frac{\lambda_i}{C_i} - \mu_{mi})y_{im}$, which yields $l_1 = 14.55$, $l_2 = 14.22$, and $l_3 = 15.93$. Notice that the budget balance is satisfied (strictly in this case), since $h_1 + h_2 = l_1 + l_2 + l_3$.

B. Simulation Results

Consider now a small market with $K = 2$ MNOs and $I = 5$ APs in a certain geographic area. Operator 1 manages 2 base stations in this area, i.e., $\mathcal{M}_1 = \{1, 2\}$, whereas operator 2 has 3 BSs, i.e., $\mathcal{M}_2 = \{3, 4, 5\}$. The study can be easily extended to a larger market. We present here results about the convergence of the bids and the market welfare to the optimal solution. Moreover, we study how certain system parameters, such as the interference parameters, affect the decisions of the broker.

The BSs' utility functions and APs' cost functions are similar as in Section V-A. That is:

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

The capacity of each AP is assumed to be $C = 15$ Mbps. Parameters ρ_{im} and θ_{mi}, $\forall m \in \mathcal{M}$, $\forall i \in \mathcal{I}$, were chosen independently and uniformly from the interval $[0.5, 1]$. We assume that the APs are closely located and induce interference to each other, and hence the interference parameters γ_{ij}'s are uniformly distributed in $[0.2, 0.4]$. Finally, the convergence parameters have been set to $\epsilon_1 = \epsilon_2 = 0.001$.

In Figure 5, we plot the social welfare achieved by the algorithm in each iteration. We observe that it gradually converges to the optimal one, i.e. to the solution of SWM (dotted line). Notice that before the convergence, the social welfare can be even higher than the respective optimal value as we have relaxed the APs' capacity constraints. By running several different experiments, e.g., with different cost/utility functions, different number of base stations and access points, we verified that the algorithm always converges as it was proved in our theoretical analysis. In Figure 6, we present the convergence of $\boldsymbol{x}$ and $\boldsymbol{y}$. Specifically, we plot the gap between $\boldsymbol{x}$ and $\boldsymbol{y}$ for the base stations and access point 1 (AP 1). We see that the distance between requested and admitted
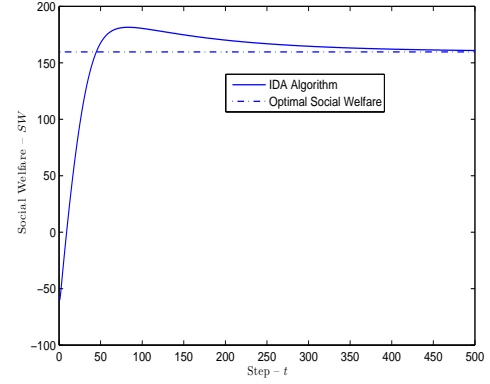

Fig. 5. Evolution of social welfare produced by the IDA, for a market with $M = 5$ base stations and $N = 5$ access points.
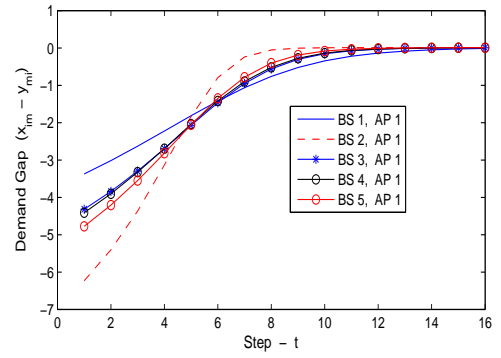

Fig. 6. Evolution of the gap between requested ($\boldsymbol{x}$) and admitted data ($\boldsymbol{y}$), for a market with $M = 5$ base stations and $N = 5$ access points. The plots refer to the 5 BSs and AP 1.

offloaded traffic gradually converges to zero. This means that the operators and the access points agree on the amount of data that should be offloaded (for each pair of BS - AP). Besides, in all the experiments that we conducted, the values of the μ_{mi} dual variables are non-zero which, due to complementary slackness conditions, validates this conclusion as well. Figures 5 and 6 imply that our IDA algorithm elicits the true hidden information (i.e., the utility and cost functions), and converges to the socially optimal solution.

