A Cheat-Proof Power Control Policy for Self-Organizing Full-Duplex Small Cells

Prabodini Semasinghe and Ekram Hossain, *Fellow, IEEE*

***Abstract*—Distributed resource allocation is considered as a significant feature of future self-organizing wireless networks. On the other hand, full-duplex transmission is an emerging technology which can theoretically double the data rate. and hence enhance the performance of wireless networks significantly. In this work, we address the problem of distributed power control in a two-tier network with full-duplex small cells underlaying one macro cell in a co-channel deployment scenario. We first formulate the corresponding distributed power control problem as a non-cooperative game and then extend it to a repeated game with imperfect public monitoring. The repeated game setting is used as it can withstand cheating. We show the existence and uniqueness of the Nash equilibrium of the formulated non-cooperative game and characterize the set of perfect public equilibrium for the repeated game. A two phase distributed algorithm is then proposed to achieve a power profile which is also a perfect public equilibrium. The power control algorithm is cheat-proof and needs only a small amount of information exchange among network nodes. The effectiveness of the algorithm is shown using numerical results. Our proposed algorithm and analysis are also valid for a half-duplex system as it is a special case of the full-duplex model presented in the paper.**

***Index Terms*—Self-organizing full-duplex small cells, distributed joint downlink-uplink power control, repeated games, public perfect equilibrium.**

I. Introduction

The demand for high speed data will be increased ever than before in future wireless communication networks. One of the most promising ways to increase data rate is to deploy small cells. Small cells can increase data rate by providing higher quality links and also by exploiting more spatial spectrum reuse. Therefore, cellular networks are becoming heterogeneous with different types of small cells (e.g., femto, micro, pico cells) underlaying the traditional macro-cellular networks. Full-duplex transmission (i.e., transmitting and receiving at the same time in the same frequency band[1]) is also another emerging technology which can theoretically double the data rate. Moreover, recent studies have shown that full-duplex technology works better for low power transmission nodes as self-interference can be reduced to the noise power level [1]. Therefore, exploring the possibilities to deploy full-duplex small cells is important.

The authors are with the Department of Electrical and Computer Engineering at the University of Manitoba, Canada (emails: semasilp@myumanitoba.ca, ekram.hossain@umanitoba.ca). The work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

[1]Simultaneous transmission and receiption in the same frequency band is referred to as in-band full-duplexing and simultaneous transmission and receiption in different frequency bands is referred to as out-of-band full-duplexing. In this work, we consider in-band full-duplex transmission.

Different from cautiously planned macro cells, the deployment of small cells is independent and unknown to the network operator. Due to this independent deployment and increased density of small cells, centralized network control and manual intervention will be extremely inefficient. Therefore, self organization is proposed as an essential feature of future wireless networks [2]. In a self-organizing network, network nodes will take individual decisions on resource management (i.e., distributed control). Moreover, backhaul capacity available for information exchange among network nodes will be still limited. Therefore, distributed resource allocation with less information exchange is a key requirement of a self organizing small cell network.

In this work, we develop a distributed power control scheme for a full-duplex small cell network underlaying a macro network. We formulate the corresponding power control problem as a repeated game. In next section, we discuss the related work and motivation for the work presented in this paper.

II. Related Work, Motivation, and Contribution

Game theory is suitable to model distributed resource allocation problems in wireless networks. Different types of games have been used to model different types of distributed resource allocation problems. Resource allocation algorithms based on non-cooperative games, which attain Nash Equilibrium (NE) as the solution, can be found in several work. In [3] and [4], the authors formulate uplink power control problem in a small cell network as a non-cooperative game. In [5], a non-cooperative game is used to model the downlink power control problem in a small cell network. The game is shown to be a supermodular game and NE is achieved distributively by implementing the best response dynamics. In [6], the authors formulate the uplink resource block allocation problem in an overlay macrocell-femtocell network as a potential game. A distributed technique is proposed using the best response dynamics which guarantees the convergence to a NE. All the above schemes are for either uplink or downlink resource allocation, i.e., for half-duplex systems.

For full-duplex systems, the previous solutions are not applicable since uplink and downlink transmissions now affect each other due to self-interference and hence cannot be decoupled. There are very few work in the literature which consider distributed resource allocation for full-duplex small cells. In [7], the authors propose a stable matching game for subcarrier allocation in full-duplex single-tier small cells. However, a theoretical analysis of the presented algorithm is absent. In [8], the joint uplink-downlink resource allocation problem for

978-1-4799-6664-6/16/$31.00

a full-duplex single cell is modeled as a non-cooperative game and solved for the NE. Different from the above mentioned work, our model considers distributed joint downlink-uplink power control for a two-tier network with multiple full-duplex small cells underlaying a macro cell.

