# Optimizing Client Association for Load Balancing and Fairness in Millimeter-Wave Wireless Networks

George Athanasiou, *Member, IEEE, ACM*, Pradeep Chathuranga Weeraddana, *Member, IEEE*, Carlo Fischione, *Member, IEEE*, and Leandros Tassiulas, *Fellow, IEEE*

***Abstract*—Millimeter-wave communications in the 60-GHz band are considered one of the key technologies for enabling multigigabit wireless access. However, the special characteristics of such a band pose major obstacles to the optimal utilization of the wireless resources, where the problem of efficient client association to access points (APs) is of vital importance. In this paper, the client association in 60-GHz wireless access networks is investigated. The AP utilization and the quality of the rapidly vanishing communication links are the control parameters. Because of the tricky nonconvex and combinatorial nature of the client association optimization problem, a novel solution method is developed to guarantee balanced and fair resource allocation. A new distributed, lightweight, and easy-to-implement association algorithm, based on Lagrangian duality theory and subgradient methods, is proposed. It is shown that the algorithm is asymptotically optimal, that is, the relative duality gap diminishes to zero as the number of clients increases.**

***Index Terms*—60-GHz wireless access networks, association control, resource allocation.**

## I. Introduction

MILLIMETER-WAVE (mmW) communications have recently attracted the interest of academia, industry, and standardization bodies, although the technology was invented and used many decades ago, especially in the context of military applications [1], [2]. mmW communications utilize the part of the electromagnetic spectrum between 30 and 300 GHz, which corresponds to wavelengths from 10 to 1 mm [3]. Several promising technologies, including silicon-germanium (SiGe) [4], are emerging as low-cost and low-power solutions for the design of 60-GHz front-end circuits.

Due to the 60-GHz great commercial potential, multiple industry-led efforts and international organizations have emerged for the standardization [5]. Examples include IEEE 802.15.3c [6], IEEE 802.11ad [7], and many others. More than 5 GHz of continuous bandwidth is available in many countries worldwide, making 60-GHz systems attractive for gigabit wireless applications such as gigabyte file transfer and uncompressed high-definition video transmission. Scenarios such as small-cells [8] and mobile data offloading [9], can be accommodated with 60-GHz radio access technology.

The 60-GHz huge bandwidth offers many benefits in terms of capacity and flexibility. For example, even with a low spectral efficiency such as 0.4 b/s/Hz, 60-GHz communication systems can provide a very high data rate of 1 Gb/s [10]. The estimated spectral efficiency for IEEE 802.11n [11] communication systems is 25 b/s/Hz for achieving 1 Gb/s, which makes their application to high-bandwidth services unacceptable in terms of cost and simple implementation, especially in very populated wireless networks [10]. Moreover, allowed transmission powers are higher at 60 GHz than in ultrawideband (UWB) systems [12] and wireless local area networks (WLANs) [13]. However, [14] suggests that 60-GHz radiation with a transmit power described in [10] and [13] does not cause significant harm. Lastly, the interference levels for 60 GHz are much lower compared to the congested 2.4- and 5-GHz bands. The exploitation of these unique characteristics is essential for efficient resource allocation.

In this paper we address the fundamental resource allocation problem of the client association in 60-GHz wireless access networks. Such a problem is more challenging in the 60-GHz band than traditional wireless networks since the wireless channel is unstable in high frequencies and several events can violate the efficient operation of the network, such as moving obstacles that can block the communication [15]. Specifically, we consider the natural situation where each client has to be associated to one of the wireless APs. This gives rise to a challenging mixed integer linear optimization problem, which is combinatorial and nonconvex in general and thus hard to solve efficiently. We show that our problem is NP-hard. Existing methods, such as solution approaches for the *generalized assignment problem* in combinatorial optimization [16, Sec. 8], cannot be used because our main goal is to *minimize the maximum access point (AP) utilization* in the network and ensure a *fair* load distribution, which cannot be modeled by such a *generalized assignment problem*. Nevertheless, based on Lagrangian duality theory [17, Sec. 5] and on subgradient methods, we develop a new solution approach and derive a distributed and iterative algorithm for client association (*DAA*). We show the asymptotic optimality of the proposed solution method by an analytical bound on the duality gap. The sensitivity of the convergence speed of *DAA* to the variation of the numbers of clients and APs is analytically investigated. Numerical simulations illustrate and compare *DAA* to benchmark algorithms.

Unlike the client association approaches in more traditional access networks [18], we take into account the load, the channel quality and the special communication characteristics

Manuscript received February 11, 2013; revised December 17, 2013; accepted February 13, 2014; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor L. Andrew. This work was supported by the Swedish Research Council and the EU project Hydrobionets.

G. Athanasiou, P. C. Weeraddana, and C. Fischione are with the Automatic Control Lab, Electrical Engineering School and Access Linnaeus Center, KTH Royal Institute of Technology, 100-44 Stockholm, Sweden (e-mail: georgioa@kth.se; chatw@kth.se; carlofi@kth.se).

L. Tassiulas is with the Computer and Communication Engineering Department, University of Thessaly, Volos 38221, Greece (e-mail: leandros@uth.gr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNET.2014.2307918

of 60-GHz radio channel and design a dynamic association mechanism that ensures balanced and fair load distribution among the APs. Global methods [19] may be employed to find the solution of the combinatorial client association problem. However, global approaches have the drawbacks of: 1) the prohibitive computational complexity, even in the case of problems with few variables; and 2) they are inherently centralized. In contrast, our proposed method is *fast* and can be implemented in a *distributed* manner. The model and the lightweight algorithm proposed in this work are general and can be applied with different existing MAC mechanisms for 60-GHz access networks, such as IEEE 802.11ad.

The rest of the paper is organized as follows. In Section II, we give a literature overview. The system model and the problem formulation are presented in Section III. In Section IV, we give the general solution approach to the client association problem by using duality theory and subgradient method. In Section V, we describe the properties of the proposed algorithm. In Section VI, numerical results are presented. Lastly, Section VII concludes the paper.

## II. Related Work

During the last decade, resource allocation and in particular association/handoff control for WLANs have been the focus of intense research. In what follows, we review representative literature related to fairness and load balancing in multicell wireless networks, where client association plays a central role. We then discuss their applicability to 60-GHz wireless access networks. A short literature overview where *unique* medium access control and resource allocation problems are studied in mmW networks is given to motivate the need of new approaches that will fully utilize the characteristics of this technology. The section ends with the a summary of our contribution.

The authors in [18] study a client association policy that ensures network-wide max-min fair bandwidth allocation to the users. They provide a rigorous formulation of the association control problem that considers bandwidth constraints of both the wireless and backhaul links. The optimal solution to the aforementioned problem is approximated by an algorithmic approach. The work in [20] presents self-configuring algorithms that provide improved client association and fair resource sharing. The presented approach is based on the Gibbs sampler and does not require explicit coordination among the wireless devices. Moreover, the "multi-homing" scenario is introduced in [21], where the traffic is split among the available APs. In this approach, the throughput is maximized by constructing a fluid model of user population that is multi-homed by the available APs in the network. In [22], the authors study the problem of jointly optimizing partial frequency reuse and load-balancing schemes in multicell networks to achieve network-wide proportional fairness. The expected throughput acts as the client association/handoff decision making metric.

Another line of research considers user service requests by readjusting the load across all APs. In [23], a dual-association approach in wireless mesh networks is presented, where the APs for unicast traffic and the APs for broadcast traffic are independently chosen by exploiting overlapping coverage and optimizing the overall network load. Moreover, in [24] and [25], dynamic association and reassociation procedures are introduced with the use of the notion of the *airtime cost*. The cross-layer extension of this mechanism considers the routing-based information from the mesh backbone. More recently, a class of novel user association schemes that achieve load balancing was proposed in [26]. The authors consider jointly cell association and resource allocation. They formulate a logarithmic utility maximization problem where the equal resource allocation is optimal and design a distributed algorithm via dual decomposition. Complementary, the authors in [27] propose an iterative distributed user association policy that adapts to spatial traffic loads and converges to a globally optimal allocation.

The previous approaches are hard to apply in 60-GHz wireless access networks due to the special characteristics of the 60-GHz channel and the obvious differences with the rest wireless access technologies that we have previously mentioned. It follows that novel mechanisms must be designed to provide optimal resource allocation. These mechanisms must take into account the characteristics of 60-GHz wireless channel such as increased path loss, short range, fragile links, etc. Unfortunately, there is not much research in this field.

Some recent interesting approaches on 60-GHz wireless personal and local area networks have appeared in the literature. The directionality and blockage problems of mmW networks are studied in [15]. A cross-layer approach is presented, where a single-hop transmission is preferred when line of sight (LOS) is available and a relay node is randomly selected as an alternative. In [28], a resource management mechanism is proposed based on the exclusive region (ER) to exploit the spatial reuse of mmW networks. The authors in [29] describe an interference analysis framework that enables a quantitative evaluation of collision loss probability for a mmW mesh network with uncoordinated transmissions, as a function of the antenna patterns and spatial density of simultaneously transmitting nodes. Concurrent transmissions in 60-GHz wireless networks are studied in [30] by exploiting the spatial reuse and time-division multiplexing gain. It is shown that the network throughput is improved compared to single-hop transmission.

The current 60-GHz standardization bodies, such as IEEE 802.11ad, adopt the received signal strength indicator (RSSI)-based mechanism as the basic association functionality. However, high RSSI values cannot univocally indicate high throughput. This is because RSSI not only depends on the distance from the APs, but also on the APs transmission powers. The accuracy of the RSSI-based technique is significantly affected by the high path loss, dispersion, and directionality of the 60-GHz wireless channel. Moreover, since the wireless channel is a shared medium, throughput depends on the population of each cell. An AP may become overloaded if a large number of clients are associated with it. Therefore, new metrics are required.

In contrast to the existing work in literature, this paper considers the special characteristics of the 60-GHz channel in an optimization problem where the objective is to *minimize the maximum AP utilization* in the network. Moreover, we design a lightweight distributed algorithm that balances the AP utilization by optimizing the client association process. We believe that this paper is the first to study such a fundamental resource allocation problem in 60-GHz wireless access networks, especially when complex scenarios with many users and high traffic demands are considered [8], [9]. We propose a simple yet efficient solution and compare it to basic association policies, al-

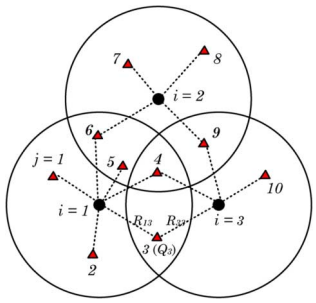



Fig. 1. Wireless access network: $\mathcal{N} = \{1, 2, 3\}$, $\mathcal{M} = \{1, \ldots, 10\}$, $\mathcal{M}_1 = \{1, \ldots, 6\}$, $\mathcal{M}_2 = \{4, \ldots, 9\}$, $\mathcal{M}_3 = \{3, 4, 9, 10\}$, $\mathcal{N}_j = \{1\}$ for $j \in \{1, 2\}$, $\mathcal{N}_j = \{2\}$ for $j \in \{7, 8\}$, $\mathcal{N}_j = \{3\}$ for $j = 10$, $\mathcal{N}_j = \{1, 2\}$ for $j \in \{5, 6\}$, $\mathcal{N}_j = \{1, 3\}$ for $j = 3$, $\mathcal{N}_j = \{2, 3\}$ for $j = 9$, $\mathcal{N}_j = \{1, 2, 3\}$ for $j = 4$. The area inside the solid-lined circles represents the transmission regions of each AP. The demanded data rate for client 3 is $Q_3$, and the offered transmission rate from AP 1 to client 3 is $R_{13}$.

ready in use in the present 60-GHz communication technologies under standardization. The work that is presented and evaluated in the forthcoming sections is complementary to the aforementioned resource management and scheduling approaches (the clients must first be assigned to the available APs, and then the scheduling of the transmission can be handled).

## III. System Model and Problem Formulation

A 60-GHz wireless access network consisting of $N$ APs and $M$ clients is considered. We denote the set of APs by $\mathcal{N} = \{1, \ldots, N\}$ and the set of clients by $\mathcal{M} = \{1, \ldots, M\}$. The set of clients that can be associated to AP $i$ is denoted by $\mathcal{M}_i$. We assume that there are no isolated clients, i.e., $\mathcal{M}_1 \cup \cdots \cup \mathcal{M}_N = \mathcal{M}$. We denote by $\mathcal{N}_j$ the set of candidate APs that client $j$ could be associated with. Fig. 1 shows an example access network, where the clients positioned inside a disc with radius $r$ (centered at the location of AP $i$) can be associated with AP $i$. However, the disk-shaped region is only used for illustrative purposes.

