A Q-learning Based Approach to Interference Avoidance in Self-Organized Femtocell Networks

Mehdi Bennis* and Dusit Niyato†
*Centre for Wireless Communications, University of Oulu, Finland
†School of Computer Engineering, Nanyang Technological University, Singapore
E-mail: bennis@ee.oulu.fi, dniyato@ntu.edu.sg

***Abstract*—The femtocell concept is an emerging technology for deploying the next generation of the wireless networks, aiming at indoor coverage enhancement, increasing capacity, and offloading the overlay macrocell traffic. Nevertheless, the detrimental factor in such networks is *co-channel* interference between macrocells and femtocells, as well as among neighboring femtocells. This in turn can dramatically decrease the overall capacity of the network. In addition, due to their non-coordinated nature, femtocells need to *self-organize* in a distributed manner not to cause interference on the macrocell, while at the same time managing interference among neighboring femtocells. This paper proposes and analyzes a Reinforcement-Learning (RL) framework where a macrocell network is underlaid with femtocells sharing the same spectrum. A *distributed* Q-learning algorithm is proposed in which each Femto Base Station/Access Point (FBS/FAP) gradually learns (by interacting with its local environment) through trials and errors, and adapt the channel selection strategy until reaching convergence. The proposed Q-learning algorithm is cast into high level and low level subproblems, in which the former finds in a decentralized way the channel allocation through Q-learning, while the latter computes the optimal power allocation. Investigations show that through learning, femtocells are not only able to self-organize with only local information, but also mitigate their interference towards the macrocell network.**

I. Introduction

Recently, a new type of indoor Base Station (BS), called femtocell, has gained the attention of the industry [1]. A femtocell is a low-cost and low-power BS deployed by the end-users, designed to extend indoor coverage. Femtocells are connected to the network of the operator over a backhaul connection such as Digital Subscriber Line (DSL) or optical fiber. Meanwhile, femtocells also provide coverage to the end-customers using a cellular network standard, e.g., Universal Mobile Telecommunication System (UMTS), Wireless Interoperability for Microwave Access (WiMAX), and Long-Term Evolution (LTE).

Femtocells are expected to benefit both end-users and network operators [1]. Users will enjoy better signal qualities due to the reduced distance between the transmitter and the receiver, resulting in more reliable communications, higher throughputs, power and battery savings. From the operators point of view, femtocells will extend the indoor coverage, enhance system capacity, and share the spectrum in a more efficient manner [6], [12]. Femtocells will also help manage the exponential growth of the traffic, thanks to the handover of the indoor traffic to the backhaul. Moreover, femtocells will reduce the system cost, since they are paid and maintained by the owners. However, these benefits are not easy to accomplish, and there are challenges that mobile operators must address before successfully deploying femtocell networks. One of the major challenges for their successful deployment is *inter-tier* interference. Due to the aggressive frequency reuse, the system throughput in the presence of femtocells is ultimately limited by the inter-tier interference between macro and femtocells. This calls for effective interference management strategies, such as distributed power allocation or interference avoidance techniques.

Interference avoidance has never been a trivial task neither in macrocell deployments nor in femtocell networks. Due to the selfish nature of femtocells and uncertainty on their number and locations, operators must use new optimal dynamic approaches rather than the classical static network planning and optimization to avoid interference. In order to successfully react to the changes of the traffic, and minimize interference in femtocell deployments, the use of sophisticated self-organization techniques is paramount. Self-organization will allow femtocells to integrate themselves into the network of the operator, learn about their environment (such as neighboring femtocells and local interference map) and tune their parameters (transmit power and carrier frequency) accordingly.

A. Related Work

Different self-organization strategies for femtocells have been introduced. In [5], a power control method was proposed for pilot and data channels in UMTS networks that ensures a constant coverage femtocell radius. Each femtocell sets the transmit power such that on average it is equal to the power received from the closest macrocell at a target femtocell radius. In [14], a method was presented for coverage adaptation for UMTS networks using information on mobility events of outdoor passing and indoor users. Each femtocell sets its power to a value that on average minimizes the total number of attempts of outdoor passing users to connect to such femtocell. In [3], a distributed utility-based SINR adaptation at femtocells was proposed in order to alleviate cross-tier interference at the macrocell from cochannel femtocells. In [2], interference avoidance using a time-hopped Code Division Multiple Access (CDMA) physical layer and sectorial antennas is investigated. These approaches are mostly based on wide-band code division multiple Access (WCDMA) networks, and do not mitigate

978-1-4244-8865-0/10/$26.00 ©2010 IEEE

interference through sub-channel allocation, which is a very important feature of current Orthogonal Frequency Division Multiple Access (OFDMA) systems. Recent works investigating interference mitigation for OFDM femtocell networks can be found in [8], [11] for frequency resource partitioning, and [6], [9] using tools from game theory.

