

RESEARCH

Open Access

A two-stage game theoretic approach for self-organizing networks

Habib BA Sidi*, Rachid El-Azouzi and Majed Haddad

Abstract

Growth of network access technologies in the mobile environment has raised several new issues due to the interference between the available access. Thanks to the currently used access methods such as the orthogonal frequency division multiple access in mobile networks and the long term evolution-advanced systems, the intra-cell interferences are avoided and the quality of service has increased. Nevertheless, the diversity and multiplicity of base stations in the network have left behind a major problem of inter-cell interferences. In this paper, we focus on the optimization of the total throughput of cellular networks using fractional frequency reuse and allowing each mobile user to individually choose its serving base station. We derive analytically the utilities related to the network manager and mobile users and develop a Stackelberg game to obtain the equilibrium. We propose a distributed algorithm that allows the base stations, using a light collaboration, to achieve an efficient utilization of the frequencies, with the optic of maximizing the total system utility. This algorithm is based on stochastic gradient descent which requires some information to be exchanged between neighboring base stations. At user association level, we propose an iterative distributed algorithm based on automata learning algorithm. Both algorithms allow the system to converge to the Stackelberg equilibrium. Furthermore, simulation results carried out based on a realistic network setting show promising results in terms of global utility and convergence issues. In this setting, we include scenarios with a varying number of users and address the problem of robustness and scalability of the proposed approach.

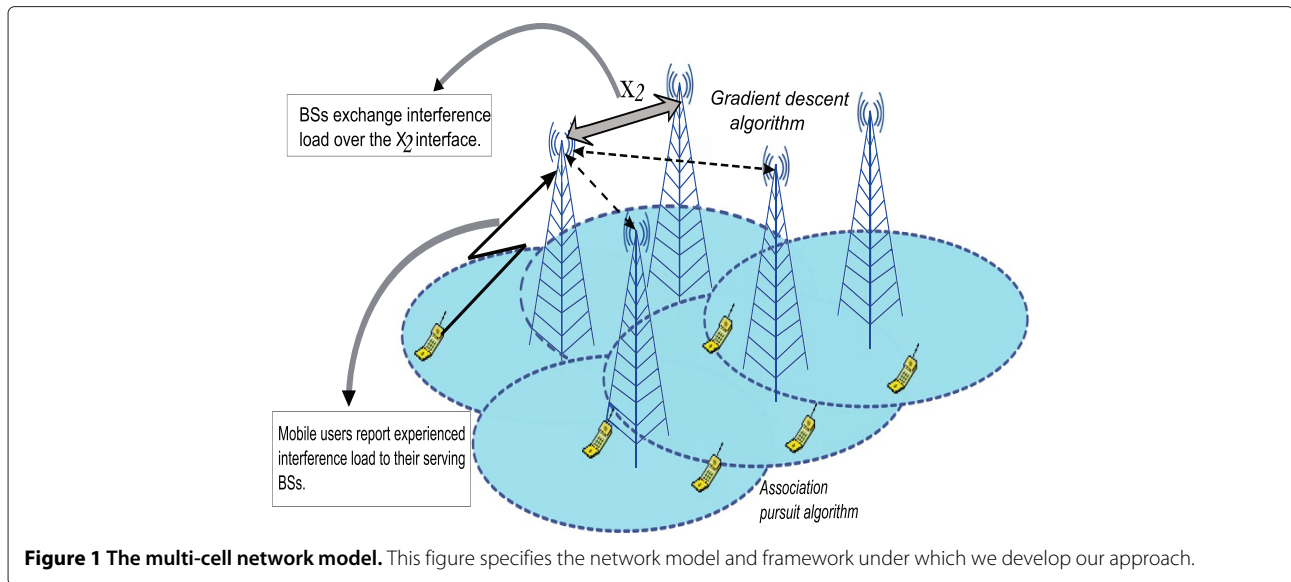
Introduction

Recently, the use of self-organizing network (SON) features in a framework of general policy management has been suggested. In such frameworks, SON entities are used as a means to enforce high-level operator policies, introduced in the management plane, and translated into low-level objectives guiding coordinated SON entities [1]. Among the most important self-optimization mechanisms in radio access networks (RAN) are interference coordination [2], mobility management, and energy saving [3]. Several such problems need further investigation to fully benefit from SON in RAN, in areas where little material has been published. Examples are autonomous cell outage management and coverage capacity optimization [4]. It is noted that the problem of coordinating simultaneous SON processes is an open and challenging problem that needs to be addressed in order to allow the deployment of SON mechanisms.

In this paper, we propose a self-optimization framework for inter-cell interference coordination in an orthogonal frequency division multiple access (OFDMA) network. Inter-cell interference can dramatically degrade cell performance and perceived quality of service (QoS), particularly at cell edge. We are interested in distributed solutions that can be implemented in a flat architecture (e.g., long term evolution (LTE)-Advanced architecture). To coordinate interference between neighboring cells, eNodeBs need to exchange information. In the case of LTE, for example, signaling between eNodeBs can be exchanged over the X2 interface (see Figure 1). Recent works such as fractional frequency reuse [5,6] and soft frequency reuse [7] allowing users in different channel conditions to benefit from different reuse patterns have been proposed. Still, all of these schemes mentioned above are static interference management approaches, where a specific reuse pattern is predetermined *a priori* by a network operator at offline.

Specifically, we assume that the fractional frequency-reuse (FFR) of a cell can be configured dynamically. In

*Correspondence: habib.sidi@etd.univ-avignon.fr
CERI/LIA, University of Avignon, Agroparc BP 1228, Avignon 84911, France



that case, some base stations (BSs or eNodeBs) would be enabled to adjust their FFR in order to provide coverage/capacity for other neighboring cells. We further model the network behavior as a Stackelberg game between the network manager and the mobile users using the game theory framework [8].

At the core lies the idea that introducing a certain degree of hierarchy in non-cooperative games not only improves the individual efficiency of all users but can also be a way of reaching a desired trade-off between the global network performance at the equilibrium and the requested amount of signaling. The proposed approach can be seen as an intermediate scheme between the totally centralized policy and the non-cooperative policy. It is also quite relevant for flexible networks where the trend is to split the intelligence between the network infrastructure and mobile users' equipments. In the Stackelberg game, the network manager is acting as the leader and mobile users as the followers. In the first stage, the leader chooses its strategy profile and announces it to the followers. Then, the followers decide their respective outcomes depending on the strategy profile of the leader. Under our scenario, the network manager maximizes the total network throughput by means of power control and announces its strategy profile to mobile users. Each mobile will decide *individually* to which of the available base stations it is best to connect according to its radio condition and the strategy profile broadcasted by the network.

We also propose a two-stage self-optimization algorithm for both the leader and the followers. The objective is to achieve dynamically an efficient frequency reuse pattern based on their past experience and their learning capabilities. The leader's algorithm is based on stochastic gradient descent algorithm which requires some

information to be exchanged between neighboring base stations. For user association, we propose an iterative distributed algorithm based on automata learning mechanisms. Both algorithms have been shown to converge to the Stackelberg equilibrium while providing substantial gain compared to optimal solution and fixed full reuse scheme.

The original contributions of our approach are threefold:

- Investigating fractional frequency reuse technique for inter-cell interference coordination in an OFDMA network
- Modeling the interaction between the network and mobiles using a Stackelberg game framework
- Proposing a hierarchical algorithm that allows convergence towards the Stackelberg equilibrium

In comparison to our previous work [9] presented at Wireless Days 2011, this paper^a extends with richer developments the materials presented before. Especially, we further explore the case when the network environment is dynamic. By dynamic, we mean that the number of users varies in time with mobiles arriving and departing the system. Through extensive simulations based on a realistic network setting, the proposed approach is shown to be robust and scalable. In this latter setting, we also give some insight on how to design a trade-off between the global network performance at the equilibrium and the requested amount of signaling. More clearly, the following contributions have been developed:

- Addressing convergence and stability properties of our distributed mechanisms with an evaluation of the computational cost.

