IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 30, 2016, accepted February 23, 2017, date of publication March 1, 2017, date of current version March 28, 2017.

*Digital Object Identifier 10.1109/ACCESS.2017.2676166*

# Distributed User Association in Energy Harvesting Dense Small Cell Networks: A Mean-Field Multi-Armed Bandit Approach

**SETAREH MAGHSUDI[1] AND EKRAM HOSSAIN[2], (Fellow, IEEE)**
[1]Yale School of Engineering and Applied Science, Yale University, New Haven, CT 06520, USA
[2]Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

Corresponding author: E. Hossain (ekram.hossain@umanitoba.ca)

This work was supported in part by the CRD Grant from the Natural Sciences and Engineering Research Council of Canada and in part by a Post-Doctoral Fellowship from the German Research Foundation (DFG) under grant MA 7111/1-1.

**ABSTRACT** The emerging ultra-dense small cell networks (UD-SCNs) will need to combat a variety of challenges. On the one hand, massive number of devices sharing the limited wireless resources renders centralized control mechanisms infeasible due to the excessive cost of information acquisition and computation. On the other hand, to reduce the energy consumption from fixed power grid and/or battery, network entities (e.g., small cell base stations and user devices) may need to rely on the energy harvested from the ambient environment (e.g., from environmental sources). However, opportunistic energy harvesting introduces uncertainty in the network operation. In this paper, we study the distributed user association problem for energy harvesting UD-SCNs. After reviewing the state-of-the-art research, we outline the major challenges that arise in the presence of energy harvesting due to the uncertainty (e.g., limited knowledge on energy harvesting process or channel profile) as well as limited computational capacities. Finally, we propose an approach based on the mean-field multi-armed bandit games to solve the uplink user association problem for energy harvesting devices in a UD-SCN in the presence of uncertainty.

**INDEX TERMS** Ultra-dense small cell networks, energy harvesting, distributed control, user association, multi-armed bandits.

## I. INTRODUCTION

Traditional cellular networks suffer from shortcomings such as limited uplink capacity, poor cell-edge coverage, and heavy loads at macro base stations (MBSs), all affecting the users' experience adversely. Thus, implementing low-cost, low-power small base stations (SBSs) in order to offload the MBSs traffic and to increase the network capacity is foreseen as a promising solution to deliver the expected services of the next generation wireless networks, i.e., to sufficiently support human-centric (human-to-human) communications as well as machine-type (machine-to-machine) interactions [1], [2]. In general, every small cell is expected to serve up to few hundreds devices; thus, an ultra-dense deployment of SBSs is necessary to technically support the massive growth of smart devices (e.g., IoT devices) in wireless networks. As an immediate consequence, it becomes imperative to search for new mathematical tools that are suitable for handling a variety of problems that arise due to an ultra-dense deployment of SBSs as well as a massive number of end-user devices. As a specific example, network management becomes very challenging due to the following reasons. On one hand, centralized methods are not suitable since they require excessive amount of information and incur large computational and signaling costs. On the other hand, traditional distributed control mechanisms, such as those adapted from game theory, either yield slow convergence or require costly data exchange among neighboring nodes. Clearly, both problems are aggravated when the number of network nodes becomes very large.

Another important concern in future ultra-dense small cell networks (UD-SCNs) is to obtain the required energy, both at SBSs- and at end-users. From one side, SBSs are irregularly deployed, so that not all of them can be connected to a power grid. From the other side, the limited battery lifetime of human-centric and machine-type devices, thus the requirement of frequent recharge, is an immediate challenge. To date, two main solutions are foreseen to mitigate energy problems:

- Wireless energy transfer, where devices are powered by a dedicated power source (power beacon) [3];
- Opportunistic ambient energy harvesting, where devices locally harvest the ambient energy, for example the solar or wind energy, if available. Moreover, ambient RF energy harvesting (for example, from the nearby television broadcast signals) [4] is a feasible solution for network entities such as sensors to reduce the dependency on batteries and other fixed energy resources so that frequent recharge becomes unnecessary. Usually, energy harvesting devices apply either harvest-and-use (HU) or harvest-store-use (HSU) strategies [5]. In the former, the energy needs to be consumed immediately after it has been harvested, whereas in the latter, the harvested energy can be stored temporarily and then reused later. Recently, harvest-use-store (HUS) strategy is being investigated as well, where the harvested energy is used immediately and the rest is stored [6]. In selecting the energy harvesting strategy, the storage efficiency/capacity, network requirements, and technical needs for implementation (e.g., hardware) should be taken into account.

Despite being reliable, wireless energy transfer might be inefficient, as it is prone to storage and transfer loss. In fact, the efficiency can be achieved only thorough directional transfer (energy beam-forming), which necessitates the acquisition of accurate channel state information (CSI) [7]. Precise CSI acquisition, however, yields excessive overhead, specifically when a large number of devices must be powered. In addition, due to hardware limitations, currently only very small amount of power can be transferred using a wireless link. Opportunistic ambient energy harvesting, in contrast, is energy-efficient and environment-friendly. Moreover, it does not rely on CSI availability; nonetheless, it is inherently a stochastic process and hence counts as a source of uncertainty in the network. More precisely, often the amount of available energy is not known prior to decision making and network planning.

One of the main goals of deploying SCNs is to reduce the MBS traffic by offloading some users' traffic to the SBSs. To this end, every user equipment becomes associated with some BS (either an SBS or an MBS), with the problem being named as *user association* or *cell association*. Although user association is a fundamental problem also in traditional cellular networks, it is significantly more challenging in energy harvesting UD-SCNs, since (i) every user might be located within the coverage area of a set of SBSs; and (ii) the amount of harvested energy at SBSs and/or the user devices as well as the channel qualities from the user devices to the SBSs are random (i.e., uncertain). *The uplink/downlink communication performance (e.g., the transmission rate) of a user device will depend on the SBS it associates to, the number of user devices served by that SBS, the transmission power of the user device/SBS (which in turn depends on the energy harvested) as well as the channel qualities*. In addition, due to the absence of a central controller, it is preferable to perform user association in a distributed manner. Moreover, due to small coverage radius of SBSs, for mobile users, handover has to be performed rather frequently. In addition, due to the absence of a central controller, it is preferable to perform user association in a distributed manner.

In this paper, we focus on the problem of distributed uplink user association of energy harvesting devices in UD-SCNs. As described before, distributed control in such networks is a twofold challenge: first due to the dense deployment of SBSs as well as the large number of devices (e.g., sensors), and second due to the uncertainty introduced by energy harvesting and channel qualities. We argue that in a UD-SCN with energy-harvesting devices, a variety of optimization problems can be formulated as distributed decision making problems in a multi-agent system, where each solution corresponds to an outcome of the interactions of a large number of agents under uncertainty. We first review the state-of-the-art and outline several open challenges and future research directions on user association in SCNs in presence of energy harvesting. Then, we review a recently-developed mathematical tool, namely mean-field multi-armed bandit games, that can be used to model and analyze distributed control problems in UD-SCNs, including the distributed user association problem with a variety of optimization objectives. We then develop a mean-field multi-armed bandit model for the uplink user association problem in a UD-SCN, where a large number of energy harvesting devices intend to select an appropriate SBS to achieve a minimum uplink transmission rate during a transmission interval, in order to guarantee the minimum required quality of service (QoS).