Existing research on Reinforcement-Learning (RL) [13] have been carried out in cognitive radio networks (e.g., see [17], [18], [7], [16]). In [17], the authors focused on the resource competition in a spectrum auction system, where the channel allocation is determined by the spectrum regulator, which is different from this paper in which no regulator exists. In [18], a distributed opportunistic spectrum access for cognitive radio using correlated equilibrium and no-regret learning was studied in which mutual communication among secondary users is assumed. Finally, a Q-learning based algorithm was investigated in [7] and [16] in the context of network selection for heterogeneous wireless networks, and channel selection in multi-user cognitive radios, respectively. Due to the fact that there is no mutual communication among different femtocells, many traditional learning techniques (e.g., fictitious play, Nash-Q learning, and evolutionary games [15]) cannot be used since they need information to be exchanged among players (e.g., exchanging their action and payoff information).

B. Contributions

The contributions of this paper can be summarized as follows:

- The coexistence between a macrocell and closed-access femtocell networks is modeled and analyzed using tools from reinforcement-learning technique. Due to the fact that femtocells are non-cooperative with no mutual communication/coordination, femtocells need to self-organize by gradually learning from their environment (through trials and errors), and adapt their strategy until reaching convergence.
- A *distributed* algorithm relying only on *local* information is proposed for femtocells to mitigate their interference towards the macrocell network.
- The Q-learning approach is compared to the random and static allocation strategies.

Unless otherwise stated, the following assumptions are made throughout the paper: (1) there is no communications or coordination among femtocells, thus excluding the possibility of cooperation, (2) channel state transition probabilities are unknown to femtocells, hence our assumption is different from traditional Markov decision processes (MDP) which requires state transition probabilities, (3) the game is imperfectly monitored, where each player does not know other players' actions.

This paper is organized as follows. In Section II, the system model which consists of an interference channel is introduced. Section III describes the reinforcement learning (Q-learning based) formulation for the macrocell and femtocell coexistence in the same spectrum band, in addition to the distributed Q-learning algorithm. Numerical results of the proposed two-level learning algorithm are provided in Section IV. Finally conclusions are given in Section V highlighting possible extensions of this work.

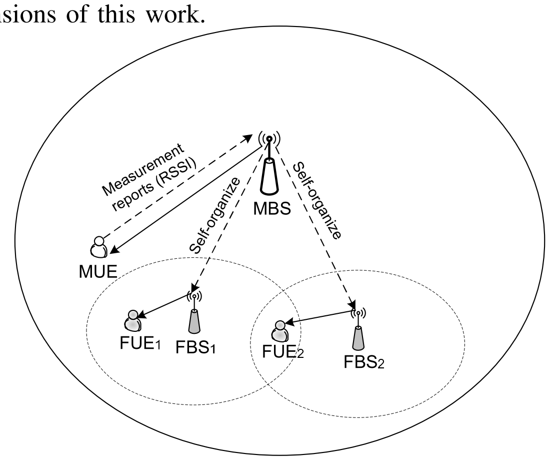

Fig. 1. Illustration of the considered interference scenario where two femtocells serve their associated femto-user equipments (FUEs) while the macro-BS serves its own macro-user equipment (MUE). MUEs periodically report their measurement reports to their MBS which in turn forces the femtocells to self-organize accordingly.

II. System Model

The system consists of one (single transmit and receive antenna) Macro Base Station (MBS) underlaid with N_f (single transmit and receive antenna) Femtocell Base Stations (FBSs) transmitting over N_{sub} sub-carriers. Orthogonal downlink signaling is assumed in each time-slot (i.e., 1 user/slot/cell).

