On Modeling Channel Selection in LTE-U as a Repeated Game

J. Pérez-Romero*, O. Sallent*, H. Ahmadi[&], I. Macaluso[+]
*Dept. of Signal Theory and Communications, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
[&] School of Electrical and Electronic Engineering, University College Dublin, Ireland
[+] CONNECT, Center for future networks and communications, Trinity College Dublin, Ireland
e-mail: jorperez@tsc.upc.edu, sallent@tsc.upc.edu, hamed.ahmadi@ucd.ie, macalusi@tcd.ie

Abstract— **This paper addresses the channel selection problem for Long Term Evolution Unlicensed (LTE-U). Channel selection is a frequency-domain mechanism that facilitates the coexistence of multiple networks sharing the unlicensed band. In particular, the paper considers a fully distributed approach where each small cell autonomously selects the channel to set-up an LTE-U carrier. The problem is modeled using a non-cooperative repeated game and the Iterative Trial and Error Learning - Best Action (ITEL-BA) learning algorithm is used to drive convergence towards a Nash Equilibrium. The proposed approach is evaluated by means of simulations in different situations analyzing both the throughput performance and the convergence behavior.**

Keywords - LTE-U; Unlicensed bands; Channel Selection, Game Theory.

I. Introduction

Long Term Evolution Unlicensed (LTE-U), also known as Licensed Assisted Access (LAA), is a promising enhancement that enables LTE to operate in unlicensed bands, with a clear focus on the 5 GHz band [1]. Although licensed spectrum remains 3GPP operators' top priority to deliver advanced services and better user experience, the use of unlicensed spectrum will be an important complement to meet the ultra-high capacity needs foreseen for 4G and beyond. In this respect, LTE-U is being currently considered for leveraging supplemental downlink capabilities that boost data rates and capacity to small cells, while at the same time licensed band LTE provides reliable connection for mobility, signaling, voice and data in both uplink and downlink.

The adoption of LTE-U brings a number of challenges to be addressed associated with the use of unlicensed spectrum. In particular, LTE-U will need to coexist with other technologies (e.g., Wi-Fi, other LTE-U networks) operating on the same band. Therefore, efficient coexistence mechanisms have to be devised to ensure a fair access of multiple LTE-U cells of the same or different operators as well as multiple Wi-Fi access points.

Channel selection (also referred to as carrier selection) is the mechanism used to decide the operating channel (i.e., center frequency and associated bandwidth) where a small cell sets up an LTE-U carrier. Dynamic channel selection enables a flexible choice of the operating channel and, therefore, it can be used as a frequency-domain coexistence mechanism to ensure that a given LTE-U small cell is a "good neighbor" to other nearby occupiers of the unlicensed band [2]. As discussed in [3] and [4], a fully distributed approach, where each small cell makes decisions on its own, would involve less demanding network coordination architectures, information exchange protocols and procedures. Besides, from a decision-making logic point of view, exploiting learning from past experience seems a pertinent principle in the LTE-U context. Each small cell may autonomously learn what channels are usually not being used by its neighbors and then tend to select such free channels. Furthermore, the adaptability of the learning-based decision-making process will provide robustness to the solution and the capability to react to changes in the scenario.

In such a fully distributed approach, the channel selection problem for LTE-U can be modeled based on game theory concepts. Game theory is a powerful mathematical tool that analyzes the strategic interactions among multiple decision makers [7]. By characterizing the channel selection problem as a game, players' behaviors and actions can be analyzed in a formalized structure that facilitates the application of the existing theoretical achievements in game theory. In particular, this paper presents a novel perspective in the formulation of the channel selection problem in LTE-U, considering a non-cooperative repeated game [8]. Regarding the learning algorithm, this paper considers the ITEL-BA (Iterative Trial and Error Learning - Best Action) proposed in [9], which was proved to converge to a Nash Equilibrium (NE) in pure strategies. ITEL-BA typically converges to an NE within a small number of iterations.

The rest of the paper is organized as follows. Section II presents the considered system model and the formulation of the channel selection in LTE-U as a game. Then, Section III presents the proposed learning strategy for the considered game. Section IV presents the simulation-based framework for performance assessment, while different performance results are presented in Section V. Finally, Section VI summarizes the main conclusions.

II. System Model and Formulation of Channel Selection as a Game

Let assume a set of S small cells (SCs) denoted as $\Sigma=\{1,..,S\}$ making use of the 5 GHz unlicensed band as a supplemental downlink for extending the available capacity in the licensed band. The total band is considered to be organized in K channels of bandwidth B. Let us denote by $A=\{1,...,K\}$ the set of available channels.