Each node (AP or client) is equipped with steerable directional antennas, and it can direct its beams to transmit or to receive [30]. Note that the antenna separation in 60-GHz multiantenna setups is on the order of millimeters, and therefore thousands of antenna elements can be fabricated in a small space [2], [31], which theoretically accounts for a huge degrees-of-freedom (DoF) gain. Furthermore, we adopt the natural assumption that AP $i$ can support all the clients in $\mathcal{M}_i$ with a separate transmit beam, e.g., multiuser spatial division multiple access (MU-SDMA) techniques [32].[1] We consider the case where receivers are using single-user detection (i.e., a receiver decodes each of its intended signals by treating all other interfering signals as noise) and assume that the achievable rate from AP $i$ to client $j \in \mathcal{M}_i$ is

$$R_{ij} = W \log_2 \left(1 + \frac{P_{ij} G_{ij}}{(N_0 + I_j) W}\right) \tag{1}$$

where $W$ is the system bandwidth, $P_{ij}$ is the transmission power of AP $i$ to client $j$, $G_{ij}$ is the power gain from AP $i$ to client $j$, $N_0$ is the power spectral density of the noise at each receiver, and $I_j$ is the interference spectral density at client $j$. All these assumptions are coherent with the literature and existing standards [6], [7]. The power gain $G_{ij}$ is modeled as in [30]. In particular, we use the Friis transmission equation together with the *flat-top* transmit/receive antenna gain model [36], where a fixed gain is considered within the beamwidth and zero gain is considered outside the beamwidth of the antenna. In addition, we consider Rayleigh small-scale fading. Thus, we have

$$G_{ij} = \frac{G_{ij}^{\mathrm{Tx}} G_{ij}^{\mathrm{Rx}} \lambda^2 \alpha_{ij}}{16\pi^2 \left(\frac{d_{ij}}{d_0}\right)^{\eta}}, \qquad i \in \mathcal{N}, j \in \mathcal{M}_i \tag{2}$$

where $G_{ij}^{\mathrm{Tx}}$ is the transmit antenna gain from AP $i$ to client $j$, $G_{ij}^{\mathrm{Rx}}$ is the receive antenna gain from AP $i$ to client $j$, $\lambda$ is the wavelength, $\alpha_{ij}$ is the *fading coefficient* that is an exponentially distributed random variable with unit mean to model the Rayleigh small-scale fading, $d_{ij}$ is the distance between AP $i$ and client $j$, $d_0$ is the *far field reference distance*, $\eta$ is the path-loss exponent,[2] and $I_j$ is the communication interference at client $j$. We capitalize on the well studied 60-GHz propagation characteristics [29], [36], such as highly directional transmissions with very narrow beamwidths and increased path losses due to the oxygen absorption, in order to assume that the communication interference $I_j$ is very small and does not affect significantly the achievable rates in the network.[3] The achievable communication rates given in (1) are used to define the AP utilizations as described in the sequel.

We denote by $Q_j$ the *demanded data rate* of client $j$. The channel utilization between AP $i$ and client $j$ is denoted by $\beta_{ij}$ and is given by the ratio of $Q_j$ and $R_{ij}$, i.e.,

$$\beta_{ij} = \frac{Q_j}{R_{ij}}. \tag{3}$$

Intuitively, the channel utilization $\beta_{ij}$ gives an indication of the communications performance, in terms of the *potential* loading of the communication channel between AP $i$ and client $j$. Thus, the sum of channel utilizations of AP $i$ (or AP $i$ utilization) is given by $\sum_{j \in \mathcal{M}_i} \beta_{ij} x_{ij}$, where $(x_{ij})_{j \in \mathcal{M}_i}$ are binary decision variables, which indicate the client association. In particular, for all $i \in \mathcal{N}$ and $j \in \mathcal{M}_i$

$$x_{ij} = \begin{cases} 1, & \text{if client } j \text{ is associated to AP } i \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

AP $i$ utilization is a metric reflecting the load. Our goal is to *minimize the maximum AP utilization*. Specifically, the problem can be formally expressed as

$$\text{minimize} \quad \max_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}_i} \beta_{ij}\, x_{ij} \tag{5a}$$

$$\text{subject to} \quad Q_j x_{ij} \leq R_{ij}, \qquad i \in \mathcal{N},\ j \in \mathcal{M}_i \tag{5b}$$

$$\sum_{i \in \mathcal{N}_j} x_{ij} = 1, \qquad j \in \mathcal{M} \tag{5c}$$

$$x_{ij} \in \{0, 1\},\ j \in \mathcal{M},\ i \in \mathcal{N}_j \tag{5d}$$

[1] Techniques, such as combined digital/analog signal processing can be used to design dramatically lower-complexity transceivers trading off the full DoF gain [13], [33]–[35]. However, transceivers to exploit full DoF gain in mmW radios can be quite challenging due to limitations in the current hardware [13], [33].

[2] $\eta \in [2, 6]$ in IEEE 802.11ad networks [37].

[3] In particular, a probabilistic analysis of the interference incurred due to uncoordinated transmissions is presented in [29] and [36]. It is shown that even uncoordinated transmission for different transmit–receive pairs leads to small collision probabilities, and therefore, the links in the network can be considered as *pseudo-wired*. That is, interference can essentially be ignored in MAC or higher layers design. Similar assumptions are also supported by using efficient channel allocation in the network [38] and efficient scheduling algorithms that support concurrent transmissions [30].

where the variable is $(x_{ij})_{i\in\mathcal{N},\, j\in\mathcal{M}_i}$. The main problem parameters are $(\beta_{ij})_{i\in\mathcal{N},\, j\in\mathcal{M}_i}$, $(Q_j)_{j\in\mathcal{M}}$, and $(R_{ij})_{i\in\mathcal{N},\, j\in\mathcal{M}_i}$.[4] The constraint (5b) assures that the demand of client $j$ is less than or equal to the achievable rate from AP $i$ to client $j$. This constraint can usually be satisfied due to the huge available bandwidth of mmW channel. The constraint (5c) ensures that client $j$ is assigned to only one AP. The constraint (5d) indicates that the decision variables are binary.

We note that our approach may be generalized for networks that operate at lower frequencies, given that the interference is constant or fixed. However, in typical lower-frequency multi-cell networks, the interference is heavily dependent on the assignment $\mathbf{x} = (x_{ij})_{i\in\mathcal{N},\, j\in\mathcal{M}_i}$, which makes the application of the proposed method not quite legitimate. Nevertheless, our problem formulation can be general when the important modification that $\beta_{ij}$ be a function of the assignment $\mathbf{x}$ would be introduced. In other words, one has to analytically describe the effect of decision variables $\mathbf{x}$ on $\beta_{ij}$ to generalize the problem formulation, i.e., use $\beta_{ij}(\mathbf{x})$ instead of $\beta_{ij}$. Such a generalization would give a new challenging optimization problem, the solution of which in its own right deserves an independent and substantial investigation.

The client association problem is combinatorial, and we have to rely on exponentially complex global methods [19] to *solve* it, unless new methods are developed. In the sequel, we present one such efficient solution approach, which, although strictly nonoptimal, is asymptotically optimal when $M$ grows.

## IV. Solution via Dual Problem

We start by equivalently reformulating problem (5) into its epigraph form [17, Sec. 4.1.3]. Without loss of generality, we can assume that $\beta_{ij} \leq 1$ for all $i \in \mathcal{N}$ and $j \in \mathcal{M}_i$ for the following reason. If $\beta_{ij} > 1$ for some $i$'s and $j$'s, this means that the data rates demanded by those clients are higher than the data rates achievable in the corresponding wireless channels. In this case, we can collect all the $i, j$ pairs for which $\beta_{ij} > 1$ and modify the corresponding sets $\mathcal{N}_j$ and $\mathcal{M}_i$, without affecting the optimal value of the original problem (5). In particular, if $\beta_{ij} > 1$, then we set $\mathcal{N}_j = \mathcal{N}_j \setminus \{i\}$ and $\mathcal{M}_i = \mathcal{M}_i \setminus \{j\}$.[5] As a result, we ensure that $Q_j \leq R_{ij}$ and hence any feasible $(x_{ij})_{i\in\mathcal{N},\, j\in\mathcal{M}_i}$ that satisfies constraints (5c)–(5d) must also satisfy constraint (5b). That is, constraint (5b) is redundant and can be dropped. Thus, standard equivalent epigraph form of problem (5) is

$$\begin{aligned}
\text{minimize} \quad & t && \text{(6a)}\\
\text{subject to} \quad & \sum_{j\in\mathcal{M}_i} \beta_{ij}\, x_{ij} \leq t, \qquad i \in \mathcal{N} && \text{(6b)}\\
& \sum_{i\in\mathcal{N}_j} x_{ij} = 1, \qquad j \in \mathcal{M} && \text{(6c)}\\
& x_{ij} \in \{0,1\}, \qquad j \in \mathcal{M},\ i \in \mathcal{N}_j && \text{(6d)}
\end{aligned}$$

where the variables are $t$ and $\mathbf{x} = (x_{ij})_{i\in\mathcal{N},\, j\in\mathcal{M}_i}$. We denote by $p^\star$ the optimal value of the problem (6).

Note also that problem (6) is a mixed integer linear program (MILP). Its complexity is established in the following proposition.

*Proposition 1:* Problem (6) is NP-hard.

*Proof:* See Appendix A. ∎

Therefore, the existing solvers are, of course, centralized and are typically based on global branch and bound algorithms, where the worst-case complexity grows exponentially with the problem sizes [17, Sec. 1.4.2]. Even small problems, with a few tens of variables, can take a very long time to be solved. Moreover, problem (6) is different from the *generalized assignment problem* [16, Sec. 8] due to its special structure. In particular, not all the variables of problem (6) are discrete as opposed to the requirement that all variables should be discrete in the case of a generalized assignment problem. Therefore, the existing solution approaches [16, Sec. 8 and 10] for the generalized assignment problem do not apply here. In the sequel, we apply Lagrangian duality to problem (6) to develop a novel solution method that is distributed and fast.

### A. Dual Problem

Let us first form the partial Lagrangian by *dualizing* the first constraints of problem (6). To do this, we introduce multipliers $\boldsymbol{\lambda} = (\lambda_i)_{i\in\mathcal{N}}$ for the first set of inequality constraints. Thus, the partial Lagrangian is given by

$$L(t, \mathbf{x}, \boldsymbol{\lambda}) = t\left(1 - \sum_{i\in\mathcal{N}} \lambda_i\right) + \sum_{j\in\mathcal{M}} \sum_{i\in\mathcal{N}_j} \beta_{ij}\lambda_i x_{ij} \qquad (7)$$

where we used the equivalence of the following two sets:[6]

$$\{(i,j) | i \in \mathcal{N}, j \in \mathcal{M}_i\} \equiv \{(n,m) | m \in \mathcal{M}, n \in \mathcal{N}_m\}\,. \qquad (8)$$

Let $g(\boldsymbol{\lambda})$ denote the dual function obtained by minimizing the partial Lagrangian (7) with respect to $t$ and $\mathbf{x}$. For notational simplicity, let us further denote by $\mathcal{X}$ the set of vectors $\mathbf{x}$ that satisfy the constraints (6c) and (6d) of problem (6). In particular, $\mathcal{X}$ can be expressed as a Cartesian product of some sets $\mathcal{X}_j \subset \mathbb{R}^{n_j},\ j \in \mathcal{M}$, i.e.,

$$\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_M \qquad (9)$$

where $\mathcal{X}_j$ is given by

$$\mathcal{X}_j = \left\{ \mathbf{x}_j = (x_{ij})_{i\in\mathcal{N}_j} \,\middle|\, \sum_{i\in\mathcal{N}_j} x_{ij} = 1,\ x_{ij}\in\{0,1\},\ i\in\mathcal{N}_j \right\}. \qquad (10)$$

Thus, the dual function is

$$\begin{aligned}
g(\boldsymbol{\lambda}) &= \inf_{\substack{t\in\mathbb{R}\\ \mathbf{x}\in\mathcal{X}}} L(t, \mathbf{x}, \boldsymbol{\lambda}) && \text{(11a)}\\
&= \begin{cases} \inf_{\mathbf{x}\in\mathcal{X}} \sum_{j\in\mathcal{M}} \sum_{i\in\mathcal{N}_j} \beta_{ij}\lambda_i x_{ij}, & \sum_{i\in\mathcal{N}} \lambda_i = 1 \\ -\infty, & \text{otherwise} \end{cases} && \text{(11b)}\\
&= \begin{cases} \sum_{j\in\mathcal{M}} \inf_{\mathbf{x}_j\in\mathcal{X}_j} \left( \sum_{i\in\mathcal{N}_j} \beta_{ij}\lambda_i x_{ij} \right), & \sum_{i\in\mathcal{N}} \lambda_i = 1 \\ -\infty, & \text{otherwise} \end{cases} && \text{(11c)}
\end{aligned}$$

[4] In general, client association affects the interference levels in wireless networks. However, when the characteristics of the 60-GHz wireless channel are considered (high oxygen absorption, etc.) we can argue that the interference can be in the order of noise [29], [36], which in turn allows us to suppress the dependence of interference on the client association and consider it fixed [30].

[5] For example, suppose that $\beta_{12} > 1$, and the corresponding sets $\mathcal{N}_2 = \{1, 2\}$ and $\mathcal{M}_1 = \{1, 2, 3\}$. Then, we simply modify $\mathcal{N}_j$ and $\mathcal{M}_i$ as follows: $\mathcal{N}_2 = \{2\}$, $\mathcal{M}_1 = \{1, 3\}$

[6] This equivalence can be visualized by using a bipartite graph, where the nodes are the elements of two disjoint sets, the set of APs (i.e., $\mathcal{N}$) and the set of clients (i.e., $\mathcal{M}$), and the edges are the potential AP-client associations.

$$= \begin{cases} \sum_{j\in\mathcal{M}} g_j(\boldsymbol{\lambda}), & \sum_{i\in\mathcal{N}} \lambda_i = 1 \\ -\infty, & \text{otherwise} \end{cases} \tag{11d}$$

where the equality (11b) follows from that the linear function $t(1-\sum_{i\in\mathcal{N}}\lambda_i)$ is bounded below only when it is identically zero, the equality (11c) follows from (9) and (10), and $g_j(\boldsymbol{\lambda})$ in (11d) is the optimal value of the problem

$$\begin{aligned} \text{minimize} \quad & \sum_{i\in\mathcal{N}_j} \beta_{ij}\lambda_i x_{ij} \\ \text{subject to} \quad & \mathbf{x}_j \in \mathcal{X}_j \end{aligned} \tag{12}$$

with the variable $\mathbf{x}_j$. Even though problem (12) is combinatorial, it has a closed-form solution given by

$$x_{ij}^\star = \begin{cases} 1, & i = \arg\min_{n\in\mathcal{N}_j} \beta_{nj}\lambda_n \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

and is computable very fast.[7] The Lagrange dual problem is

$$\text{maximize} \quad g(\boldsymbol{\lambda}) = \sum_{j\in\mathcal{M}} g_j(\boldsymbol{\lambda}) \tag{14a}$$

$$\text{subject to} \quad \sum_{i\in\mathcal{N}} \lambda_i = 1 \tag{14b}$$

$$\lambda_i \geq 0,\ i \in \mathcal{N} \tag{14c}$$

where the variables is $\boldsymbol{\lambda}$. We denote by $d^\star$ the optimal value of the problem (14) that will be useful later. Note that the Lagrange dual problem (14) is a convex optimization problem, even though the primal problem (6) is *not* convex (see [17, Sec. 5.2]). Let us next focus on the dual problem (14) and its solution method, which allows us to find a good feasible solution to the original problem (6).

