

Dynamic Resource Allocation in Heterogeneous Wireless Networks

Jayant Thatte [*jayantt@stanford.edu*]

1 Introduction

As the wireless technology and standards continue to evolve, wireless communication systems and standards continue to promise to support an even higher number of mobile users, new applications, higher data rates, better coverage and more stringent quality of service (QoS) standards. Users need to be provided with high data rates and reliable service irrespective of their mobility or location. Heterogeneous Networks (HetNets) is one possible network architecture that can help in meet these challenging demands. Deployment of small cells not only helps in achieving better area coverage in wireless networks, but also helps in increasing network capacity through efficient spectrum reuse. Small cells can even reduce the burden on the traditional macrocell-based cellular network by off-loading some of the traffic onto small cell network. The future of wireless networks is thus expected to be highly heterogeneous, with WiFi hot-spots, user-deployed femtocells and other types of small cells coexisting within the macrocell network.

As the density of small cells in a network increases, measures need to be taken in order to ensure that the QoS is not degraded for the macrocell users as well as for the users of nearby small cells. Interference management in HetNets is, therefore, of critical importance. This is typically achieved through intelligent resource allocation schemes for the small cells. In HetNets, the mobile network is constructed with layers of small and large cells. This architecture is faced with the task of resource allocation (power, channel, time) for small cells in order to ensure reliable and high quality service to both primary (macrocell) users as well as secondary (femtocell) users. Furthermore, since the small cells are usually user-deployed, the locations and number of small cells in a HetNet is not fixed. This calls for dynamic and intelligent resource allocation algorithms for these networks. Various methods have been utilized for control of femtocell resources: open vs closed access and centralized control vs distributed coordination.

Within the scope of this project, the small cells are assumed to follow closed-access policy (typical for user-deployed cells) and the resource allocation is assumed to be achieved through a distributed control paradigm. Distributed control implies that each small cell independently chooses an action based of the information that it has and although each agent has bounded rationality, the network collectively evolves to a more optimal state. Game theory and reinforcement learning are two of the popular disciplines that tackles problems of this nature. In the project, dynamic resource allocation schemes based on reinforcement learning and game theory, respectively, are compared based on their assumptions, objectives, computational complexity of the associated algorithms and the degree of information exchange required among the small cells.

2 Accelerated Reinforcement Learning Approach

2.1 Q-Learning and Docation

Reinforcement learning is a branch of machine learning where the agent with bounded rationality learns the optimal policy in order to maximize the cumulative reward from the environment. Q-learning (QL) is a type of reinforcement learning in which an agent learns an optimal control policy from delayed rewards acquired through interaction with an environment through a Markov Decision Process. The problem is modeled as a set of states. The agent observes the current state s of the environment and then takes an action a which makes the environment transition to a new state s' . The agent maintains a table (called Q-table) of the optimal action to be taken in each of the possible states, which it updates over time as it learns the optimal policy. The Q-table is updated in the following manner.

$$Q^{(t)}(s, a) \leftarrow (1 - \eta)Q^{(t-1)}(s, a) + \eta[r + \gamma \max_{a'} Q^{(t-1)}(s', a')] \quad (1)$$

where $Q^{(t)}(s, a)$ denotes the cumulative reward that the agent would get if the agent is to start from state s , take action a and thereafter take actions that maximize the Q-value at each state. The term r denotes the reward (feedback) received from the environment as a result of taking action a from state s .

Docition (literally, teaching) is a method of speeding up QL in a multi-agent environment, where a more "experienced" agent shares its Q-table (knowledge learnt so far), with a less experienced agent. Docition works when the state structure is defined such that the optimal action in any given state is the same for all agents. The learnt Q-table can be shared among all the agents and thus, the learning process takes place much faster than if each agent were to find the optimal policy independently. This is a pronounced advantage of QL over game theory techniques. However, if the state structure is defined such that the rewards for different agents are maximized by taking different actions when in the same state, then docition can actually slow down the learning process.