The channel selection problem consists of the decision making process individually undertaken by each SC to decide the operating channel where it will set up an LTE-U carrier. Globally, the process involves conflict among intelligent,

This work is supported by the Spanish Research Council and FEDER funds under RAMSES grant (ref. TEC2013-41698-R), and Science Foundation Ireland under grants No. 10/IN.1/I3007, 10CE/I1853 and 13/RC/2077.

rational and interactive decision-makers and it can be modeled as a repeated game [8], where each SC is a player in the game.

Without loss of generality, time is assumed to be organized in generic units denoted as "time steps" that specify the instants when the game is played. At the beginning of every time step, each player performs an action that consists of the selection of a channel to set-up an LTE-U carrier. Action $a_i(t)\in A$ denotes the channel selected by SC i in time step t.

At the end of a time step, each SC obtains a reward or payoff as a result of the selections made by all the SCs. We define the reward of SC i as the normalized average throughput achieved in the selected channel by the SC:

$$r_i\left(a_i\left(t\right),\mathbf{a}_{\mathbf{-i}}\left(t\right)\right)=\overline{R_i\left(a_i\left(t\right),\mathbf{a}_{\mathbf{-i}}\left(t\right)\right)}/R_{\max} \quad (1)$$

where $\overline{R_i\left(a_i\left(t\right),\mathbf{a}_{\mathbf{-i}}\left(t\right)\right)}$ is the average throughput obtained by the i-th SC when it is operating in channel $a_i(t)$ and the rest of SCs are operating in the channels given by $\mathbf{a}_{\mathbf{-i}}(t) = [a_1(t), ..., a_{i-1}(t), a_{i+1}(t),..., a_S(t)]$, and R_{max} is a normalization factor. Note that $\overline{R_i\left(a_i\left(t\right),\mathbf{a}_{\mathbf{-i}}\left(t\right)\right)}$ is a monotonically increasing function of the Signal to Interference and Noise Ratio (SINR) experienced by the users served by the i-th SC. In turn, $\overline{R_i\left(a_i\left(t\right),\mathbf{a}_{\mathbf{-i}}\left(t\right)\right)}$ is a non-increasing function of the number of SCs that are using the same channel $a_i(t)$.

Each player selects an action with the objective of optimizing its own reward. In order to achieve this, each player will apply a learning technique to the action selection decision-making during the game. A fully distributed situation is assumed in which each SC learns its own action selection strategy without having explicit knowledge on the strategy followed by the other SCs. Then, the learning is achieved through the interaction with the environment, so that the learner discovers which actions yield the highest reward by trying them. Furthermore, the learning process may lead the game to a state where none of the SCs can improve its throughput by unilaterally changing its selection. This state is called a Nash Equilibrium [10].

III. LEARNING STRATEGY

We propose the use of the ITEL-BA algorithm for channel selection in LTE-U. ITEL-BA is an extension of the Interactive Trial and Error Learning (ITEL) of [11], a simple but effective learning algorithm where players can only observe the results of their own actions. ITEL-BA was proposed in [9] in the context of a carrier aggregation scenario. This approach significantly reduces the convergence time to an NE, under the assumption that players are able to gather additional information. In particular, we assume that players are able to measure interference in all the channels.

From a general perspective, the learning is an iterative process where each iteration performed in a time step can be broadly divided into three phases: (i) selection of a new action according to a certain strategy; (ii) observation of the environment by measuring the obtained reward resulting from the selected action, which gives the players an idea of how well they played; (iii) improvement of the action selection strategy based on the current observation.

In the particular case of the ITEL-BA algorithm, each player retains a benchmark action and the corresponding benchmark reward as a reference to evolve the action selection strategy. Let denote as $a_{Bi}(t)$ and $r_{Bi}(t)$, respectively, the benchmark action and benchmark reward of player i at the beginning of time step t. At this time, the operation of the ITEL-BA algorithm considers that the player i selects an action $a_i(t)$, which can be the benchmark action or a different action within the set A. The action is chosen depending on the so-called *mood* of the player, which basically captures the degree of satisfaction of the player with the current benchmark action and benchmark reward. The mood of player i at the beginning of time step t is denoted as $m_i(t)$ and it can be *content*, *discontent*, *hopeful* or *watchful*.

As a result of the action selected by the i-th player, $a_i(t)$, and the actions selected by the other players, $\mathbf{a}_{\mathbf{-i}}(t)$, player i will measure the obtained reward $r_i(a_i(t),\mathbf{a}_{\mathbf{-i}}(t))$ at the end of time step t. The comparison between the obtained reward and the benchmark reward will in turn be used to update the mood, the benchmark action and benchmark reward for the next time step, respectively $m_i(t+1)$, $a_{Bi}(t+1)$, $r_{Bi}(t+1)$. Fig. 1 shows graphically the detailed action selection process and update rules of the mood and benchmark action/reward for the ITEL-BA algorithm executed by player i in each time step t. At initialization (t=0), the mood of player i is set to *discontent*, and its benchmark action $a_{Bi}(0)$ is selected randomly from the set $A=\{1,...,K\}$.