- Exploring the robustness of the proposed approach with time-varying number of users, thus simulating a seamless dynamic environment.
- Giving some insight on the ways of finding a desired trade-off between the desired global network performance and the amount of control feedback.
- At the equilibrium, our mechanisms achieve up to 90% of the optimal association policy, with similar results with partial exchange of information or perturbed environment.

The paper is organized as follows: The system model is exposed in the ‘The system model’ section. The ‘Network resources’ section provides a description of the network scenario adopted throughout the paper. In the ‘Hierarchical game formulation’ section, we present the game theoretic framework and propose formally how the network manager and mobile users can obtain their respective equilibria by means of a Stackelberg formulation. In the ‘Learning for optimal decision’ section, the proposed hierarchical algorithm is investigated for both the leader and the followers. In the ‘Implementation and validation’ section, simulation results under realistic wireless network settings are shown to exhibit interesting features in terms of self-optimizing deployment for inter-cell interference coordination. The ‘Conclusion’ section concludes the paper.

Scenario description

The system model

Consider the downlink of a multi-cell system, operating in an OFDMA context giving rise to an inter-cell interference phenomenon. Power control is used by the base stations in an effort to preserve power and to limit interference and fading effects. With the same goal of maximizing their payoff, mobile users try to connect to the best serving cell. Specifically, we consider $\mathcal{M} = \{1, \dots, M\}$ as the set of all possible serving base stations (or cells) within the network and $\mathcal{K} = \{1, \dots, K\}$ as a set of K mobile users randomly distributed over the network. Each cell operates in a multi-band context with N physical resource blocks (PRB). Let $\mathcal{N} = \{1, \dots, N\}$ be the set of N PRBs per cell. Mobile users strategies s_k are the choice of a PRB n at a given BS j , i.e., $s_k = (j, n)$. Hence, the signal received by a mobile user k using strategy s_k depends not only on the BS transmit power but also on the interferences introduced by the other cells. The signal-to-interference-plus-noise ratio (SINR) measured at the user k associated with BS j can be expressed for all $j \in \mathcal{M}$ and $k \in \mathcal{K}$ as follows:

$$\text{SINR}_{j,k} = \frac{h_{j,k} \cdot P(s_k)}{\sigma^2 + \sum_{l \neq k} h_{j,k} \cdot P(s_l) \cdot f(s_k, s_l)}, \quad (1)$$

where $h_{j,k}$ is the block fading process measured at user k associated with BS j , $P(s_l)$ is the power received from BS i at PRB n' for mobile user l with $s_l = (j', n')$, and σ^2 is the noise variance. The interference function $f(s_k, s_l)$ is defined as follows:

$$f(s_k, s_l) = \begin{cases} 1, & \text{if } n = n' \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Network resources

A key example of dynamic resource allocation is that of power control, which serves as means for both battery savings at the mobile as well as interference management in the network. Formally, in this work, we assume that the network manager optimizes its global utility by means of power control optimization. Let \mathbf{P} be the $(M \times N)$ power control matrix whose element $P(j, n)$ represents the power received from BS $j \in \mathcal{M}$ at PRB $n \in \mathcal{N}$. Given these optimized power levels \mathbf{P} , mobile users choose the association actions that optimize their individual utilities. Notice that the maximization of the total throughput by the network manager is based on information sent by mobile users on interferences experienced from neighboring cells. We further assume that each base station can allocate a PRB to only one mobile user at a given time slot.

Hierarchical game formulation

We make use of a hierarchical equilibrium solution concept, i.e., the Stackelberg game, where the network manager is acting as the leader and mobile users are the followers. In view of maximizing its utility, the leader enforces its strategy on the followers that react rationally to this enforcement. A mobile user can decide to either transmit data or stay silent depending on its utility^b We assume that each mobile k has a target SINR noted by η_k which reflects its required QoS. Let $\mathcal{H}_k \subset \mathcal{M}$ be the set of base stations within a radius of r from user k such that $r \leq (\frac{P_{\max}}{\eta_k \cdot \sigma^2})^{1/\beta}$ where β is the path loss coefficient and P_{\max} is the maximum power at each base station. The motivation behind doing so is that, for computation purpose, one may only consider the subset \mathcal{H}_k rather than the original set \mathcal{M} . Let w_k be the user’s strategy when the user decides to stay silent on that specific slot. Hence, the set of mobile user actions is $\Omega_k = \mathcal{H}_k \times \mathcal{N} \cup w_k$. Mobile user utility function for each choice of strategy is the following:

$$v_k(s_k) = \begin{cases} (R_{j,k} + \epsilon) \mathbb{I}_{\{\text{SINR}_{j,k} > \eta_k\}} - \epsilon, & \text{if } s_k \neq w_k \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where ϵ is a small positive value and $R_{j,k} = \log(1 + \text{SINR}_{j,k})$ is the throughput of user k associated to BS j . This means

that if a mobile user decides to transmit, it obtains either a utility equal to its transmission rate ($R_{j,k}$) or a negative utility ($-\epsilon$) depending on its SINR. Otherwise, the mobile user decides to stay silent ($v(s_k) = 0$). As a result, this tends to lead users who do not contribute enough utility to outweigh the interference degradation and remain silent. In order to provide a right balance between efficiency and fairness between cells, one possible remedy would be to use the so-called α -fairness [10]. This guarantees that any point in which one BS is shut down cannot be a local maxima. The global utility can be expressed as follows:

$$U = \begin{cases} \frac{1}{1-\alpha} \sum_j U_j^{1-\alpha}, & \text{if } \alpha \neq 1 \\ \sum_j \log(U_j), & \text{if } \alpha = 1 \end{cases} \quad (4)$$

where $U_j = \sum_k R_{j,k}$ and α is the fairness parameter. The network manager is assumed to perfectly know the set of strategies and the utilities of the K mobile users. Similarly, it is guaranteed under this setting that the followers can observe the actions of the leader through the broadcast channel. Accordingly, the Stackelberg game can be formulated as follows:

$$\begin{aligned} \mathbf{P}^{\text{SE}} &= \arg \max_{\mathbf{P}} U(\mathbf{P}(\mathbf{s}^{\text{NE}})) \\ \text{s.t.} \quad & \sum_{n=1}^N P(j, n) \leq P_{\max} \quad ; \forall j \in \mathcal{M} \end{aligned} \quad (5)$$

where \mathbf{s}^{NE} is a Nash equilibrium among K mobiles considering the strategy of the leader.

Let $\mathcal{S} = \{\Omega_1 \times \dots \times \Omega_K\}$ be the strategy space of our one shot game and $s = (s_k, s_{-k})$ a strategy profile in the game.

Mathematically, the Nash equilibrium can be expressed by the following inequality for all association strategies $\mathbf{s} \in \mathcal{S}$:

$$v_k(s_k, s_{-k}) \geq v_k(r_k, s_{-k}); \quad \forall k = 1, \dots, K \quad (6)$$

for every $r_k \in \Omega_k$ and $s_{-k} \in \Omega_{-k}$ where $\Omega_{-k} = \{\Omega_1 \times \dots \times \Omega_{k-1} \times \Omega_{k+1} \times \dots \times \Omega_K\}$ is the joint feasible strategy space of all users but the k th one.