2.2 The Model

QL-based resource allocation is discussed in [1]. Accelerated QL using docition is discussed in [2]. The approach discussed in these papers is outlined below.

Agents: Each small cell acts as an agent.

States: The states need to be defined in a way such that docition is possible. The state of a small cell i at time t is defined as

$$s_i^{(t)} = \{I_i^{(t)}, M_i^{(t)}, B_i^{(t)}\} \quad (2)$$

where I takes value of 1 if the capacity of the most affected macrocell user falls below a certain predefined threshold and is 0 otherwise. B captures the distance region of the macro basestation from the i^{th} small cell and M captures the distance region of the closest macrocell user from the i^{th} small cell. $B, M \in \{0, 1, 2, 3\}$, with 0 indicating smallest distance region (corresponding to the largest interference) and 1, 2 and 3 represent decreasing level of interference.

Actions: It is assumed that each small cell can transmit with a power chosen from a finite universe of power values. Then, the action for each small cell, is the resource allocation that it chooses.

Reward Function:

$$r_i^{(t)} = \begin{cases} \frac{d_{MUE,i}^{(t)}}{d_{th}} C_i^{(t)} - \frac{d_{th}}{d_{MUE,i}^{(t)}} (C_{MUE}^{(t)} - \Gamma_{MUE})^2 & \text{if } C_{MUE}^{(t)} \geq \Gamma_{MUE} \\ \frac{d_{MUE,i}^{(t)}}{d_{th}} C_i^{(t)} - \frac{d_{th}}{d_{MUE,i}^{(t)}} K_p & \text{else} \end{cases} \quad (3)$$

where $d_{MUE,i}^{(t)}$ is the distance of the i^{th} small cell from the closest macrocell user, d_{th} is some distance threshold for normalization, $C_{MUE}^{(t)}$ and $C_i^{(t)}$ are the capacities of the closest macrocell user and the small cell user respectively, Γ_{MUE} is the capacity threshold for the macrocell user that the small cells must maintain and K_p is the penalty term if this threshold is violated. The reward function is proximity dependent, that is, it penalizes/rewards the small cells closer to a macrocell user more than those farther away.

2.3 Simulation Scenario

The simulation scenario consists of one macrocell with a number of small cells deployed within the coverage area of the macrocell. Each small cell is assumed to have one user. An urban path loss model is used to compute path loss based on the simulation parameters and assumptions from [3]. The simulation was run several times with random layouts, in a Monte-Carlo fashion. For the purpose of simulation, only one macrocell user is assumed. This does not change the analysis since each small cell only considers the macrocell user closest to it.

Table I. SIMULATION PARAMETERS

Channel	Equation
BS to MUE	$PL(dB) = 15.3 + 37.6 \log_{10} d$
BS to FBS	$PL(dB) = 15.3 + 37.6 \log_{10} d + WP_{out}$
FBS to associated FUE	$PL(dB) = 38.46 + 20 \log_{10} d + 0.7d_{indoor} + w * WP_{in}$
FBS to MUE	$PL(dB) = \max(15.3 + 37.6 \log_{10} d, 38.46 + 20 \log_{10} d) + 0.7d_{indoor} + w * WP_{in} + WP_{out}$
FBS to alien FUE	$PL(dB) = \max(15.3 + 37.6 \log_{10} d, 38.46 + 20 \log_{10} d) + 0.7d_{indoor} + w * WP_{in} + 2 * WP_{out}$

Table II. SIMULATION PARAMETERS

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
P_{min}	-20 dBm	P_{max}	25 dBm	N_{power}	31	Step Size	1.5 dBm
d_{M1}	15 m	d_{M2}	50 m	d_{M3}	125 m	Γ_{MUE}	3.1623 bps/Hz
d_{B1}	50 m	d_{B2}	350 m	d_{B3}	600 m	p_{BS}	37 dBm
d_{th}	25 m	K_p	1000	σ^2	-120 dBm		

Figure 1: Simulation Parameters

2.4 Results

A typical simulation layout is shown in the figure below. The small cells are randomly deployed within a fixed radius from the macrocell.