In Figure 7, we consider a market with $M = 5$ base stations and $N = 5$ access points, and we plot the evolution of the social welfare for different values of the step size $s^{(t)}$ that is used for the update of the Lagrange multipliers (line 12 of Algorithm 1). For this experiment, we set $C = 30$ Mbps and $\epsilon_1 = \epsilon_2 = 0.05$. We can see how the step size affects the convergence speed of the algorithm. Namely, for small step sizes the algorithm needs more slots in order to converge to the optimal solution compared to the cases that we use larger step sizes. For example, when $s^{(t)} = 0.05$ the mechanism converges after approximately 35 iterations, while for $s^{(t)} = 0.25$ it needs only 7 iterations. Nevertheless, if we further increase the step, we will see that the mechanism does not converge to the optimal solution.

Finally, in Table III we present the average number of iterations (and the standard deviation) for the convergence of the algorithm, for different network sizes, where we used a step
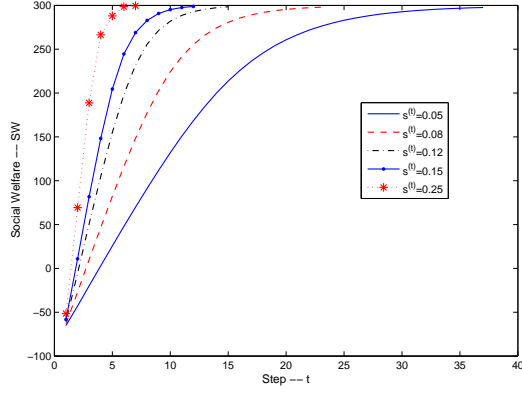
Fig. 7. Impact of step size on the convergence speed of the algorithm for a market with $M = 5$ base stations (one for each operator), and $N = 5$ access points.

Fig. 8. Impact of interference parameters on total amount that each AP offloads, and on the dual variables λ_1 and λ_2.

TABLE III
CONVERGENCE OF ALGORITHM 1

MNOs	APs	Average Iterations	Std
M=4	N=4	10.4	0.49
M=5	N=5	12.8	0.44
M=6	N=6	14.7	0.49
M=7	N=7	16.3	0.57
M=8	N=8	17.7	0.47
M=9	N=9	18.9	0.37

$s^{(t)} = 0.12$ and run each experiment 20 times. Notice that the required number of iterations until the convergence increases slightly with the network size, i.e., the number of base stations and/or access points. Considering that such mechanisms can be programmed or hardwired at the BSs and APs (similarly to the respective algorithms that are implemented in TCP [26]), the algorithm can converge very fast in practice.

Figure 8 illustrates how the interferences among APs affect the optimal offloading decisions. For this experiment, we consider a smaller market with 3 base stations and 3 access points. We assume that all the APs are in range and induce to each other interference which varies. Among various entries of the interference matrix γ_{ij} for $i, j \in \{1, 2, 3\}$, $\gamma_{ii} = 0$ for $i \in \{1, 2, 3\}$, γ_{31} and γ_{13} are fixed, and we only decrease γ_{21} (which is equal to γ_{12}) and increase γ_{23} (which is equal to γ_{32}), hence we have an increase of the ratio γ_{23}/γ_{21} (which is equal to γ_{32}/γ_{12}). This yields a decreasing total interference for AP 1 (both to other APs and from other APs, as the interference matrix is symmetric), and an increasing interference for AP 2. As it is shown in the upper plot of Figure 8, the higher is the interference, the less is the traffic that the AP offloads. Additionally, the solution of the BAP problem, and hence the respective dual variables, change as well, as it is shown in the lower plot of this figure. In particular, the higher is the interference experienced by an AP i, the larger is the value of the dual variable λ_i that is related to the capacity constraint.