Learning for optimal decision

The interaction between the leader and the followers provides a potential incentive for both agents to make a decision process based on their respective perceived payoff. This section focuses on how to reach the Stackelberg equilibrium for both the leader and the followers. To accomplish the task of global optimization problem, a two-stage optimization algorithm is proposed. One difficulty in our context is that mobile users do not know the payoffs (thus the strategy) of each other at each stage. Thus, the environment of each mobile user, including its opponents, is dynamic and may not insure convergence of the algorithm. In [11], authors develop a Nash-Stackelberg fuzzy Q-learning in a heterogeneous cognitive network. As an

alternative way, we adopt a hierarchical algorithm. The proposed approach requires neighboring base stations to exchange load (or interference) information experienced at user level on regular intervals. Consequently, the hierarchical algorithm is performed based on a coordination on both local (user level) and global scope (network level), which could scale accordingly.

As far as the two-stage learning algorithm is concerned, this can be conducted in the following steps: First, every user reports to its serving base station the experienced interference from neighboring cells. Then, the interference information is exchanged between base stations over the X2 interface while trying to optimize the global network utility by means of power control. Based on these power levels (broadcasted by BSs), each user checks *distributively* whether the serving BS is still the best choice according to its utility. Otherwise, it can perform a handover to the other RANs after checking that it could be admitted on it. As a result, this approach tends to substantially reduce signaling overhead from the base stations.

Leader: gradient computation mechanism

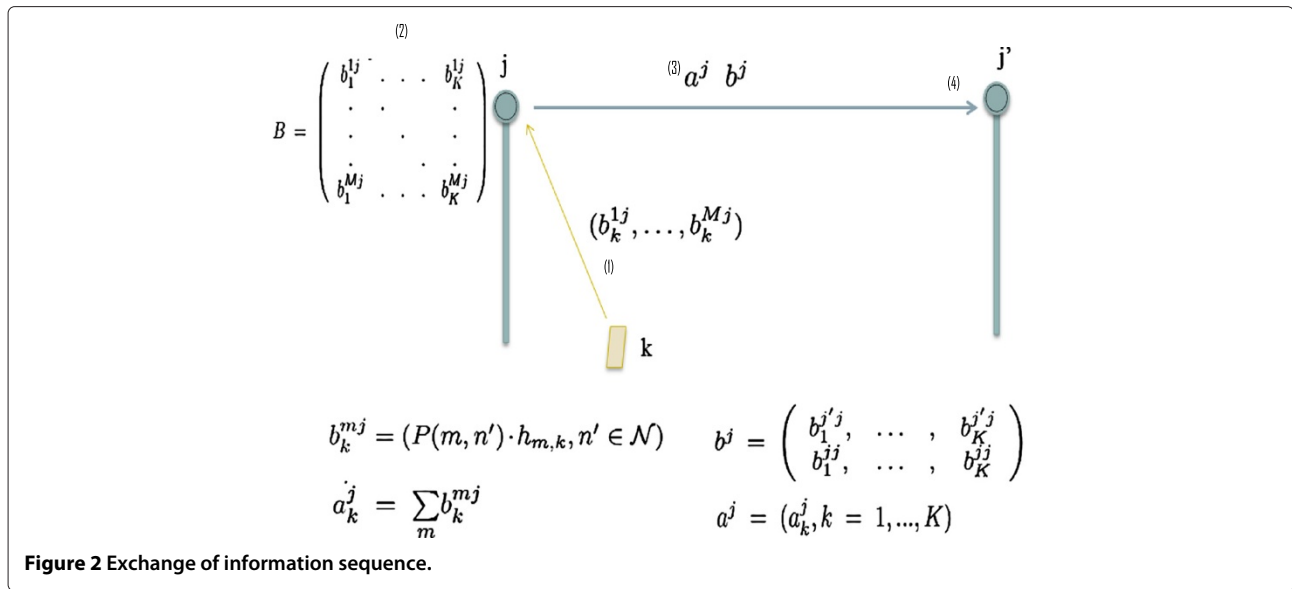
In this section, we propose an operational way of computing the derivative of the global utility in a distributive fashion. At each time epoch, consider that each mobile user k reports to its base station j the matrix $(b_k^{1j}, \dots, b_k^{Mj})$, where $b_k^{mj} = (P(m, n') \cdot h_{m,k}, n' \in \mathcal{N})$ is the vector of interferences perceived by user k from base station m or its signal strength on sub-band n when $m = j$. In the scope of this paper, and without loss of generality, we consider that only one user can interfere per base station with another user from a neighboring base station if they use corresponding channels. Hence, adjacent channel interferences are not included, and we assume that users are not allocated more than one PRB at the time. The vector b_k^{mj} restricts then to a single interference value for each m . Base station j will then be able to build the hyperma-

trix $B = \begin{pmatrix} b_1^{1j} & \dots & b_K^{1j} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ b_1^{Mj} & \dots & b_K^{Mj} \end{pmatrix}$ and send the matrices $a^j = (a_k^j,$

$k = 1, \dots, K)$, with $a_k^j = \sum_m b_k^{mj}$ and $b^j = \begin{pmatrix} b_1^{jj} & \dots & b_K^{jj} \\ b_1^{jj} & \dots & b_K^{jj} \end{pmatrix}$

to each base station j' . To illustrate our mechanism, we depicted in Figure 2 the exchange flow of information. The exchange of information works as follows:

- (1) The UE collects interference information from each neighbor over every allocated RBs and forwards the obtained vector to its serving eNodeB.
- (2) Once information is received at the eNodeB from all the attached users, an interference matrix is built that



contains interference information from each neighbor (M neighbors on the figure).

- (3) Interference load is then put on format for each neighbor.
- (4) Finally, each neighbor receives on the X2 interface the interference information concerning all the covered mobiles.

The derivative of the utility from base station j computed here below is then obtained in base station j' for sub-band n by

$$\begin{cases} \frac{\partial U_j(P(j', n))}{\partial P(j', n)} = \sum_{i=1}^K \left(\frac{b_i^{j'j}/P(j', n)}{\sigma^2 + a_i^j} - \frac{b_i^{j'j}/P(j', n)}{\sigma^2 + a_i^j - b_i^{jj}} \right) \text{ and} \\ \frac{\partial U_{j'}(P(j', n))}{\partial P(j', n)} = \sum_{i=1}^K \left(\frac{h_{j',k}}{\sigma^2 + a_i^{j'}} \right). \end{cases} \quad (7)$$

So far, $\frac{dU}{dP(j', n)} = U_{j'}^{-\alpha} \sum_j \frac{\partial U_j}{\partial P(j', n)}$. We then need to express $\frac{\partial U_j}{\partial P(j', n)}$ for all j , assuming we are considering cell j' . Because the derivative goes the same for every sub-band, we will focus only on one particular sub-band n . It can be easily shown that $\frac{\partial U_j}{\partial P(j', n)}$ is given by

$$\begin{cases} \left(\frac{h_{j',k}}{\sigma^2 x + \sum_m P(m, n) \cdot h_{j,m}} - \frac{h_{j',k}}{\sigma^2 + \sum_{m \neq j} P(m, n) \cdot h_{j,m}} \right); & \text{if } j \neq j' \\ \left(\frac{h_{j,k}}{\sigma^2 + \sum_m P(m, n) \cdot h_{j,m}} \right); & \text{if } j = j'. \end{cases}$$

The pseudo-code for the proposed gradient descent approach is given in *Algorithm 1*.

Algorithm 1

In each base station j'

1. Find the value of $\frac{dU}{dP(j', n)}(P_t(j', n))$ at time t , given the current power level $P_t(j', n)$.
2. Express $P_{t+1}(j', n) = P_t(j', n) + \gamma_t \frac{\partial U}{\partial P(j', n)}(P_t(j', n))$ to have new values of the power vector at time $t + 1$.
3. Allocate the powers to each sub-band, and let the users associate (see **Algorithm 2**).
4. Receive feedback from users and build the hypermatrix B .
5. Send $a^{j'}$ and $b^{j'}$ to each base station j .
6. If $\left(\max_n (P_{t+1}(j', n) - P_t(j', n)) \right) \leq \epsilon$ stop; else go to 1.