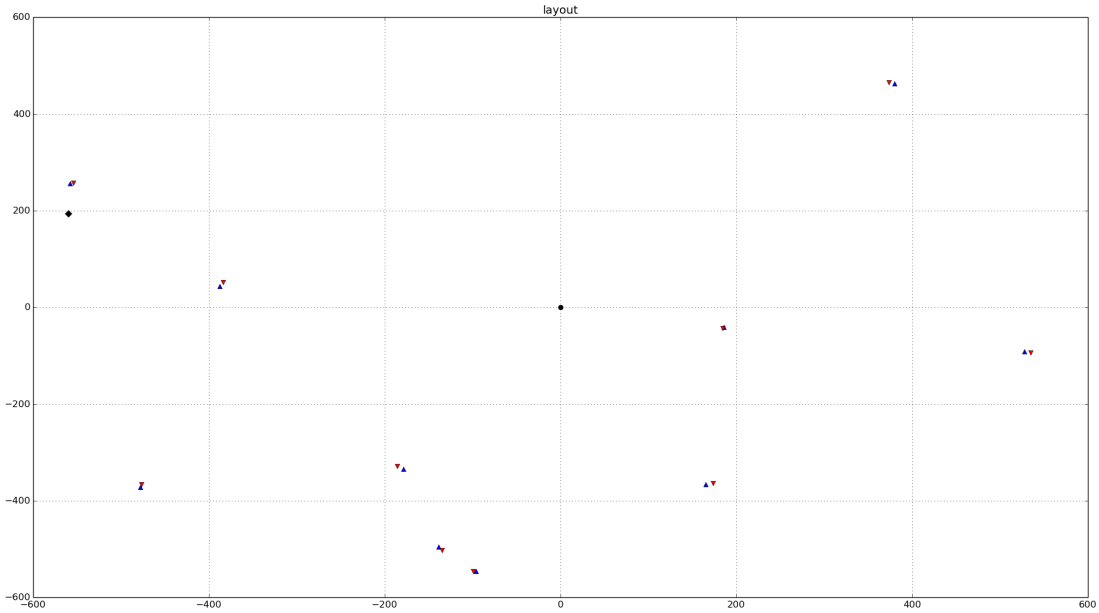


Figure 2: Simulation layout: The black dot is the base station, located at (0,0), the black diamond is macrocell user, red upward triangle are small cells and blue downward triangles are small cell users.

Evolution of capacities is for the above layout is shown in the figure below. The small cells are allowed to increase their power while ensuring that macrocell user capacity does not fall below a certain threshold. From the graphs, it can be seen that the capacities for small cell users increase as the small cells become more intelligent through reinforcement learning. The capacity of the macrocell user reduces, but while still remaining above the acceptable threshold.