VI. RELATED WORK

Several recent studies (e.g., [27]) have quantified the benefits of macrocellular data offloading to Wi-Fi networks. For
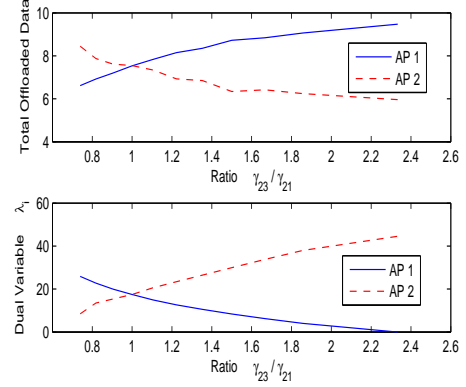
example, studies in [28] showed that approximately 80% of mobile data traffic is generated and consumed indoors, and hence can be offloaded to APs. The authors in [29], [30] studied optimal offloading strategies for delay tolerant applications that take into account the delay constraints of users. Clearly, the offloading benefits depend on the availability of APs that are accessible and have extra capacity to offload cellular traffic. Interestingly though, the problem of incentivizing Wi-Fi open access for offloading has received very little attention until today.

Another option for offloading is to utilize femtocell access points (FAPs). 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 compensations [31]. This compensation can be either a price discount [32], or a direct payment from the operator. A few related works in this area studied monopolistic markets (i.e., with one operator) [33], [52] and often did not consider the challenges $(I1)-(I5)$ for the offloading market [34], [43].

Several studies have proposed more sophisticated architectures and offloading schemes. For example, in [48] the authors studied a dynamic traffic offloading framework, in which multi-mode small cell base stations can route their traffic either through cellular or Wi-Fi networks (i.e., operate in both modes). Their decision depends on the type and volume of the traffic, the quality-of-service (QoS) requirements of the users, and network conditions such as the network load and interference. Another interesting case is to offload cellular traffic by leveraging device-to-device communications [49], [50], [51]. However, these approaches are valid only when the users are interested in the same content (as it often happens in online social networks), and when users are in proximity so as to be within effective communication range.

The market mechanism that we consider in this work differs substantially from other related economic schemes. For example, in our previous work [35], we studied the interaction of multiple MNOs and APs in offloading markets, under the assumption of complete information for all the participants. Such competition models give significant insights for the market operation, and has been extensively used for wireless

network markets, e.g. [36]. We have also analyzed the monopolistic setting when one MNO leases multiple APs (e.g., those owned by the MNO's subscribers) [42]. However, the problem becomes substantially more challenging when the network information about the buyers and sellers is incomplete.

Markets with many buyers and sellers under incomplete information are usually cleared through double auctions. One prominent class of double auctions is the VCG mechanism, which however exhibits a very high computational complexity and can yield a budget imbalanced outcome [14]. Hence, VCG double auctions are not suitable for the problem under consideration, as it may require the broker to fund the market operation.

Another prominent scheme is the McAfee mechanism [13], which has recently been proposed for spectrum allocation in secondary spectrum markets [44] and for traffic relaying [45]. 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 [44], [45]. Besides, in all the previous studied cases, the outcome is inefficient, which is an inherent characteristic of McAfee auction. Clearly, such an auction scheme is not suitable for the problem under consideration, as it cannot capture the particular aspects and realistic issues of the offloading problem, i.e., $(I1) - (I5)$.

In this paper, we adopted a different approach and used a market mechanism based on the NUM framework [15], and, in particular, one that is based on the bandwidth allocation scheme for networks [16]. 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 that the bidders have limited market information, e.g., about the number of participants and their resources, and hence cannot estimate their impact [17]. For the offloading market under consideration, price-taking behavior arises as there are many AP owners who are not aware of each other's capacity and communication needs (hence their cost functions), and multiple base stations with no information about each other's demands.