### *B. Solving the Dual Problem via Projected Subgradient Method*

The objective function $g(\boldsymbol{\lambda})$ is, in general, a nonsmooth (therefore nondifferentiable) concave function. A common approach to handle such nondifferentiable functions is the subgradient method [39] because gradient-based algorithms cannot be applied. Therefore, the projected subgradient method [39], [40] is used to solve the dual problem (14).

First, we denote by $\mathbf{u}$ a subgradient of $-g$ at a feasible $\boldsymbol{\lambda}$, where $\mathbf{u} = (u_i)_{i\in\mathcal{N}}$. Specifically

$$u_i = -\sum_{j\in\mathcal{M}_i} \beta_{ij} x_{ij}^\star \tag{15}$$

where $x_{ij}^\star$ for all $j \in \mathcal{M}$ and $i \in \mathcal{N}_j$ is obtained as the solution of problem (12) for all $j \in \mathcal{M}$. Thus, the projected subgradient method is given by

$$\boldsymbol{\lambda}^{(k+1)} = P\left(\boldsymbol{\lambda}^{(k)} - \alpha_k \mathbf{u}^{(k)}\right) \tag{16}$$

where $k$ is the current iteration index of the subgradient method, $\alpha_k > 0$ is the $k$th step size,[8] and $P$ is Euclidean projection onto the unit simplex $\Pi = \{\boldsymbol{\lambda} \mid \sum_{i\in\mathcal{N}}\lambda_i = 1,\ \lambda_i \geq 0\}$ (see [40, Exercise 2.1.12]). By employing (16) in an iterative manner, we can *solve* the dual problem (14). However, recovering a primal feasible solution is nontrivial because the original problem (6) is nonconvex. A discussion of these nontrivial issues and how to find a good feasible solution is deferred to Section V-B to maintain a cohesive presentation. Let us next describe how the computation of the solution of problem (14) is performed in a distributed manner.

### *C. Distributed Algorithm for Client Association*

Recall that the dual function $g(\boldsymbol{\lambda})$ is separable among the clients $j \in \mathcal{M}$; see the objective function (14a) of problem (14). Therefore, the subgradient components (15) for the subgradient method (16) can be computed by coordinating the problem (12) for all $j \in \mathcal{M}$. This suggests *DAA*, presented as follows.

**_DAA_: Distributed algorithm for client association**

1. Initialization: The local channel utilizations, i.e., $(\beta_{ij})_{j\in\mathcal{M}_i}$ are given, at every AP $i$. Set subgradient iteration index $k = 1$. Each AP $i$ broadcasts the initial *feasible* prices $\lambda_i^{(k)}$ to its local clients $j \in \mathcal{M}_i$.
2. Every client $j$ sets $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(k)}$ and locally determines its association by solving problem (12). Denote by AP $i_j$ the AP for which $x_{ij} = 1$.
3. Client $j(\in \mathcal{M})$ signals *only* to AP $i_j$ and does not send any signaling to other AP $i$'s, where $i \in \mathcal{N}_j \setminus \{i_j\}$.
4. Every AP $i$ computes $u_i$ by summing $\beta_{ij}$ over the clients $j \in \mathcal{M}_i$, who had signaled in step 3, see (15).
5. Subgradient iteration: APs communicate and form $\mathbf{u}^{(k)}$ by combining each $u_i$ and by performing (16) to compute $\boldsymbol{\lambda}^{(k+1)}$.
6. Stopping criterion: if the stopping criterion is satisfied, STOP. Otherwise, set $k = k+1$, each AP $i$ broadcasts the *feasible* prices $\lambda_i^{(k)}$ to its local clients $j \in \mathcal{M}_i$ and go to step 2.

The first step initializes *DAA*. Step 2 represents the optimization performed in a decentralized fashion by each client for fixed $\boldsymbol{\lambda}$. The optimization at each client $j$ is a very simple operation and is to find the AP $i_j$, where, $i_j = \arg\min_{i\in\mathcal{N}_j}\beta_{ij}\lambda_i$ [see (13)]. Step 3 requires signaling between clients and APs. In particular, each client $j$ signals only to AP $i_j$. This signaling process can be performed very efficiently, e.g., binary signaling. As a result, we have a light protocol between clients and APs. In Step 4, each AP $i$ locally computes $u_i$ (see 15), which is the summation of $\beta_{ij}$ over the client who signaled the AP. Step 5 requires coordination of all APs, which is efficiently accomplished by using high-speed wired connection to the Internet or to an enterprise local network.[9] In particular, APs coordinate to perform (16), which is the projection of a point onto the unit simplex. The result of this operation is given by the solution to a convex optimization problem, which can be carried out efficiently. Step 6 is the stopping criterion for the algorithm. If the stopping criterion is satisfied, *DAA* terminates. Otherwise, the

[7] If $\mathcal{I} = \arg\min_{n\in\mathcal{N}_j}\beta_{nj}(\lambda_n + \mu_{nj})$ is not a singleton, then an arbitrary $i \in \mathcal{I}$ is chosen.

[8] We chose *square summable but not summable* step size (e.g., $\alpha_k = a/k$, where $0 < a < \infty$), which guarantees the asymptotic convergence of the subgradient method [39].

[9] A *fully distributed protocol*, which relies only on local message exchanges, can be designed in a straightforward manner to perform step 5 of *DAA* in a fully distributed manner. In particular, one can employ state-of-the-art alternating direction method of multipliers to perform (16), where the associated solution approach rely on the concepts of sharing [41, Sec. 7.3] combined with extensively studied consensus algorithms for averaging.

algorithm continues in an iterative manner. In practice, a natural stopping criterion would be running it for a fixed number of iterations. It is worth pointing out that, in practice, abrupt changes can occur in the performance of the 60-GHz channel between an AP and an associated clients during *DAA* iterations or after the termination of *DAA* (due to blockage or poor channel quality). As a result, $R_{ij}$'s will be changed, which accounts for changes in $\beta_{ij}$'s. In such situations, the algorithm can continue by using the current $\boldsymbol{\lambda}$ as the prices. Such initializations are known as warm-start strategies in the mathematical optimization community. In this way, we avoid the reexecution of the entire *DAA*.

### D. Distributed Implementation Over Existing Standards

Let us discuss now how the actual implementation of the proposed *DAA* algorithm could be achieved on top of the existing standards, IEEE 802.15.3c and IEEE 802.11ad. The distributed algorithm is performed periodically in the system to ensure the balanced operation of the network. The period of the execution is given by the control messages established by the medium access control protocol, as we see in detail below.

The iterative association algorithm does not have to be executed every time a client initiates an association or a handoff process. We assume that the *newcomer* client follows the association mechanism that IEEE 802.15.3c and IEEE 802.11ad define, based on the RSSI. Then, our algorithm is periodically executed to *correct* possible suboptimal client associations in the network by reallocating the available resources. As mentioned before, the distributed nature of the association algorithm is crucial, in the direction of offloading the APs, and makes good use of the small computational resources that the clients may provide. Both 802.15.3c and 802.11ad define control frames (denoted as beacon frames) that are periodically broadcast by the APs in the network. The APs can utilize these frames to trigger the initialization of *DAA* and carry the required information to the clients. The APs inform their clients about the initialization of *DAA* by setting a special bit into the beacon frame. Thus, the clients are ready to cooperate toward the optimal resource allocation in the network. The information required by the algorithm can be carried in the control frames or piggybacked to the data frames that the APs send to the clients [38]. Moreover, the clients are piggybacking the information in the data frames sent to APs. Thus, the algorithm is executed in perfect harmony with the networking protocols, without interrupting the actual network operation (data communication) and causing extra delays.

## V. Algorithm Properties

In this section, we first show the convergence performance for the proposed algorithm. Then, we show how to recover a *good* primal feasible solution. Next, we highlight some sufficient conditions under which strong duality holds for the MILP (6) followed by a couple of examples. Finally, we show analytically the *asymptotic optimality* of the algorithm, where the *relative duality gap* diminishes to zero as the number of clients in the system grows.

Recall that $p^\star$ is the optimal value of the original MILP (6) and $d^\star$ is the optimal value of the associated dual problem (14). We refer to $p^\star$ as the *primal optimal value*, $d^\star$ as the *dual optimal value*, $(p^\star - d^\star)$ as the *optimal duality gap*, and $(p^\star - d^\star)/p^\star$ as the *optimal relative duality gap*, which are useful in the rest of the paper.

### A. Convergence

*DAA* essentially solves the dual problem (14) by using the projected subgradients method, and the convergence of the algorithm is established by the following proposition.

*Proposition 2:* Let $g_{\text{best}}^{(k)}$ denote the *best* dual objective value found after $k$ subgradient iterations, i.e., $g_{\text{best}}^{(k)} = \max\{g(\boldsymbol{\lambda}^{(1)}), \ldots, g(\boldsymbol{\lambda}^{(k)})\}$. Then, $\forall \epsilon > 0\ \exists n \geq 1$ such that $\forall k\ k \geq n \Rightarrow (d^\star - g_{\text{best}}^{(k)}) < \epsilon$.

*Proof:* The proof is built on the material presented in [39, Sec. 3.2] and [40]. Let us denote by $\boldsymbol{\lambda}^\star$ the optimal solution of dual problem (14). Thus, we have

$$\begin{aligned}
&\left\|\boldsymbol{\lambda}^{(k+1)} - \boldsymbol{\lambda}^\star\right\|_2^2 \\
&= \left\| P\left(\boldsymbol{\lambda}^{(k)} - \alpha_k \mathbf{u}^{(k)}\right) - \boldsymbol{\lambda}^\star \right\|_2^2 && \text{(17a)}\\
&\leq \left\| \left(\boldsymbol{\lambda}^{(k)} - \alpha_k \mathbf{u}^{(k)}\right) - \boldsymbol{\lambda}^\star \right\|_2^2 && \text{(17b)}\\
&= \left\|\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}^\star\right\|_2^2 - 2\alpha_k \mathbf{u}^{(k)T}(\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}^\star) + \alpha_k^2 \left\|\mathbf{u}^{(k)}\right\|_2^2 && \text{(17c)}\\
&\leq \left\|\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}^\star\right\|_2^2 - 2\alpha_k \left(g(\boldsymbol{\lambda}^\star) - g(\boldsymbol{\lambda}^{(k)})\right) + \alpha_k^2 \left\|\mathbf{u}^{(k)}\right\|_2^2 && \text{(17d)}\\
&= \left\|\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}^\star\right\|_2^2 - 2\alpha_k \left(d^\star - g(\boldsymbol{\lambda}^{(k)})\right) + \alpha_k^2 \left\|\mathbf{u}^{(k)}\right\|_2^2 && \text{(17e)}
\end{aligned}$$

where (17a) follows from (16), (17b) follows from that the projection onto unit simplex $\Pi$ always decrease the distance of a point to every point in $\Pi$ and in particular to the optimal point $\boldsymbol{\lambda}^\star$, (17d) follows from the definition of subgradient, i.e., $-g(\boldsymbol{\lambda}^\star) \geq -g(\boldsymbol{\lambda}^{(k)}) + \mathbf{u}^{(k)\mathrm{T}}(\boldsymbol{\lambda}^\star - \boldsymbol{\lambda}^{(k)})$, and (17e) follows from that $d^\star = g(\boldsymbol{\lambda}^\star)$. Recursively applying (17e) and rearranging the terms, we get

$$\begin{aligned}
&2\sum_{l=1}^{k} \alpha_l \left(d^\star - g\left(\boldsymbol{\lambda}^{(l)}\right)\right) \\
&= -\left\|\boldsymbol{\lambda}^{(k+1)} - \boldsymbol{\lambda}^\star\right\|_2^2 + \left\|\boldsymbol{\lambda}^{(1)} - \boldsymbol{\lambda}^\star\right\|_2^2 + \sum_{l=1}^{k} \alpha_l^2 \left\|\mathbf{u}^{(k)}\right\|_2^2 && \text{(18a)}\\
&\leq R^2 + G^2 \sum_{l=1}^{k} \alpha_l^2 && \text{(18b)}
\end{aligned}$$

where (18b) follows from that $\|\boldsymbol{\lambda}^{(k+1)} - \boldsymbol{\lambda}^\star\|_2 \geq 0$, $\|\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^\star\|_2 \leq R = \sqrt{2}$ for any $\tilde{\boldsymbol{\lambda}} \in \Pi$, and the norm of any subgradient $\mathbf{u}$ of $-g$ at any $\tilde{\boldsymbol{\lambda}} \in \Pi$ is bounded, see (15), i.e.,

$$\|\mathbf{u}\|_2 \leq G = \sqrt{\sum_{i \in \mathcal{N}} \left(\sum_{j \in \mathcal{M}_i} \beta_{ij}\right)^2}. \tag{19}$$

Moreover, clearly we have

$$d^\star - g_{\text{best}}^{(k)} \leq d^\star - g\left(\boldsymbol{\lambda}^{(l)}\right), \qquad l = 1, \ldots, k. \tag{20}$$

Thus, from (18b), (20), and by noting that step size $\alpha_l = a/l$, $0 < a < \infty$ is *square summable* (i.e., $\sum_{l=1}^{\infty} \alpha_l^2 = a^2\pi/6$), we obtain an upper bound on $d^\star - g_{\text{best}}^{(k)}$ as

$$d^\star - g_{\text{best}}^{(k)} \leq \frac{R^2 + G^2 \sum_{l=1}^{k} \alpha_l^2}{2\sum_{l=1}^{k} \alpha_l} \leq \frac{\frac{R^2}{2} + \frac{a^2 G^2 \pi}{12}}{\sum_{l=1}^{k} \alpha_l}. \tag{21}$$

Since $\sum_{l=1}^{k} \alpha_l = a \sum_{l=1}^{k} (1/l)$ is strictly monotonically increasing in $k$ (it grows without bound as $k \to \infty$), for any $\epsilon > 0$ we can always find a integer $n \geq 1$ such that $\sum_{l=1}^{k} \alpha_l >$

$(R^2/2 + a^2G^2\pi/12)/\epsilon$ if $k \geq n$. For example, by noting that $\sum_{l=1}^{k}(1/l) > \log(k+1)$ for all $k = 1, 2, \ldots$, we can compute $n = \exp((R^2/2 + a^2G^2\pi/12)/a\epsilon) - 1$. ■

The bound derived in (21) allows us to predict some key behaviors of the convergence of the proposed algorithm. To see this, we note from (19) that the numerator of the bound depends on $(\beta_{ij})_{i\in\mathcal{N},\ j\in\mathcal{M}_i}$ for fixed $a$. Now suppose that the number of clients increases. This forces $G$ to increase as well. As a result, the corresponding total iterations to reach the given accuracy $\epsilon$ will also grow. Roughly speaking, this means that, for fixed number of APs, the larger the number of clients is, the larger the total number of iterations required for the convergence of *DAA*. On the other hand, suppose that the user distribution is such that $\sum_{j\in\mathcal{M}_i}\beta_{ij}$ is roughly the same for each AP $i$. Thus, if the total number of APs is increased, then $G$ will become larger, and as a result, the total number of iterations to convergence is increased as well. These algorithm behaviors are numerically illustrated in Section VI.