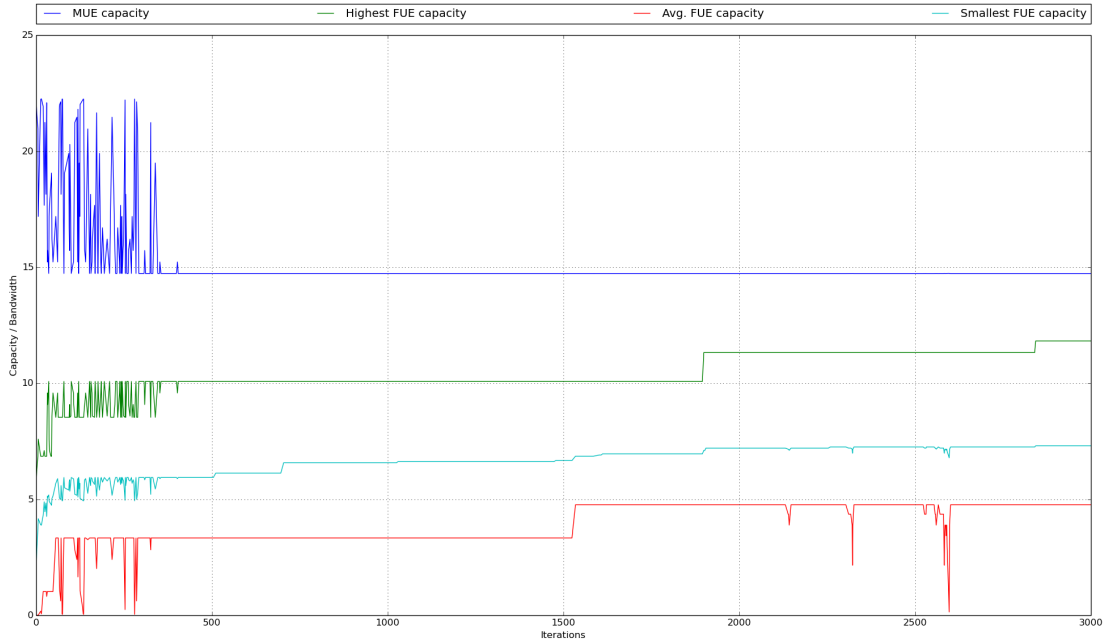


Figure 3: Evolution of capacities as optimal policy is learnt

3 Evolutionary Game Theory Approach

3.1 Evolutionary Game Theory

In game theory, each agent chooses an action independently and rationally and the overall system evolves to settle into one of the Nash Equilibria, if one exists. Evolutionary game theory (EGT) focuses on the dynamics of the attained equilibrium. EGT looks at how robust a certain strategy is, if a small number of players not following the said strategy are introduced in a population following the strategy. Similar to the reinforcement learning, EGT also consists of a set of players and a set of actions to choose from, for each player.

Consider a population with K agents and A be the set of pure strategies. Let x_a be the fraction of the population choosing a strategy $a \in A$. Trivially, we know that $\sum_{a \in A} x_a = 1$. The strategy adaptation process of players in an evolutionary game can be modeled using a set of ordinary differential equations called replicator dynamics [4].

$$\dot{x}_a = x_a(\pi_a(t) - \bar{\pi}(t)) \tag{4}$$

$\pi_a(t)$ is some sort of utility (reward) for choosing action a and $\bar{\pi}(t)$ is the average utility defined as

$$\bar{\pi} = \sum_{a \in A} \pi_a x_a \tag{5}$$

Equilibrium is reached when all the rates go to zero.

3.2 The Model

EGT-based dynamic resource allocation is proposed in [4]. In this model, a single macrocell is considered. Small cells are distributed within the coverage of the macrocell using Poission Point Process (PPP) with parameter λ in two dimensions. Each small cell can choose from a fixed set of powers and frequencies. Let N be set of orthogonal sub-carrier frequencies in downlink transmission and L be the set of valid power levels. TH is a minimum capacity

threshold for macrocell users. Each small cell is allowed to use a certain frequency and power only if doing so would not violate the capacity threshold for the most affected macrocell user. For the purpose of simulation, a single macrocell user is assumed per frequency band. This does not affect the algorithm since each small cell only looks at the most affected macrocell user in the frequency band.

$$SINR_{n,k} = \frac{g_{k,k}p_{n,k}}{N_0 + g_{m,k}p_m + \sum_{l=1, l \neq k}^K g_{l,k}p_{n,l}} \quad (6)$$

where m is subscript corresponding to macrocell basestation, $g_{i,j}$ is the channel gain from i to j including exponential path loss and Rayleigh fading, and $p_{n,k}$ is the transmit power of k^{th} small cell in frequency band n . $p_{n,k}$ is non-zero if the small cell has chosen to transmit in band n and is zero otherwise. Each small cell transmits in only one band at any instant of time.