At the expense of this assumption, we proposed an iterative double auction mechanism which satisfies all the desirable economic properties. Our work substantially departs from the algorithm in [16]. Namely, the resource sellers (the APs) in our work, unlike the respective sellers (the links) in [16], try to maximize their own net payoffs 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 [16]. A similar approach was followed in our previous work [46] for bandwidth allocation in peer-to-peer networks, however the objectives and the solution method were significantly different.

VII. CONCLUSIONS

Nowadays it is clear that data offloading can alleviate congestion of 2G/3G cellular networks, and also serve as a low-cost auxiliary technology for the emerging 4G/5G networks [5]. Offloading techniques can actually complement similar methods that try to address the capacity crunch, such as MNOs' cooperative networking solutions, and cognitive networks [36], [37], [38], [39], [40], [41]. In this paper, we studied a market where MNOs lease third-party owned Wi-Fi or femtocell APs to offload their mobile data traffic on demand. This is a promising solution for increasing the user perceived network capacity in a dynamic and scalable fashion, with low costs for cellular networks. The technologies to implement such solutions are already in place, e.g., secure and seamless offloading methods [5], [8].

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. Our analysis explicitly takes into account the particular characteristics and challenges of wireless networks, such as the interference of APs and different offloading benefits of mobile users depending on their locations. An interesting further extension of this work is to study market mechanisms considering the price anticipating behaviors of the MNOs and APs, and characterize the (possible) performance degradation due to this effect.

REFERENCES

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

[2] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017", Feb. 2013.

[3] Ericsson Press Release, "Ericsson Mobility Report", June 2014.

[4] Bloomberg: "AT&T Pays 2 Billions to Qualcom for Spectrum", 2009.

[5] European Commission, Directorate-General for Communications Networks, Content and Technology, "Study on Impact of Traffic off-loading and Related Technological Trends on the Demand for Wireless Broadband Spectrum", *ISBN: 978-92-79-30575-7*, 2013.

[6] AT&T Press Release, "AT&T Expands Wi-Fi Hotzone Pilot Project to Additional Cities", July, 2010.

[7] BT Wifi Press Release, "O2 Brings 3,000 New Wi-Fi Hotspots to iPhone Customers", July, 2008.

[8] Cisco, "Making Wi-Fi as Secure and Easy to Use as Cellular", 2012.

[9] Republic Wireless, http://www.republicwireless.com/ Nov. 2012.

[10] Victoria W. Kipp, "The Battle of NIMBY", *PRIMEDIA Business Magazines & Media Inc*, 2002.

[11] Spectrum bridge, http://www.spectrumbridge.com/.

[12] R. B. Myerson, and M. A. Satterthwaite, "Efficient Mechanisms for Bilateral Trading", *Journal of Econ. Theory*, vol. 29, no. 2, 1983.

[13] R. P. McAfee, "A Dominant Strategy Double Auction", *Journal of Economic Theory*, vol. 56, no. 2, 1992.

[14] P. Maille, and B. Tuffin, "Why VCG Auctions can Hardly be Applied to the Pricing of Interdomain and Ad hoc Networks", *in Proc. of NGI*, 2007.

[15] D. P. Palomar, and M. Chiang, "A Tutorial on Decomposition Methods for Network Utility Maximization ", *IEEE JSAC*, vol. 24, no. 8, 2006.

[16] 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, 1998.

[17] L. Johansen, "Price-Taking Behavior", *Econometrica*, vol. 4, no. 7, 1977.

[18] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power Consumption Modeling of Different Base Station Types in Heterogeneous Cellular Networks", *in Proc. ICT MobileSummit*, 2010.