Importantly, most of the existing distributed resource allocation algorithms are based on different types of non-cooperative *one-shot games*, i.e., the game is played only once. However, the resource allocation process such as a power control process is a repeated process. Therefore, it is more natural to model a resource allocation problem as a repeated game [9] than as a one-shot game. In a repeated game, a base game (called the *stage game*) is played a finite number of times or infinitely many times. When the game is played several times, each player can take decisions based on previous actions of other players. Therefore, decisions produced by repeated games can fundamentally differ from those of one-shot games. Power control and spectrum sharing problems of wireless systems are modeled as repeated games in [10], [11] and [12]. However, those work rely on the availability of perfect information of each player's history (*Perfect Monitoring*). We relax this perfect monitoring assumption in our model and assume imperfect monitoring of a public signal (*Imperfect Public Monitoring*) to make it more appropriate for a self-organizing system. In [13], the authors also use repeated games with imperfect public monitoring to develop a TDMA-based spectrum sharing scheme for a cognitive radio network.

The main contributions of this paper are listed below.

1) Formulation of the joint downlink-uplink power control problem for a co-channel deployed two-tier network with full-duplex small cells underlaying a macro cell as a non-cooperative game. The formulation considers power control for both macro and small cell tiers.
2) Analysis of the existence and uniqueness of the NE of the formulated non-cooperative stage game.
3) Extension of the formulated non-cooperative game to a repeated game to obtain a solution which Pareto dominates the NE of the stage game.
4) Theoretical characterization of perfect public equilibrium[2] payoff set for the repeated game formulation with imperfect public monitoring.
5) A distributed algorithm based on the theory of perturbed Markov Chains to obtain a PPE power control policy.

III. System Model

We consider K small cells with full-duplex users and base stations underlaying a half-duplex MBS. Each of the base stations serves a single user at a given time instant. Each base station and its serving user are considered as one entity and denoted by set $\mathcal{K} = \{0, 1, 2, ..., K\}$, where 0 denotes the macro cell and $1, 2, ..., K$ denote the small cells. Co-channel deployment is considered, i.e., all base stations including the MBS transmit on the same channel. Downlink and uplink transmit powers of each cell $k \in \{0, 1, 2, ..., K\}$ are denoted by p_k^{UL} and p_k^{DL}. SBSs can operate in any of the three modes (i.e., full-duplex mode when $p_k^{UL}, p_k^{DL} > 0$, half-duplex mode when either p_k^{UL} or p_k^{DL} is zero, OFF mode when $p_k^{UL} = p_k^{DL} = 0$). It is also assumed that the half-duplex MBS always transmits in the downlink (i.e., $p_0^{UL} = 0$). For simplicity, only long-term signal attenuation due to path-loss is considered. However, it is straightforward to include fading and the analysis presented in the next sections will still be valid.

The downlink SINR at the user and the uplink SINR at the base station of a generic cell k is given by (1) and (2), respectively, as follows:

$$\mathrm{SINR}_k^{DL} = \frac{p_k^{DL} r_{k_b,k_u}^{-\alpha}}{N_0 + I_{k_u}^{UL} + I_{k_u}^{DL} + \mathcal{C}(p_k^{UL})}, \tag{1}$$

$$\mathrm{SINR}_k^{UL} = \frac{p_k^{UL} r_{k_u,k_b}^{-\alpha}}{N_0 + I_{k_b}^{UL} + I_{k_b}^{DL} + \mathcal{C}(p_k^{DL})}, \tag{2}$$

where N_0 is variance of noise power, α is the path-loss exponent, k_u and k_b denote the user and base station of cell k, respectively, $r_{i,j}$ denotes the distance between network node[3] i and j. Also,

- $I_l^{UL} = \sum_{j \in \mathcal{K}/k} p_j^{UL} r_{j_u,l}^{-\alpha}$ is the interference caused by the uplink transmission of all the other users at the network node l,
- $I_l^{DL} = \sum_{j \in \mathcal{K}/k} p_j^{DL} r_{j_b,l}^{-\alpha}$ is the interference caused by the downlink transmission of the other base stations at the network node l,
- $\mathcal{C}(p)$ is the self-interference at a network node which transmits at power p.

The above equations are valid for the macro cell as well. In particular, $\mathrm{SINR}_0^{UL} = 0$ and $\mathcal{C}(p_k^{UL}) = 0$ since $p_k^{UL} = 0$. It is also assumed that each network node selects its transmit power from a pre-defined finite set of values.

The MBS is capable of measuring the interference caused by the small cells at its user. After each transmission, the MBS broadcasts a public message based on this measured interference to all SBSs on a delay-free channel. It is also assumed that all users have a perfect delay-free feedback channel to their base stations. Thus, users can update the base stations about their performances at each step.

IV. Game Formulation and Analysis

A. Stage Game, $\mathcal{G}_s$

We first formulate the stage game $\mathcal{G}_s$.