The general idea of the action selection strategy shown in Fig. 1 is that a *content* player will be selecting the benchmark action $a_{Bi}(t)$ most of the time, and will occasionally experiment with new actions according to a probability $\varepsilon<<1$ called exploration rate. In the latter case, it will change the benchmark action/reward if the new action is better than the old. Instead, a *discontent* player will try out new actions frequently, eventually becoming *content* with a probability ϕ that depends on how well the selected action is performing in terms of reward.

The *hopeful* and *watchful* moods correspond to transitional situations, triggered by changes in the behavior of other players (or in the environment). Specifically, if a *content* player selects its *benchmark* action and receives a different reward than the benchmark reward (e.g. because some other player changed its action), then the player becomes *hopeful* if the reward increased and *watchful* if it decreased. If the player is *hopeful* and the reward of the benchmark action stays up for one more time step, the player will become *content* again and will update the benchmark reward to the new value. In turn, if the player is *watchful* and the reward stays down for one more time step, the player will become *discontent*.

Whenever a *discontent* player selects an action or when a *content* player experiments with new actions different from the benchmark, the action selection criterion used by the ITEL-BA algorithm is defined as:

$$a_i*\left(t\right)=\arg\max_{a_i'\in A}\hat{r}_i\left(a_i',\mathbf{a}_{\mathbf{-i}}\left(t-1\right)\right) \quad (2)$$

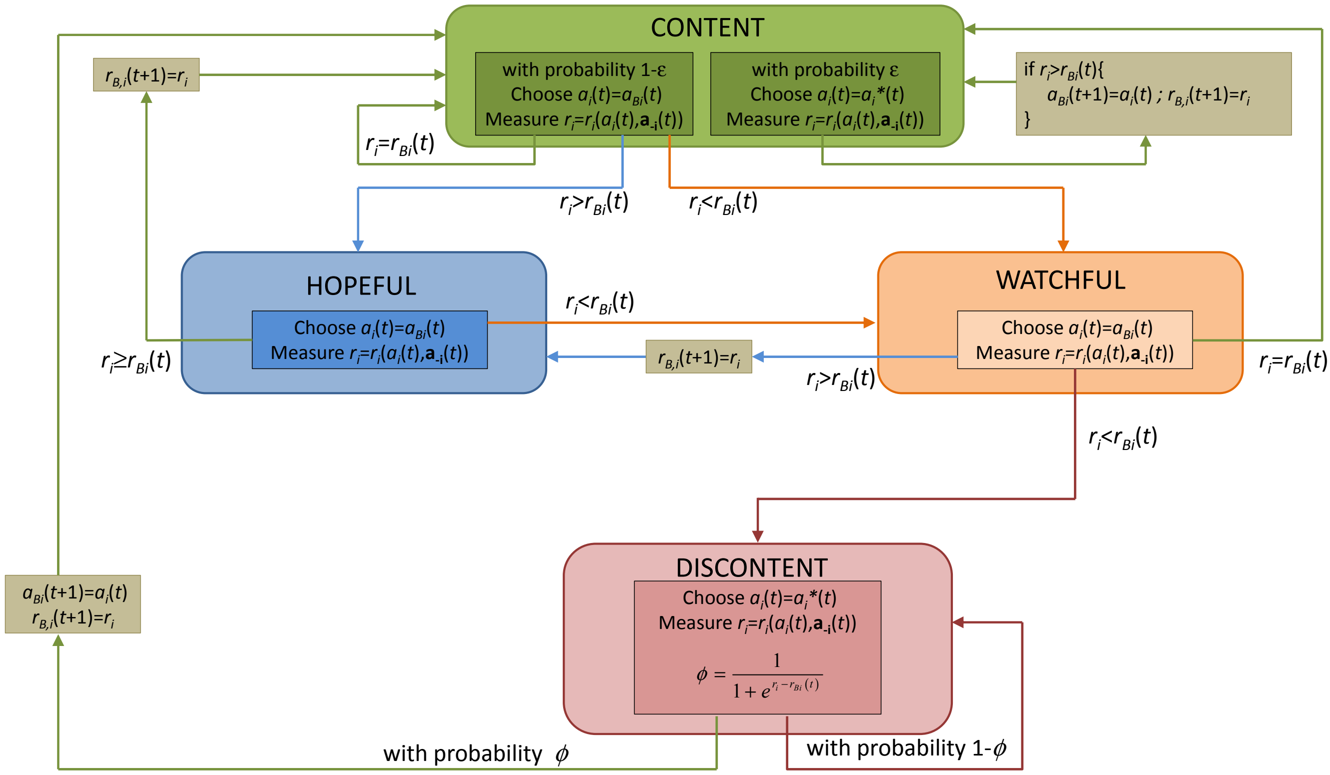

Fig. 1. Action selection process and update rules of the mood and benchmark action/reward for the ITEL-BA algorithm.