For the k^{th} small cell, the utility is defined as

$$u_k(a_k) = \log_2(1 + SINR_k) \frac{\exp(TH - I_{k,m})}{1 + \exp(TH - I_{k,m})} \quad (7)$$

The utility is the capacity of the small cell user with an exponential penalty for having interference $I_{k,m}$ caused by the small cell to the nearest macrocell user.

The key idea followed by the authors while developing this model is to ensure minimum information exchange between different agents. As such, any small cell, only need to know the average utility of the small cells in order to determine the next action. To achieve this, the utility u_k given above is averaged with respect to the location of the small cell, the location of the small cell user and over Rayleigh shadowing.

$$u_k = E_{r,h,I}[\ln(1 + SINR)] \quad (8)$$

where expectation is over the distance r of the small cell from the macrocell basestation, shadowing h and accumulated interference I caused by other agents to the k^{th} small cell in the frequency in which it is transmitting. Thus,

$$u_k = \frac{2}{R^2} \int_{r=0}^R r \int_{t=0}^{\infty} Pr \left(h > \frac{r^\alpha (e^t - 1)(I_n + N_0)}{p_k} \right) dt dr \quad (9)$$

where R is the radius of the macrocell, I_n is the interference, p_k is the transmit power of the small cell, α is the pathloss exponent, h is the shadowing with exponential distribution with parameter μ . Let the interference I_n have a distribution $f(i_n)$. Then, averaging over interference gives us

$$u_k = \frac{2}{R^2} \int_{r=0}^R r \int_{t=0}^{\infty} \int_{i_n=0}^{\infty} \exp \left(-\frac{\mu r^\alpha (e^t - 1)(i_n + N_0)}{p_k} \right) f(i_n) di_n dt dr \quad (10)$$

The interference i_n is going to come from other small cells transmitting in the same band. The distribution of frequency bands over small cells is going to be random, especially in a dense small cell network. This means that the geographical distribution of small cells transmitting in a frequency band n is also going to follow two dimensional PPP with parameter $\lambda_n = x_n \lambda$ where x_n is the fraction of small cells transmitting in frequency band n at any given time.

Using this assumption, the above triple integral can be simplified to

$$u_k = \frac{2}{R^2} \int_{r=0}^R r \int_{t=0}^{\infty} \exp \left(-\frac{\mu r^\alpha (e^t - 1)N_0}{p_k} \right) \exp \left\{ -\pi x_n \lambda E_p \left[p^{\left(\frac{2}{\alpha} \right)} \right] E_h \left[h^{\left(\frac{2}{\alpha} \right)} \right] s^{\left(\frac{2}{\alpha} \right)} \Gamma \left(1 - \frac{2}{\alpha} \right) \right\} dt dr \quad (11)$$

This equation was then solved numerically in Python. Solving this expression gives us the utility for a small cell in terms of $E_p \left[p^{\left(\frac{2}{\alpha} \right)} \right]$ and λ_n which only depend on the fraction of the small cell population using each of the powers in set L and on the fraction of the small cell population transmitting in frequency band n , respectively. This

minimizes the required information exchange between the small cells.

Simulation: To begin with, all small cells choose random power and frequency. Then each small cell computes its own utility. Following this the average utility of the femtocell network is computed (at some central computing resource) and is broadcast to all the small cells. Each small cell then compares its utility with the average utility. If its utility is smaller than average, the small cell increases its transmit power. If the transmit power is already at a maximum, then the femtocell chooses another frequency, either randomly or by sensing its environment. Another option is, that some central unit could allocate frequency to the small cell based on the local spectrum usage. This is lead to a partially distributed control paradigm. Because the end state is when all small cells have utility close to the average utility of the network, this scheme explicitly ensures fairness in resource allocation

3.3 Results

EGT-based resource allocation proposed in [4] does not require the knowledge of the exact layout of the system. As such, during simulations, it is not required to have the exact layout. Having just the distances to the closest macrocell user and the parameter λ is sufficient. For the purpose of simulation, 50 small cells were used. It can be seen from the simulations that the capacity of the femtocells rises as the population evolves. What is interesting to note is that, unlike in QL-based approach, the worst capacity of the femtocells rises significantly as the population evolves and the best and the worst capacities of the femtocells are within a few percent of each other. Thus, EGT-based approach also ensures fairness, in addition to improving the overall capacity.