At a given time-slot, let a Macro User Equipment (MUE) denote the scheduled macro-user connected to its serving MBS. Designate MBS's transmit power on a given sub-carrier to be $p_0^{(n)}$. Let $|h_{0,0}^{(n)}|^2$ denote the channel gain between the MBS and its associated MUE in sub-carrier n. Likewise, $|h_{i,j}^{(n)}|^2$ denotes the channel gain between transmitter i and receiver j on sub-carrier n. Moreover, let σ_n^2 be the variance of Additive White Gaussian Noise at MUE, which is assumed to be constant over all sub-carriers for simplicity. The transmit power of FBS i on sub-carrier n is denoted by $p_i^{(n)}$ and the set of all feasible power allocations is:

$$\mathcal{P} = \left\{ p_i^{(n)} \geq 0, \sum_{n=1}^{N_{sub}} p_i^{(n)} \leq \bar{P}_i, 1 \leq n \leq N_{sub}, 1 \leq i \leq N_f \right\} \tag{1}$$

with $\bar{P}_i$ being the total power constraint per femtocell i.

The achievable rate of MBS at MUE (assuming Gaussian signalling) is given as:

$$C_0 = \sum_{n=1}^{N_{sub}} \log_2 \left(1 + \frac{|h_{0,0}^{(n)}|^2 p_0^{(n)}}{\sigma^2 + \underbrace{\sum_i |h_{i,0}^{(n)}|^2 p_i^{(n)}}_{\text{femtocells}}} \right) \tag{2}$$

where at every time-instant, the MUE is guaranteed a Quality-of-Service (QoS) requirement such that $C_0 \geq \Gamma^0$, where Γ^0 is the quality-of-service (QoS) performance threshold[1].

The achievable rate for FBS_i, $i \in \{1, ..., N_f\}$, serving its femto-user FUE_i is given as:

$$C_i = \sum_{n=1}^{N_{sub}} \log_2 \left(1 + \frac{|h_{i,i}^{(n)}|^2 p_i^{(n)}}{\sigma^2 + \underbrace{|h_{0,i}^{(n)}|^2 p_0^{(n)}}_{\text{macrocell}} + \underbrace{\sum_{j \neq i} |h_{j,i}^{(n)}|^2 p_j^{(n)}}_{\text{femtocells}}} \right). \tag{3}$$

To mitigate interference inflicted by femtocells to the macrocell, the following protocol is considered: a macro-user periodically reports measurement results, such as the received signal strength indicator (RSSI), to its serving macro base station, based on which the macro base station sends (through X_2 interface, or over-the-air) a self-organization message to the interfering femtocells. This is shown in Figure 1 for the case of *one* macrocell and *two* interfering femtocells. Assuming a closed-access policy for femtocells, the macro base station forces FBS_1 and/or FBS_2 to switch their transmisions to other carriers and/or adjust their power levels (power control).

III. Reinforcement-Learning Framework for Macro-Femtocell Coexistence

In this section, we present the process associated with the learning procedure and more specifically the application of Q-learning model to the macro-femtocell coexistence scenario.

A. Reinforcement-Learning (RL)

The Q-learning model consists of a set of states $s \in \mathcal{S}$ and actions $a \in \mathcal{A}$ aiming at finding a policy that maximizes the observed rewards over the interaction time of the agents/players (i.e., femtocells). Every FBS_i, $\forall i \in \{1, ..., N_f\}$ explores its environment, observes its current state s, and takes a subsequent action a, according to a decision policy $\pi : s \rightarrow a$. With their ability to learn, the knowledge about other players' strategies is not explicitly needed by FBS_i. Instead, a Q-function maintains the knowledge about other players in the network locally, based on which decisions are made. Several authors have shown that Q-learning converges to optimal values in Markov decision process environment [4]. Thus, the goal of the agent is to find an optimal policy $\pi^*(s)$ for each state s, which maximizes a cumulative measure of the rewards over time.