Set of players ($\mathcal{K}$): Each base station and its serving user are considered as one player of $\mathcal{G}_s$. The set of players is denoted by $\mathcal{K} = \{0, 1, 2, ..., K\}$.

Set of actions for player k, ($\mathcal{A}_k$): Downlink and uplink transmit powers of each base station-user pair k is drawn from a finite set of power levels given by

$$\mathcal{P}_k^{UL} = \{0, p_k^{UL,1}, p_k^{UL,2}, ..., p_k^{UL,max}\}$$

[2]The perfect public equilibrium (PPE) is equivalent to the concept of subgame perfect equilibrium in repeated games with perfect history. The formal definition of PPE will be given in Section IV.

[3]A network node is either a user or a base station.

and

$$\mathcal{P}_k^{DL} = \{0, p_k^{DL,1}, p_k^{DL,2}, ..., p_k^{DL,max}\},$$

respectively. The action of player k (i.e., $\boldsymbol{a}_k$) is the combination of uplink and downlink transmit powers. This is also called the transmit power profile (i.e., $\boldsymbol{p}_k$) of player k. The action of player k is denoted by

$$\boldsymbol{a}_k = \boldsymbol{p}_k = \begin{bmatrix} p_k^{UL} & p_k^{DL} \end{bmatrix}^T. \tag{3}$$

$\boldsymbol{a}_{-k}$ represents the actions of all players other than k. The action profile of the system is given by: $\boldsymbol{a} = \boldsymbol{p} = [\boldsymbol{a}_0, \boldsymbol{a}_1, ..., \boldsymbol{a}_K] \in \mathcal{A} = \mathcal{A}_0 \times \mathcal{A}_1 \times ... \times \mathcal{A}_K$. The size of the action set of each player is given by: $|\mathcal{A}_k| = |\mathcal{P}_k^{UL}||\mathcal{P}_k^{DL}|$. Also, note that for the MBS (i.e., player 0), $\mathcal{P}_0^{UL}$ is a null-set as it only transmits in the downlink. Therefore, player 0's action set is only composed of its possible downlink transmit power levels. We use the symbols $\boldsymbol{a}$ and $\boldsymbol{p}$ interchangeably in the rest of the paper.

Payoff function of player k, π_k: Traditionally, when the SBSs are underlaid with a macro cell network, the payoff function for small cells has the following form ([2]):

$$\begin{aligned} \text{Payoff}_k = &\mathcal{F}\left(\text{SINR}_k^{UL}, \text{SINR}_k^{DL}\right) \\ &- c \times \text{Interference at macro user}, \end{aligned} \tag{4}$$

where $\mathcal{F}\left(\text{SINR}_k^{UL}, \text{SINR}_k^{DL}\right)$ is the gained utility which is a function of SINR at the receivers and c is a constant. Then there is a negative utility or a penalty for causing interference to the macro network which is denoted by the 2^{nd} term in (4). However, this penalty is virtual and does not cause any actual performance reduction in small cells. Therefore, dishonest SBSs may keep on increasing their utility (which is the first term in (4)) neglecting the penalty. This can reduce the performance of the macro network and also the performance of other small cells as well. Therefore, instead of considering a payoff function of the above format, we make the macro cell also a player of the game and each player's payoff is only based on the utility gained due to the SINR value at the receiver. The payoff of player k is calculated as follows:

$$\pi_k = \log\left(\text{SINR}_k^{UL}\right) + \log\left(\text{SINR}_k^{DL}\right). \tag{5}$$

This payoff function is equivalent to the rate if SINR $>> 1$. In the following theorem, we comment on the existence and the uniqueness of the NE of the above stage game $\mathcal{G}_s$.

Theorem IV.1. *Stage game $\mathcal{G}_s$ has a unique NE when the following conditions are satisfied:*

- *Self-interference is a linear function of the transmit power, and*
- *Interference cancellation factor is less than the path-loss between the transmitter and the intended receiver.*

Moreover, the equilibrium power profile is given by $\boldsymbol{p}_{k,NE} = \left(p_k^{UL,max}, p_k^{DL,max}\right)$ for all $k \in \mathcal{K}$.

Proof: The proof is ommited due to limited space. ■

The only NE for the above game is all base stations transmit in their maximum possible power. Therefore, at the NE, the players also experience the maximum interference. Thus, the NE may not be the Pareto optimal power profile. It is possible to obtain a solution, which Pareto dominates $\boldsymbol{p}_{k,NE}$, by formulating a repeated game as shown in next section.

B. Repeated Game, $\mathcal{G}_r$

We consider repeating the game $\mathcal{G}_s$ infinitely many times and the corresponding repeated game is denoted by $\mathcal{G}_r$. The main components of $\mathcal{G}_r$ are given below.