## II. USER ASSOCIATION IN UD-SCNs: STATE-OF-THE-ART AND OPEN ISSUES

### A. STATE-OF-THE-ART

In the existing literature, *user association in SCNs is performed with a variety of objectives* [8], including interference mitigation, capacity maximization and energy efficiency, as most important examples. Nonetheless, regardless of the specific objective, *only a small minority of research works address the association problems under the assumption that SBSs and/or users rely on the opportunistic ambient energy harvesting as the source of power. Also, majority of them are centralized solutions which may not be suitable for an ultra-dense deployment of SCNs*. Below we briefly review the state-of-the-art.

#### 1) INTERFERENCE MANAGEMENT THROUGH USER ASSOCIATION

In a dense SCN, the spectrum shall be reused by closely located small cells. Thus, while the majority of users might enjoy a strong signal from the corresponding SBS, the inter-cell interference is strong as well, especially for cell-edge users. Smart interference cancellation through user association is investigated in [9]. Stochastic geometry is used to model and solve the user association problem. Similarly, interference mitigation through power control and user association is studied in [10]. The problem is formulated as an integer program and a heuristic algorithm is proposed to solve it. Reference [11] develops an algorithm to associate users to small cells, aiming at maximizing the sum utility of average achievable rates within the macro cell. Small cell group muting (SCGM) is applied to mitigate the interference among adjacent small cells. Zhou *et al.* [12] use a model based on Stackelberg game for distributed user association and interference coordination in order to improve the performance of macro cell as well as small cells, through minimizing the interference. None of these works, however, considers energy harvesting along with interference mitigation and user association.

#### 2) CAPACITY-MAXIMIZING USER ASSOCIATION

Similar to any other network architecture, capacity maximization is a fundamental problem in UD-SCNs. Kim *et al.* [13] suggest a user association scheme which maximizes the sum-rate via dual connectivity. Reference [14] considers an overlay small cell network with backhaul constraint, and optimizes a weighted sum-rate with carrier aggregation. In both works energy source is fixed and deterministic. Maghsudi and Hossain [15] investigate a distributed user association problem in the downlink of an SCN where SBSs obtain the required energy through ambient energy harvesting. The SCN is modeled as a competitive market with uncertainty, where SBSs and users are represented as consumers and commodities, respectively. Based on this model, a distributed user association method is developed. As another example, in [16], downlink user association is performed in a distributed manner at end-users, where every user aims at maximizing the probability of successful transmission. Note that both works investigate the *downlink* association problem where the SBSs harvest the energy. Consequently, the proposed solutions are not suitable for the uplink association problem with energy harvesting devices. The main reason is that in downlink association, multiple users are selected by an SBS, while in uplink association, each user has to select one SBS. Moreover, for downlink transmission, power allocation and scheduling are usually investigated as well.

#### 3) ENERGY-EFFICIENT USER ASSOCIATION

It is known that in UD-SCNs, energy efficiency and capacity are improved when small cells are turned on and off using a sleep-awake scheduling method [17], called *cell planning* or *cell scheduling*. It is however clear that turning SBSs on and off has to be performed in combination with a sophisticated user association mechanism, in a way that changing SBS density does not degrade the user satisfaction level. Dong *et al.* [18] optimize the SBS density for energy efficiency in cellular networks by using stochastic geometry, and optimize the user association matrix by using quantum particle swarm optimization. Similarly, in [19], two heuristic algorithms are proposed to jointly optimize base station operation and user association in heterogeneous networks. Joint energy and spectrum efficiency is studied in [20]. Taking the quality of service into account, the authors develop a low-complexity algorithm to solve the formulated optimization problem approximately. However, none of these works considers energy harvesting in the system model.

### B. OPEN ISSUES

Despite being under intensive investigation, many features of energy harvesting UD-SCNs remain unexplored. In particular, *enhancing a network with energy harvesting impacts its characteristics dramatically due to uncertainty, so that the conventional solutions may no longer be applicable*. For example, although the problems of interference management, power control, joint cell planning and user association in SCNs have been already investigated in the literature, enabling SBSs and/or devices to harvest the ambient energy renders the existing solutions non-applicable. In particular, to solve the aforementioned problems efficiently, the availability of transmission energy (which is uncertain) should be taken into consideration. Moreover, emerging networking concepts (such as caching and full-duplexing), can be efficiently implemented only in conjunction with intelligent user association. Hence, one might need to consider energy harvesting in UD-SCN scenarios, in order to reduce the grid and/or battery power usage. Therefore, new objectives will need to be examined when performing the user association. In the following, we outline some promising research directions.

#### 1) UPLINK DISTRIBUTED USER ASSOCIATION FOR ENERGY HARVESTING DEVICES IN SCNs

Conventionally, cellular systems are designed to perform the transmission of large data streams in the downlink, in order to serve human-driven service demands. Nevertheless, by emerging networking concepts such as IoT-driven UD-SCNs, the next generation of wireless networks are required to also support a variety of machine-type communications, for instance information transmission in wireless sensor networks used for site surveillance [21]. Such data streams are small individually but create a heavy traffic in the uplink when they superimpose. Moreover, closely located devices that transmit in the same frequency band cause interference to each other. Changing traffic pattern and mitigating the interference can be realized by intelligent user association. Such association schemes need to be designed to achieve some desired performance objective considering uncertainties in energy harvesting as well as channel qualities.

Again, most of the existing methods for user association rely on a central controller, and/or demand excessive amount of information at SBSs or user devices. Thus, we need to develop distributed association methods which are capable of dealing with very large network size (number of SBSs and/or users) as well as information shortage. Moreover, within the scope of energy harvesting networks, an efficient solution to this problem ought to address the uncertainty imposed by random energy arrivals and/or intensities. *Later in this paper,* *we will present such a method for distributed uplink user association for energy harvesting devices in a UD-SCN based on the theory of mean-field multi-armed bandit games.*

#### 2) JOINT USER ASSOCIATION AND MODE SELECTION IN DEVICE-TO-DEVICE (D2D)-ENHANCED SCNs

In heterogeneous SCNs, nearby users might be able to establish direct (i.e., device-to-device) communication links as an alternative to transmitting the data via an SBS. Thus, enhancing the network with the possibility of direct transmission gives rise to the problem of transmission mode selection. In UD-SCNs, where every user might be in the coverage area of multiple SBSs, mode selection has to be performed jointly with user association, since the performance of network-assisted D2D communication can vary with user association (e.g., due to different number of channels available for D2D communication in different small cells). Moreover, in order to reduce the overhead of information acquisition and computational cost, the selection is preferably performed by devices, i.e., in a distributed manner. Similar to other problems, the uncertainty of energy harvesting and lack of channel knowledge complicates the problem, since the (a priori unknown) available energy as well as the device's channel matrix play an important role in selecting the optimal transmission mode and (possibly) the subsequent user association. Maghsudi and Stanczak [22] propose a distributed mode selection method in heterogeneous networks by using multi-armed bandit theory; however, energy harvesting is not included in the system model.