where $\hat{r}_i\left(a_i', \mathbf{a}_{-i}(t-1)\right)$ is an estimation of the hypothetical reward that the i-th SC would obtain by transmitting in channel a_i' during time step t assuming that the other SCs keep the same channels $\mathbf{a}_{-i}(t-1)$ that they selected in the previous time step t-1. To compute this estimation the i-th SC needs to measure the existing interference in all the available channels A during time step t-1. These interference measurements will capture the usage of the channels done by the other SCs according to their selected actions $\mathbf{a}_{-i}(t-1)$. Whenever there are multiple actions with the same maximum value of $\hat{r}_i\left(a_i', \mathbf{a}_{-i}(t-1)\right)$ the selected action $a_i^*(t)$ will be chosen randomly among these ones. It must be noted that the i-th SC needs to compute $a_i^*(t)$ only when it wants to change its action. This means that the SC will not measure the existing interference in all the available channels A all the time.

IV. Simulation Framework

The considered scenario to evaluate the performance of the proposed approach is based on the indoor scenario for LTE-U coexistence evaluations defined in the 3GPP Study Item [1]. It consists of a single floor building where two operators deploy 4 small cells (SCs) each. SCs are equally spaced and centered along the shorter dimension of the building, as depicted in Fig. 2. Small cells SC1 to SC4 are owned by operator 1 (OP1), while SC5 to SC8 are owned by operator 2 (OP2). A total of 10 User Equipments (UEs) per operator are randomly distributed inside the building. Each UE is associated to the SC of its own operator that provides the highest received power. Small cells are deployed at height 6m while the antenna height of the UEs is 1.5m. The SC-to-UE and SC-to-SC path loss and shadowing are computed using the ITU InH model in [12].

The 5 GHz unlicensed band is considered, organized in K channels of bandwidth B=20 MHz, numbered as k=1,..,K. channels. Each SC is configured to exploit one channel as supplemental downlink for extending the available capacity in the licensed band. The transmit power in one LTE-U carrier is 15 dBm. Omnidirectional antenna patterns are assumed with a total antenna gain plus connector loss of 5 dB. The terminal noise figure is 9 dB. The exploration rate of the ITEL-BA algorithm is ε=0.01.

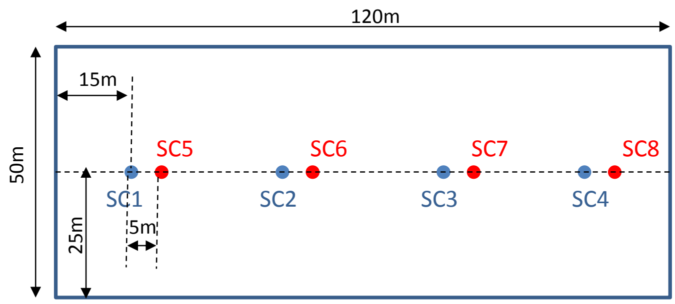

Fig. 2. Layout of the floor building

The computation of the average throughput $\overline{R_i\left(a_i(t), \mathbf{a}_{-i}(t)\right)}$ obtained by the i-th SC when it is operating in channel $a_i(t)$ and the other SCs operate in channels $\mathbf{a}_{-i}(t)$ is done based on the model described in [3]. The model captures the use of a Listen-Before-Talk (LBT) scheme that regulates the transmissions of multiple SCs working in the same channel, as required by the regulation of some markets like Europe and Japan [13]. In turn, the normalization factor R_{max}

corresponds to the maximum throughput that an SC can achieve in an LTE-U channel. It is defined as $R_{max}=B\cdot S_{\max}\cdot(1-\theta_{idle})$ where S_{max} is the maximum spectral efficiency in b/s/Hz that the technology can achieve (assumed here to be 4.4 b/s/Hz [14]) and θ_{idle} is the fraction of time associated with the idle periods imposed by the LBT strategy, assumed here θ_{idle}=0.05 [13].

The computation of the estimated hypothetical reward $\hat{r}_i\left(a_i',\mathbf{a}_{\text{-i}}\left(t-1\right)\right)$ assumes that the i-th SC is able to measure the interference in each channel. Based on this, the SINR and the opportunities of transmission given by the LBT scheme are estimated and the throughput is computed following [3].