End algorithm.

Note that the implementation of gradient-like algorithms is familiar in optimization problems. The convergence of such algorithms has been shown in [12], under some specific conditions such that the derivative of objective function is Lipschitz continuous which is satisfied here, and for an accurate choice of γ_t .

Proposition 1. *The derivative of our utility function is Lipschitz continuous.*

Proof. If $s_k \neq w_k$, the utility function is given by

$$U = \begin{cases} -\epsilon & \text{if } \text{SINR}_{jk} > \eta_k \\ R_{jk} + \epsilon & \text{otherwise.} \end{cases}$$

Again, $\text{SINR}_{jk} > \eta_k \implies U = R_{jk} + \epsilon$. Let us show that ∇U is Lipschitz continuous. We have $\frac{\partial U_j}{\partial P(j',n)} =$

$$\begin{cases} \frac{h_{j'k}}{\sigma^2 + \sum_m P(m,n)h_{m,k}} - \frac{h_{j'k}}{\sigma^2 + \sum_{m \neq j'} P(m,n)h_{m,k}} & \text{if } j \neq j' \\ \frac{h_{jk}}{\sigma^2 + \sum_m P(m,n)h_{m,k}} & \text{if } j = j' \end{cases}$$

Then, for another value $P_1(j', n)$ of power level on channel $P_2(j', n)$, we obtain

$$\begin{aligned} & \left\| \frac{\partial U_j}{\partial P_1(j', n)} - \frac{\partial U_j}{\partial P_2(j', n)} \right\| \\ &= \begin{cases} \left\| \frac{h_{j'k}}{\sigma^2 + \sum_{m \neq j'} P(m,n)h_{m,k} + P(j,n)h_{j,k} + P_1(j',n)h_{j',k}} \right. \\ \quad - \frac{h_{j'k}}{\sigma^2 + \sum_{m \neq j'} P(m,n)h_{m,k} + P_1(j',n)h_{j',k}} \\ \quad - \frac{h_{j'k}}{\sigma^2 + \sum_{m \neq j'} P(m,n)h_{m,k} + P(j,n)h_{j,k} + P_2(j',n)h_{j',k}} \\ \quad \left. + \frac{h_{j'k}}{\sigma^2 + \sum_{m \neq j'} P(m,n)h_{m,k} + P_2(j',n)h_{j',k}} \right\| & \text{if } j \neq j' \\ \left\| \frac{h_{jk}}{\sigma^2 + \sum_{m \neq j} P(m,n)h_{m,k} + P_1(j,n)h_{j,k}} \right. \\ \quad \left. - \frac{h_{jk}}{\sigma^2 + \sum_{m \neq j} P(m,n)h_{m,k} + P_2(j,n)h_{j,k}} \right\| & \text{if } j = j' \end{cases} \end{aligned}$$

if $j = j'$, then

$$\begin{aligned} & \left\| \frac{\partial U_j}{\partial P_1(j', n)} - \frac{\partial U_j}{\partial P_2(j', n)} \right\| = \left\| \frac{P_2(j,n) - P_1(j,n)}{(C + P_1(j,n))(C + P_2(j,n))} \right\|; \\ & C = \sigma^2/h_{j',k} + 1/h_{j',k} \sum_{m \neq j'} P(m,n)h_{m,k} \\ & \leq \frac{1}{C^2} \|P_2(j,n) - P_1(j,n)\|. \end{aligned}$$

We have

$$\begin{aligned} C &= \sigma^2/h_{j,k} + 1/h_{j,k} \sum_{m \neq j} \\ & \quad \times P(m,n)h_{m,k}, \text{ since } h_{j,k} \text{ is an attenuation factor} \\ & \geq \sigma^2 + \sum_{m \neq j} P(m,n)h_{m,k} \\ & \geq \sigma^2 \end{aligned}$$

so that

$$\left\| \frac{\partial U_j}{\partial P_1(j', n)} - \frac{\partial U_j}{\partial P_2(j', n)} \right\| \leq \frac{1}{\sigma^4} \|P_2(j,n) - P_1(j,n)\|$$

Similarly, if $j \neq j'$, then

$$\begin{aligned} & \left\| \frac{\partial U_j}{\partial P_1(j', n)} - \frac{\partial U_j}{\partial P_2(j', n)} \right\| \\ &= \left\| \frac{P_2(j',n) - P_1(j',n)}{(C + P_1(j',n) + \frac{P(j,n)h_{j,k}}{h_{j',k}})(C + P_2(j',n) + \frac{P(j,n)h_{j,k}}{h_{j',k}})} \right. \\ & \quad \left. - \frac{P_2(j',n) - P_1(j',n)}{(C + P_1(j',n))(C + P_2(j',n))} \right\| \\ &= \left\| \frac{\frac{P(j,n)h_{j,k}}{h_{j',k}} [C + P_1(j',n) + P_2(j',n) + \frac{P(j,n)h_{j,k}}{h_{j',k}}]}{(C + P_1(j',n))(C + P_2(j',n))(C + P_1(j',n) + \frac{P(j,n)h_{j,k}}{h_{j',k}})(C + P_2(j',n) + \frac{P(j,n)h_{j,k}}{h_{j',k}})} \right\| \\ & \quad \times \|P_2(j',n) - P_1(j',n)\| \\ & \leq \left\| \frac{\frac{P(j,n)h_{j,k}}{h_{j',k}} [C + P_1(j',n) + P_2(j',n) + \frac{P(j,n)h_{j,k}}{h_{j',k}}]}{C^4} \right\| \|P_2(j',n) - P_1(j',n)\| \\ & \leq \frac{P_{\max}(C + 3P_{\max})}{C^4} \|P_2(j',n) - P_1(j',n)\| \\ & \leq \frac{P_{\max}(\sigma^2 + (M + 2)P_{\max})}{\sigma^8} \|P_2(j',n) - P_1(j',n)\| \end{aligned}$$

This ends the proof. \square

In our computations, we use the implementation of WOLFE linear search to find an appropriate value of γ_t at each iteration.

On another hand, at each iteration, the gradient algorithm delivers the values of the power vector for each base station that can go out of the bounds of the allowed space. To handle this problem, we implement a computational mechanism to satisfy the power constraint in (5). We define the constraint $c(P(j,n), P_{\max})$ to relax the problem, where c is built as follows:

Define $\xi = \{n \in \mathcal{N} \text{ s.t. } P(j,n) > p_{th}\}$, where p_{th} is a threshold value for every PRB n s.t. $\sum_n p_{th} = P_{\max}$. Let $\delta =$

$\left(P_{\max} - \sum_{\mathcal{N} \setminus \xi} P(j,n) - \sum_{\xi} p_{th} \right)$; if $\exists k \in \xi$ and $\sum_n P(j,n) \geq P_{\max}$, set the values of each $P(j,n)$, using the projection $\bar{P}(j,n) = \min\left(\frac{\delta}{|\xi|} + p_{th}, P(j,n)\right)$. It is to say that the remaining power on each BS power budget, if any, is evenly shared among the channels requiring a power level above the threshold value.

Followers: pursuit algorithm

At the user level of our Stakelberg framework, we use the pursuit algorithm as a tool to allow users to reach iteratively and individually a Nash equilibrium. The pursuit algorithm is a distributed association algorithm proposed in [13] allowing each individual in a set of players to

select a given strategy, among several others, that will best maximize its utility within a limited number of iterations.