#### 3) USER ASSOCIATION FOR FULL-DUPLEX TRANSMISSION IN SCNs

Full-duplex transmission[1] is an emerging technology to increase the spectral efficiency. Recent studies show that in-band full-duplex technology exhibits better performance for low-power transmissions, which makes it suitable for UD-SCNs. However, it becomes inefficient to perform the uplink and downlink user association separately, since uplink and downlink transmissions are mutually dependent due to self-interference. More precisely, even an efficient user association method in the downlink (uplink) in conventional half-duplex networks might cause high interference in the uplink (downlink) in an in-band full-duplex communication scenario. While user association in full-duplex SCNs has been addressed previously (see [23], for example), investigating the effect of energy harvesting is an open issue. Taking the uncertainty of energy harvesting into consideration makes the problem more difficult, since in this case the set of users that can be successfully served by an SBS (in uplink and downlink) is non-deterministic. While this problem is already challenging to be solved by a central controller, it becomes aggravated if all network nodes take active roles in selecting their correspondence. In this case, it is vital to search for distributed decision making methods that converge to an efficient equilibrium.

#### 4) USER ASSOCIATION FOR SMALL CELL CACHING AND MULTICAST IN PRESENCE OF ENERGY HARVESTING

In the past few years, small cell caching has attracted an ever-increasing attention as a solution, to reduce the backhaul traffic. The basic idea is to save popular files in SBSs, instead of fetching the data frequently from the core network. In a densely deployed SCN, every small cell has a small coverage area and serves mainly its nearby users. Meanwhile, it is noticeable that popular contents are generally requested by multiple users simultaneously. Thus, a feasible solution to reduce the SBS traffic, also to offload the wired backhaul, is to cache the popular files and then to employ multicasting for data transmission. This concept itself gives rise to context-aware user association, where users are clustered and associated based on the possibility of demanding similar files, which can be predicted using machine learning methods and historical requested data. Although cache-aware user association has been briefly investigated in the literature so far [24], energy harvesting (at the SBSs and/or the user devices) shall also be considered in the user association problem, by which the problem formulation will be influenced.

## III. MEAN-FIELD BANDIT GAMES

Although a large body of literature propose centralized approaches to solve the resource allocation problems including the user association, in UD-SCNs, such methods require very costly information acquisition at the central controller and incur heavy computations cost, thus are not practical. Hence, for energy-harvesting UD-SCNs, it is beneficial to formulate the control problem as distributed optimization problem under uncertainty, which can be even multi-objective. In most cases, the formulated optimization problem can be afterward modeled as distributed decision making problem is a large multi-agent system with uncertainty.

However, naive learning methods such as regret matching or conventional solution concepts based on game, auction or contract theory are not sufficient to deal with such problems in large and non-deterministic systems. Such solutions mostly (i) require unaffordable prior knowledge at least at some agents, or heavy information exchange among agents; (ii) are not able to deal with the uncertainty; (iii) converge slowly for medium/large number of actions and agents, if at all. In the following, we describe an efficient mathematical model to analyze UD-SCNs that does not suffer from such shortcomings, namely *mean-field multi-armed bandit game* [25]. As we see later, in this model, users do not need to have any prior information about channel quality, network traffic, and/or energy harvesting profile. Moreover, we show that the complexity is low even for very large number of users. We also establish that a convergence to equilibrium is guaranteed.

To this end, we first provide a brief tutorial of conventional bandit games, paving the path to describe the mean-field

[1] Simultaneous transmission and reception in the same/different frequency band are respectively referred to as in-band/out-band full-duplexing.

approximation of such games when the number of agents grows large. *To our knowledge, bandit games in conjunction with mean-field approximation has not been applied in wireless communications literature so far, rendering an introductory tutorial advantageous.* Afterward, in Section IV, we use this approach to solve the uplink distributed user association problem with energy harvesting devices. At the same time, we would like to emphasize that, this mathematical tool can be applied in a wide variety of resource allocation problems beyond user association.

### A. SINGLE-AGENT MULTI-ARMED BANDITS

Multi-armed bandits (MAB) is a class of sequential optimization problems, where given a set of arms (actions), a player pulls an arm at successive trials to receive some a priori unknown reward. Upon pulling an arm, the player observes only the reward of the played arm, and not those of other arms. Due to the lack of information, there might be some difference between the maximum reward (achievable by pulling the optimal arm), and the reward of the actual played arm. This difference is referred to as *regret*. The player decides which arm to pull in a sequence of trials so as its average regret is minimized, or its (discounted) aggregate reward is maximized. The basic problem is to deal with the famous *exploration-exploitation dilemma*, i.e., to find a balance between receiving immediate rewards (exploitation) and gathering information to achieve large rewards in the future (exploration).

Bandit games are classified based on (i) the random nature of the arms' reward processes, (ii) density and type of agents, (iii) availability of side-information, (iv) randomness in action availability, (v) number of actions and agents, and so on. To date, a variety of algorithmic solutions have been developed to solve different types of single-agent bandit problems. A well-known policy, which we later use in this paper, is the upper confidence bound (UCB) strategy. The seminal UCB policy is designed specifically for stochastic stationary bandit problems, where the set of instantaneous rewards of each arm are independent and identically-distributed (i.i.d.) random variables. At every round of selection, the UCB policy estimates an upper-bound of the mean reward of each arm $m \in \mathcal{M}$ at some fixed confidence level. The arm with the highest estimated bound is then played, and bounds are updated after observing the reward [26]. The seminal UCB policy is summarized in Algorithm 1. For a survey on bandit problems see [27] and [28].

---

**Algorithm 1** Upper Confidence Bound Selection Policy [26]

**Initialization:** Play each arm $m \in \mathcal{M}$ once.

**Loop** for $t = M + 1, M + 2, \ldots$

- Calculate the index of every arm $m \in \mathcal{M}$, denoted by $I_{m,t}$ at round $t$, as
$$I_{m,t} = \bar{u}_{m,t-1} + \sqrt{\frac{2\ln(t)}{T_{m,t-1}}}, \tag{1}$$
where $T_{m,t-1}$ is the total number of rounds arm $m$ is selected, and $\bar{u}_{m,t-1}$ is the average reward of arm $m$, both up to round $t-1$.
- Select the arm with the largest index. Formally,
$$a_t = \arg\max_{m \in \mathcal{M}} I_{m,t} \tag{2}$$

---

### B. MULTI-AGENT MULTI-ARMED BANDITS

When multiple agents are involved in a bandit game, the agents affect each other, in the sense that the reward achieved by every agent is determined not only through its own actions, but also through the joint action profile of other agents. In other words, the payoff of every arm to every agent depends not only on the type (or ability) of that specific agent, but also on the set of agents selecting that arm. For instance, in a congestion model, the individual rewards might decrease in case multiple agents select an arm (negative externalities), whereas in a coordination model, the reward might increase (positive externalities). In addition to minimizing the regret, in a multi-agent setting, it is important to reach some sort of steady state or equilibrium.