V. RESULTS

To assess the behavior of the proposed approach, the SCs of OP1 execute the ITEL-BA strategy. Regarding the operation of the SCs of OP2, three different possibilities are considered: (1) they are inactive, which enables the study of the game when only the SCs of OP#1 are players and there is no external influence to the game (subsection A), (2) they use a fixed channel assignment, which enables the study of the game when the SCs of OP#1 are players and are influenced by some fixed external constraints (subsection B) and, (3) when they also execute the ITEL-BA strategy, which enables the study of the game when the SCs of OP#1 are players but there are also other players in the game (subsection C). Finally, subsection D presents a comprehensive analysis on the algorithm's convergence time.

A. Behavior of the ITEL-BA algorithm with no external influence

To illustrate the behavior of the ITEL-BA algorithm let us consider first a situation with K=4 channels and in which the SCs of OP2 are inactive. Fig. 3 illustrates an example of the time evolution of the benchmark actions $a_{Bi}(t)$ identified by the SCs 1 to 4 of OP1 for one realization of the simulation, and Fig. 4 presents the corresponding evolution of the mood of each SC. As seen in Fig. 3 the initial benchmark actions, which are randomly selected by each SC, happen to be the vector $\mathbf{a_B}(0)$=(2,3,1,2). After some initial exploration, and based on the random behavior of the algorithm while the SCs are *discontent*, at t=5 time steps all the SCs become *content* with benchmark actions $\mathbf{a_B}(5)$=(4,2,4,3). For this benchmark action Table I shows the values of the rewards $r_i(a_i,\mathbf{a}_{\mathbf{B}\text{-i}})$ that each SC i can achieve with each possible action a_i while the rest of SCs different from i keep their benchmark actions $\mathbf{a}_{\mathbf{B},\text{-i}}$. It is observed that the reward that SC1 achieves by selecting a_{B1}=4 is only 0.5. The same occurs for SC3, which achieves a reward of 0.5 with its benchmark action a_{B3}=4. Clearly, this is not a good situation because SC3 and SC4 are using the same channel. Indeed, the benchmark action set $\mathbf{a_B}$=(4,2,4,3) does not represent an NE, because both SC1 and SC3 can achieve a better reward by deviating from the benchmark action, as shown in Table I.

Nevertheless, as seen in Fig. 3 and Fig. 4, the benchmark action $\mathbf{a_B}$=(4,2,4,3) is maintained until t=109 time steps and all the SCs remain *content*. Then, at t=109 SC1 randomly decides to explore other options and applies the criterion (2) when selecting the channel. As it can be seen in Table I, when SC1 applies the criterion (2) while the rest of SCs keep their benchmark action, the result is that SC1 will select channel a_1*=1, as this channel provides the maximum reward. Therefore, the new benchmark actions after t>109 time steps become $\mathbf{a_B}$=(1,2,4,3), which actually is an NE because no SC can unilaterally improve its payoff by choosing a different action. Consequently, after t>109 the benchmark action vector will remain the same, because any of the SCs that randomly decides to explore will end up with the selection of the same benchmark action. Then, it can be concluded that in this specific example the ITEL-BA strategy has reached convergence to an NE after 109 time steps.

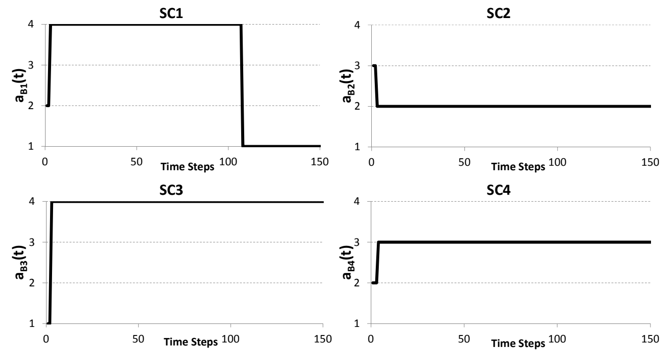

Fig. 3 Example of the time evolution of the benchmark actions of the SCs of OP1.

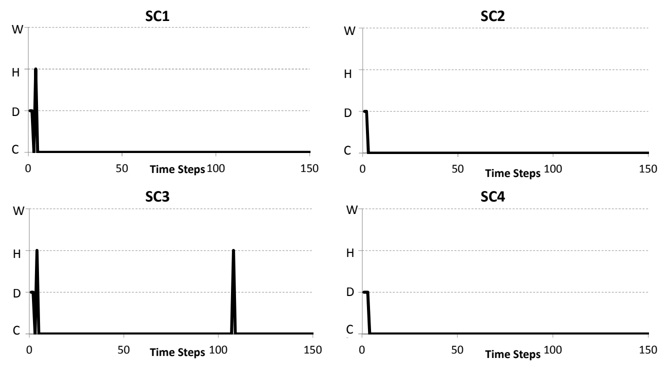

Fig. 4. Example of the time evolution of the mood of the SCs of OP1 (C: content, D: discontent, H: hopeful, W: watchful).