The *expected* discounted reward over an infinite horizon is given by:

$$V^{\pi}(s) = \mathbb{E}\left\{\gamma^t \times r(s_t, \pi^*(s_t)) | s_0 = s\right\}, \tag{4}$$

where $0 \leq \gamma \leq 1$ is a discount factor and r is the agent's reward at time t. Furthermore, Equation (4) can be rewritten as:

$$V^{\pi}(s) = R(s, \pi^*(s)) + \gamma \sum_{v \in S} P_{s,v}(\pi(s)) V^{\pi}(v), \tag{5}$$

where $R(s, \pi^*(s)) = \mathbb{E}\{r(s, \pi(s))\}$ is the mean value of reward $r(s, \pi(s))$, and $P_{s,v}$ is the transition probability from state s to v. Moreover, the optimal policy π^* satisfies the optimality criterion:

$$V^*(s) = V^{\pi^*}(s) = \max_{a \in \mathcal{A}} \left(R(s,a) + \gamma \sum_{v \in \mathcal{S}} P_{s,v}(a) V^*(v) \right), \tag{6}$$

It is generally difficult to explicitly calculate the reward $R(s, a)$ and transition probability $P_{s,v}(a)$. However, through Q-learning, the knowledge of these values can be gradually learnt and reinforced with time. For a given policy π, define a Q-value as:

$$Q^*(s,a) = R(s,a) + \gamma \sum_{v \in S} P_{s,v}(a) V^{\pi}(v), \tag{7}$$

which is the expected discounted reward when executing action a at state s and then following policy π thereafter.

In this work, we use the Q-learning algorithm to iteratively approximate the state-action value function $Q(s, a)$. The agent keeps trying all actions in all states with non-zero probability and must sometimes explore by choosing at each step a random action with probability $\epsilon \in (0, 1)$, and the greedy action with probability $(1 - \epsilon)$. This is referred to as ϵ-greedy exploration [4]. Another option is to use the Boltzmann exploration strategy with temperature parameter κ, where the action a in state s is taken with a probability $P(a|s)$, and the femtocell receives a reinforcement r. The actions are chosen according to their Q-values as:

$$p(a|s) = \frac{e^{Q(s^i,a)/\kappa}}{\sum_{a' \neq a} e^{Q(s^i,a')/\kappa}}. \tag{8}$$

B. Q-learning formulation

In our problem formulation, we divide the problem into *carrier* and *power* allocation sub-problems. These two sub-problems are inter-related in which the femtocell first selects a set of sub-carriers, after which the transmit power is allocated (two-stage decision making process). In the upper level, a Q-learning formulation is used to solve the carrier selection sub-problem, while in the lower-level a convex optimization problem is formulated using the gradient method to calculate the power allocation. In this case, the reward of channel allocation (Q-learning) will be the result of the power allocation sub-problem.

The agent, state, and perceived reward associated to the Q-learning game are defined as follows:

- **Agent** i: FBS_i, $\forall 1 \leq i \leq N_f$.
- **State** $s_t^i = (s_t^{i,n}, n \in \{1, ..., N_{sub}\})$ being the composite state of every FBS_i at time t in sub-carrier n where $s_t^{i,n} = (\mathcal{I}^{i,n}, \mathcal{M}^{i,n}, \mathcal{F}^{i,n})$

[1]Should the QoS of the macrocell user be violated, the macrocell sends a message to interfering femtocells, for instance through the X_2 interface

- $\mathcal{I}^{i,n} \in \{0,1\}$ indicates whether FBS$_i$ generates interference towards MUE or not, in sub-carrier n where the QoS of MUE is violated, (i.e., $C_0 < \Gamma^0$).
- $\mathcal{M}^{i,n}$: the number of interfering macrocell users camping near femtocell i transmitting on sub-carrier n.
- $\mathcal{F}^{i,n}$: the number of interfering (neighboring) femtocells of femtocell i transmitting in sub-carrier n.

- **Action** $a_t^i = (a_t^{i,n}, n \in \{1, ..., N_{sub}\})$ where $a_t^{i,n} \in \{0,1\}$. The action of FBS$_i$ at time t is to transmit over a set of sub-carriers $\mathcal{C} \subseteq \{1, ..., N_{sub}\}$.
- **Reward** $r_t^i = (r_t^{i,n}, n \in \{1, ..., N_{sub}\})$ where $r_t^{i,n} = \log_2(1 + SINR_{i,n})$ is the perceived reward of femtocell i transmitting in sub-carrier n at time t. Its solution is given by:

$$\begin{aligned} r_t^i = \quad & \arg\max_{\mathbf{p}_i} C_i && (9)\\ s.t. \quad & C_0 \geq \Gamma^0 && (10)\\ & \textstyle\sum_{n=1}^{N_{sub}} p_i^{(n)} \leq \bar{P}_i, \quad \mathbf{p}_i \geq 0 \end{aligned}$$

where C_i is defined in (3) and $\mathbf{p}_i = [p_i^{(1)}, ..., p_i^{(N_{sub})}]$.