### B. Recovering a Feasible Primal Point

As we discussed in Section V-A, we can *solve* the dual problem to any given accuracy to yield the dual optimal value $d^\star$ and the dual optimal solution $\boldsymbol{\lambda}^\star$. If the primal problem is convex, from $d^\star$ and $\boldsymbol{\lambda}^\star$, we can usually obtain the primal optimal value $p^\star$ and primal optimal solution $(t^\star, \mathbf{x}^\star)$ [17, Sec. 5.5.5]. However, recall that the original MILP (6) is a *nonconvex* problem. Therefore, unlike convex problems, there is *no guarantee* that from $d^\star$ and $\boldsymbol{\lambda}^\star$, we obtain $p^\star$ and $(t^\star, \mathbf{x}^\star)$.[10] Nevertheless, in the case of MILP (6), a *primal feasible point* is obtained during each iteration $k$ of the algorithm (see step 2 of the algorithm). Thus, it is natural to go for the best choice, among all the primal feasible points obtained so far. For example, a good approximation for the primal optimal value would be

$$p_{\text{best}}^{(k)} = \min\left\{t^{(1)}, \ldots, t^{(k)}\right\} \tag{22}$$

where $(t^{(k)},\ \mathbf{x}^{(k)})$ is the primal feasible point in iteration $k$.[11] A good feasible point would be $(t_{\text{best}}^{(k)}, \mathbf{x}_{\text{best}}^{(k)})$, which is the primal feasible point that corresponds to $p_{\text{best}}^{(k)}$.

Even though the value $p_{\text{best}}^{(k)}$ is not usually as good as the primal optimal value $p^\star$, Monte Carlo simulations show that it is a good approximate value with $k$ in the order of hundreds or more, e.g., $k \geq 100$ (see Section VI). There are no clear analytical explanations of these fortuitous encounters, especially because the original problem (6) is nonconvex [40, Sec. 6.3].

### C. Duality Gap

The *duality gap* $(p^\star - d^\star)$ is one of the important metrics that can be used to quantify the performance of the proposed *DAA*. Note that, in general, the duality gap for MILP (6) is not zero because the problem is nonconvex and is in fact NP-hard. Therefore, it is not surprising that deriving general conditions under which the strong duality for MILP (6) is very difficult. Nevertheless, we first provide some examples to highlight sufficient conditions for strong duality for MILP (6). The latter part of this section derives an analytical bound on the duality gap. Moreover, we show the *asymptotic optimality* of the algorithm, where the *relative duality gap* $(p^\star - d^\star)/p^\star$ diminishes to zero as the number of clients in the system grows. Such asymptotic results are important from a theoretical and from a practical perspective; see for example the duality results associated with the well known Knapsack problem [40].

The following proposition establishes a simple result, which is instrumental to study zero duality.

*Proposition 3:* Let $p_{\text{relax}}^\star$ denote the optimal value of the linear programming (LP) relaxation of problem (6), i.e.,

$$\begin{aligned}
\text{minimize} \quad & t && && \text{(23a)}\\
\text{subject to} \quad & \sum_{j\in\mathcal{M}_i}\beta_{ij}\,x_{ij} \leq t, && i\in\mathcal{N} && \text{(23b)}\\
& \sum_{i\in\mathcal{N}_j} x_{ij} = 1, && j\in\mathcal{M} && \text{(23c)}\\
& 0 \leq x_{ij} \leq 1,\ j\in\mathcal{M}, i\in\mathcal{N}_j && && \text{(23d)}
\end{aligned}$$

with variables $t$ and $\mathbf{x} = (x_{ij})_{i\in\mathcal{N},\ j\in\mathcal{M}_i}$. Then, $d^\star = p_{\text{relax}}^\star$.

*Proof:* Note that problem (6) is always feasible. The proof is based on two key results: (a) for problem (23), we always have strong duality, i.e., $p_{\text{relax}}^\star = d_{\text{relax}}^\star$; and (b) the dual of problem (23) is identical to problem (14), and therefore $d_{\text{relax}}^\star = d^\star$, where $d_{\text{relax}}^\star$ denotes the dual optimal value of problem (23). In particular, (a) is guaranteed from strong duality results for linear programs; see [17, Sec. 5.2.3]. To prove (b), we first note the following: The partial Lagrangian obtained by dualizing first constraint of problem (23) is identical to (7), where we consider same notations for the dual variables. This is a slight abuse of notation, but it helps in the clarity of the exposition. Let $h(\boldsymbol{\lambda})$ denote the dual function obtained by minimizing the partial Lagrangian [compare to (11)]

$$h(\boldsymbol{\lambda}) = \begin{cases} \sum\limits_{j\in\mathcal{M}} h_j(\boldsymbol{\lambda}), & \sum_{i\in\mathcal{N}}\lambda_i = 1 \\ -\infty, & \text{otherwise} \end{cases} \tag{24}$$

where $h_j(\boldsymbol{\lambda})$ is the optimal solution of a problem very similar to (12), except that the constraint $\mathbf{x}_j \in \mathcal{X}_j$ is replaced by $\mathbf{x}_j \in \text{conv}(\mathcal{X}_j)$. Note that $\text{conv}(\mathcal{X}_j) \in \mathbb{R}^{n_j}$ is a *unit simplex*, and therefore the optimal value $h_j(\boldsymbol{\lambda})$ is attained at one of the vertices $\mathbf{x}_j = (x_{ij})_{i\in\mathcal{N}_j}$ of the unit simplex [42, Corollary 32.3.4]. Specifically, the components of the vertex $\mathbf{x}_j$ are identically given by (13). As a result, $h_j(\boldsymbol{\lambda}) = g_j(\boldsymbol{\lambda})$ for all $j \in \mathcal{M}$ and $h(\boldsymbol{\lambda}) = g(\boldsymbol{\lambda})$. ■

Note that if problem (23), the LP relaxation of problem (6), has integer solutions, then we can easily show that the optimal value $p^\star$ of the original MILP (6) is equal to the optimal value $p_{\text{relax}}^\star$, i.e., $p^\star = p_{\text{relax}}^\star$. Therefore, from Proposition 3, we have $p^\star = d^\star$. In other words, if problem (23) has integer solutions, then *strong duality* holds for the original MILP (6). Thus, it is natural to investigate the conditions, under which the LP (23) has integer solutions. Roughly speaking, there are not many results that establish conditions on LPs, beyond *total unimodularity* [43, Sec. 9] of the associated problem matrices, under which they have integer solutions. Unfortunately, particularized to LP (23), the related matrices are not totally unimodular, and therefore the theoretical implications of total unimodularity does not apply to [43, Sec. 9]. Nevertheless, Proposition 3 allows us to imagine special cases of MILP (6), where we have strong duality. We now present two examples.

[10] The first component $t^\star$ of the primal optimal solution of MILP (6) and the primal optimal value $p^\star$ are clearly the same.

[11] APs can compute $(t^{(k)},\ \mathbf{x}^{(k)})$ easily by using the client signaling they received at step 2 and the AP coordination at step 5. For example, $t^{(k)} = \|-\mathbf{u}^{(k)}\|_\infty$.

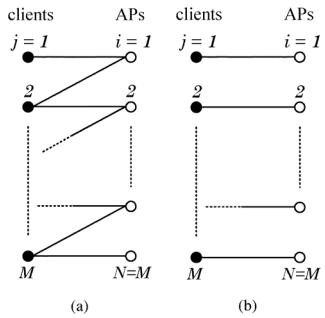



Fig. 2. Examples of network topology 1. (a) Initial communication graph. (b) Optimal association.

*Example 1:* Suppose that the initial clients and APs communication is as shown in Fig. 2(a). In particular, there are $M$ clients and $N = M$ APs. Each client $j \in \{2, 3, \ldots, M\}$ can be associated to either AP $j-1$ or $j$. However, the first client can be associated only to the AP 1. Moreover, suppose that $\beta_{jj} = \beta \in (0, 1]$ for all $j \in \{1, \ldots, M\}$ and the remaining $\beta_{ij}$s can be arbitrary values from the range $(0, 1]$.

We can easily show that the optimal client association (i.e., the optimal solution of MILP (6)) corresponds to Fig. 2(b). Moreover, we have $p^\star = \beta$. Let us now focus to the LP relaxation (23) applied to the network in Fig. 2(a). In this case, every client $j$, except the client 1, can be associated to both AP $j-1$ *and* AP $j$. However, we can prove by contradiction: The solution of problem (23) corresponds to the optimal association depicted in Fig. 2(b). Thus, we have $p^\star = p^\star_{\text{relax}}$. From Proposition 3, we have strong duality for MILP (6) (i.e., $p^\star = d^\star$), and the proposed algorithm achieves $d^\star$.

*Example 2:* Consider the network shown in Fig. 3(a). There are $N$ APs and two types of clients connected to APs. The Type-1 clients can communicate only with a *single* AP. For example, each client $j \in \{1, \ldots, J_1\}$ can communicate only with AP 1, and therefore, they must be assigned to AP 1. On the other hand, the Type-2 clients can communicate with all the APs, e.g., clients $J_N + 1, \ldots, M$. As a result, Type-2 clients can be associated to any AP. Now suppose that $\beta_{ij}$ values associated with Type-1 clients are such that $\sum_{j=1}^{J_1} \beta_{1j} = \sum_{j=J_1+1}^{J_2} \beta_{2j} \cdots = \sum_{j=J_{N-1}+1}^{J_N} \beta_{Nj} = B \in \mathbb{R}_+$ and $\beta_{ij}$ values associated with Type-2 clients are all equal to $\beta \in (0, 1]$. Moreover, suppose that the number of Type-2 clients is a multiple of $N$, i.e., $M - J_N = mN$ for some $m \in \mathbb{Z}_+$.

The optimal client association [i.e., the optimal solution of MILP (6)] corresponds to Fig. 3(b), where the Type-2 clients are equally distributed among the APs. In particular, each AP is associated with $m$ Type-2 clients, and we have $p^\star = B + m\beta$. Let us now consider the solution given by the LP relaxation (23). Note that there is no choice for Type-1 clients, other than associating them to the only AP they can communicate. From the symmetry, we can easily see that associating each Type-2 client $j$ among all the APs $i = 1, \ldots, N$ with equal shares is a particular solution to the LP relaxation (23), i.e., for every Type-2 client $j$, $x_{ij} = (1/N)$, $i = 1, \ldots, N$. Thus, at each AP $i$, the utilization corresponding to Type-1 clients is $B$ and the utilization corresponding to Type-2 clients becomes $(M - J_N)(1/N)$, which is identical to $m$ (recall $M - J_N = mN$). Therefore, we have $p^\star_{\text{relax}} = (B + m) = p^\star$, and strong duality holds for MILP (6).

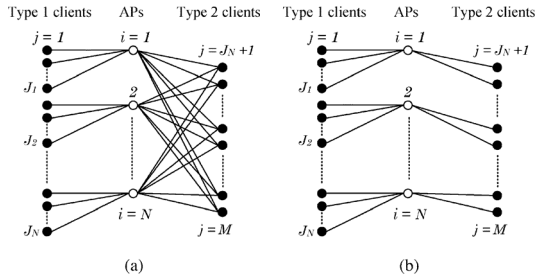



Fig. 3. Examples of network topology 2. (a) Initial communication graph. (b) Optimal association.

The examples above give insights into the algorithm's behavior in spacial cases. However, they can be used to build intuitive ideas of general networks. It is indeed important to analyze the proposed algorithm's behavior in general as well. In the sequel, we provide theoretical substantiation that allows us to predict the general algorithm properties in terms of the optimal duality gap and the relative duality gap.

The following theorem formally establishes a bound on the duality gap and the asymptotic optimality of *DAA*.