For small number of agents, there are some results that connect multi-agent multi-armed bandit games with correlated and Nash equilibria (see [28] for a review). Perfect Bayesian equilibrium is another notion of equilibrium that is widely-used in conjunction with learning games. For multi-armed bandit games with *large* number of agents, however, such equilibrium notions are not practical since they yield excessive complexity and long convergence time, already for a moderate number of agents. For example, for a multi-armed bandit game to converge to correlated equilibrium, every agent has to observe the joint action profile of all other agents (full monitoring) and forecast their future moves, which is clearly a highly-involved task from the computational aspects, even when only few agents compete with each other. Thus it is imperative to search for new frameworks and solution concepts that are suitable to deal with large number of players.

### C. MEAN-FIELD MULTI-AGENT MULTI-ARMED BANDITS

In general, games with (very) large number of agents are analyzed using mean-field approximation. In mean-field approximation, every agent regards the rest of the world as being stationary, considering individual moves of agents as unimportant details. In other words, mean-field analysis restates game theory as an interaction of each individual with the mass of others. While mean-field analysis for games with perfect information is well-established, applying this concept to multi-armed bandit games is a recently-emerging research direction, initiated by Gummadi *et al.* in [25], [29]. In what follows, we describe mean-field multi-armed bandit games briefly, without getting too much into mathematical details. To show an application, in Section IV, we return to energy-harvesting UD-SCNs and model the uplink cell assignment problem by mean-field multi-armed bandits.

Consider a multi-agent multi-armed bandit game $\mathfrak{G}$ that consists of a set $\mathcal{M}$ of $M$ arms and a set $\mathcal{N}$ of $N$ agents. At every round $t$, each agent $n \in \mathcal{N}$ selects an action, denoted by $a_{n,t}$, from the predefined action set $\mathcal{M}$, and receives some a priori unknown reward. Each agent $n \in \mathcal{N}$ is characterized by some type $\boldsymbol{\theta}_n \in \mathcal{A} = [0,1]^M$ and some state $\mathbf{Z}_n$. The type of each agent $n \in \mathcal{N}$ is a random variable with $M$ components, sampled from some distribution $W$ at each round of the game. The $m$-th component of $\boldsymbol{\theta}_n$ captures the parameters that influence the reward of arm $m \in \mathcal{M}$ to agent $n \in \mathcal{N}$. In accordance with the general bandit setting, the type and its distribution are unknown to the agent. Let $f_{m,t}$ denote the fraction of agents that pull arm $m \in \mathcal{M}$ at round $t$. Then the reward of arm $m \in \mathcal{M}$ to an agent $n \in \mathcal{N}$ is a Bernoulli random variable with some parameter $Q(f_{m,t}, \boldsymbol{\theta}_{n,t})$. The state of each agent is the collection of its past actions and rewards. More formally, $\mathbf{Z}_{n,t} = \left(w_{1,t}, l_{1,t}, \ldots, w_{M,t}, l_{M,t}\right)$, with $w_{m,t}$ and $l_{m,t}$, $m \in \mathcal{M}$, being the total number of successes and failures of arm $m$ up to round $t$, when selected by agent $n \in \mathcal{N}$. Hence, unlike the type, the state is known to the agent. At every round $t$, every agent $n \in \mathcal{N}$ uses some (randomized) selection policy (for example, the UCB algorithm, described in Section III-A) to map $\mathbf{Z}_{n,t-1}$ to some action $a_{n,t}$.

Every agent regenerates after some random time which follows a geometric distribution with parameter $1-\alpha$, $\alpha \in [0,1)$. Physically, the regeneration corresponds to the dynamicity of the network, in the sense that any regenerating agent quits the game and a new agent takes its place. Clearly, the regeneration can also capture the changes in the type for each agent. Let us refer to $\mathbf{f}_t = \left(f_{1,t}, f_{2,t}, \ldots, f_{M,t}\right)$ as population profile at time $t$. The population profile evolves as the agents select their actions over time, according to the *mean-field dynamics*. For an arbitrary agent $n \in \mathcal{N}$, the *mean-field dynamics* works as follows: If a trial is a regeneration trial (by chance), the agent's type $\boldsymbol{\theta}_{n,t}$ is sampled from distribution $W$; Moreover, the state $\mathbf{Z}_{n,t}$ is reset to zero. Otherwise (in case of no regeneration), the type remains unchanged and a (randomized) selection policy (for example, the UCB algorithm, described in Section III-A) is used to map $\mathbf{Z}_{n,t-1}$ to some action $a_{n,t}$. It is assumed that all agents apply the same policy $\delta$ throughout the game. As mentioned before, the agent receives some random reward following a Bernoulli distribution with parameter (success probability) $Q(f_{m,t}, \boldsymbol{\theta}_{n,t})$, and the state vector $\mathbf{Z}_{n,t}$ is updated. The dynamics is summarized in Algorithm 2. Roughly-speaking, we say that a *mean-field equilibrium (MFE)* is achieved when in the mean-field dynamics, the population profile $\mathbf{f}$ remains fixed, so that for every agent, the system is stationary. For details see [25] and [29].

To our knowledge, the mean-field multi-armed bandit model is so far only studied for Bernoulli reward process, which restricts the applicability. Thus a future line of research is to extend the analysis (including mean-field dynamics, existence and uniqueness of mean-field equilibrium) to other models of rewards' randomness.