In essence, at every time instant t, every FBS$_i$ is in a given state $s_t^{i,n}$, builds its own local interference map in every carrier n, takes action $a_t^{i,n}$, and receives an immediate reward $r_t^{i,n}$. To gain insight on the Q-learning problem formulation, Figure 2 shows an illustration where a macrocell is underlaid with three femtocells and sharing four sub-carriers, in which the occupied frequency is gray-colored. FBS$_1$ has one MUE and one neighboring femtocell interfering on its downlink. FBS$_2$ has only one neighboring interfering femtocell FBS$_1$. Finally, FBS$_3$ has no interference at all in its vicinity. Moreover, at time instant t, the local state information of FBS$_1$ is given as: $s_t^{1,1} = \{1,1,1\}$, $s_t^{1,2} = \{1,1,1\}$, $s_t^{1,3} = \{0,0,1\}$, and $s_t^{1,4} = \{1,1,1\}$. For the agents' actions at time t, FBS$_1$ can only transmit in the third carrier ($a_t^1 = \{0,0,1,0\}$), while FBS$_2$ may transmit in all carriers ($a_t^2 = \{1,1,1,1\}$), and likewise for FBS$_3$ ($a_t^3 = \{1,1,1,1\}$). The distributed Q-algorithm performed by every femtocell i is shown in Algorithm 1.

C. Q-value

The Q-function is defined as the expected reward for femtocell i in state $s_t^{i,n}$ over time and is further defined as:

$$Q(s^{i,n}, a^{i,n}) = r(s^{i,n}, a^{i,n}) + \gamma \sum_{s \to v} p_{s,v} Q(v^{i,n}, b^{i,n}). \quad (11)$$

D. Updating Q-values

The Q-learning process aims at finding $Q(s^{i,n}, a^{i,n})$ in a recursive manner using available information $(s^{i,n}, a^{i,n}, v, r_t^{i,n})$ where the update equation is given as [4]:

$$\begin{aligned} Q_t(s^{i,n}, a^{i,n}) = (1-\alpha) Q_{t-1}(s^{i,n}, a^{i,n}) + \quad (12)\\ \alpha \left[r_t(s^{i,n}, a^{i,n}) + \gamma V_{t-1}(v_t) \right] \end{aligned}$$

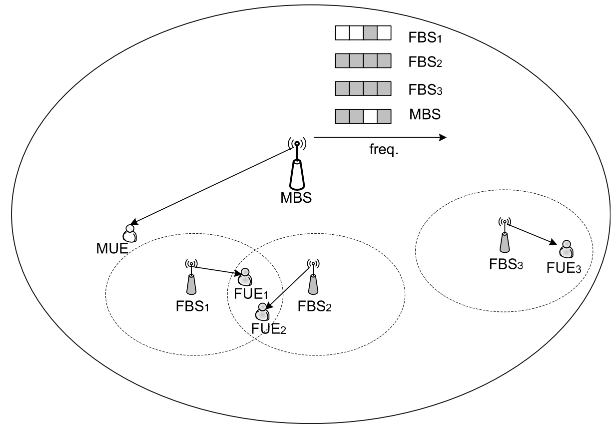

Fig. 2. Illustration of the considered network topology with one macrocell and $N_f = 3$ femtocells sharing $N_{sub} = 4$ sub-carriers.

where $V_{t-1}(v_t) = \max_{b^{i,n} \neq a^{i,n}} Q_{t-1}(v^{i,n}, b^{i,n})$ and α is the learning rate.

Algorithm 1 Q-learning based algorithm for macro-femtocell coexistence

Input: Number of sub-carriers N_{sub}, Number of femtocells N_f;
Init: $Q_0(0,0) = 0$ initialize Q-value for each femtocell i
for all sub-carrier n **do**
  for time t **do**
    if $(rand(.) < \epsilon)$ **then**
      select action randomly [exploration step]
    else
      choose action $a_t^{i,n} = \arg\max_{a_t^{i,n}} Q(s_t^{i,n}, a_t^{i,n})$ [exploitation step]
    end if
    receive immediate reward for femtocell i at time $t+1$: r_i^{t+1},
    observe new state $s_{t+1}^{i,n}$
    update Q-table as given in (12)
  end for
end for

IV. Numerical Results

In this section, we evaluate the performance of the proposed Q-learning algorithm by means of numerical results. One macrocell is considered with $N_f = 2$ uniformly distributed femtocells sharing a spectrum band composed of $N_{sub} = 5$ sub-carriers (co-channel interference scenario) where 1 macro/femto user per macro/femto is considered[2]. The channels gains are generated as Rayleigh variables with unit variance. Moreover, the decentralized Q-learning algorithm is implemented distributively by each femtocell with a learning rate of $\alpha = 0.5$, discount factor of $\gamma = 0.9$, and a macrocell target rate of $\Gamma^0 = 2$ [bps/Hz].