Set of players: Same set of players as in $\mathcal{G}_s$, $\mathcal{K} = \{0, 1, 2, ..., K\}$.

Action set of a player k: The action set of a generic player k at each step of $\mathcal{G}_r$ is same as the action set of that player in $\mathcal{G}_s$ which is denoted by $\mathcal{A}_k$. The action of a generic player k at t^{th} step of the game $\mathcal{G}_r$ is now denoted by $\boldsymbol{a}_k(t) = \boldsymbol{p}_k(t)$.

Payoff of player k, v_k: Payoff of player k is

$$v_k = (1-\delta) \lim_{T\to\infty} \sum_{t=0}^{T} (\delta)^t \pi_k(t), \tag{6}$$

where δ is the discount factor. The discount factor can also be seen as the probability of not ending the game at a certain time step. Also, $\boldsymbol{v}$ denotes the payoff vector of all players, $\boldsymbol{v} = \begin{bmatrix} v_1 & v_2 & ... & v_K \end{bmatrix}^T$.

History at step t, $\boldsymbol{h}(t)$: History of the game $\mathcal{G}_r$ at step t is composed of actions played by all the players until step $(t-1)$. The action profile of the system at step t is given by,

$$\boldsymbol{a}(t) = [\boldsymbol{a}_0(t), \boldsymbol{a}_1(t), ..., \boldsymbol{a}_K(t)].$$

Therefore, history at step t is given by

$$\boldsymbol{h}(t) = [\boldsymbol{a}(1), \boldsymbol{a}(2), ..., \boldsymbol{a}(t-1)]. \tag{7}$$

However, for each player to have knowledge of $\boldsymbol{h}(t)$, the players need to exchange information about their actions after each step of the game. This would result in huge amount of information exchange and hence not be realistic for a system with limited backhaul. Therefore, we assume that the MBS measures the interference at its user and broadcasts a public message based on that measurement after each step of the game. We denote the MBS's measurement on interference at its user at time t by $I\left(\boldsymbol{p}_{-0}(t)\right)$ which can be written as

$$I\left(\boldsymbol{p}_{-0}(t)\right) = \sum_{k\in\mathcal{K}/0} \left(I_{0_u}^{UL} + I_{0_u}^{DL}\right) + \eta, \tag{8}$$

where η is the estimation error and $\eta \sim \mathcal{N}\left(\mu, \sigma^2\right)$. Based on $I\left(\boldsymbol{p}_{-0}(t)\right)$, the MBS transmits a public message $m(t)$ which is drawn from a finite set of messages given by $\mathcal{M} \in \{m_1, m_2, ..., m_M\}$. The public history at step t can be written as $\boldsymbol{h}_{pub}(t) = [m(1), m(2), ..., m(t-1)]$.

Public strategy of player k, $\boldsymbol{s}_k$: The public strategy of player k is a mapping of any possible public history to an action. Game $\mathcal{G}_r$ has an infinite number of possible public histories as it is a repeated game with infinite time horizon. Therefore, the public strategy space ($\mathcal{S}_k$) of each player k is also infinite. Also, the public strategy profile of the system is given by $\boldsymbol{s} = [\boldsymbol{s}_0, \boldsymbol{s}_1, ..., \boldsymbol{s}_K]$.

The equilibrium concept for a repeated game with public history and public strategies is called perfect public equilibrium (PPE). The formal definition of PPE is given below.

Definition IV.2. Perfect public equilibrium (PPE): A public strategy profile $\bar{\boldsymbol{s}} = [\bar{\boldsymbol{s}}_0, \bar{\boldsymbol{s}}_1, ..., \bar{\boldsymbol{s}}_K]$ is a perfect public equilibrium if for any public history $\boldsymbol{h}(tnb)$, the continuation public strategy given by $\bar{\boldsymbol{s}}|h(t)$ is a Nash equilibrium of the continuing subgame, for all players, i.e.,

$$v_k(\bar{\boldsymbol{s}}|h_{pub}(t)) \geq v_k(\hat{\boldsymbol{s}}_k, \bar{\boldsymbol{s}}_{-k}|h_{pub}(t)), \ \ \forall \hat{s}_k, \ \forall k. \quad (9)$$

Before characterizing the set of PPEs of game $\mathcal{G}_r$, we define the following terms.

Definition IV.3. Minmax payoff: The minmax payoff is defined for the stage game. The minmax payoff of player k ($\pi_{k,minmax}$) is the maximum payoff that player k can achieve when everybody else is trying to minimize the payoff of k, i.e.,

$$\pi_{k,minmax} = \min_{\boldsymbol{a}_{-k} \in \mathcal{A}_{-k}} \max_{\boldsymbol{a}_k \in \mathcal{A}_k} \pi_k\left(\boldsymbol{a}_k, \boldsymbol{a}_{-k}\right). \quad (10)$$

Note that, the Nash equilibrium of game $\mathcal{G}_s$ gives the minmax payoffs for all players.