Besides the description of the above realization, which provides insight into the step-by-step operation of the algorithm, 10^5 different random realizations of the simulation have been conducted to draw a more comprehensive understanding of the ITEL-BA operation. Each realization is characterized by different outcomes of the random processes involved in the algorithm of Fig. 1, which lead to different realizations of the game. The results show that in all the cases the SCs converge to one out of 24 possible channel selection combinations with equal probability, as shown in Fig. 5. A detailed analysis of all the possible combinations of channels and SCs in this scenario reveals that these 24 combinations are actually all the NE that exist in this scenario. Each NE corresponds to a combination in which each SC uses a different channel (i.e. the 24 combinations correspond to the 4!=24 permutations of 4 channels). In all the NE the achievable normalized throughput is 1. Therefore, ITEL-BA converges to any of the possible NE with the same probability.

Table I. Achieved rewards $r_i(a_i,\mathbf{a}_{B,-i})$ for different benchmark actions $\mathbf{a}_B$. Highlighted cells in grey show the reward for the players in correspondence to the benchmark actions.

	$\mathbf{a}_B$=(4,2,4,3)				$\mathbf{a}_B$=(1,2,4,3)			
$r_i(a_i,\mathbf{a}_{B,-i})$	a_i=1	a_i=2	a_i=3	a_i=4	a_i=1	a_i=2	a_i=3	a_i=4
SC1	1.0	0.5	0.83	**0.5**	**1.0**	0.5	0.83	0.5
SC2	1.0	**1.0**	0.5	0.33	0.5	**1.0**	0.5	0.5
SC3	1.0	0.5	0.5	**0.5**	0.5	0.5	0.5	**1.0**
SC4	1.0	0.5	**1.0**	0.39	0.79	0.5	**1.0**	0.5

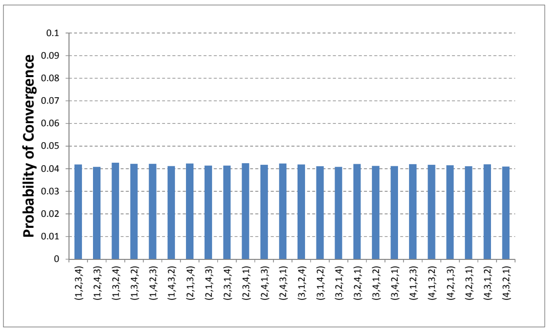

Fig. 5. Probability of convergence to the different NE.

B. Behavior of ITEL-BA in the presence of a fixed external influence

In the following we analyze the situation when the SCs of OP2 are active in the scenario with K=4 channels and have a fixed assignment of channels. In particular, SCs 5,6,7,8 use channels 1,2,3,4, respectively. At time t=0 when the SCs of OP1 become active and start the game, the SCs of OP2 are already active using their assigned channels, so they impact on the interference measurements made by the SCs of OP1.

A detailed analysis of this scenario considering all the possible choices made by the SCs 1 to 4 of OP1 reveals that there are a total of 16 NE. It is worth noting that, given that the SCs of OP2 are using some specific channels, not all the 24 combinations in which SCs 1 to 4 use a different channel are NE and, therefore, this scenario leads to a reduced number of NE compared to the previous case where OP2 was inactive.

Fig. 6 plots the probability of convergence to each NE after executing a total of 10^5 random realizations of this scenario. Each NE is denoted as a vector with the channels used by each SC 1 to 4. It can be observed that all the NE are characterized by the fact that each SC of OP1 uses a different channel. However, not all NE correspond to the same throughput, as it will be dependent on the interference generated by the SCs of OP2. In this respect, Fig. 6 also plots the normalized average throughput per SC of OP1 that can be obtained in each NE. It is worth noting that in general the system tends to converge with higher probability to the NE that provide the higher throughput values.

C. Behavior when both operators apply ITEL-BA

Finally, let us consider the case with K=4 channels in which the SCs of both OP1 and OP2 apply the ITEL-BA strategy and all of them start to play the game at t=0. In this case there are a total of 1368 NE considering all the possible channels used by each SC 1 to 8. Fig. 7 plots the probability of convergence to each NE (identified by a numeric index) after a total of 10^5 random realizations as well as the normalized average throughput per SC of both OP1 and OP2 in each NE. Similarly to Fig. 6, it is observed that the system tends to converge with higher probability to the NE that correspond to the higher throughput values, which reveals the good behavior of the ITEL-BA strategy: even though each player solely intends to maximize its own reward, it happens that the algorithm tends to favor NE that provide also good performance from a global perspective.