*Theorem 1:* The optimal duality gap of mixed integer linear program (6) is bounded as follows:

$$p^\star - d^\star \leq (N+1)\left(\varrho + \max_{j \in \mathcal{M}} \varrho_j\right) \tag{25}$$

where $\varrho = \max_{i \in \mathcal{N}, j \in \mathcal{M}_i} \beta_{ij}$ and $\varrho_j = \min_{i \in \mathcal{N}_j} \beta_{ij}$. Moreover, the *relative* duality gap diminishes to 0 as $M \to \infty$.

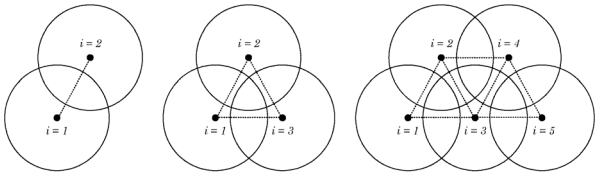
*Proof:* See Appendix B. ■

The theorem suggests that the duality gap is always bounded by a *constant* that does not depend on the number of clients in the system. Note that the bound grows like $N$, and therefore we can expect an increase in the duality gap for larger $N$. The theorem states that for larger $M$, the relative duality gap become almost zero.

These are very important to get an insight of the behavior of the proposed algorithm in general networks. For example, when $M$ is sufficiently large enough (though not infinity), Theorem 1 can still be of interest, and its importance can be justified as follows. The antenna separation in 60-GHz multiantenna setups is in the order of millimeters, and therefore thousands of antenna elements can be fabricated in a small space [31]. Hence, there is a huge degrees-of-freedom gain that can be achieved by employing multiantennas at access points. Moreover, state-of-the art multiple access schemes such as space-division multiple access (SDMA) can be readily applied for gigabits-per-second client–access point communication. In this context, Theorem 1 can be gracefully used to predict the algorithm's behavior; see Section VI for numerical examples.

## VI. Numerical Examples

In this section, we present the numerical evaluation of the proposed algorithm in a multiuser multicell environment. We compare *DAA* to: 1) random association policy; 2) RSSI-based policy, which is the association mechanism used in standards;




(a) (b) (c)

Fig. 4. Example simulation topologies: (a) 2 APs; (b) 3 APs; (c) 5 APs.

and 3) optimal solution of the optimization problem (6) using IBM CPLEX optimizer [44].

We define the SNR operating point at a distance $d$ [distance units] from any AP as

$$\mathtt{SNR}(d) = \begin{cases} \frac{P_0 \lambda^2}{(16\pi^2 N_0 W)}, & d \le d_0 \\ P_0 \lambda^2 / (16\pi^2 N_0 W) \cdot \left(\frac{d}{d_0}\right)^{-\eta}, & \text{otherwise.} \end{cases}$$

Circular cells as depicted in Fig. 1 are considered, where the radius of each cell $r$ is chosen such that $\mathtt{SNR}(r) = 10$ dB. The APs are located such that the distance between any consecutive APs is $D = 1.1r$. For example, Fig. 4(a) shows the case for $N = 2$ APs, Fig. 4(b) shows the case for $N = 3$, and Fig. 4(c) shows the case for $N = 5$. The clients are uniformly distributed among the circular cells and the potential AP–client association (i.e., $\mathcal{M}_i$ and $\mathcal{N}_j$); see Fig. 1.

We set $\lambda = 5$ mm, $N_0 = -134$ dBm/MHz, $W = 1200$ MHz, and $d_0 = 1$ m; see (1) and (2). Moreover, for all $j \in \mathcal{M}$, we set $I_j = 0$, and for all $i \in \mathcal{N}$, $j \in \mathcal{M}_i$, we set $P_{ij} = P_0 = 0.1$ mW and $G_{ij}^{\mathrm{Tx}} = G_{ij}^{\mathrm{Rx}} = 1$. In order to check the average performance of the algorithms, we consider $\bar{T} = 1000$ time-slots, where the fading coefficients $\alpha_{ij}$ for all $i \in \mathcal{N}$, $j \in \mathcal{M}_i$ and the demanded data rates $Q_j$ for all $j \in \mathcal{M}$ are constant during each time-slot $T \in \{1, \ldots, \bar{T}\}$ and independently change from one slot to another. In particular, the exponential random variables $\alpha_{ij}$ with unit mean are independent and identically distributed over the time-slots. Moreover, we assume that $Q_j$ are uniformly distributed on [0,400] Mb/s and independent and identically distributed over the time-slots.

To see the average convergence behavior of the proposed algorithm, we consider the average *primal* objective value of problem (6) obtained by *DAA*. In particular, the average primal objective value from *DAA* after $k$ subgradient iterations is defined as $\mathtt{P}^{(k)} = (1/\bar{T}) \sum_{T=1}^{\bar{T}} p_{\mathrm{best}}^{(k)}(T)$, where $p_{\mathrm{best}}^{(k)}(T)$ is the best *primal feasible* objective value of problem (6) after $k$ iterations at time-slot $T$; see (22).[12] The average objective values from benchmark algorithms, random association policy, RSSI policy, and the optimal policy are defined in a similar manner and are denoted by $\mathtt{P}_{\mathtt{rand}}$, $\mathtt{P}_{\mathtt{RSSI}}$, and $\mathtt{P}^{\star}$, respectively.[13] Moreover, the average dual optimal value obtained by *DAA* is $\mathtt{D}^{\star} = (1/\bar{T}) \sum_{T=1}^{\bar{T}} d^{\star}(T)$, where $d^{\star}(T)$ is the optimal objective value of dual problem (14) at time-slot $T$.

Fig. 5 shows $\mathtt{P}^{(k)}$ versus subgradient iteration $k$, for the cases where $N = 5$, $M = 100$ [Fig. 5(a)] and $N = 5$, $M = 200$ [Fig. 5(b)]. Results show that there is a noticeable effect of varying $M$ (the number of clients in the system) on convergence time. In particular, the convergence is faster for smaller $M$. This observation is consistent with our analytical study presented in Section V-A. Proposed *DAA* clearly outperforms the RSSI policy used in 802.11, 802.15.3c, and 802.11ad, as well as the random policy. For example, *DAA* yields a performance improvement of about 20% compared to RSSI policy in both considered cases. The gap between $\mathtt{P}^{\star}$ and $\mathtt{P}^{(k)}$, even after a relatively larger number of subgradient iterations (e.g., $k = 300$) is indeed expected due to the nonconvexity of the original problem (5); see Section V-B. Nevertheless, average dual optimal value $\mathtt{D}^{\star}$ from *DAA* is almost equal to the optimal $\mathtt{P}^{\star}$.

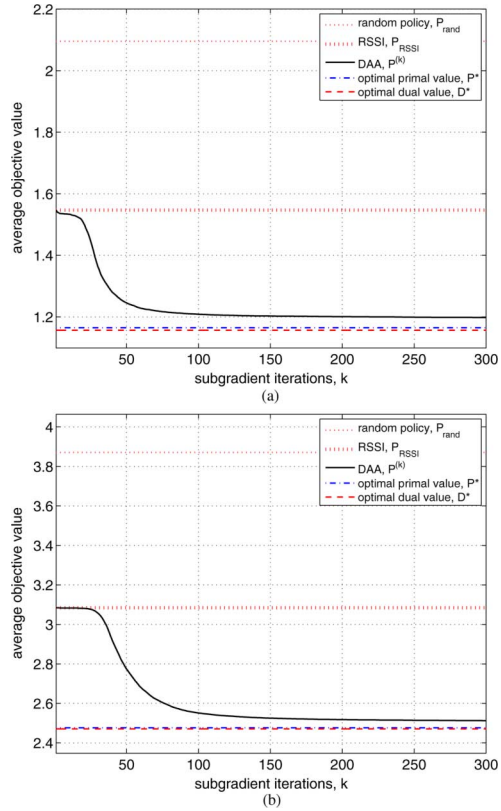



Fig. 5. Influence of the number of clients on the convergence. (a) Average objective value $\mathtt{P}^{(k)}$ versus iterations $k$, 5 APs, 100 clients. (b) Average objective value $\mathtt{P}^{(k)}$ versus iterations $k$, 5 APs, 200 clients.

Fig. 6 shows $\mathtt{P}^{(k)}$ versus subgradient iteration $k$, for the cases where $N = 3$, $M = 30$ [Fig. 6(a)[ and $N = 10$, $M = 100$ [Fig. 6(b)]. Here, the clients density or the number of clients per AP is roughly the same (i.e., 10). Results show that there is a clear effect of varying $N$ (while keeping client density fixed) on convergence. In particular, the convergence is faster for smaller $N$. This observation is in line with our analytical study presented in Section V-A. The performances of other benchmark algorithms are very similar to those in Fig. 5.

Fig. 7 shows the average objective from *DAA* after $K = 1000$ subgradient iterations, $\mathtt{P}^{(K)}$ versus the number of clients $M$ for the cases $N = 2$ [Fig. 7(a)] and $N = 10$ [Fig. 7(b)]. Plots for the benchmark algorithms are also depicted. Results show that the average objective values associated with each algorithm increase as $M$ increases. This is intuitively expected because APs become more loaded as the number of clients grows. Results further show that the $\mathtt{P}^{(K)}$ that came from *DAA* are very close to the optimal $\mathtt{P}^{\star}$ in both cases, and the performance gap is not sensitive to changes in $M$. Note that the performance of the random and the RSSI policies are substantially low, and their performance degradation becomes even noticeable for larger $M$; see

[12]Due to the nonconvexity of the original problem (5), $p_{\mathrm{best}}^{(k)}(T)$ does not necessarily achieve the optimal value even when $k \to \infty$.

[13]Since the benchmark algorithms do not depend on subgradient iteration $k$, like $\mathtt{P}^{(k)}$, there is no superscript $(k)$ required for $\mathtt{P}_{\mathtt{rand}}$, $\mathtt{P}_{\mathtt{RSSI}}$, and $\mathtt{P}^{\star}$.

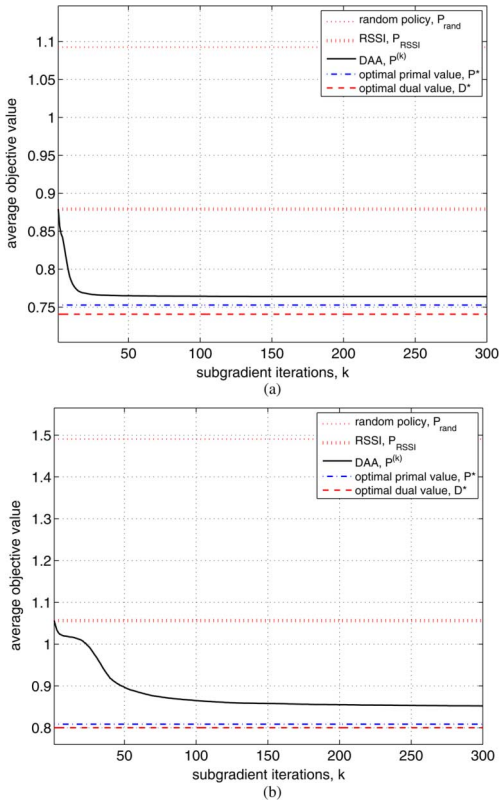



Fig. 6. Influence of the number of APs on the convergence. (a) Average objective value $\mathtt{P}^{(k)}$ versus iterations $k$, 3 APs, 30 clients. (b) Average objective value $\mathtt{P}^{(k)}$ versus iterations $k$, 10 APs, 100 clients.

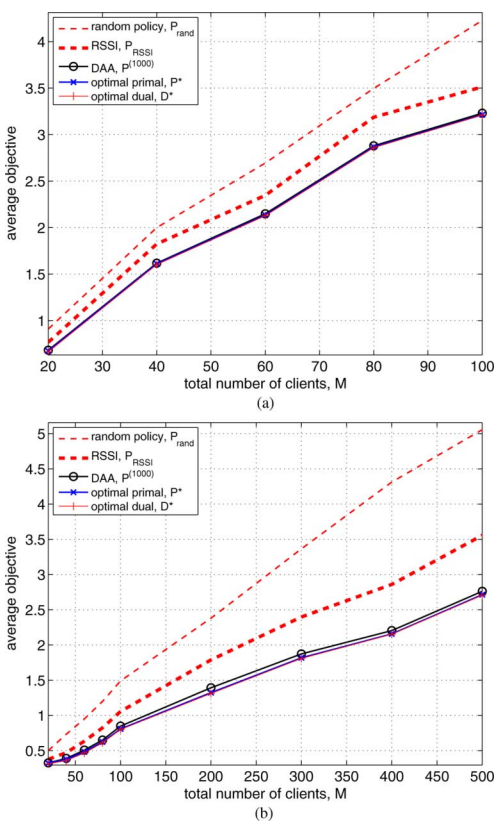



Fig. 7. Average objective values $\mathtt{P}_{\mathtt{rand}}, \mathtt{P}_{\mathtt{RSSI}}, \mathtt{P}^{\star}, \mathtt{P}^{(1000)}$ and average dual optimal $\mathtt{D}^{\star}$ versus the number of users $M$. (a) 2 APs. (b) 10 APs.