**Algorithm 2** Mean-Field Dynamics for Multi-Armed Bandit Games [25]

**for** $t = 1, 2, \ldots$ **do**
  **if** $t$ is a regeneration trial, **then**
    The agent's type $\boldsymbol{\theta}_{n,t+1}$ is sampled from some distribution $W$.
    The state $\mathbf{Z}_{n,t+1}$ is reset to zero.
  **else**
    Use a (randomized) selection policy $\delta$ to map $\mathbf{Z}_{n,t}$ to some action $a_{n,t}$. The mapping $\delta$ can be any standard (optimal) bandit policy such as the upper confidence bound (UCB) policy, summarized in Algorithm 1.
    Observe the reward.
    Update $\mathbf{Z}_{n,t}$ to $\mathbf{Z}_{n,t+1}$.
  **end if**
**end for**

## IV. DISTRIBUTED UPLINK USER ASSOCIATION FOR ENERGY HARVESTING DEVICES IN SCNs

### A. SYSTEM MODEL

We consider a dense small cell network that consists of a set $\mathcal{M}$ of $M$ cells[2] and a set $\mathcal{N}$ of $N$ devices. Every device $n \in \mathcal{N}$ intends to transmit $J_n \leq J$ data packets in the uplink direction in successive transmission rounds.[3] At every transmission round $j$, each device transmits one data packet to an SBS of its choice, implying that the association is performed in a distributed manner. In what follows we omit the time notion $j$ for the simplicity of notation unless an ambiguity arises. By $\mathcal{N}_{m,j}$ we denote the set of $N_{m,j}$ devices to be served by SBS $m \in \mathcal{M}$ at round $j$. Every device obtains the energy through ambient energy harvesting by applying a harvest-use strategy; that is, for every transmission round, it harvests the energy and then uses all of it for transmission. Note that due to hardware limitation, small wireless devices are usually only able to harvest small amounts of energy, especially in short time. Without loss of generality, we assume that the energy equals power.

Since energy harvesting is opportunistic, for every device $n \in \mathcal{N}$ and at every round $j$, the amount of harvested energy, denoted by $P_{n,j}$, is *unknown* a priori. We assume that $P_{n,j}$, $j = 1, \ldots, J_n$, are i.i.d. random variables following half-normal distribution with parameter $\sigma_n^2 > 0$. This assumption is not restrictive since the half-normal distribution can be replaced by any other distribution, without affecting the solution approach.

The $N_{m,j}$ devices which select any SBS $m \in \mathcal{M}$ share the available spectrum resources equally in an orthogonal manner. For each small cell $m \in \mathcal{M}$, the inter-cell inter-

[2]We do not explicitly consider the macro cell which could be one of these $M$ cells. We use the term SBS generically to denote a BS in one of these cells.

[3]If a device quits transmission, it is replaced by another device so that the number of devices is always equal to $N$. As described before, this corresponds to *regeneration* in mean-field game models which follows a geometric distribution.

ference experienced by every device $n \in \mathcal{N}_{m,j}$, denoted by $I_{nm,j} \geq 0$, is regarded as noise. Treating interference as noise is commonly used for SCNs, for instance in [30], [31], and [32], among many others. Transmissions are also corrupted by zero-mean additive white Gaussian noise with variance $N_0$. At round $j$, the real-valued channel gain between device $n \in \mathcal{N}_{m,j}$ and small cell $m \in \mathcal{M}$ is denoted by $h_{nm,j}$. We assume frequency non-selective block fading channel model, where the random variable $h_{nm}$ follows a Rayleigh distribution with parameter $\frac{1}{\sqrt{2\beta_{nm}}}$, and remains constant during the transmission of every packet $j = 1, ..., J_n$ for all $n \in \mathcal{N}$ and $m \in \mathcal{M}$, and changes from one transmission round to another. The random channel power gain $h'_{nm} = h^2_{nm}$ thus follows an exponential distribution with parameter $\beta_{nm}$. At every round $j$, the *type* of every device $n \in \mathcal{N}$, denoted by $\boldsymbol{\theta}_{n,j} \in (0, 1]^M$, is defined as the collection

$$\boldsymbol{\theta}_{n,j} = \left(\theta_{n1,j}, \theta_{n2,j}, ..., \theta_{nM,j}\right), \tag{3}$$

where for $m \in \mathcal{M}$,

$$\theta_{nm,j} = \frac{h'_{nm,j}}{I_{nm,j} + N_0}. \tag{4}$$

Note that devices do not have any prior knowledge on channel quality and/or interference level. In other words, the *type* is unknown a priori.

Let $f_{m,j} = \frac{N_{m,j}}{N}$ denote the fraction of devices that select SBS $m$ at round $j$. Thus, for each $n \in \mathcal{N}_{m,j}$ and for transmitting every data packet $j$, the achievable transmission rate is given by

$$r_{nm,j} = \frac{W_m}{Nf_{m,j}} \log\left(1 + \frac{P_{n,j}h'_{nm,j}}{I_{nm,j} + N_0}\right), \tag{5}$$

where $W_m$ is the available bandwidth at SBS $m \in \mathcal{M}$. Moreover, as stated before, $P_{n,j}$ is the transmit power of device $n$ at trial $j$, which *equals the amount of energy harvested at that trial*. For transmission of every data packet, every device $n \in \mathcal{N}$ requires a specific QoS that is expressed in terms of a minimum data rate $r_{n,\min}$. Hence, for any device $n \in \mathcal{N}$, at every transmission round $j$, we define the reward of selecting SBS $m \in \mathcal{M}$ as

$$u_{n,j}(m) = \begin{cases} 1, & \text{if } r_{nm,j} \geq r_{n,\min}, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The success probability of user $n$ when selecting SBS $m$ at every transmission round $j$ is then given as

$$p^{(s)}_{nm,j} = \Pr\left[r_{nm,j} \geq r_{n,\min}\right], \tag{7}$$

and the failure probability yields $p^{(f)}_{nm,j} = 1 - p^{(s)}_{nm,j}$. Thus, successful transmission is a Bernoulli random variable with parameter $p^{(s)}_{nm}$.

From (5), it can be easily concluded that given $r_{n,\min}$, for any specific $f_{m,j}$, $h'_{nm,j}$, and $I_{nm,j}$,

$$P_{n,j,\min} = \frac{I_{nm,j} + N_0}{h'_{nm,j}} \left(e^{\frac{Nf_{m,j}r_{n,\min}}{W_m}} - 1\right) \tag{8}$$

yields $r_{nm,j} \geq r_{n,\min}$. Thus we have $p^{(s)}_{nm,j} = \Pr\left[P_{n,j} \geq P_{n,j,\min}\right]$. By the half-normal assumption on the harvested energy we can conclude

$$p^{(s)}_{nm,j} = 1 - \operatorname{erf}\left[\frac{P_{n,j,\min}}{\sqrt{2}\sigma_n}\right], \tag{9}$$

so that

$$p^{(s)}_{nm,j} \propto \frac{\theta_{nm,j}}{f_{m,j}}, \tag{10}$$

which, as intuitively expected, corresponds to a congestion model or game with negative externalities.

### B. MEAN-FIELD EQUILIBRIUM

According to the system model described before, the *type* of each device $n \in \mathcal{N}$ is the collection of channel qualities (including interference), given by (3) and (4). Upon selecting any SBS $m \in \mathcal{M}$ at round $j$, the device transmits successfully (receives reward) with probability $p^{(s)}_{nm,j}$. The success probability depends on the *type* of user $n$ as well as the fraction of devices that select SBS $m$, $f_{m,j}$, as declared by (8) and (9). Prior to selecting an SBS, however, the users do not have any information about channel gains, interference, and/or user traffic, since such knowledge is very costly to acquire, if possible at all. Thus, the uplink user association can be cast as a multi-armed bandit game. Since the number of users is large in UD-SCNs, the mean-field approximation can be used. The following proposition describes the characteristics of mean-field equilibrium in the multi-armed bandit game model of user association problem.