Algorithm 2

At each iteration t

1. Select a strategy $s_k \in \Omega_k$ according to the current powers level \mathbf{P} .
2. Update the vector of average utilities $u_{\text{avg},k}$, using the chosen strategy s_k ($u_{\text{avg},k}$) and provided utility $v_k(s_k)$.
3. Find the strategy $s_k = \arg \max(u_{\text{avg},k})$ that gives the best average.
4. Make
$$\begin{cases} p_{s_k}(t+1) = p_{s_k}(t) + \delta(1 - p_{s_k}(t)) \\ p_{s_i}(t+1) = p_{s_i}(t) - \delta p_{s_i}(t), \quad i \neq k \end{cases}$$
5. if $\max(|p_{s_i}(t+1) - p_{s_i}(t)| < \epsilon)$ stop, else go to 1.

End Algorithm

It has been proven in [13] that the pursuit algorithm always converges under some specific conditions on the step size parameter. They show that when the step size parameter is very small, the game converges to a stable equilibrium for the learning automata game. This algorithm has the property to converge to an extremum of the game when there exists a pure equilibrium. To reach mixed equilibrium, the authors in [14] present a distributed algorithm that can be used in such situations. However, mixed equilibria are not efficient in our context since it will lead mobile users to process continuously handovers between base stations. To avoid mixed equilibria, we introduce a cost of handovers in the utility function to give more incentive to mobile users in reaching pure equilibria.

Discussion on cost of handover

As stated in the previous paragraph, a major weakness of the learning algorithm in mobile networks is the number of handovers, especially when the algorithm converges to mixed equilibria. We try to tackle this issue by introducing a cost of handover as a reward in the utility function to users who are not operating handovers. User utility function is given by, $v_k(s_k) =$

$$\begin{cases} (R_{j,k} + \epsilon) \mathbb{I}_{\{\text{SINR}_{j,k} > \eta_k\}} - \epsilon + \alpha_h, & \text{if } s_k \neq w_k \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where $\alpha_h = \beta_h(1 - \mathbb{I}_{\{\text{handoff}\}})$ and β_h is a small positive value. In Figure 3, we compare the number of handovers with and without the defined handover control. On the figure, we can see that the handover control policy decreases considerably the number of handovers and that the system remains stable after a few iterations. Interestingly, we noticed in our simulations that users are more motivated in following the handover control when the control is a reward rather than a penalty as we suggested in the first place. A trade-off on using such control can be seen at the utility side. As shown in Figure 4, the gap in utility between the two policies can be marginal.

Implementation and validation

To go further with the analysis, we resort to realistic network simulations. We consider a cellular radio network as described in Figure 1 where users are attempting to communicate during a downlink transmission, subject to mutual inter-cell interferences. Specifically, a hexagonal cellular system functioning at 1.8 GHz where the cell radius is equal to $R = 200$ m is considered. Note that this

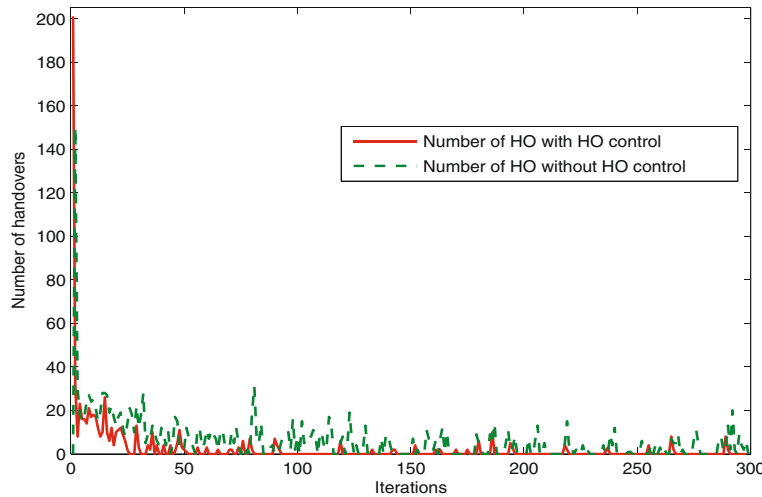


Figure 3 Comparison of the number of handover. We compare two different approaches where mobile users use or do not use a penalty factor to control the switching rate. The number of handovers operated over iterations is plotted in this figure for the two approaches.

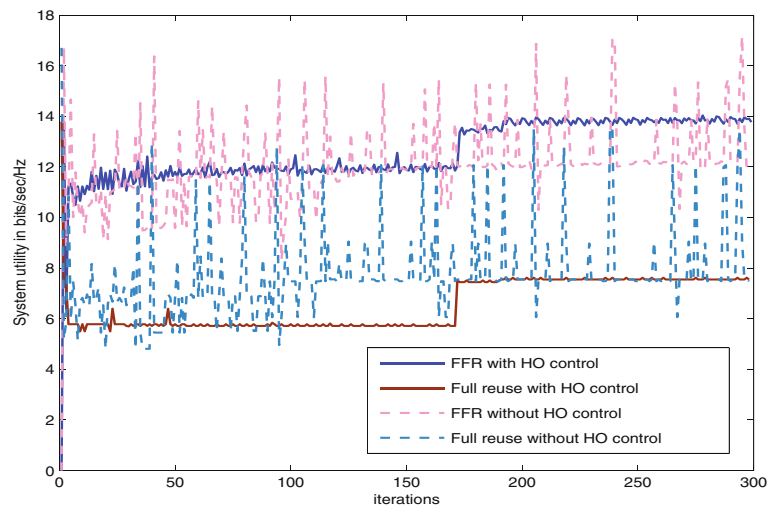


Figure 4 Comparison of network utilities. This figure shows the performances in terms of overall network utility, when mobile users control or do not control the switching rate.

radius only stands for geographical positions in the network. It does not prevent users located out of this area to connect with another base station in case of significant connection opportunity. Channel gains are based on the COST-231 path loss model [15] including log-normal shadowing with a standard deviation of 10 dB, plus fast fading assumed to be i.i.d. circularly symmetric with distribution $\mathcal{CN}(0, 1)$. The peak power constraint is given by $P_{\max} = 100$ mW. We evaluate under those settings the joint processing of the gradient descent algorithm with the pursuit algorithm. Without loss of generality, we assume that every cell has the same number of users randomly positioned inside the cell. We consider a cluster of seven interfering cells, featured with ten PRBs each. The values of the other parameters are set in Table 1.

The iteration scale parameter in Table 1 traduces how frequently BSs update the gradient algorithm and set new values of powers. By tuning this parameter, one can

Table 1 Simulations settings

Parameter descriptions	Values
Number of cells	7
Number of PRBs per cell	10
Number of users per cell	4
Outer radius of hexagonal cells	200m
Distance to insure target SINR	300m
Fairness parameter	$\alpha = 1$
Iterations scale ^a	1 for 30

This table shows the values of different parameters used for the simulation scenarios. In this table, we basically specify the structure of the network used for our simulations. ^aThis is the scale of iterations between the gradient and pursuit algorithms.

control the amount of signalization between BSs. We consider in our simulations that users run 30 iterations of the association algorithm for 1 iteration of the gradient. We first build the framework for a fairness parameter $\alpha = 1$ which represents the proportional fairness algorithm and then extended it to different values of α .