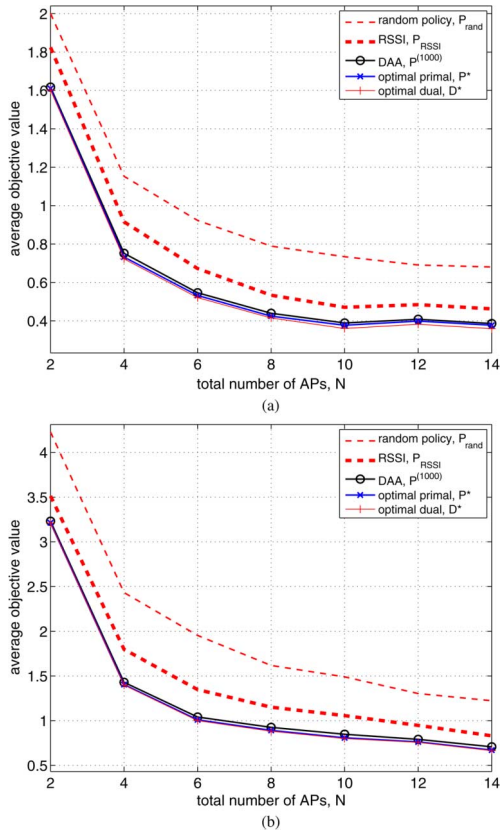



Fig. 8. Average objective value $\mathtt{P}_{\mathtt{rand}}, \mathtt{P}_{\mathtt{RSSI}}, \mathtt{P}^{\star}, \mathtt{P}^{(1000)}$ and average dual optimal $\mathtt{D}^{\star}$ versus the number of APs $N$. (a) 40 users. (b) 100 users.

Fig. 7(b). As expected $\mathtt{D}^{\star}$ provides a global lower bound on the performance and is hardly distinguishable from $\mathtt{P}^{\star}$.

Fig. 8 shows the average objective value versus the number of APs $N$ for the cases where $M = 40$ [Fig. 8(a)] and $M = 100$ [Fig. 8(b)]. The performance ranking of the algorithms is very similar to Fig. 7. Results show that the average objective values decrease as $M$ increases. This is intuitively explained by noting that the larger the $N$ is, the smaller the client density is, and therefore the smaller the average objective values of each algorithm becomes. Results again show that *DAA* performs close to the optimal and outperforms the random and RSSI policies.

To see the effect of the increasing number of clients on the *relative duality gap* (see Section V-C), we define the metric *average relative duality gap*, $\mathtt{Ave-RDG}$. In particular, $\mathtt{Ave-RDG} = (1/\bar{T})\sum_{T=1}^{\bar{T}}(p^{\star}(T) - d^{\star}(T))/p^{\star}(T)$, where $p^{\star}(T)$ is the optimal value of primal problem (6) and $d^{\star}(T)$ is the optimal value of dual problem (14), at time-slot $T$. Moreover, we denote by $\mathtt{Ave-RDG-best-achieved}$, a related metric defined similar to $\mathtt{Ave-RDG}$, except that $p^{\star}(T)$ is replaced with $p_{\text{best}}^{(K)}(T)$, i.e., the best *primal feasible* objective value achieved from *DAA* after $K$ iterations at time-slot $T$.

Fig. 9(a) measures the percentage $\mathtt{Ave-RDG}$ versus $M$ for different $N$'s. For all considered $\mathtt{Ave-RDG}$, $N$ approaches to zero as $M$ increases. This is consistent with our analytical results established by Theorem 1; see Section V-C. Fig. 9(b) shows that the plots of percentage $\mathtt{Ave-RDG-best-achieved}$ versus $M$ are very similar to those in Fig. 9(a). This behavior is not surprising because the performance of *DAA* is close to the optimal performance; see Figs. 7 and 8.

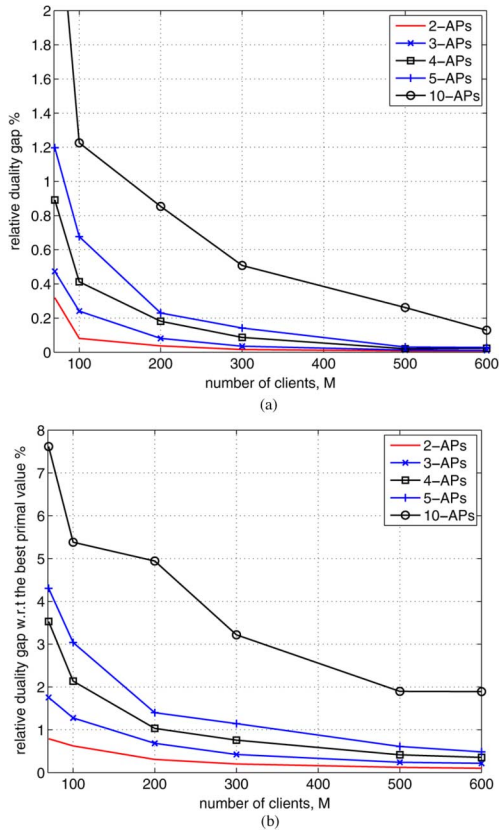



Fig. 9. (a) Average relative duality gap `Ave−RDG` versus the number of clients $M$. (b) Average best achieved relative duality gap `Ave−RDG−best−achieved` versus the number of clients $M$.

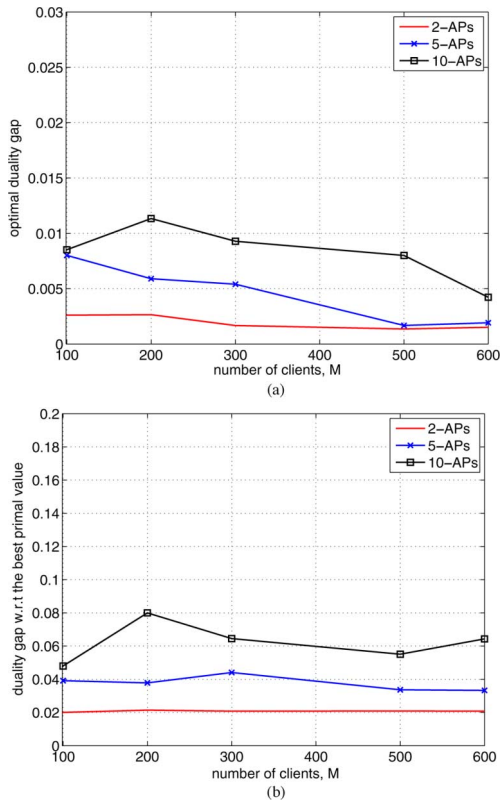



Fig. 11. (a) Average optimal duality gap `Ave−DG` versus the number of clients $M$. (b) Best achieved average duality gap `Ave−DG−best−achieved` versus the number of clients $M$.

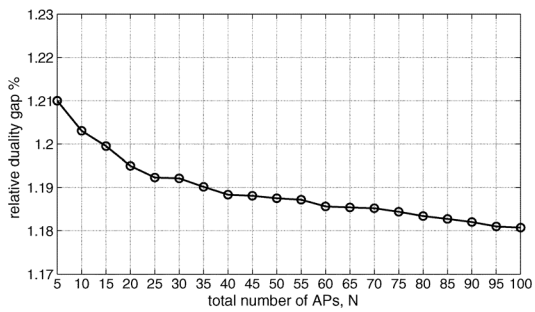



Fig. 10. Average relative duality gap `Ave−RDG` versus the number of APs $N$, where the ratio $M/N = 10$ is kept fixed.

Fig. 10 shows the percentage `Ave−RDG` versus $N$, while the ration $M/N$ is kept fixed. Results show that `Ave−RDG` decreases as the number of nodes in the network increases. Even though the behavior cannot be analytically substantiated, it is intuitively expected because the larger the size of the network, the greater the number of possible client–AP assignment options, which in turn can be exploited for better performance.

Fig. 11 depicts the dependence of the average duality gap on $M$. In particular, we define the *optimal average duality gap* as $\mathtt{P}^\star - \mathtt{D}^\star$, and we plot `Ave−DG` versus $M$ as shown in Fig. 11(a). Moreover, we define the *best achieved average duality gap* `Ave−DG−best−achieved` as $\mathtt{P}^{(K)} - \mathtt{D}^\star$. The corresponding plots are shown in Fig. 11(b). In both cases, results show that there is no apparent effect of the varying $M$ on the duality gap. Nevertheless, as we discussed in Section V-C [see (25)], the average duality gap grows when $N$ increases.

To examine the *fairness* of the final client association among the APs, we consider the well-known Jain's fairness index [45] as the fairness metric. We denote by $J^{(k)}(T)$ the fairness level resulted by the proposed *DAA* at time-slot $T$ after $k$ iterations. In particular, we define $J^{(k)}(T) = \left(\sum_{i\in\mathcal{N}} Y_i^{(k)}(T)\right)^2 / (N \sum_{i\in\mathcal{N}} Y_i^{(k)}(T)^2)$, where $Y_i^{(k)}(T) = \sum_{j\in\mathcal{M}_i} \beta_{ij} x_{ij}^{(k)}(T)$, with $x_{ij}^{(k)}(T)$ being the best feasible solution resulted from *DAA* at time-slot $T$ and after $k$ iterations. The average fairness index $J^{(k)}$ resulted from *DAA* after $k$ iterations is simply defined as $J^{(k)} = (1/\bar{T}) \sum_{T=1}^{\bar{T}} J^{(k)}(T)$. The average fairness indexes resulted from the benchmark algorithms, the random association policy, the RSSI policy, and the optimal policy are defined in a similar manner and are denoted by $\mathtt{J}_{\mathtt{rand}}$, $\mathtt{J}_{\mathtt{RSSI}}$, and $\mathtt{J}^\star$, respectively.[14]

Fig. 12 depicts $J^{(k)}$ versus $k$ compared to the benchmark fairness indexes $\mathtt{J}_{\mathtt{rand}}$, $\mathtt{J}_{\mathtt{RSSI}}$, and $\mathtt{J}^\star$ for the case where $N = 5$ and $M = 100$. Note that the fairness index ranges from $1/N$ (worst performance) to 1 (best performance). As expected, the optimal association provides the best performance. Results show that within a few hundreds of iterations, *DAA* achieves a fairness level very close to the optimal. Results further show that *DAA* significantly outperforms the random and the RSSI policies.

[14]Benchmark algorithms do not depend on subgradient iteration $k$. Therefore, like $\mathtt{J}^{(k)}$, there is no superscript $(k)$ required for $\mathtt{J}_{\mathtt{rand}}$, $\mathtt{J}_{\mathtt{RSSI}}$, and $\mathtt{J}^\star$.

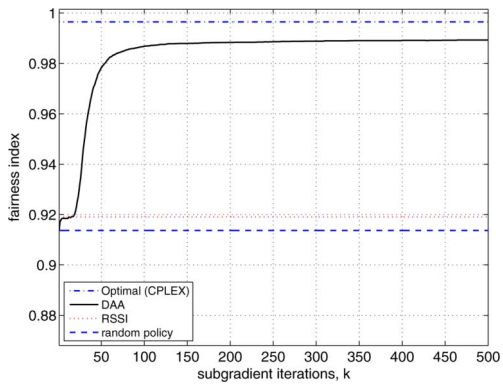



Fig. 12. Average fairness index versus the iterations $k$, 5 APs, 100 clients.

In order to provide a statistical description of the speed of the proposed algorithm, we consider empirical the cumulative distribution function (CDF) plots. Specifically, for each time-slot $T \in \{1, \ldots, \bar{T}\}$, we store the total CPU time required for *DAA* to find $p^{(K)}_{\text{best}}(T)$. For comparison, we use the total CPU time required to find the *optimal* value $p^{\star}(T)$. Fig. 13(a) shows the empirical CDF plots of the number of iterations for $M = 100, 200, 300$, with $N = 10$. In the case of *DAA*, the effect of changing the problem size by increasing $M$ on the CDF plots is almost indistinguishable. However, in the case of optimal method, there is a prominent increase in the time required to compute $p^{\star}(T)$. Fig. 13(b) depicts the average time required by *DAA* and the optimal method versus $M$. Results show that the average time required by *DAA* to find possibly a suboptimal solution is not sensitive to the variation of $M$ and is almost zero. However, the average time required by the optimal method to find the optimal solution grows approximately exponentially with $M$. This is certainly expected because problem (6) is combinatorial, and therefore the worst-case complexity of the global method (CPLEX) grows exponentially with the problem size [17, Sec. 1.4.2]. Thus, there is naturally a tradeoff between the optimality and the efficiency of the algorithms. Nevertheless, Figs. 7, 8, and 13 and the asymptotic results in Fig. 9 indicate that *DAA* yields a good tradeoff between the optimality and the efficiency, which is favorable for practical implementation. To quantify this tradeoff, Table I presents the tradeoff related to optimality and speed of *DAA* when compared to CPLEX solver (average values are presented). It is observed that while the population of the clients grows, the deviation of the optimal solution is becoming lower. In parallel, the speed gain increases.

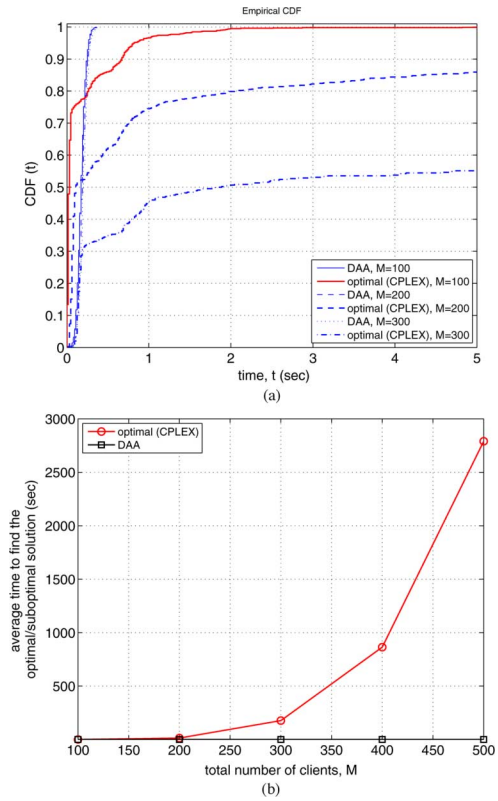



Fig. 13. (a) Empirical CDF plots of total CPU time, $N = 10$. (b) Average time to find optimal/suboptimal solution, $N = 10$.