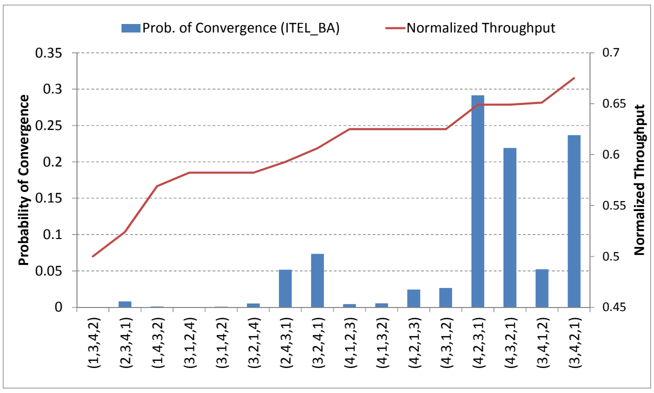

Fig. 6. Probability of convergence to the different NEs and normalized throughput of each NE for the case K=4 when the SCs of OP2 have a fixed channel assignment.

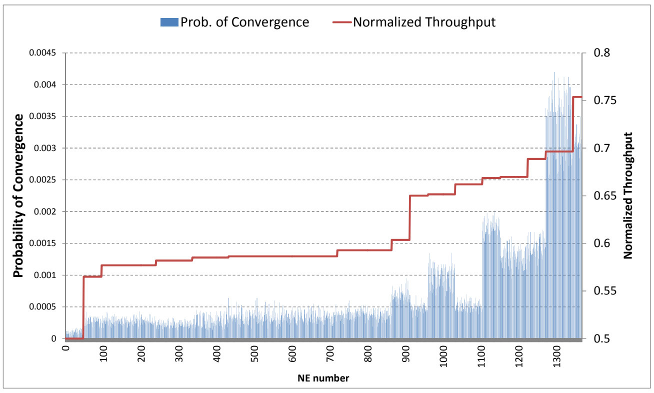

Fig. 7. Probability of convergence to the different NEs and normalized throughput of each NE for the case K=4 when the SCs of both OP1 and OP2 apply ITEL-BA.

D. Convergence time analysis

In the following we analyze the time needed by ITEL-BA to reach convergence to an NE. The cases K=4 and K=8 channels are considered. For each case, three possibilities are considered for the SCs of OP2, namely when they are inactive (denoted as No OP2), when they apply a fixed assignment (denoted as OP2 Fixed) and when they apply the ITEL-BA algorithm (denoted as OP2 ITEL-BA). For each case, 10^5 random realizations have been conducted. Fig. 8 plots the average convergence time for each of these possibilities. In addition, as a relevant metric, Fig. 8 also presents for each case the percentage of NE with respect to the total number of possible combinations of channels to be selected by each SC applying ITEL-BA. As an example, in the case K=8 channels and when the SCs of OP2 are not active, there are a total of 8^4=4096 possible combinations, and 1680 of them are NE, so the fraction of NE is 41%.

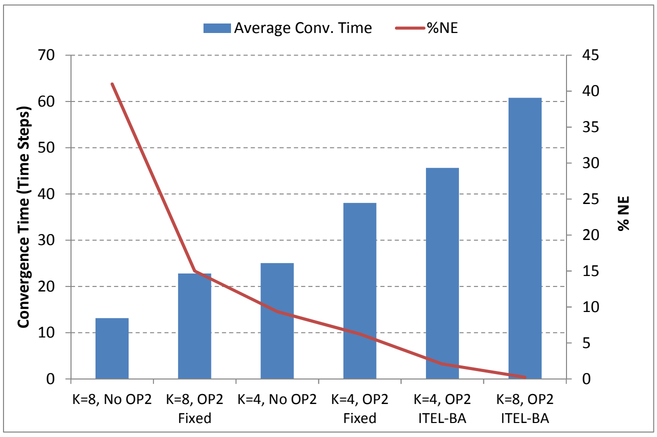

Fig. 8. Average convergence time of the ITEL-BA algorithm under different conditions.

It is observed in Fig. 8 that the average convergence time is in the order of some tenths of time steps depending on the case. In addition, there is a clear inverse correlation between the fraction of NE and the convergence time, so that the lower the fraction of NE is for a given scenario, the higher the convergence time will be. The rationality behind this result is that the total number of combinations represents the overall solution search space on which an NE has to be found. Then, if the number of NE represents a small fraction of the total search space, it is likely that the system would need to explore and discard more combinations before reaching an NE, so the convergence time increases.