Definition IV.4. Enforceability: A payoff profile $\boldsymbol{v}$ of the repeated game $\mathcal{G}_r$ is ***enforceable***, if $v_k \geq \pi_{k,minmax}$ for all k. (Strictly enforceable if $v_k > \pi_{k,minmax}$.)

Definition IV.5. Feasibility: A payoff vector $\boldsymbol{v}$ of the repeated game $\mathcal{G}_r$ is ***feasible***, if there exist rational, non-negative values of $\beta_{\boldsymbol{a}}$ for all $\boldsymbol{a} \in \mathcal{A}$ such that v_k for all $k \in \mathcal{K}$ can be expressed as

$$\sum_{\boldsymbol{a} \in \mathcal{A}} \beta_{\boldsymbol{a}} \pi_k\left(\boldsymbol{a}\right) \quad \text{with} \quad \sum_{\boldsymbol{a} \in \mathcal{A}} \beta_{\boldsymbol{a}} = 1. \quad (11)$$

We consider a simplified version of game $\mathcal{G}_r$ (denoted by $\bar{\mathcal{G}}_r$) with $\mathcal{A}_k = \{a_1, a_2\}$ $\forall k$, $\mathcal{M} = \{m_1, m_2\}$ and $\eta \sim \mathcal{N}(0,1)$. The MBS sets $m(t) = m_1$ if $(I(\boldsymbol{p}(t)) + \eta) \leq I_{th}$ and $y(t) = m_2$ otherwise. I_{th} is a pre-defined threshold. Let $\Pi_k\left(\boldsymbol{a}_{-k}\right)$ be a matrix of size $|\mathcal{A}_k| \times |\mathcal{M}|$. For an arbitrary player k and a fixed $\boldsymbol{a}_{-k}$, $\Pi_k\left(\boldsymbol{a}_{-k}\right)$ is defined as

$$\Pi_k\left(\boldsymbol{a}_{-k}\right) = \begin{pmatrix} \Pr(m_1/\boldsymbol{a}_1, \boldsymbol{a}_{-k}) & \Pr(m_2/\boldsymbol{a}_1, \boldsymbol{a}_{-k}) \\ \Pr(m_1/\boldsymbol{a}_2, \boldsymbol{a}_{-k}) & \Pr(m_2/\boldsymbol{a}_2, \boldsymbol{a}_{-k}). \end{pmatrix}$$

An action profile $\boldsymbol{a}$ has ***individual full rank*** for player k if $\Pi_k\left(\boldsymbol{a}_{-k}\right)$ has rank $|\mathcal{A}_k|$. Now we state the following lemma for game $\bar{\mathcal{G}}_r$.

Lemma IV.6. *For any user $k \in \mathcal{K}$ of game $\bar{\mathcal{G}}_r$, any pure action profile $\boldsymbol{a} \in \mathcal{A}$ has individual full rank.*

Proof: For any arbitrary user k for fixed $\boldsymbol{a}_{-k}$, let $\Pr(m_1/\boldsymbol{a}_1, \boldsymbol{a}_{-k}) = \theta_1^k$ and $\Pr(m_1/\boldsymbol{a}_2, \boldsymbol{a}_{-k}) = \theta_2^k$. Then,

$$\Pi_k\left(\boldsymbol{a}_{-k}\right) = \begin{pmatrix} \theta_1^k & 1-\theta_1^k \\ \theta_2^k & 1-\theta_2^k \end{pmatrix}.$$

The probability of getting m_1 for any $\boldsymbol{i} \in \mathcal{A}_k$ (i.e., θ_i^k) can be derived as, $\Pr(m_1/\boldsymbol{i}, \boldsymbol{a}_{-k})$

$$\begin{aligned} &= \Pr\left(I(\boldsymbol{i}, \boldsymbol{a}_{-k}) + \epsilon \leq I_{th}\right) = \Pr\left(\epsilon \leq I_{th} - I(\boldsymbol{i}, \boldsymbol{a}_{-k})\right) \\ &= 0.5\left[1 + \operatorname{erf}\left(\frac{I_{th} - I(\boldsymbol{i}, \boldsymbol{a}_{-k})}{\sqrt{2}}\right)\right] = \theta_i^k. \end{aligned} \quad (12)$$

Since $\theta_1^k \neq \theta_2^k$, according to (12) the determinant of Π_k is non-zero. Therefore, Π_k has full rank which equals to $|\mathcal{A}_k|$ for any k, i.e., for any user k any pure action profile $\boldsymbol{a} \in \mathcal{A}$ has individual full rank. ■

Now, we present the theorem below to characterize the set of PPEs in game $\bar{\mathcal{G}}_r$.