TABLE I
OPTIMALITY AND SPEED OF *DAA*

| Number of clients when $N = 10$ | Deviation from optimal (%) | Speed (faster than CPLEX) |
|---|---|---|
| 100 | 4.67% | ×28 |
| 200 | 3.63% | ×42 |
| 300 | 3.42% | ×68 |
| 400 | 2.98% | ×115 |
| 500 | 2.51% | ×252 |

## VII. CONCLUSION

In this paper, we considered the problem of optimizing the allocation of the clients to the APs in 60-GHz wireless access networks. The objective in our problem formulation was to *minimize the maximum AP utilization*. The optimization problem was combinatorial. Thus, we proposed a distributed association algorithm (*DAA*) based on Lagrangian duality theory and subgradient methods. *DAA* is compliant with the existing WiFi and 60-GHz standards, and it can be easily implemented on top of the MAC mechanisms that they define. We studied the behavior of *DAA* through theoretical analysis, where we proved its asymptotic optimality properties. Moreover, we presented a numerical analysis, where *DAA* was compared to other association policies in realistic scenarios. We tested convergence, scalability, time efficiency, and fairness. Our results indicate that the proposed solution could be well applied in the forthcoming 60-GHz wireless access networks.

## APPENDIX A
## PROOF OF PROPOSITION 1

It is sufficient to show that deciding whether there exists a feasible objective value $t$ of problem (6) no more than a given value $\kappa \in \mathbb{R}$ is NP-hard. We show this by polynomial time reduction from the 3-partitioning (3-PART) problem, which is known to be NP-complete [46, p. 96]. Consider the 3-PART instance: $3m+1$ numbers $a_1, \ldots, a_{3m}, \kappa$ such that $\sum_{j=1}^{3m} a_j = m\kappa$ and $\kappa/4 < a_j \leq \kappa/2$ for all $j$. The 3-PART problem poses the following question: Is there a partition of $\{1, \ldots, 3m\}$ into sets $\mathcal{S}_1, \ldots, \mathcal{S}_m$ such that $\sum_{j \in \mathcal{S}_i} a_j = \kappa \, \forall i = 1, \ldots, m$? Given this instance of 3-PART, we define an instance of problem (6) having $\mathcal{N} = \{1, \ldots, m\}$, $\mathcal{M} = \{1, \ldots, 3m\}$, $\mathcal{N}_j = \mathcal{N}$ for all $j \in \mathcal{M}$, $\mathcal{M}_i = \mathcal{M}$ for all $i \in \mathcal{N}$, and $\beta_{ij} = a_j$ for all $i \in \mathcal{N}$. Now note that deciding whether there exists a feasible objective value $t$ of problem (6) no more than $\kappa$ is an NP-hard problem

since the answer is affirmative if and only if the answer to the corresponding instance of 3-PART is affirmative.

## APPENDIX B
## PROOF OF THEOREM 1

The proof is based on a proposition from [47], which we restate here for clarity and for simplifying the presentation.

*Proposition 4:* Consider the possibly *nonconvex* problem

$$\begin{aligned} \text{minimize} \quad & \sum_{j\in\mathcal{J}} f_j(\mathbf{y}_j) && \text{(26a)}\\ \text{subject to} \quad & \mathbf{y}_j \in \mathcal{Y}_j,\ j\in\mathcal{J} && \text{(26b)}\\ & \sum_{j\in\mathcal{J}} \mathbf{h}_j(\mathbf{y}_j) \le \mathbf{b} && \text{(26c)} \end{aligned}$$

where the variables are $\mathbf{y}_j \in \mathbb{R}^{y_j}$. The problem parameters $\mathcal{J} = \{1,\dots,J\}$, $\mathbf{b}$ is a given vector in $\mathbb{R}^Q$, $\mathcal{Y}_j$ is a subset of $\mathbb{R}^{y_j}$, and $f_j : \text{conv}(\mathcal{Y}_j) \to \mathbb{R}$ and $\mathbf{h}_j : \text{conv}(\mathcal{Y}_j) \to \mathbb{R}^Q$ are functions defined on the convex hull of $\mathcal{Y}_j$. The following assumptions hold for the primal problem (26).

*Assumption 1:* There exists at least one feasible solution of problem (26).

*Assumption 2:* For each $j$, the subset of $\mathbb{R}^{y_j+Q+1}$

$$\{(\mathbf{y}_j, \mathbf{h}_j(\mathbf{y}_j), f_j(\mathbf{y}_j)) \mid \mathbf{y}_j \in \mathcal{Y}_j\} \tag{27}$$

is compact.

*Assumption 3:* For each $j$, given any vector $\tilde{\mathbf{y}}$ in $\text{conv}(\mathcal{Y}_j)$, there exists $\mathbf{y} \in \mathcal{Y}_j$ such that $\mathbf{h}_j(\mathbf{y}) \le \tilde{\mathbf{h}}_j(\tilde{\mathbf{y}})$, where $\tilde{\mathbf{h}}_j : \text{conv}(\mathcal{Y}_j) \to \mathbb{R}^Q$ is the *convexified* version of $\mathbf{h}_j$ on $\text{conv}(\mathcal{Y}_j)$ and the notation "$\le$" here means the component-wise inequality. In particular, for all $\tilde{\mathbf{y}} \in \text{conv}(\mathcal{Y}_j)$

$$\tilde{\mathbf{h}}_j(\tilde{\mathbf{y}}) = \inf\left\{ \sum_{k=1}^{y_j+1} \alpha^k \mathbf{h}_j(\mathbf{y}^k) \,\middle|\, \tilde{\mathbf{y}} = \sum_{k=1}^{y_j+1} \alpha^k \mathbf{y}^k, \ \mathbf{y}^k \in \mathcal{Y}_j, \sum_{k=1}^{y_j+1} \alpha^k = 1, \alpha^k \ge 0 \right\}. \tag{28}$$

Moreover, consider the dual problem of (26), i.e.,

$$\begin{aligned} \text{maxmize} \quad & d(\boldsymbol{\nu}) = \inf_{\substack{\mathbf{y}_j\in\mathcal{Y}_j \\ j\in\mathcal{J}}} \left\{ \sum_{j\in\mathcal{J}} \left[f_j(\mathbf{y}_j) + \boldsymbol{\nu}^{\mathrm{T}} \mathbf{h}_j(\mathbf{y}_j)\right] - \boldsymbol{\nu}^{\mathrm{T}}\mathbf{b} \right\} \\ \text{subject to} \quad & \boldsymbol{\nu} \ge \mathbf{0} \end{aligned}$$

with variables $\boldsymbol{\nu} = (\nu_1,\dots,\nu_Q) \in \mathbb{R}^Q$. Then, we have

$$P^\star - D^\star \le (Q+1) \max_{j\in\mathcal{J}} \rho_j \tag{29}$$

where $P^\star$ denotes the optimal value of problem (26), $D^\star$ denotes the optimal value of the dual problem (29), and $\rho_j$ is a nonnegative scalar such that

$$\rho_j \le \sup_{\mathbf{y}_i\in\mathcal{Y}_j} f_j(\mathbf{y}_j) - \inf_{\mathbf{y}_i\in\mathcal{Y}_j} f_j(\mathbf{y}_j). \tag{30}$$

*Proof:* We do not reproduce the proof here, but refer the interested reader to [47, Sec. 5.6.1, pp. 371–376] for a rigorous proof and to [40, Sec. 5.1.6] for an intuitive explanation. ■

Now we rely on Proposition 4 above to prove Theorem 1. The key steps of the proof are: (a) equivalently reformulating MILP (6) in the form (26); and (b) showing that the Assumptions 1–3 of Proposition 4 hold for this equivalent problem.

Let us start by considering the following problem that is closely related to MILP (6):

$$\begin{aligned} \text{minimize} \quad & \sum_{j\in\mathcal{M}} t_j \\ \text{subject to} \quad & \sum_{j\in\mathcal{M}} \bar{\beta}_{ij}\, x_{ij} \le \sum_{j\in\mathcal{M}} t_j, \qquad i\in\mathcal{N} \\ & \sum_{i\in\mathcal{N}} \gamma_{ij} x_{ij} = 1, \qquad j\in\mathcal{M} \\ & x_{ij} \in \{0,1\},\ j\in\mathcal{M}, i\in\mathcal{N} \\ & 0 \le t_j \le t^{\max} + \bar{\beta}_{n_j j} x_{n_j j},\ j\in\mathcal{M} \end{aligned} \tag{31}$$

where the variables are $\mathbf{t} = (t_1,\dots,t_M)$ and $\mathbf{x} = (x_{ij})_{i\in\mathcal{N},\, j\in\mathcal{M}}$. The problem $\bar{\beta}_{ij}$ and $\gamma_{ij}$ are defined as

$$\bar{\beta}_{ij} = \begin{cases} \beta_{ij}, & i\in\mathcal{N},\ j\in\mathcal{M}_i \\ 0, & \text{otherwise} \end{cases} \qquad \gamma_{ij} = \begin{cases} 1, & j\in\mathcal{M},\ i\in\mathcal{N}_j \\ 0, & \text{otherwise} \end{cases}$$

$n_j = \arg\min_{i\in\mathcal{N}_j} \bar{\beta}_{ij}$, and $t^{\max} < \infty$ is an upper bound on $t_j^\star$, the optimal solution component of problem (31) that corresponds to $t_j$. For example, we use $t^{\max} = \max_{i\in\mathcal{N}, j\in\mathcal{M}} \bar{\beta}_{ij}$, throughout this paper. We can easily show that problem (31) is equivalent to original MILP (6) and the optimal value $P^\star$ is equal to the optimal value $p^\star$ of MILP (6), i.e.,

$$P^\star = p^\star. \tag{32}$$

We refer to problem (31) as the *modified MILP*, which is in the form (26), where:

1) $\mathcal{J} = \mathcal{M}$ and $J = M$;
2) $\mathbf{y}_j = (\mathbf{z}_j, t_j) \in \mathbb{R}^{y_i}$, with $\mathbf{z}_j = (x_{ij})_{i\in\mathcal{N}}$ and $y_j = N+1$;
3) $f_j(\mathbf{y}_i) = t_j$;
4) $\mathcal{Y}_j = \{((x_{ij})_{i\in\mathcal{N}}, t_j) \mid \sum_{i\in\mathcal{N}} \gamma_{ij} x_{ij} = 1,\ x_{ij} \in \{0,1\},\ i\in\mathcal{N},\ t_j \in [0, t^{\max} + \bar{\beta}_{n_j j} x_{n_j j}]\}$;
5) $\mathbf{h}_j(\mathbf{y}_j) = ((\bar{\beta}_{1j} x_{1j} - t_j), \dots, (\bar{\beta}_{Nj} x_{Nj} - t_j)) \in \mathbb{R}^Q$, with $Q = N$;
6) $\mathbf{b} = \mathbf{0}$.

Let us now show that the Assumptions 1–3 of Proposition 4 hold for the modified MILP (31). It is straightforward to see that Assumptions 1 and 2 hold. Checking whether Assumption 3 holds is less trivial as we show next.

Let $\tilde{\mathbf{y}}$ be any given vector in $\text{conv}(\mathcal{Y}_j)$. From the definition of $\tilde{\mathbf{h}}_j(\tilde{\mathbf{y}})$ [see (28)], we can express it as

$$\tilde{\mathbf{h}}_j(\tilde{\mathbf{y}}) = \sum_{k=1}^{y_j+1} \alpha^k \mathbf{h}_j(\mathbf{y}^k) \tag{33}$$

for some $\mathbf{y}^k \in \mathcal{Y}_j$ and $\alpha^k$ such that $\tilde{\mathbf{y}} = \sum_{k=1}^{y_j+1} \alpha^k \mathbf{y}^k$, $\sum_{k=1}^{y_j+1} \alpha^k = 1$, $\alpha^k \ge 0$. Particularized to the modified MILP (6), we can write

$$\begin{aligned} &\tilde{\mathbf{h}}_j(\tilde{\mathbf{y}}) \\ &= \sum_k \alpha^k \left(\bar{\beta}_{1j} x_{1j}^k - t_j^k, \dots, \bar{\beta}_{n_j j} x_{nj}^k - t_j^k, \right. \\ &\qquad \left. \dots, \bar{\beta}_{Nj} x_{Nj}^k - t_j^k\right) && \text{(34a)} \\ &= \sum_k \left(\bar{\beta}_{1j} \alpha^k x_{1j}^k - \alpha^k t_j^k, \dots, \bar{\beta}_{n_j j} \alpha^k x_{nj}^k - \alpha^k t_j^k, \right. \\ &\qquad \left. \dots, \bar{\beta}_{Nj} \alpha^k x_{Nj}^k - \alpha^k t_j^k\right) && \text{(34b)} \end{aligned}$$

$$= \Bigg( \bar{\beta}_{1j} \sum_k \left(\alpha^k x_{1j}^k\right) - \sum_k \left(\alpha^k t_j^k\right), \ldots, \bar{\beta}_{n_j j} \sum_k \left(\alpha^k x_{n_j j}^k\right) - \sum_k \left(\alpha^k t_j^k\right), \ldots, \bar{\beta}_{Nj} \sum_k \left(\alpha^k x_{Nj}^k\right) - \sum_k \left(\alpha^k t_j^k\right) \Bigg) \tag{34c}$$

$$\geq \Bigg( \bar{\beta}_{1j} \sum_k \left(\alpha^k x_{1j}^k\right) - \sum_k \left(\alpha^k t_j^k\right), \ldots, \bar{\beta}_{n_j j} \sum_k \left(\alpha^k x_{n_j j}^k\right) - \sum_k \left(\alpha^k \left(t^{\max} + \bar{\beta}_{n_j j} x_{n_j j}^k\right)\right), \ldots, \bar{\beta}_{Nj} \sum_k \left(\alpha^k x_{Nj}^k\right) - \sum_k \left(\alpha^k t_j^k\right) \Bigg) \tag{34d}$$

$$\geq \left(\bar{\beta}_{1j} 0 - t_c^{\max}, \ldots, -t^{\max}, \ldots, \bar{\beta}_{Nj} 0 - t_c^{\max}\right) \tag{34e}$$

$$= \left(-t_c^{\max}, \ldots, -t^{\max}, \ldots, -t_c^{\max}\right) \tag{34f}$$

$$= \mathbf{h}_j(\mathbf{y}) \tag{34g}$$

where $t_c^{\max} = t^{\max} + \bar{\beta}_{nj}$ and $\mathbf{y} = (0, 0, \ldots, 1, 0, \ldots, 0, t_c^{\max})$, which is feasible, i.e., $\mathbf{y} \in \mathcal{Y}_j$ and Assumption 3 holds. Note that the first three equalities (34a)–(34c) follow from straightforward manipulations, inequality (34d) follows from that $t_j^k \leq t^{\max} + \bar{\beta}_{n_j j} x_{n_j j}^k$ for all $k$, inequality (34e) follows from that $\sum_k (\alpha^k x_{1j}^k) \geq 0$ for all $k$, and the last two equalities (34f) and (34g) follow from straightforward manipulations.