Dynamic fractional frequency reuse

In Figure 5, we illustrate the snapshot of the dynamic fractional frequency reuse pattern at the equilibrium. The small colored disks indicate the positions of users inside the cells, and the face colors are the frequencies used by those users. Disks are indexed with a couple of values (BS, power) where the first value represents the base station to which this user is connected and the second the power level assigned by the base station on that frequency. As expected, users close to each other are attributed different frequencies, and power levels are set accordingly to avoid a high level of interferences. From the same figure, we also have an overview on user-network association. Indeed, many cases appear where users would rather associate in a neighboring cell rather than in the cell where they are positioned due to the influence of path loss and/or interference impairments. For instance, in Figure 5, the user indexed (2, 7.5) in cell 2 is connected to BS 2 and is assigned frequency F_2 with a high power level. This reflects the maximization goal of the gradient algorithm since frequency F_2 is reused only once by a user far away in cell 7 at a low power level.

Utility maximization

In Figure 6, we compare the proposed FFR algorithm with traditional fixed reuse patterns, namely, the full reuse. The

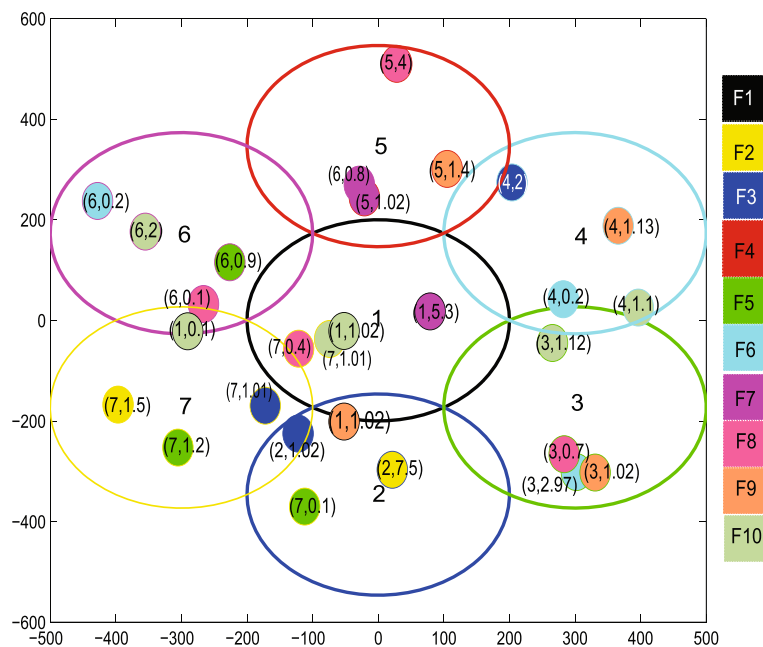


Figure 5 Snapshot of the dynamic fractional frequency reuse pattern at the equilibrium for $\alpha = 1$. The figure shows the achieved FFR pattern in the network when our learning mechanism attains the equilibrium for a fairness factor $\alpha = 1$.

exhaustive search algorithm, shown as a dashed line, considers all possible combinations of PRB selection given the power level of the gradient algorithm. This will thus serve as an *optimal* association solution for users and will demonstrate just how much gain may theoretically be exploited through the pursuit algorithm. It clearly appears that the joint gradient and pursuit algorithms perform better than the full reuse and reduce considerably the gap

with the exhaustive search. As shown in Figure 6, we reach up to 90% of the overall network throughput compared to the exhaustive (optimal) association search.

Fairness issues

In this section, we intend to show the impact of fairness on the global utility maximization by simulating different values of the α -fairness parameter. For $\alpha = 0$

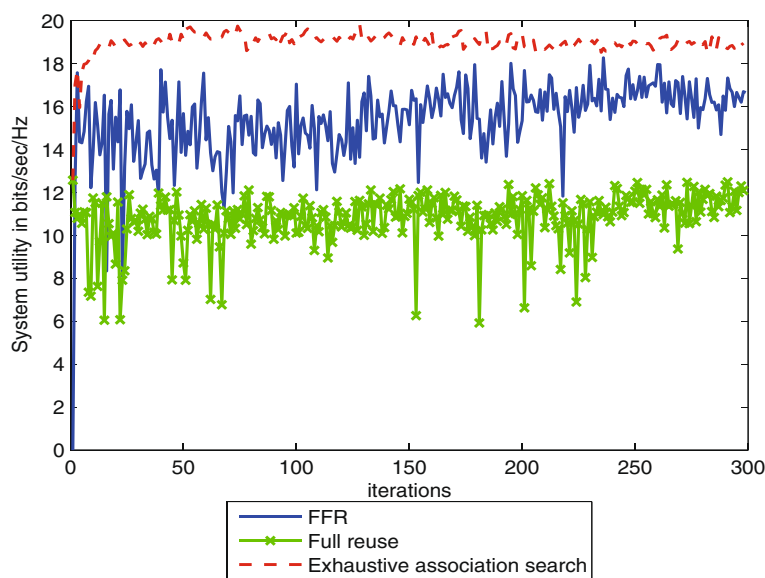


Figure 6 Network utility for $\alpha = 1$. In this figure, we plot the performances achieved by our mechanism in comparison with the global optimum and a worst-case scenario for a fairness factor $\alpha = 1$.

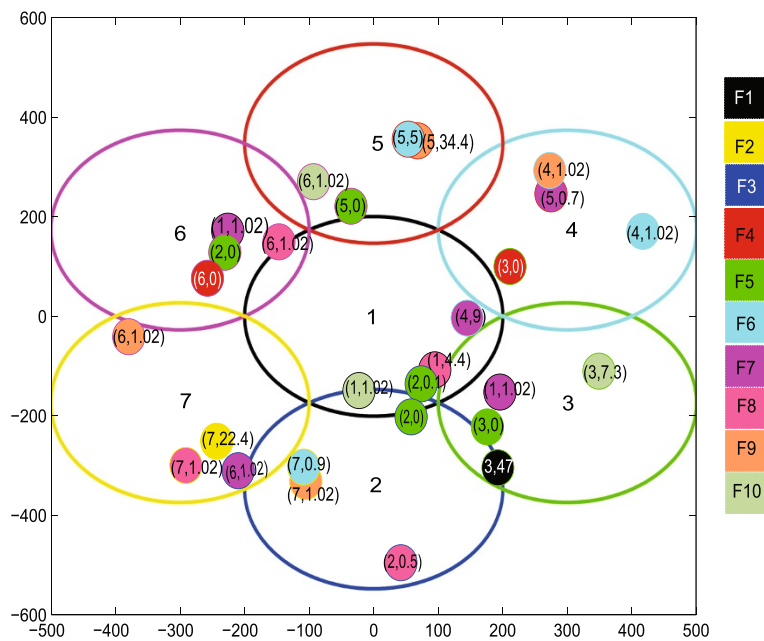


Figure 7 Snapshot of the FFR system pattern for $\alpha = 0$. The figure shows the achieved FFR pattern in the network when our learning mechanism attains the equilibrium for a fairness factor $\alpha = 0$.

(the maximum throughput algorithm), we can see from Figure 7 that some BSs (BS 6 for instance) are set to idle. Several other channels in the network are also switched off, while a few number of users are attributed very high levels of power. This behavior was somehow expected since using $\alpha = 0$ means that the major goal of the network is to maximize the overall network utility.

Nevertheless, as shown in Figure 8, this policy does not help to tighten the gap to the exhaustive association search (71%) as much as using a value of $\alpha = 1$.