*Proposition 1: In the mean-field multi-armed bandit game model for the uplink user association problem,*

1) *There exists a mean-field equilibrium;*
2) *Let* $a_{nm} = \sqrt{\frac{2}{\pi}} \frac{I_{nm}+N_0}{h'_{nm}\sigma_n}$ *and* $b_{nm} = \frac{Nr_{n,\min}}{W_m}$. *Moreover, assume that for all* $n \in \mathcal{N}$ *and* $m \in \mathcal{M}$, *the regeneration parameter* $\alpha$ *(see Section III) satisfies* $\alpha \leq \frac{1}{1+a_{nm}b_{nm}e^{b_{nm}}}$. *Then the mean-field equilibrium is unique and the mean-field dynamics converges to it from any initial point.*

*Proof:* See **Appendix**. ∎

### C. COMPLEXITY AND OVERHEAD

The mean-field dynamics can be implemented fully distributively. The time and space complexity depend on the policy $\delta$, which maps the state into the action in case no regeneration takes place. The UCB policy, for example, calculates an index for each action corresponding to some confidence bound. The arm with the largest index is then played (see **Algorithm 1**). Thus, the time and space complexity is polynomial in the number of actions. Note that as the state, the dynamics only saves the total number of successes and failures, not the actual strings. In fact, the most important metric to observe is the convergence to mean-field equilibrium, where the population profile $\mathbf{f}$ remains almost fixed.

### D. NUMERICAL PERFORMANCE EVALUATION

We consider an SCN where the devices apply the mean-field dynamics described in Section III, with the policy $\delta$ being the UCB strategy. The goal of every device is to successively decide for the SBS which, with the highest probability, yields a reward; that is, the arm with the largest success probability, $p_{nm}^{(s)}$. Note that the reward is defined in terms of successful transmission. Without loss of generality and for simplification, we choose $W_m = N$, $r_{n,\min} = 0.75$, and $\sigma_n = 1$, for all $n \in \mathcal{N}$ and $m \in \mathcal{M}$. For every device $n \in \mathcal{N}$, the collection of channel gains $\mathbf{h}'_n = \left(h'_{n1}, \ldots, h'_{nM}\right)$ is selected at random. Note that for every device $n \in \mathcal{N}$, and for every SBS $m \in \mathcal{M}$, $h'_{nm}$ are independent but non-identically distributed. The state $\mathbf{Z}_n$ is initialized randomly.

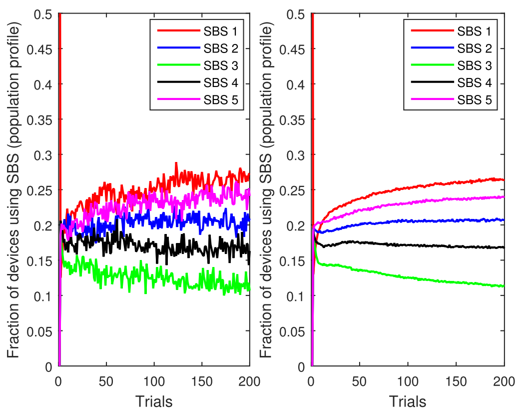


**FIGURE 1. The effect of the number of devices on the performance of mean-field dynamics (left: $10^3$ devices, right: $10^5$ devices). In the right fluctuations are less, meaning that the system settles easier at equilibrium.**

First we investigate the convergence and equilibrium performance of mean-field multi-armed bandit dynamics. To this end, we simulate the number of end-user devices (agents) which select each SBS (arm) for transmission (population profile, see Section III). Fig. 1 shows the effect of the number of devices on the equilibrium performance, where the number of SBSs is selected as $M = 5$. As expected, the mean-field dynamics performs better for larger number of devices. That is, with larger number of devices, fluctuations are less, meaning that the system settles at some point. The reason is as follows: In mean-field dynamics, every device observes the rest of the world as stationary stochastic, ignoring individual moves by considering them as unimportant details. This assumption is specifically valid for large number of devices, since in a small system individual moves might have great impacts and thus cannot be neglected. Note that although the fluctuations are larger for smaller number of devices (here $10^3$), the performance is still acceptable.

In Fig. 2, we select the number of devices $N = 10^5$ and investigate the effect of the number of SBSs (arms) on the dynamics' equilibrium performance. The figure shows that the convergence of the mean field dynamics to equilibrium is not adversely affected by the number of SBSs.

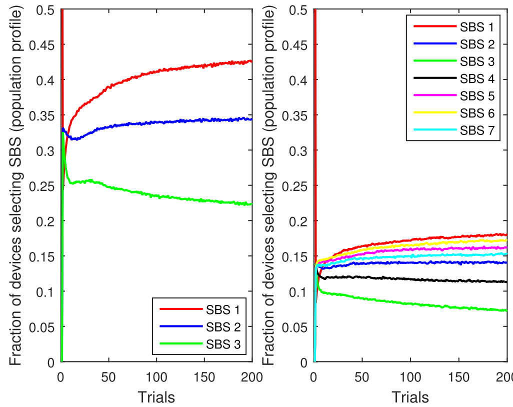


**FIGURE 2. The effect of the number of SBSs on the performance of mean-field dynamics (left: 3 SBSs, right: 7 SBSs). The dynamic converges to equilibrium regardless of the number of SBSs.**

Next we investigate the throughput performance of mean-field multi-armed bandit games, where the total number of successful transmissions is simulated. We select the number of devices and SBSs as $N = 10^3$ and $M = 3$. We compare the performance of mean-field multi-armed bandit, in terms of aggregate average throughput (average number of successful transmissions), with those of the following assignment schemes:

- *Optimal (centralized-informed) association*: In this scenario, user association is performed by a central unit given global information. By means of exhaustive search, every device is assigned to the SBS to which it has the maximum likelihood of successful transmission.
- *$\epsilon$-Greedy*: At each trial, with probability $\epsilon \in (0, 1)$, the user selects an SBS uniformly at random (exploration), whereas with probability $1 - \epsilon$, the user selects the best SBS so far (exploitation).
- *Explore-then-commit*: At first, for a specific number of trials, SBSs are selected in a round-robin manner (exploration). Afterward, the best SBS in terms of average reward (successful transmission) is selected constantly (exploitation).
- *Random association*: Every device selects an SBS simply at random.