Figure 3 shows the convergence rate of the Q-learning algorithm for both femtocells. It is noted that at the beginning ($t = 1$), both femtocells violate the macrocell constraint given in (10). As a result, a self-organization mechanism is triggered where FBS$_1$ and FBS$_2$ learn from their local environment, build their Q-table, and adapt their strategies until reaching convergence. Convergence is obtained when both femtocells are satisfied with their payoff, while satisfying the QoS requirement of macrocell (i.e., $C_{MBS} \geq \Gamma^0$). Nevertheless, as

[2]This boils down to an interference channel (IFC) setting with $1 + N_f$ transmitters and receivers.

can be clearly noticed, since femtocells are assumed to be non-cooperative with no information exchange among them, their final outcome can be unsatisfactory, which sparks incentives for cooperation. One way would be through exchanging *public* information such as femtocells Q-tables. This is however beyond the scope of this paper.

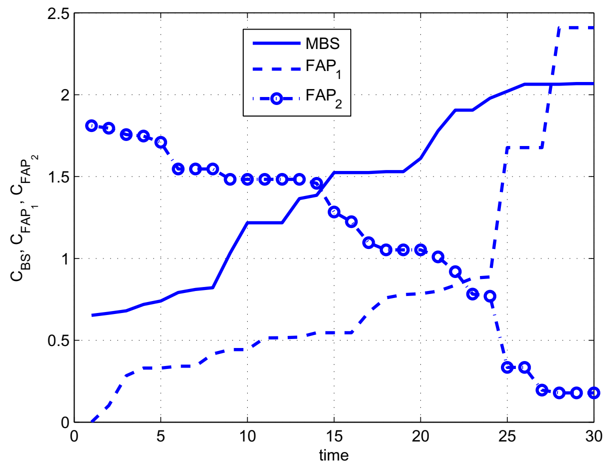

Fig. 3. Dynamics of the proposed Q-learning algorithm with $N_f = 2$ femtocells sharing a set of $N_{sub} = 5$ carriers.

Figure 4 shows the macrocell capacity as a function of the femtocell density N_f. The Q-learning approach is compared to both the *random* and *static* allocation strategies. Clearly, the Q-learning approach outperforms the random approach. Also, for $N_f = 4$ femtocells, the minimum required macrocell rate is halved with the random strategy, whereas the static approach still guarantees the minimum required macrocell rate with three femtocells after which that condition is violated with an increasing number of femtocells.

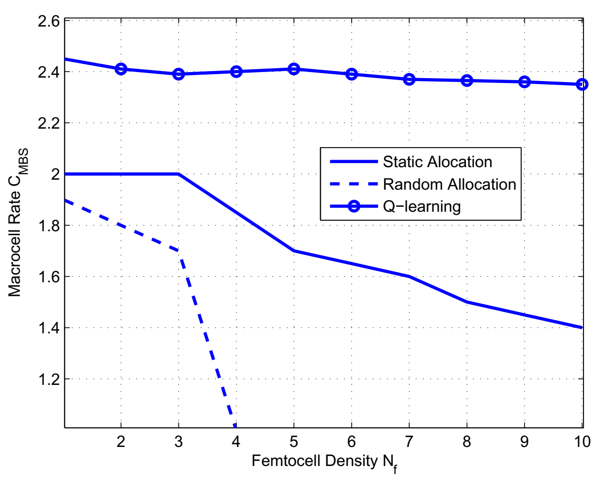

Fig. 4. Obtained macrocell capacity C_{MBS} as a function of the femtocell density N_f with different allocation strategies.