Further analysis of the max-min fairness policy ($\alpha \rightarrow \infty$) shows that most of the BSs are set to idle, and only a few channels are activated. Being too fair leads then the network to follow the policy of highly loaded BSs, thus

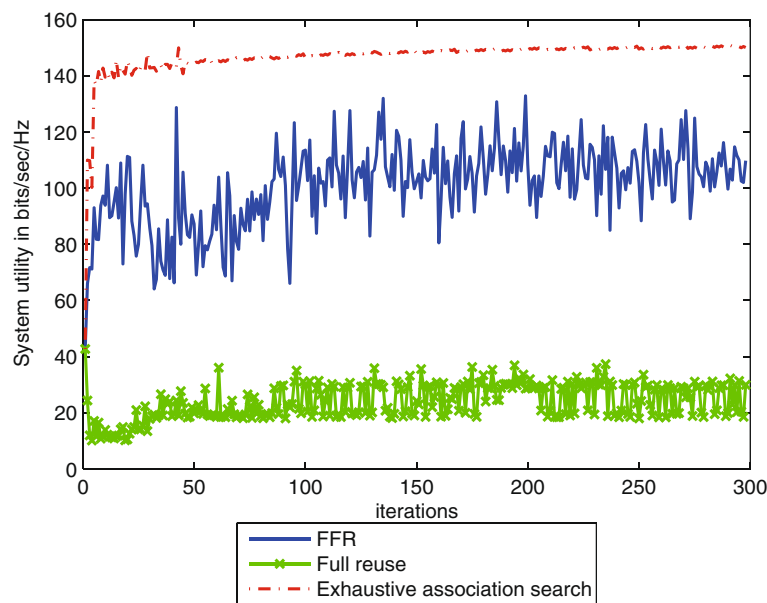


Figure 8 Network utility for $\alpha = 0$. In this figure, we plot the performances achieved by our mechanism in comparison with the global optimum and a worst-case scenario for a fairness factor $\alpha = 0$.

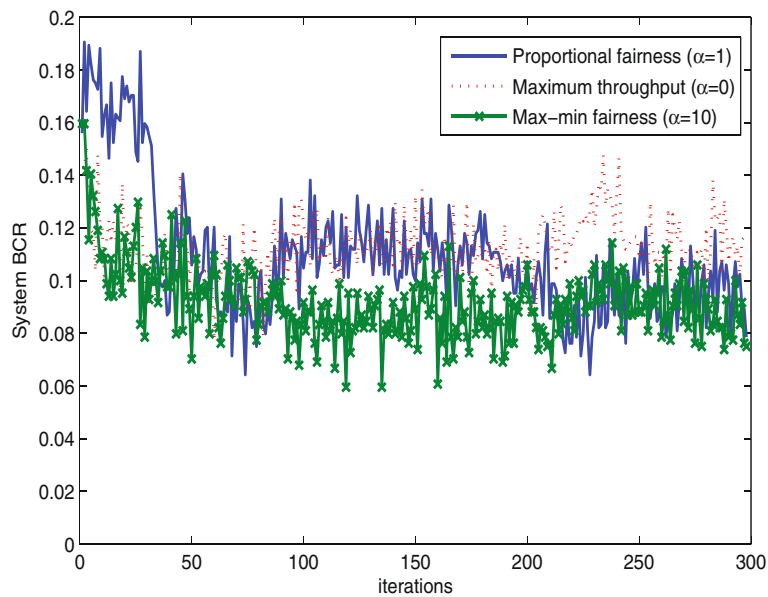


Figure 9 Block call rate for different values of α . The figure shows the blocking probability for the different levels of fairness considered.

providing an overall network utility which is almost null. Finally, we plot in Figure 9 the network block call rate (BCR) for the increasing number of iterations obtained when users follow the strategy corresponding to the Stackelberg equilibrium. We can observe that the BCR can be substantially reduced as the number of iteration increases. Moreover, the fairness policy has a negligible influence on the BCR which remains less than 10% for the different fairness policies.

Robustness and scalability

Next, we evaluate in this section a seamless adaptation of our algorithms to a dynamic environment. We simulate a discrete time system over several iterations and generate a burst of user arrival at a specific time instant during the simulation time. Indeed, while new arrivals generally occur every minute in the cellular systems, our association algorithm converges at the order of a few milliseconds. This speed of reactivity and adaptation shows

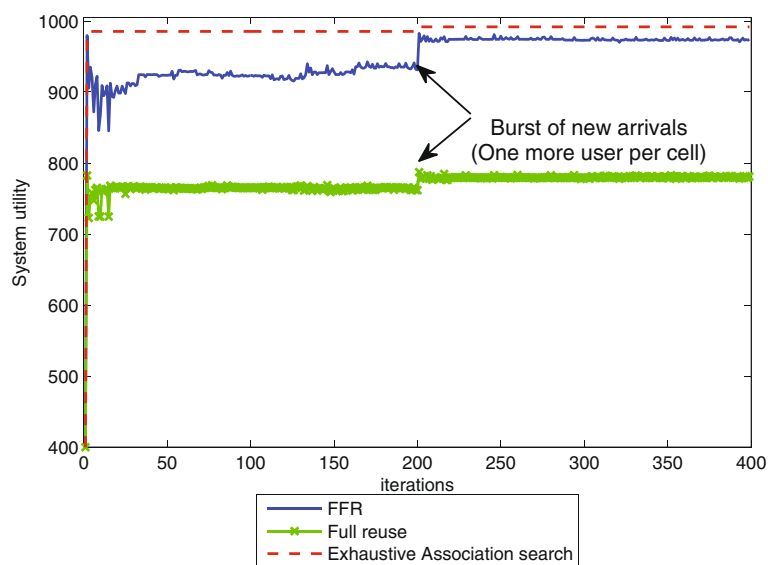


Figure 10 Network utility for $\alpha = 0.98$ with the first-ring interference information. In this figure, we analyze the performances of our mechanisms under a perturbed environment with a variation of the number of mobile users, where limited information is exchanged with only the first-ring neighbors.

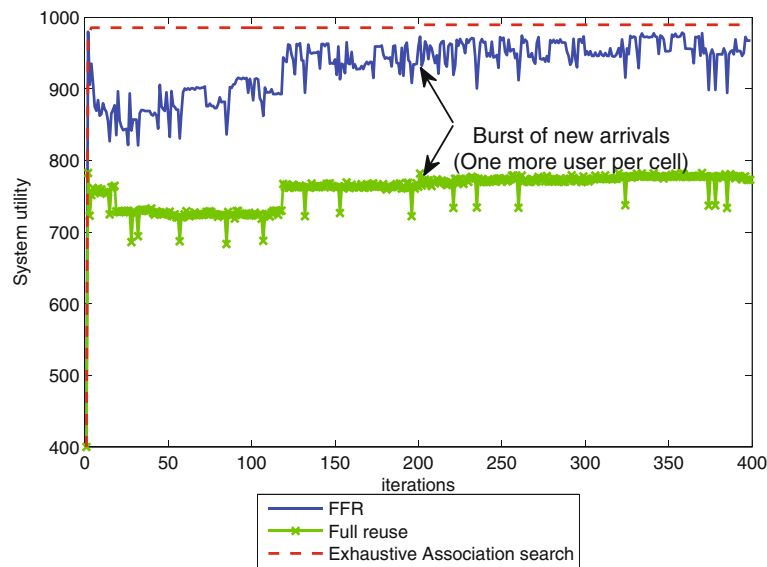


Figure 11 Network utility for $\alpha = 0.98$ with two neighbor interference information. In this figure, we analyze the performances of our mechanisms under a perturbed environment with a variation of the number of mobile users, where limited information is exchanged with only two of the closest neighboring sites.

improved performances of our hierarchical algorithm and is traduced in Figures 10, 11, and 12. We consider two rings of a small cell network where each base station is featured with four PRBs and contains three users each, randomly positioned inside the cell at the beginning of the simulation. In this new setting, we assume that the algorithm iteration scale is 1 for 100. For each of the simulated schemes, we assume that each BS can exchange

interference information only with a subset of all the interfering neighbors. This consideration helps to understand how the proposed scheme reacts when the amount of information exchanged between BSs is limited.