The results are shown in Fig. 3. It can be observed that given enough time to converge, the mean-field multi-armed bandit performs well in comparison to other approaches. We should emphasize that despite its slightly better performance, centralized-informed assignment yields excessive overhead for information acquisition as well as large complexity. Hence it is impracticable, in particular for networks with very large number of devices. It is also worth noting that our work is the first user association model that combines the uncertainty of (both) energy harvesting and wireless channel with the high density of SBSs and devices in SCNs using the mean-field multi-armed bandit model. Thus, the performance of no other user association method can be fairly compared with ours. Finally, in Fig. 4, we depict the average wasted energy over time. At every harvesting round followed by a

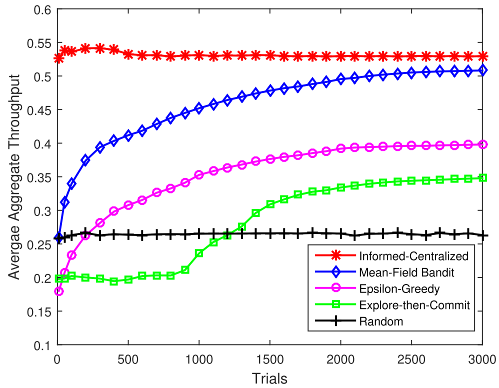


**FIGURE 3. The throughput performance of mean-field multi-armed bandits model compared with centralized association given global information and few other approaches.**

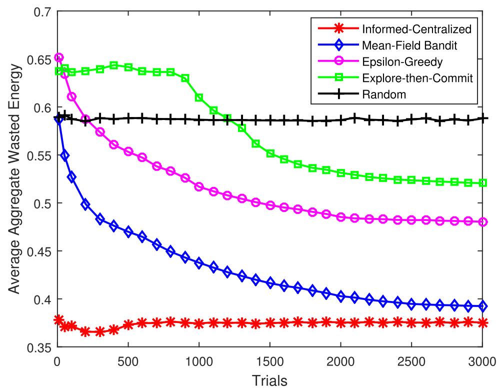


**FIGURE 4. The energy performance of mean-field multi-armed bandits model compared with centralized association given global information and few other approaches.**

transmission, the energy is wasted if the transmission is not successful; that is, if the required QoS is not achieved due to inappropriate SBS selection. As it can be seen from the figure, the average wasted energy reduces over time since the users learn and make better choices. As expected, the bandit approach performs well. Note that this figure confirms the results of Fig. 3, since in case of successful transmission the energy is not wasted.

## V. CONCLUSION

We have studied the user association problem in energy harvesting ultra-dense small cell networks, where densely-deployed SBSs serve a large number of users, and the required energy of SBSs and/or users is obtained through local ambient energy harvesting. Due to its opportunistic nature, energy harvesting introduces some uncertainty to the network, in addition to the random wireless channel quality. We reviewed state-of-the-art as well as future research directions. We further described mean-field multi-armed bandit game model, in which a large number of agents with limited information sequentially select an action from a finite action set, thereby affecting the welfare of each other. Due to the mean-field analysis, this model is particularly appropriate for analyzing the interactions of energy harvesting devices in dense networks with strictly-limited prior information, where users contribute to the network management by playing active roles in decision making. With energy harvesting at the devices, we have modeled the uplink user association problem by a mean-field multi-armed bandit game. Theoretical results establish that a unique mean-field equilibrium exists to which the mean-field dynamics converges from any initial condition. Moreover, numerical results show the applicability of the model to hyper-dense networks.

## APPENDIX
## PROOF OF PROPOSITION 1

According to the mean-field multi-agent multi-armed bandit model described in Section III-C, assume that upon pulling arm $m \in \mathcal{M}$, agent $n \in \mathcal{N}$ receives a random reward which follows a Bernoulli distribution with success probability $p_{nm}^{(s)} = Q(f_m, \theta_{nm})$. First we state the following definition as well as two theorems [25].[4]

*Definition 1 (Lipschitz Function): A function $g : \mathcal{X} \to \mathcal{Y}$ is called Lipschitz if there exists a real constant $L \geq 0$ such that, for all $x_1, x_2 \in \mathcal{X}$,*

$$d_Y(f(x_1), f(x_2)) \leq L d_X(x_1, x_2). \tag{11}$$

*For example, any function with a bounded first derivative is Lipschitz.*

*Theorem 1: If $Q$ is continuous in $f$, then a mean-field equilibrium exists.*

*Theorem 2: Suppose that $Q$ is Lipschitz in $f$ with Lipschitz constant $L$. Then if $\alpha(1 + L) < 1$, then there exists a unique mean-field equilibrium, and the mean-field dynamics converges to it from any initial condition.*

The first part of Proposition 1 immediately follows from Theorem 1, since by (10) (also see (8) and (9)), the success probability $p_{nm}^{(s)}$ is continuous in $f_m$.

To proof the second part, we first substitute (8) in (9), so that we have

$$p_{nm}^{(s)} = 1 - \operatorname{erf}\left[\frac{I_{nm} + N_0}{\sqrt{2}\sigma_n h'_{nm}}\left(e^{\frac{N f_m r_{n,\min}}{W_m}} - 1\right)\right]. \tag{12}$$

Then, $\frac{\partial p_{nm}^{(s)}}{\partial f_m}$ can be calculated as

$$\left|\frac{\partial p_{nm}^{(s)}}{\partial f_m}\right| = a_{nm} b_{nm} e^{-\frac{\pi}{4} a_{nm}^2 \left(-1 + e^{b_{nm} f_m}\right)^2 + b_{nm} f}, \tag{13}$$

with $a_{nm}$ and $b_{nm}$ being as defined in the proposition. Since $0 \leq f_m \leq 1$, we have

$$\frac{\partial p_{nm}^{(s)}}{\partial f_m} \leq a_{nm} b_{nm} e^{b_{nm}}. \tag{14}$$

Thus $p_{nm}^{(s)}$ is Lipschitz in $f_m$ with Lipschitz constant being $a_{nm} b_{nm} e^{b_{nm}}$. Thus by Theorem 2, if $\alpha(1 + a_{nm} b_{nm} e^{b_{nm}}) < 1$, the mean-field equilibrium is unique, and the mean field dynamics converges to it from any initial point. Hence the result follows.

[4] The full pre-print can be found at http://dx.doi.org/10.2139/ssrn.2045842.