Theorem IV.7. *Let V^* denote the set of feasible and enforceable payoffs of game $\bar{\mathcal{G}}_r$. For any smooth subset W in the interior of V^*, there exists a $\hat{\delta}$ such that for all $\delta \in (\hat{\delta}, 1)$ each point in W is a PPE.*

Proof: The above theorem is derived from the folk theorem for repeated games with imperfect monitoring in the seminal paper by Fudenberg [14]. The folk theorem states that for any smooth subset W in the interior of V^*, there exists a $\hat{\delta}$ such that for all $\delta \in (\hat{\delta}, 1)$ each point in W is a PPE if the following two conditions hold:

1) Any pure action profile has individual full rank, and
2) The number of public messages should exceed the value $|\mathcal{A}_k|-1$.

$\bar{\mathcal{G}}_r$ satisfies the above conditions, hence the set of PPEs can be characterized as in the above theorem. ■

According to theorem IV.7, any point in V^* can be obtained as a PPE for patient players, i.e., when $\delta \longrightarrow 1$. Thus, there can be infinitely many PPEs. Our objective is to find a desired operationg point for the system which is also a PPE. The minmax action profile can then be used as a punishment to motivate players to stay at the desired point.

V. Repeated Game-Based Power Control Algorithm

The proposed algorithm is composed of two phases, namely, learning phase and operation phase.

A. Learning Phase: Finding a Pareto Optimal Operating Point

During the learning phase, the players distributively learn an operating point. In [15], authors propose a distributed algorithm based on the theory of perturbed Markov chains to achieve a social optimal action profile for a system with a finite number of players each with finite set of actions. We modify this algorithm in order to find an operating point for the repeated game $\mathcal{G}_r$.

Each player k is with a state given by $y_k = [\bar{\boldsymbol{a}}_{\boldsymbol{k}}, \mathfrak{m}_k]^T$ where $\bar{\boldsymbol{a}}_{\boldsymbol{k}} \in \mathcal{A}_k$. $\mathfrak{m}_k$ is called the *mood* of player k which can take two values called as *Content* (C) or *Discontent* (D). All players individually update their current actions and states according to the learning rules given in **Algorithm 1**. This algorithm does not need any information exchange among players and decisions are taken solely based on individual payoff. After sufficient number of iterations each player would select an action profile which provides the social optimal payoff with a probability close to 1 ([15]).

At the end of the learning phase, the MBS observes the value of interference received at its user I_{max} and sets the interference threshold $I_{th} = I_{max} + \zeta$, where ζ is some predefined value to compensate the measurement errors.

Algorithm 1 Learning Phase

Initialization: $Iter_{max}$= maximum number of iterations, select $\bar{\boldsymbol{a}}_{\boldsymbol{k}}$ randomly, $\mathfrak{m}_k = D$, $iteration = 1$, $c = |\mathcal{K}|+1$

repeat

$t = iteration$, $\epsilon = \frac{1}{\sqrt{t}}$

Step 1: Selecting $\boldsymbol{a}_k(t)$

if $\mathfrak{m}_k = C$ **then**

Select $\boldsymbol{a}_k(t)$ according to following PDF:

$$p(d) = \begin{cases} 1 - \epsilon^c, & \text{if } d = \bar{\boldsymbol{a}}_k \\ \frac{1-\epsilon^c}{|\mathcal{A}_k|-1}, & \text{otherwise} \end{cases}$$

else

Select $\boldsymbol{a}_k(t)$ according to following PDF:

$$p(d) = \left\{ \frac{1}{|\mathcal{A}_k|}, \qquad \forall d \in \mathcal{A}_k \right.$$

end if

Calculate the normalized instantaneous payoff $u_k(t)$

Step 2: Update state $[\bar{\boldsymbol{a}}_k \;\; \mathfrak{m}_k]$

if $\mathfrak{m}_k = C$ **then**

if $\boldsymbol{a}_k(t) = \bar{\boldsymbol{a}}_k$ **then**

$\bar{\boldsymbol{a}}_k = \boldsymbol{a}_k(t)$, $\mathfrak{m}_k = C$

else

$$[\bar{\boldsymbol{a}}_k \;\; \mathfrak{m}_k] = \begin{cases} [\boldsymbol{a}_k(t) \;\; C], & \text{with prob. } \epsilon^{(1-u_k(t))} \\ [\boldsymbol{a}_k(t) \;\; D], & \text{with prob. } 1 - \epsilon^{(1-u_k(t))} \end{cases}$$

end if

else

$$[\bar{\boldsymbol{a}}_k \;\; \mathfrak{m}_k] = \begin{cases} [\boldsymbol{a}_k(t) \;\; C], & \text{with prob. } \epsilon^{(1-u_k(t))} \\ [\boldsymbol{a}_k(t) \;\; D], & \text{with prob. } 1 - \epsilon^{(1-u_k(t))} \end{cases}$$