[19] K. Son, H. Kim, Y. Yi, B. Krishnamachari, "Toward Energy-Efficient Operation of Base Stations in Cellular Wireless Networks", *Book chapter of Green Communications: Theoretical Fundamentals, Algorithms, and Applications*, CRC Press, Taylor - Francis, Sep. 2012.

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

[21] Y. Kishiyama, A. Benjebbour, H. Ishii, and T. Nakamura, Evolution Concept and Candidate Technologies for Future Steps of LTE-A, *in Proc. of IEEE ICCS*, 2012

[22] W. Cheung, T. Quek, and M. Kountouris, "Throughput Optimization, Spectrum Allocation, and Access Control in Two-Tier Femtocell Networks", *IEEE JSAC*, vol. 30, no. 3, 2012.

[23] A. Karnik, A. Iyer, and C. Rosenberg, "What is the right model for wireless channel interference?", *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, 2009.

[24] D. Bertsekas, "Nonlinear Programming", *Athena Scientific,* 1999.

[25] J. G. Andrews, "Seven ways that HetNets are a cellular paradigm shift", *IEEE Communications Magazine*, vol. 51, no. 3, 2013.

[26] S. Shakkottai, and R. Srikant, "Network Optimization and Control", *Foundations and Trends in Networking*, vol. 2, no. 3, 2007.

[27] 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.

[28] Informa Telecoms & Media, "Femtocell Market Status", Feb. 2011.

[29] M. Cheung, and J. Huang, "Optimal Delayed Wi-Fi Offloading", *in Proc. of WiOpt*, 2013.

[30] M. Cheung, R. Southwell, and J. Huang, "Congestion-Aware Network Selection and Data Offloading", *in Proc. of CISS*, 2014.

[31] H. S. Jo, P. Xia, and J. G. Andrews, "Downlink Femtocell Networks: Open or Closed?", *In Proc. of IEEE ICC*, 2011.

[32] S. Y. Yun, Y. Yi, D. H. Cho, and J. Mo , "The Economic Effects of Sharing Femtocells", *IEEE JSAC,* vol. 30, no. 3, 2012.

[33] 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.

[34] 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, 2012.

[35] L. Gao, G. Iosifidis, J. Huang, and L. Tassiulas, "Economics of Mobile Data Offloading", *in Proc. of IEEE SDP Workshop*, 2013.

[36] D. Niyato, E. Hossain, and Z. Han, "Dynamics of Multiple-Seller and Multiple-Buyer Spectrum Trading in Cognitive Radio Networks: A Game-Theoretic Modeling Approach", *IEEE Trans. on Mob. Comput.*, vol. 8, no. 8, 2009.

[37] J. Huang, R. Berry, and M. L. Honig, "Distributed Interference Compensation for Wireless Networks", *IEEE JSAC*, vol. 24, no. 5, 2006.

[38] J. Huang, R. Berry, and M. L. Honig, "Auction-based Spectrum Sharing", *ACM /Springer Mobile Networks and Applications*, vol. 11, no. 3, 2006.

[39] Q. Liang, S. Han, F. Yang, G. Sun, and X. Wang, "A Distributed-Centralized Scheme for Short- and Long-Term Spectrum Sharing with a Random Leader in Cognitive Radio Networks", *IEEE JSAC*, vol. 30, no. 11, 2012.

[40] L. Gao, Y. Xu, and X. Wang, "MAP:Multi-Auctioneer Progressive Auction in Dynamic Spectrum Access", *IEEE Trans. on Mob. Comp.*, vol. 10, no. 8, 2011.

[41] L. Gao, X. Wang, Y. Xu, and Q. Zhang, "Spectrum Trading in Cogntive Radio Networks: A Contract-theoretic Modeling Approach", *IEEE JSAC*, vol. 29, no. 4, 2011.

[42] L. Gao, G. Iosifidis, J. Huang, L. Tassiulas, and D. Li, "Bargaining-based Mobile Data Offloading", *IEEE JSAC*, vol. 32, no. 6, 2014.