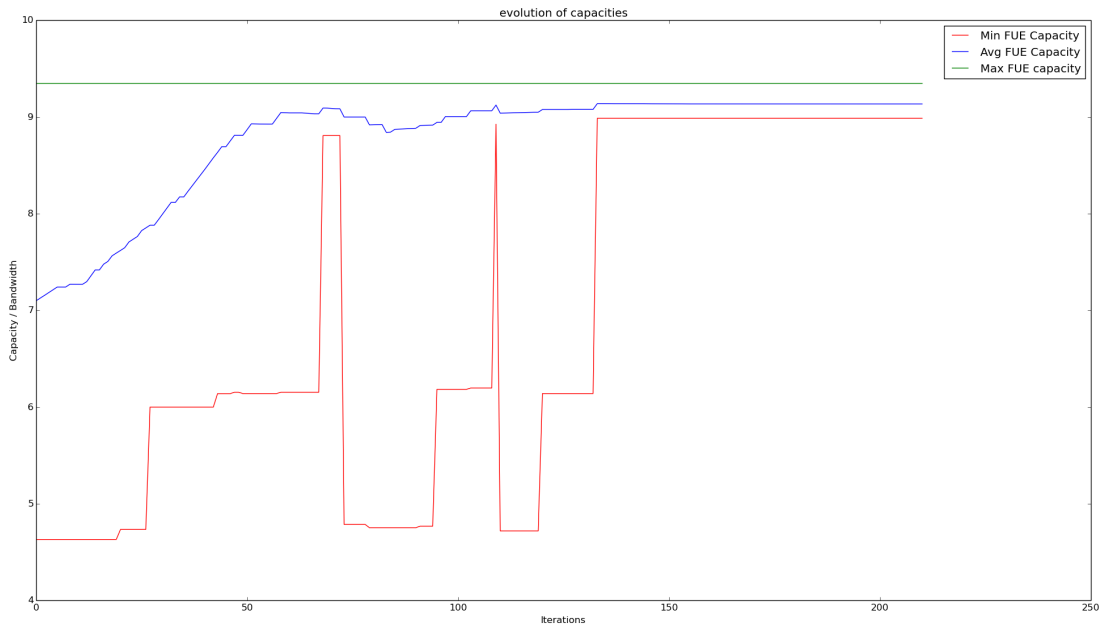


Figure 4: Evolution of capacities as optimal policy is learnt

4 Comparison

4.1 Assumptions

EGT-based approach assumed that the small cell network is random enough so that the small cells using any particular frequency band are still distributed in PPP manner within some radius of the macrocell basestation. This may not, in general be a good assumption, especially, if the femtocell network is only moderately dense. EGT-based

model is more like a statistical model that holds when the network is dense enough.

In QL-based approach, there are several parameters which need to be hard-coded and it is not clear which values of the parameters should give the best results. For example, QL-based approach divides the space into different concentric circles around the macrocell basestation and the “B”-state of a small is derived based on the region in which the cell is located. Similar division into zones is also done to determine how far the closest macrocell user is from a small cell. It is not clear how one would choose the critical distances which act as boundaries between regions, so as to get an overall optimal solution.

4.2 Objectives

Both approaches have similar primary objective: Obtain higher capacity for small cell users without significantly degrading the capacity of the nearby macrocell users. However, the two schemes follow different philosophies in achieving this. QL-based approach focuses on faster convergence through doction, whereas EGT-based approach focuses on minimal information exchange and on ensuring fairness in resource allocation.