end if

$iteration = iteration + 1$

until $iteration \geq Iter_{max}$

$\boldsymbol{p}_{k,opt} = \bar{\boldsymbol{a}}_k$,

B. Operation Phase

In phase 2, all players transmit using the learned power profile during phase 1. At the end of each step t, the MBS measures the received interference at its user, given by $I(t)$. MBS broadcast the public message $m(t) = 0$ if $I(t) \leq I_{th}$ and $m(t) = 1$ otherwise. If $I(t) > I_{th}$, it means some player has cheated by transmitting a higher power than the power profile learned at the learning phase. Then, to punish the cheating player, all players start transmitting at maximum power both in uplink and downlink which is also the NE power profile. This is also called the grim trigger strategy which is as follows:

$$\boldsymbol{p}_k(t+1) = \begin{cases} \boldsymbol{p}_{k,opt}, & \text{if} \quad h(t) = [0, 0, ..., 0] \\ \boldsymbol{p}_{k,NE}, & \text{otherwise.} \end{cases}$$

If for all k, $\pi_k(\boldsymbol{p}_{opt}) > \pi_k(\boldsymbol{p}_{NE})$, the grim trigger strategy will cause a reduction of long-term payoff for any cheating player. Therefore, if all players are rational they have no incentive to deviate from $\boldsymbol{p}_{opt}$ and grim trigger strategy (i.e, the grim trigger strategy is stable).

VI. Simulation Results and Discussion

The values of the main simulation parameters are given in Table I. We consider a network with two small cells and one macro cell.

A. Learning Phase

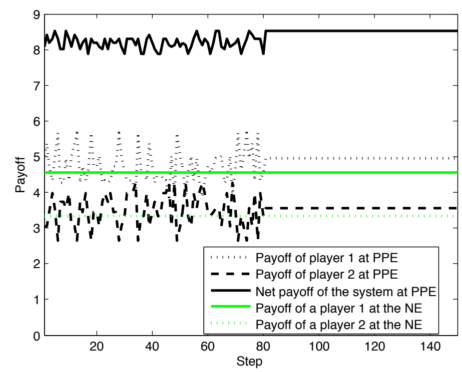

Fig. 1. Learning phase: A system with two small cell players.

Fig. 1 shows the payoff of each player during the learning phase when only small cells play the power control game. The MBS transmits at a uniform transmit power all the time. It can be seen that the learning phase converges to an operating point which gives higher payoffs than the NE for both players.

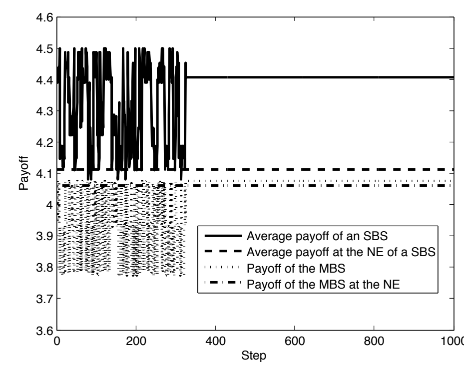

Fig. 2. Learning phase: A system with two small cells and one macro cell.

Then, we observe the learning phase when the macro cell also participates in the game with the two small cells. Fig. 2 plots the average payoff of a small cell player and the payoff of the macro player during the learning phase. The learning algorithm converges to a point which gives a better payoff than NE for all players in this case as well.

Fig. 3 compares the interference at the macro user during the learning phase and at the NE. The macro user experiences the highest interference when the system is at the NE, because all small cell nodes transmit at the maximum transmit power

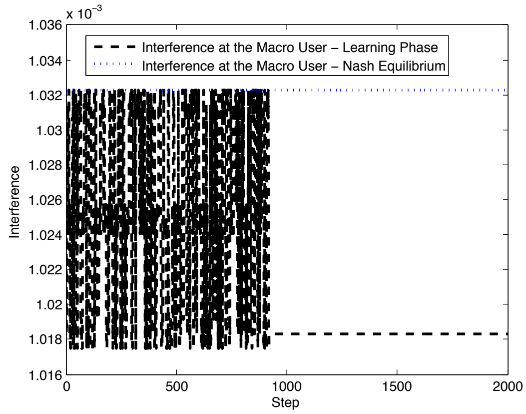

Fig. 3. Learning phase: Interference at the macro user.

at the NE. It can be seen from Fig. 3 that the interference at the macro user is reduced at the operating point learned at the learning phase.