[43] S. Hua, X. Zhuo, and S. Panwar, "A Truthful Auction Based Incentive Framework for Femtocell Access", *in Proc. of IEEE WCNC*, 2013.

[44] H. Xu, J. Jin, and B. Li, "A Secondary Market for Spectrum", *in Proc. of IEEE Infocom*, 2010.

[45] D. Yang, X. Fang, and G. Xue, "Truthful Auction for Cooperative Communications", *in Proc. of ACM Mobihoc*, 2011.

[46] G. Iosifidis, and I. Koutsopoulos, "Double Auction Mechanisms for Resource Allocation in Autonomous Networks", *IEEE JSAC*, vol.28, no.1, 2010.

[47] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, "Algorithmic Game Theory", *Cambridge University Press*, 2007.

[48] M. Simsek, M. Bennis, M. Debbah, and A. Czylwik, "Rethinking Offload: How to intelligently combine Wi-Fi and Small Cells", *in Proc. of ICC*, 2013.

[49] X. Wang, M. Chen, Z. Han, D. Wu, and T. Kwon, "TOSS: Traffic Offloading by Social Network Service-Based Opportunistic Sharing in Mobile Social Networks", *in Proc. of IEEE Infocom*, 2014.

[50] T. Yu, Z. Zhou, D. Zhang, X. Wang, Y. Liu, and S. Lu, "INDAPSON: An Incentive Data Plan Sharing System Based on Self-Organizing Network", *in Proc. of IEEE Infocom*, 2014.

[51] I. Koutsopoulos, E. Noutsi, and G. Iosifidis, "Dijkstra Goes Social: Social-Graph-assisted Routing in Next Generation Wireless Networks", *in Proc. of European Wireless Conf.,* 2014.

[52] S. Paris, F. Martignon, I. Filippini, and L. Chen, "A Bandwidth Trading Marketplace for Mobile Data Offloading", *in Proc. of IEEE Infocom*, 2013.

George Iosifidis obtained the Diploma in Electronics and Telecommunications Engineering from the Greek Air Force Academy, in 2000, and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Thessaly, Greece, in 2007 and 2012, respectively. Currently he is a Post-doc Researcher at the University of Thessaly and CERTH, Greece. His research interests lie in the broad area of network optimization and network economics.

Lin Gao is a Postdoctoral Researcher in the Department of Information Engineering at the Chinese University of Hong Kong. He received the M.S. and Ph.D. degrees in Electronic Engineering from Shanghai Jiao Tong University in 2006 and 2010, respectively. His research interests lie in the field of wireless communications and networking with emphasis on the economic incentives in various communication and network scenarios, including cooperative communications, dynamic spectrum access, cognitive radio networks, cellular-WiFi internetworks, and user-provided networks.

Jianwei Huang (S'01-M'06-SM'11) is an Associate Professor in the Department of Information Engineering at the Chinese University of Hong Kong. He is the recipient of 7 Best Paper Awards in leading international journal and conferences, including the 2011 IEEE Marconi Prize Paper Award in Wireless Communications, and he is the co-author of three recent monographs in wireless network optimization. He is the Editor of IEEE JSAC-Cognitive Radio Series and IEEE Transactions on Wireless Communications, and Chair of IEEE Communications Society Multimedia Communications Technical Committee.

Leandros Tassiulas (S'89-M'91-SM'06-F'07) obtained the Diploma in Electrical Engineering from the Aristotelian University of Thessaloniki, Greece in 1987, and the M.S. and Ph.D. degrees in Electrical Engineering from University of Maryland in 1989 and 1991, respectively. He is Professor in the Dept. of ECE, University of Thessaly, since 2002. Dr. Tassiulas received an NSF Research Initiation Award in 1992, an NSF CAREER Award in 1995, an Office of Naval Research Award in 1997, and a Bodosaki Foundation award in 1999. He also received the INFOCOM 2007 achievement award.