V. Conclusions and Future Work

In this paper, we have analyzed a Q-learning framework for macrocell and femtocell coexistence. The femtocells are able to self-organize in a decentralized manner, learn from their local environment with no information exchange, and make strategic decisions. Numerical results validate the theoretical claims in which femtocells coexist with the macrocell network while mitigating their interference towards the macro base station. Future work will look at the prospect of further improving the learning algorithms by means of public information exchange among femtocells (also known as networked femtocells). In addition, an interesting trade-off will be investigated where a faster learning convergence time may be reached at the expense of more information exchange at every femtocell.

References

[1] V. Chandrasekhar and J. G. Andrews, "Femtocell networks: A survey," IEEE Communication Magazine, 46(9):5967, September 2008.

[2] V. Chandrasekhar and J. Andrews, "Uplink capacity and interference avoidance for two-tier femtocell networks," IEEE Transactions on Wireless Communications, vol. 8, no. 7, July 2009.

[3] V. Chandrasekhar, T. Muharemovic, Z. Shen, J. G. Andrews and A. Gatherer, "Power Control in Two-Tier Femtocell Networks," IEEE Transactions on Wireless Communications, October 2008.

[4] M. E. Harmon and S. S. Harmon, "Reinforcement learning: A tutorial," 2000.

[5] H. Claussen, L. T. W. Ho, and L. G. Samuel, "An overview of the femtocell concept. Bell Labs Technical Journal," Wiley, 3(1): 221-245, May 2008.

[6] D. Niyato and E. Hossain, "Competitive spectrum sharing in cognitive radio networks: A dynamic game approach," IEEE Transactions on Wireless Communications, vol. 7, no. 7, 2651-2660, July 2008.

[7] D. Niyato and E. Hossain, "Dynamic of network selection in heterogeneous wireless networks: an evolutionary game approach," IEEE Transactions on Vehicular Communications, vol. 58, no. 4, 2651-2660, May 2009.

[8] Z. Bharucha, H. Haas, G. Auer, and I. Cosovic, "Femtocell resource partitioning," IEEE Global Telecommunications Workshops (GLOBECOM'09), Honolulu, USA, November 2009.

[9] M. Bennis and M. Debbah, "On spectrum sharing with underlaid femtocell networks," to appear at the Second IEEE International Workshop on Indoor and Outdoor Femto Cells (IEEE PIMRC), Istanbul, 26 September, 2010.

[10] S. Guruacharya, D. Niyato, E. Hossain, and D. I. Kim, "Hierarchical competition in femtocell-based cellular networks," to be presented in IEEE GLOBECOM'10, Miami FL USA, 6-10 December 2010.

[11] I. Guvenc, M. R. Jeong, F. Watanabe, and H. Inamura, "A Hybrid Frequency Assignment for Femtocells and Coverage Area Analysis for Co-Channel Operation", IEEE Commun. Lett., vol. 12, no. 12, pp. 880-882, Dec. 2008.

[12] M. Bennis, S. Lasaulce and M. Debbah, "Inter-operator spectrum sharing from a game theoretical perspective," EURASIP Journal of Advances in Signal Processing (JASP), Special issue on dynamic spectrum access for wireless networking, Aug. 2009, 1-12, Article ID 295739, doi:10.1155/2009/29573.

[13] L. Busoniu, R. Babuska and B. D. Schutter, "A comprehensive survey of multiagent reinforcement learning," IEEE Trans. Systems, Man and Cybernetics, Part C, vol.38, no.2, pp.156-172, March 2008.

[14] H. Claussen, L. T. W. Ho, and L. G. Samuel, "Self-optimization coverage for femtocell deployments," In Wireless Telecommunications Symposium, 24-26, pages 278 285, April 2008.

[15] D. Fudenberg and D. K. Levine, "The Theory of Learning in Games," The MIT Press, Cambridge, MA, 1998.

[16] H. Li, "Multi-agent Q-Learning of Channel Selection in Multi-user Cognitive Radio Systems: A Two by Two Case", IEEE Conference on System, Man and Cybernetics (SMC), 2009.

[17] F. Fu and M. van der Schaar, "Learning to compete for resources in wireless stochastic games," submitted to IEEE Trans. Veh. Technol., 2009.

[18] Z. Han, C. Pandana and K. Liu, "Distributive opportunistic spectrum access for cognitive radio using correlated equilibrium and no-regret learning," in Proc. of IEEE Wireless Communications and Networking Conference (WCNC), 2007.