## REFERENCES

[1] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 118–127, Jun. 2014.
[2] G. Fodor *et al.*, "Design aspects of network assisted device-to-device communications," *IEEE Commun. Mag.*, vol. 50, no. 3, pp. 170–177, Mar. 2012.
[3] R. Zhang and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1989–2001, May 2013.
[4] P. Kamalinejad, C. Mahapatra, Z. Sheng, S. Mirabbasi, V. C. M. Leung, and Y. L. Guan, "Wireless energy harvesting for the internet of things," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 102–108, Jun. 2015.
[5] T. Li, P. Fan, Z. Chen, and K. B. Letaief, "Optimum transmission policies for energy harvesting sensor networks powered by a mobile control center," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6132–6145, Sep. 2016.
[6] F. Yuan, Q. T. Zhang, S. Jin, and H. Zhu, "Optimal harvest-use-store strategy for energy harvesting wireless systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 698–710, Feb. 2015.
[7] S. Lee and R. Zhang, "Distributed wireless power transfer with energy feedback," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1685–1699, Apr. 2017.
[8] D. Liu *et al.*, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart., 2016.
[9] Y. Li, "Spatial interference cancellation in heterogeneous networks with biased cell association," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–7.
[10] P.-H. Chiang, P.-H. Huang, S.-S. Sun, W. Liao, and W.-T. Chen, "Joint power control and user association for traffic offloading in heterogeneous networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 4424–4429.
[11] Y.-P. Zhang, S. Feng, and P. Zhang, "Adaptive cell association and interference management in LTE-A small-cell networks," in *Proc. IEEE 78th Veh. Technol. Conf. (VTC Fall)*, Sep. 2013, pp. 1–6.
[12] X. Zhou, S. Feng, Z. Han, and Y. Liu, "Distributed user association and interference coordination in HetNets using Stackelberg game," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 6431–6436.
[13] M. Kim, S. Y. Jung, and S. L. Kim, "Sum-rate maximizing cell association via dual-connectivity," in *Proc. Int. Conf. Comput., Inf. Telecommun. Syst. (CITS)*, Jul. 2015, pp. 1–5.
[14] Z. Cui and R. Adve, "Joint user association and resource allocation in small cell networks with backhaul constraints," in *Proc. Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2014, pp. 1–6.
[15] S. Maghsudi and E. Hossain. (Nov. 2015). "Distributed user association in energy harvesting small cell networks: A competitive market model with uncertainty." [Online]. Available: https://arxiv.org/abs/1511.01976
[16] S. Maghsudi and E. Hossain, "Distributed user association in energy harvesting small cell networks: A probabilistic bandit model," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1549–1563, Mar. 2017..
[17] C.-Y. Chang, K.-L. Ho, W. Liao, and D.-S. Shiu, "Capacity maximization of energy-harvesting small cells with dynamic sleep mode operation in heterogeneous networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 2690–2694.
[18] L. Dong, G. Wu, Z. Xu, and S. Li, "Energy efficient pico base station switching-on/off in heterogeneous cellular network with minimum rate requirement," in *Proc. 16th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2014, pp. 1–6.
[19] S. Kim, S. Choi, and B. G. Lee, "A joint algorithm for base station operation and user association in heterogeneous networks," *IEEE Commun. Lett.*, vol. 17, no. 8, pp. 1552–1555, Aug. 2013.
[20] A. Mesodiakaki, F. Adelantado, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Energy and spectrum efficient user association in 5G heterogeneous networks," in *Proc. IEEE 27th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–6.
[21] H. Shariatmadari *et al.*, "Machine-type communications: Current status and future perspectives toward 5G systems," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 10–17, Sep. 2015.
[22] S. Maghsudi and S. Stanczak, "Transmission mode selection for network-assisted device to device communication: A levy-bandit approach," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 7009–7013.
[23] S. Sekander, H. Tabassum, and E. Hossain, "A matching game for decoupled uplink-downlink user association in full-duplex small cell networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–6.
[24] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Match to cache: Joint user association and Backhaul allocation in cache-aware small cell networks," in *Proc. IEEE ICC*, London, U.K., Jun. 2015, pp. 3082–3087.
[25] R. Gummadi, R. Johari, and J. Y. Yu, "Mean field equilibria of multi armed bandit games," in *Proc. 50th Annu. Allerton Conf. Commun., Control Comput.*, Oct. 2012, p. 1110.
[26] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multi armed bandit problem," *Mach. Learn.*, vol. 47, no. 2, pp. 235–256, 2002.
[27] S. Bubeck and N. Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Found. Trends Mach. Learn.*, vol. 5, no. 1, pp. 1–122, 2012.
[28] S. Maghsudi and E. Hossain, "Multi-armed bandits with application to 5G small cells," *IEEE Wireless Commun.*, vol. 23, no. 3, pp. 64–73, Jun. 2016.
[29] R. Gummadi, R. Johari, S. Schmit, and J. Y. Yu, "Mean field analysis of multi-armed bandit games," in *Proc. SSRN*, Apr. 2013, pp. 1–28.
[30] D. Liu, Y. Chen, K.-K. Chai, and T. Zhang, "Two dimensional optimization on user association and green energy allocation for HetNets with hybrid energy sources," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4111–4124, Nov. 2015.
[31] H. Kim, G. de Veciana, X. Yang, 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.
[32] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, "Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1525–1536, Sep. 2011.

**SETAREH MAGHSUDI** received the Ph.D. (*summa cum laude*) degree in electrical engineering from the Technical University of Berlin, Germany, in 2015. From 2015 to 2016, she was a Post-Doctoral Research Associate with the Department of Electrical and Computer Engineering, University of Manitoba, Canada. She is currently a Post-Doctoral Fellow with the School of Engineering and Applied Science, Yale University. Her research interests lie in the area of network analysis and optimization under uncertainty, mainly using game- and bandit theory. In particular, her work is focused on distributed resource management in large-scale networks, such as the Internet of Things and cyber-physical systems. She was a recipient of the German Research Foundation (DFG) Postdoctoral Fellowship.

**EKRAM HOSSAIN** (F'15) received the Ph.D. degree in electrical engineering from the University of Victoria, Canada, in 2001. He has been a Professor with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Canada, since 2010. He is a member (Class of 2016) of the College of the Royal Society of Canada. His current research interests include design, analysis, and optimization of wireless/mobile communications networks, cognitive radio systems, and network economics. He has authored/edited several books in these areas. He is also a member of the IEEE Press Editorial Board. He has received several research awards, including the IEEE Vehicular Technology Conference (VTC - Fall 2016) Best Student Paper Award as a Co-Author, the IEEE Communications Society Transmission, Access, and Optical Systems Technical Committee's Best Paper Award in the IEEE Globecom 2015, University of Manitoba Merit Award in 2010, 2014, and 2015 (for Research and Scholarly Activities), the 2011 IEEE Communications Society Fred Ellersick Prize Paper Award, and the IEEE Wireless Communications and Networking Conference 2012 Best Paper Award. He was elevated to an IEEE Fellow for spectrum management and resource allocation in cognitive and cellular radio networks. He was a Distinguished Lecturer of the IEEE Communications Society from 2012 to 2015. He is currently a Distinguished Lecturer of the IEEE Vehicular Technology Society. He is a registered Professional Engineer in the province of Manitoba, Canada. He serves as an Editor of the IEEE WIRELESS COMMUNICATIONS. He served as the Editor-in-Chief of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS from 2012 to 2016, an Area Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS in the area of Resource Management and Multiple Access from 2009 to 2011, an Editor of the IEEE TRANSACTIONS ON MOBILE COMPUTING from 2007 to 2012, and an Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS - Cognitive Radio Series from 2011 to 2014.