Finally, let us show that $D^\star = d^\star$, where $D^\star$ is the dual optimal value of the associated dual problem [compare to (29)] of the modified MILP (31). We denote by $P_{\text{relax}}^\star$ the optimal value of the LP relaxation of (31). Thus, we have

$$D^\star = P_{\text{relax}}^\star = p_{\text{relax}}^\star = d^\star \tag{35}$$

where the first equality follows from a similar approach as described in the proof of Proposition 3, the second equality follows from that the LP relaxations of problem (31) and of problem (6) [see problem (23)] have the same optimal value, and the last equality follows from Proposition 3.

By using (32), (34), and that Assumptions 1–3 hold for the MILP (6) together with Proposition 4, we have

$$\begin{aligned} p^\star - d^\star &\leq (N+1) \max_{j \in \mathcal{M}} \rho_j \\ &\leq (N+1) \max_{j \in \mathcal{M}} (t^{\max} + \bar{\beta}_{n_j j}) \\ &= (N+1) \left(\varrho + \max_{j \in \mathcal{M}} \varrho_j\right) \end{aligned} \tag{36}$$

where the first inequality follows from (29), the second inequality follows from (30), $\sup t_j = t^{\max} + \bar{\beta}_{n_j j}$, and $\inf t_j = 0$, and the last equality follows from that $t^{\max} = \varrho$ and $\bar{\beta}_{n_j j} = \varrho_j$, which yields (25).

Inequality (36) together with our initial assumption $\beta_{ij} \leq 1$ for all $i, j$ ensure that the numerator of the relative duality gap is bounded above by a fixed number, which is *independent* of the total number $M$ of clients. Moreover, we note that $p^\star \to \infty$ as $M \to \infty$ [see the objective function of original problem (6)]. Thus, we conclude that the *relative* duality gap $(p^\star - d^\star)/p^\star$ diminishes to zero as $M \to \infty$.

## ACKNOWLEDGMENT

The authors would like to thank T. Charalambous for his valuable comments and suggestions to improve the quality of the paper.

## REFERENCES

[1] F. Giannetti, M. Luise, and R. Reggiannini, "Mobile and personal communications in the 60 GHz band: A survey," *Wireless Pers. Commun.*, vol. 10, no. 2, pp. 207–243, Jul. 1999.

[2] P. Smulders, H. Yang, and I. Akkermans, "On the design of low-cost 60GHz radios for multigigabit-per-second transmission over short distances," *IEEE Commun. Mag.*, vol. 45, no. 12, pp. 44–51, Dec. 2007.

[3] A. D. Oliver, "Millimeter wave systems—Past, present and future," *IEE Proc. F, Radar Signal Process.*, vol. 136, no. 1, pp. 35–52, Feb. 1989.

[4] P. Smulders, "Exploiting the 60 GHz band for local wireless multimedia access: Prospects and future directions," *IEEE Commun. Mag.*, vol. 40, no. 1, pp. 140–147, Jan. 2002.

[5] Y. P. Zhang and D. Liu, "Antenna-on-chip and antenna-in-package solutions to highly integrated millimeter-wave devices for wireless communications," *IEEE Trans. Antennas Propag.*, vol. 57, no. 10, pp. 2830–2841, Oct. 2009.

[6] *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs) Amendment 2: Millimeter-Wave-Based Alternative Physical Layer Extension*, IEEE 802.15.3c Part 15.3, 2009.

[7] *Wireless Lan Medium Access Control (MAC) and Physical Layer (PHY) Specifications—Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band*, IEEE 802.11ad. Part 11, 2012.

[8] J. Hoydis, M. Kobayashi, and M. Beddah, "Green small-cell networks," *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37–43, Mar. 2011.

[9] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: How much can WiFi deliver?," *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 536–550, 2013.

[10] R. C. Daniels, J. N. Murdock, T. S. Rappaport, and R. W. Heath, "60 GHz wireless: Up close and personal," *IEEE Microw. Mag.*, vol. 11, no. 7, pp. 44–50, Dec. 2010.

[11] *Wireless LAN Medium Access Control and Physical Layer Specifications: Enhancements for Higher Throughput*, IEEE P802.11n Part 11, 2009.

[12] L. Yang and G. B. Giannakis, "Ultra-wideband communications—An idea whose time has come," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 26–54, Nov. 2004.

[13] C. H. Doan, S. Emami, D. A. Sobel, A. M. Niknejad, and R. W. Brodersen, "Design considerations for 60 GHz CMOS radios," *IEEE Commun. Mag.*, vol. 42, no. 12, pp. 132–140, Dec. 2004.

[14] M. Zhadobov, C. N. Nicolaz, R. Sauleau, F. Desmots, D. Thouroude, D. Michel, and Y. Le Drean, "Evaluation of the potential biological effects of the 60-GHz millimeter waves upon human cells," *IEEE Trans. Antennas Propag.*, vol. 57, no. 10, pp. 2949–2956, Oct. 2009.

[15] S. Singh, F. Ziliotto, U. Madhow, E. M. Belding, and M. J. W. Rodwell, "Millimeter wave WPAN: Cross-layer modeling and multihop architecture," in *Proc. IEEE INFOCOM*, Anchorage, AK, USA, May 2007, pp. 2336–2240.

[16] D. P. Bertsekas, *Network Optimization Continuous and Discrete Models*. Belmont, MA, USA: Athena Scientific, 1998.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[18] Y. Bejerano, S. Han, and L. Li, "Fairness and load balancing in wireless LANs using association control," *IEEE Trans. Netw.*, vol. 15, no. 3, pp. 560–573, Jun. 2007.

[19] R. Horst, P. Pardolos, and N. Thoai, *Introduction to Global Optimization*, 2nd ed. Dordrecht, The Netherlands: Kluwer, 2000, vol. 48.

[20] B. Kauffmann, F. Baccelli, A. Chaintreau, K. Papagiannaki, and C. Diot, "Measurement-based self organization of interfering 802.11 wireless access networks," in *Proc. IEEE INFOCOM*, Anhcorage, AK, USA, May 2007, pp. 1451–1459.

[21] S. Shakkottai, E. Altman, and A. Kumar, "The case for non-cooperative multihoming of users to access points in IEEE 802.11 WLANs," in *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 2006, pp. 1–12.

[22] K. Son, S. Chong, and G. Veciana, "Dynamic association for load balancing and interference avoidance in multi-cell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3566–3576, Jul. 2009.

[23] D. Lee, G. Chandrasekaran, M. Sridharan, and P. Sinha, "Association management for data dissemination over wireless mesh networks," *Comput. Netw.*, vol. 51, no. 15, pp. 4338–4355, Oct. 2007.

[24] G. Athanasiou, T. Korakis, O. Ercetin, and L. Tassiulas, "Dynamic cross-layer association in 802.11-based mesh networks," in *Proc. IEEE INFOCOM*, Anchorage, AK, USA, May 2007, pp. 2090–2098.

[25] G. Athanasiou, T. Korakis, O. Ercetin, and L. Tassiulas, "A cross-layer framework for association control in wireless mesh networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 1, pp. 65–80, Jan. 2009.

[26] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, Jun. 2013.

[27] K. Hongseok, G. de Veciana, Y. Xiangying, and M. Venkatachalam, "Distributed $\alpha$-optimal user association and cell load balancing in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 1, pp. 177–190, Feb. 2012.

[28] L. X. Cai, L. Cai, X. Shen, and J. W. Mark, "Rex: A randomized exclusive region based scheduling scheme for mmwave WPANs with directional antenna," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 113–121, Jan. 2010.

[29] S. Singh, R. Mudumbai, and U. Madhow, "Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control," *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1513–1527, Oct. 2011.

[30] J. Qiao, L. X. Cai, X. S. Shen, and J. W. Mark, "Enabling multi-hop concurrent transmissions in 60 GHz wireless personal area networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3824–3833, Nov. 2011.

[31] F. Mohamadi, "Build a phased array on a wafer to boost antenna performance," *Electron. Design* vol. 54, no. 20, p. 69, 2006 [Online]. Available: http://electronicdesign.com/communications/build-phased-array-wafer-boost-antenna-performance

[32] M. X. Gong, D. Akhmetov, R. Want, and M. Shiwen, "Multi-user operation in mmwave wireless networks," in *Proc. IEEE ICC*, 2011, pp. 1–6.

[33] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[34] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[35] H.-H. Lee and Y.-C. Ko, "Low complexity codebook-based beamforming for MIMO-OFDM systems in millimeter-wave WPAN," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3607–3612, Nov. 2011.

[36] R. Mudumbai, S. Singh, and U. Madhow, "Medium access control for 60 GHz outdoor mesh networks with highly directional links," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009, pp. 2871–2875.

[37] S. Y. Geng, J. Kivinen, X. W. Zhao, and P. Vainikainen, "Millimeter-wave propagation channel characterization for short-range wireless communications," *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 3–13, Jan. 2009.

[38] G. Athanasiou, I. Broustis, T. Korakis, and L. Tassiulas, "LAC: Load-aware channel selection in 802.11 WLANs," in *Proc. IEEE PIMRC*, Cannes, France, Sep. 2008, pp. 1–6.

[39] S. Boyd, "Subgradient methods," 2007 [Online]. Available: http://www.stanford.edu/class/ee364b/lectures/subgrad_method_slides.pdf

[40] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 1999.

[41] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[42] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1970.

[43] K. Truemper, *Matroid Decomposition*, revised ed. Plano, TX, USA: Leibniz, 1998.

[44] IBM, Armonk, NY, USA, "IBM ILOG CPLEX optimizer," [Online]. Available: http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/

[45] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[46] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, 19th ed. New York, NY, USA: Freeman, 1997.

[47] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Method*. Belmont, MA, USA: Athena Scientific, 1998.


**George Athanasiou** (S'05–M'10) received the Diploma, M.Sc., and Ph.D. degrees from the University of Thessaly, Volos, Greece, in 2005, 2007, and 2010 respectively, all in electrical and computer engineering.

Currently, he is a Research Scientist with the Department of Automatic Control, Electrical Engineering School and ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden. He is also co-founder and CIO of Aukoti AB, a Swedish startup on sensor networking and building automation. He has authored numerous publications in these areas in international journals and refereed conferences. His research interests include the design and performance evaluation of wireless networks, resource management, service and network management, cognitive networking, and optimization techniques.

Dr. Athanasiou is a member of the Association for Computing Machinery (ACM) and the Technical Chamber in Greece.

**Pradeep Chathuranga Weeraddana** (S'08–M'11) received the M.Eng. degree in telecommunication from the Asian Institute of Technology, Khlong Luang, Thailand, in 2007, and the Ph.D. degree in telecommunications engineering from the University of Oulu, Oulu, Finland, in 2011.

He is currently working as Postdoctoral Researcher with the Automatic Control Lab, Electrical Engineering Department and ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include application of optimization techniques in various application domains, such as signal processing, wireless communications, smart grids, privacy, and security.

**Carlo Fischione** (M'02) received the Dr.Eng. degree in electronic engineering (Laurea, *summa cum laude*) and the Ph.D. degree in electrical and information engineering from the University of L'Aquila, L'Aquila, Italy, in 2001 and 2005, respectively.

He is a tenured Associate Professor with the Automatic Control Lab, Electrical Engineering and ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden. He held research positions with the University of California, Berkeley, CA, USA, and the Royal Institute of Technology. His research interests include optimization and parallel computation with applications to wireless sensor networks, networked control systems, and wireless networks.

Dr. Fischione is an Ordinary Member of the academy of history Deputazione Abruzzese di Storia Patria (DASP). He received a number of awards, including the Best Paper Award from the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS in 2007, the Best Paper awards at the IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS) in 2005 and 2009

**Leandros Tassiulas** (S'89–M'91–SM'06–F07) received the Diploma from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1987, and the M.S. and Ph.D. degrees from the University of Maryland, College Park, MD, USA, in 1989 and 1991, respectively, all in electrical engineering.

He has been a Professor with the Department of Computer and Telecommunications Engineering, University of Thessaly, Volos, Greece, since 2002. He has held positions as Assistant Professor with the Polytechnic Institute of New York University, Brooklyn, NY, USA, from 1991 to 1995, Assistant and Associate Professor with the University of Maryland from 1995 to 2001, and Professor with the University of Ioannina, Ioannina, Greece, from 1999 to 2001. His research interests are in the field of computer and communication networks with emphasis on fundamental mathematical models, architectures and protocols of wireless systems, sensor networks, high-speed Internet, and satellite communications.

Dr. Tassiulas received a National Science Foundation (NSF) Research Initiation Award in 1992, an NSF CAREER Award in 1995, an Office of Naval Research Young Investigator Award in 1997, and a Bodosaki Foundation Award in 1999. He also received the IEEE INFOCOM 1994 Best Paper Award and the IEEE INFOCOM 2007 Achievement Award.