For the first (Figure 10), second (Figure 11), and third (Figure 12) scenarios, we consider that each BS exchanges data, respectively, with the first-ring neighbors, only with the two first closest neighbors, and finally with all the

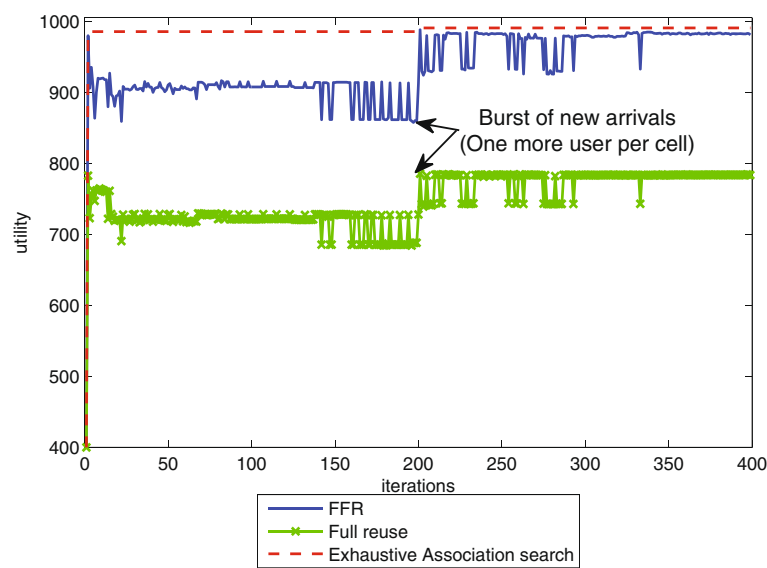


Figure 12 Network utility for $\alpha = 0.98$ with all neighbor interference information. In this figure, we analyze the performances of our mechanisms under a perturbed environment with a variation of the number of mobile users, where full information is exchanged with all other sites in the considered network area.

interfering neighbors from the two considered rings. By comparison of the different scenarios, it is observed that more exchanged information lead, as one can intuitively expect, to an increased outcome in utility. However, although this can be imputed to randomness, when comparing Figures 10 and 12, we see that the system stability is not necessarily insured by an increase of exchanged information rate. On another hand, even with a very few amount of exchanged information, our algorithm preserves a convergence to 97% of the exhaustive association search utility. From the same figures, we address the scalability of our algorithms, with the introduction of a burst of new arrivals in the system at iteration 200. Although this event is not clearly captured in the case of Figure 11, when less information is exchanged, we can observe from Figures 10 and 12 that our mechanism adapts very fast to the system evolution in order to reach the new point of convergence.

Conclusion

In this paper, we have investigated the idea of a hierarchical learning game for fractional frequency reuse in an OFDMA network. In this framework, both the network manager and mobile users learn to reach an equilibrium that optimizes the global network utility while ensuring individual utility optimization for mobile users. We have first proposed formally a game model to define how the network manager and mobile users can obtain their respective equilibria by means of a Stackelberg formulation. Then, we have presented a two-stage learning algorithm for finding a Stackelberg equilibrium and the corresponding mobiles' association strategies. Practical directions for implementability of our solution are also presented. We have shown using several numerical examples the efficiency of the obtained equilibrium compared to the exhaustive (optimal) solution and a fixed full frequency reuse pattern. In particular, in the case of proportional fair policy, the proposed FFR approach offers approximately 90% of the optimal association policy and 40% of gain with respect to the fixed full reuse. Indeed, for implementation purposes and in order to adapt to the dynamic of the mobile environment, the number of iterations before convergence should remain in the order of a few tens. Eventually, we have addressed interesting issues such as fairness, robustness, and scalability and offered insights into how to design such scenario in a wireless network environment.

Endnotes

^aThe work reported herein was partially supported by the projects Ecocells.

^bThough some users stay silent, they may be active during the next scheduling period.

Competing interests

The authors declare that they have no competing interests.

Received: 10 August 2012 Accepted: 15 March 2013
Published: 2 May 2013

References

- 3rd Generation Partnership Project (3GPP), in *TR 36.902*, Evolved Universal Terrestrial Radio Access Network (EUTRAN), Self-configuring and self-optimizing network (SON) use cases and solutions Dec 2008. http://www.etsi.org/deliver/etsi_tr/136900_136999/136902/09.03.01_60/tr_136902v090301p.pdf. Accessed May 2011
- AL Stolyar, H Viswanathan, in *Proc. IEEE INFOCOM*. Self-organizing dynamic fractional frequency reuse for best-effort traffic through distributed inter-cell coordination (Rio de Janeiro, April 2009), pp. 19–25
- NGMN Alliance, ed. by Lehser F. NGMN Recommendation on SON and O&M Requirements, Frankfurt, (5 Dec 2008). http://www.ngmn.org/uploads/media/NGMN_Recommendation_on_SON_and_O_M_Requirements.pdf Accessed 23 April 2013
- R Combes, Z Altman, M Haddad, E Altman, in *IEEE ICC 2011 Workshop on Planning and Optimization of Wireless Communication Networks*. Self-optimizing strategies for interference coordination in OFDMA networks (Kyoto, Japan, June 2011), pp. 5–9
- R Giuliano, C Monti, P Loreti, in *IEEE ICC 2011 Workshop on Planning and Optimization of Wireless Communication Networks*, vol. 15. WiMAX fractional frequency reuse for rural environments (IEEE Washington, 2008), pp. 60–65
- K Son, S Chong, G de Veciana, in *IEEE ICC 2011 Workshop on Planning and Optimization of Wireless Communication Networks*, vol. 8. Dynamic association for load balancing and interference avoidance in multi-cell networks (IEEE Washington, 2009), pp. 3566–3576
- R1-050507, in *Huawei, 3GPP TSG RAN WG1 Meeting*. 'Soft Frequency Reuse Scheme for UTRAN LTE' (Athens, Greece, 2005), pp. 9–13
- D Fudenberg, J Tirole, *Game Theory*. (MIT Press, Cambridge, 1991)
- HBA Sidi, R El-Azouzi, M Haddad, in *Wireless Days (WD), 2011 IFIP*. Fractional frequency reuse stackelberg model for self-organizing networks (Ontario, Canada, 2011), pp. 1–6. doi:10.1109/WD.2011.6098171
- J Mo, J Walrand, Fair end-to-end window-based congestion control. *IEEE/ACM Trans. Netw.* **8**, 556–567 (2000). doi:10.1109/90.879343
- M Haddad, Z Altman, SE Elayoubi, E Altman, in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*. A Nash-Stackelberg Fuzzy Q-Learning Decision Approach in Heterogeneous Cognitive Networks (Miami, 2010), pp. 1–6. doi:10.1109/GLOCOM.2010.5684318
- DP Bertsekas, JN Tsitsiklis, Parallel and distributed computation: numerical methods. *GLOBECOM*. <http://dspace.mit.edu/handle/1721.1/3719> Accessed 23 April 2013
- M Thathachar, P Sastry, in *GLOBECOM*. Network of Learning Automata: Techniques for Online Stochastic Optimization (Kluwer Academic New York, 2004)
- Y Xing, R Chandramouli, Stochastic learning solution for distributed discrete power control game in wireless data networks. *IEEE/ACM Trans. Netw.* **16**(4), 932–944 (2008). <http://doi.acm.org/10.1145/1453698.1453713>
- EURO-COST Std. 231, in *European Cooperation in the Field of Scientific and Technical Research*. Urban transmission loss models for mobile radio in the 900 and 1800 MHz bands (Commission of European Communities Luxembourg, 1991)

doi:10.1186/1687-1499-2013-119

Cite this article as: Sidi et al.: A two-stage game theoretic approach for self-organizing networks. *EURASIP Journal on Wireless Communications and Networking* 2013 **2013**:119.