B. Operation Phase

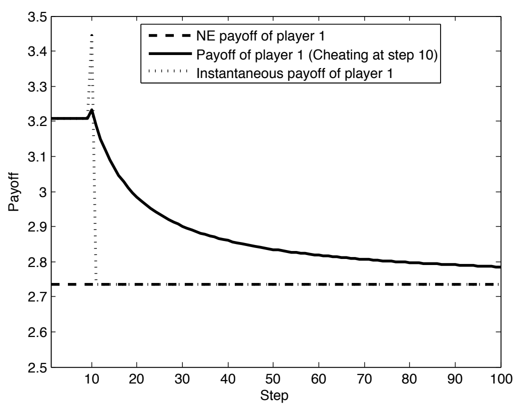

Fig. 4. Operation phase: Payoff of a cheating player.

Fig. 4 plots the variation of the payoff of a player who cheats at the operation phase. Player 1 cheats (i.e., transmits at the maximum possible power) at the 10^{th} step of the operation phase. By cheating, the player can obtain a considerably higher instantaneous payoff. The actual payoff, which is the average payoff over time (as in (6)) also increases. However, in the next step cheating is detected and grim trigger is pulled, i.e., all players start transmitting at the maximum power to punish the cheating player. This drops the cheated player's instantaneous payoff to its minmax payoff and the average payoff starts to decline. However, the objective of a player is to maximize her average payoff over time. Therefore, a rational player will not cheat. Thus, the power control algorithm is cheat-proof.

VII. CONCLUSION

We have proposed a distributed cheat-proof power control algorithm for a self-organizing heterogeneous network. The power control problem of a co-channel deployed two-tier network with small cells underlaying a macro cell has been formulated as a repeated game with imperfect public monitoring. We have characterized the perfect public equilibrium payoff set of the formulated repeated game and proposed a distributed algorithm to obtain a proper operating point for the system. The simulation results verify that the algorithm is cheat-proof if all players are rational.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
$\mathcal{P}_k^{DL}$ for small cells	$\{10, 15, 20\}$ W
$\mathcal{P}_0^{DL}$ for MBS	$\{30, 35, 40\}$ W
$\mathcal{P}_k^{UL}$ for small cells	$\{10\}$ W
N_0,α,δ,γ	0.001 W , 4, 1, 0.001

REFERENCES

[1] K. M. Thilina, H. Tabassum, E. Hossain, and D. I. Kim, "Medium access control design for full duplex wireless systems: challenges and approaches," *IEEE Communications Magazine*, vol. 53, no. 5, pp. 112–120, 2015.

[2] P. Semasinghe and E. Hossain, "Downlink power control in self-organizing dense small cells underlaying macrocells: A mean field game," *IEEE Transactions on Mobile Computing*, 2015, to appear.

[3] W. Ma, H. Zhang, W. Zheng, and X. Wen, "Differentiated-pricing based power allocation in dense femtocell networks," in *15th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Sept 2012, pp. 599–603.

[4] V. Chandrasekhar, J. Andrews, T. Muharemovict, Z. Shen, and A. Gatherer, "Power control in two-tier femtocell networks," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 4316–4328, August 2009.

[5] E. J. Hong, S. Y. Yun, and D.-H. Cho, "Decentralized power control scheme in femtocell networks: A game theoretic approach," in *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2009, pp. 415–419.

[6] I. Mustika, K. Yamamoto, H. Murata, and S. Yoshida, "Potential game approach for self-organized interference management in closed access femtocell networks," in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, May 2011, pp. 1–5.

[7] B. Di, S. Bayat, L. Song, and Y. Li, "Radio resource allocation for full-duplex ofdma networks using matching theory," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2014*, April 2014, pp. 197–198.

[8] M. Al-Imari, M. Ghoraishi, P. Xiao, and R. Tafazolli, "Game theory based radio resource allocation for full-duplex systems," in *IEEE 81st Vehicular Technology Conference (VTC Spring)*, May 2015, pp. 1–5.

[9] G. J. Mailath and L. Samuelson, *Repeated games and reputations*. Oxford University Press, 2006, vol. 2.

[10] Y. Xiao, J. Park, and M. van der Schaar, "Repeated games with intervention: Theory and applications in communications," *IEEE Transactions on Communications*, vol. 60, pp. 3123–3132, October 2012.

[11] M. Le Treust and S. Lasaulce, "A repeated game formulation of energy-efficient decentralized power control," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 2860–2869, September 2010.

[12] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 517–528, April 2007.

[13] Y. Xiao and M. van der Schaar, "Dynamic spectrum sharing among repeatedly interacting selfish users with imperfect monitoring," *IEEE Journal on Selected Areas in Communications,*, vol. 30, pp. 1890–1899, November 2012.

[14] D. Fudenberg, D. LEVINE, and E. MASKIN, "The folk theorem with imperfect public information," *Econometrica*, vol. 62, pp. 997–1039, 1994.

[15] J. R. Marden, H. P. Young, and L. Y. Pao, "Achieving pareto optimality through distributed learning," *SIAM Journal on Control and Optimization*, vol. 52, no. 5, pp. 2753–2770, 2014.