4.3 Running time

The main aim and advantage of QL-based approach is doction, in which small cells can share what they have learnt with each other and this is known to significantly reduce the convergence time of the algorithm. Thus the algorithm converges in significantly smaller time than EGT-based approach. Averaged over a number of simulations, with different number of small cells in the network, EGT-based algorithm takes well over twice the amount of time take by Q-Learning.

EGT required a longer run time also because, in order to ensure minimal exchange of exact state information, in EGT-based approach each small cell has to compute the complicated double integral to estimate the average utility of a particular choice of frequency and power.

4.4 Complexity

For QL-based approach, the complexity of each iteration is linear in terms of the number of states and the number of actions. However, due to doction, the number of iterations required until convergence will rise slower than linearly with respect to the number of small cells in the network. This is because, the Q-table is optimized collectively and hence each new small cell acts like an agent to collect experience.

For EGT-based approach, the main computing task is evaluation of the complicated integral. And this needs to be done by each small cell. In addition, as the number of all cells increase, the number of iterations required for convergence will also increase. Hence, the computing time for EGT increases much faster than linearly with respect to the number of small cells.

4.5 Information Exchange

QL-based approach requires heavy information exchange among femtocells. For doction, the entire Q-table (number of states times number of actions number of entries) must be passed on from the teaching agent to the less experienced agent. EGT-based approach is more of a statistical model and hence the only information exchange required is each small cell sending out utility information and some central node broadcasting back the average utility of the network. In addition, in both approaches, the small cell needs to estimate the degree of interference which it is causing to the nearest macrocell user.

4.6 Fairness

In QL-based approach, each small cell does what is best for itself in any given state without causing degradation to the macrocell user. However, there is no explicit cooperation between the small cells. This means that it is possible that the highest femtocell user capacity can be significantly different from that of the femtocell user who has the worst capacity. QL-based approach simply optimizes for the sum of all capacities of the femtocell users.

In EGT-based approach, convergence is achieved when all of the small cell capacities are as close as possible to the average capacity of the small cell network. This automatically ensures fairness in resource allocation.

5 Conclusion

QL-based resource allocation scheme is significantly faster in terms of computation time and the computational complexity of the algorithm rises in a much more gradual fashion than that for EGT-based approach. In addition, in QL-based approach, when an agent experiences a new situation (change in location of a small cell, change in distance to the closest macrocell user) or when a new femtocell is deployed to the network (a new agent with no prior experience), the agent doesn't need to learn the optimal policy from scratch. It can simply borrow the Q-table from the closest located small cell. This means that the network is much more responsive to unseen circumstances.

EGT-based approach is better when it is not possible to determine the exact channel state or when the network is dense enough to use the statistical assumptions used in EGT-based scheme. EGT-based resource allocation required minimal information exchange between small cells and basestation and hence, the load on the backhaul network is very small, when for a network with large number of small cells. Finally, EGT strictly imposes fairness whereas in QL, the best and the worst capacities may be significantly different provided the sum capacity of the network, as a whole, increases.

The Python scripts used for simulations can be found at <http://www.stanford.edu/~jayantt/ee359/>

References

- [1] J. R. Tefft and N. J. Kirsch, *Accelerated Learning in Machine Learning-based Resource Allocation Methods for Heterogenous Networks*. Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, The 7th IEEE International Conference on, Sept. 2013
- [2] J. R. Tefft and N. J. Kirsch, *A proximity-based Q-learning reward function for femtocell networks* IEEE Vehicular Technology Conference, Sept. 2013
- [3] 3GPP, *3GPP R4-092042 TSG RAN WG4 (radio) meeting 51: Simulation assumptions and parameters for FDD HeNB RF requirements* Tech. Rep., May 2009
- [4] Semasinghe, P.; Kun Zhu; Hossain, E., *Distributed resource allocation for self-organizing small cell networks: An evolutionary game approach* Globecom Workshops (GC Wkshps), 2013 IEEE , vol., no., pp.702,707, 9-13 Dec. 2013