As an additional result, Fig. 9 plots the Cumulative Distribution Function (CDF) of the convergence time for the different cases under analysis. It is observed that in all the cases the 90-th percentile of the convergence time is approximately in the order of 3 times the average convergence time.

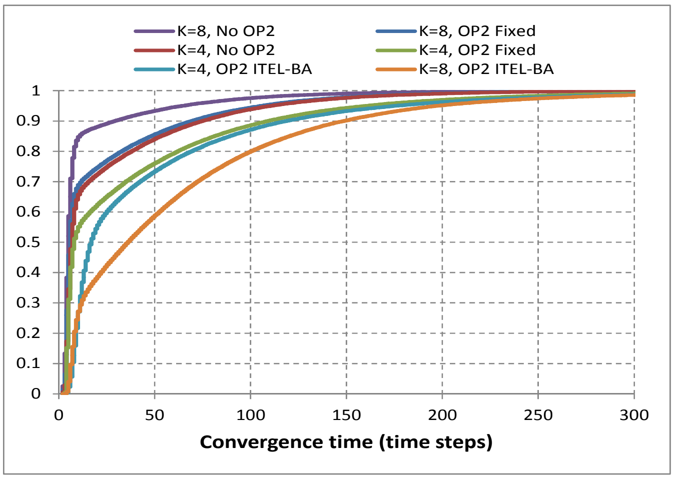

Fig. 9. CDF of the convergence time of the ITEL-BA algorithm under different conditions.

VI. Conclusions and Future Work

This paper has focused on the channel selection problem for LTE-U as a mechanism that enables the coexistence of multiple networks using the same unlicensed band. A fully distributed channel selection approach has been considered where each small cell autonomously chooses the channel to set-up an LTE-U carrier for supplemental downlink. The problem has been modeled as a non-cooperative repeated game and the ITEL-BA strategy has been proposed as the learning algorithm that drives the system towards NE.

The proposed framework has been evaluated in an indoor scenario under different conditions regarding the number of players and presence of external influences. Results have revealed the capability of the proposed framework to converge to NE. Besides, when there exist multiple NE that offer different throughput performance, it has been observed in the analyzed cases that the system tends to converge with higher probability to the NE that provide the higher throughput values. The required convergence time has been studied, obtaining an inverse correlation between the convergence time and the fraction of NE with respect to the overall solution search space.

As future work, it is planned to perform an in-depth comparison between Q-learning (as considered in [3] and [4]) and ITEL-BA in terms of achieved throughput performance, convergence time as well as implementation complexity, which requires a detailed definition of the channel measurement procedures for each solution approach.

References

[1] 3GPP TR 36.889, "Study on Licensed-Assisted Access to Unlicensed Spectrum (Release 13)", November, 2014.

[2] Qualcomm, "LTE in Unlicensed Spectrum: Harmonious Coexistence with Wi-Fi", June 2014.

[3] O. Sallent, J. Pérez-Romero, R. Ferrús, R. Agustí, "Learning-based Coexistence for LTE Operation in Unlicensed Bands", IEEE ICC 2015 - Workshop on LTE in Unlicensed Bands: Potentials and Challenges, June, 2015.

[4] J. Pérez-Romero, O. Sallent, R. Ferrús, R. Agustí, "A Robustness Analysis of Learning-based Coexistence Mechanisms for LTE-U Operation in Non-Stationary Conditions", IEEE Vehicular Technology Conference (VTC Fall), September, 2015.

[5] F. Bernardo, R. Agustí, J. Pérez-Romero, O. Sallent, "An Application of Reinforcement Learning for Efficient Spectrum Usage in Next Generation Mobile Cellular Networks", IEEE Transactions on Systems, Man and Cybernetics - Part C, Vol. 40, No. 4, pp. 477-484, July, 2010.

[6] N. Vucevic, J. Pérez-Romero, O. Sallent, R. Agustí "Reinforcement Learning for Joint Radio Resource Management in LTE-UMTS Scenarios", Computer Networks, Elsevier, May, 2011, Vol. 55, No. 7, pp. 1487-1497.

[7] B. Wang, Y. Ku, K.J. Ray Liu, "Game theory for cognitive radio networks: an overview", Computer Networks, Elsevier, Vol. 54, 2010.

[8] S. Lasaulce, H. Tembine, *Game Theory and Learning for Wireless Networks. Fundamentals and Applications*, Academic Press, Elsevier, 2011.

[9] H. Ahmadi, I. Macaluso, L.A. DaSilva, "Carrier Aggregation as a Repeated Game: Learning Algorithms for Efficient Convergence to a Nash Equilibrium", IEEE GLOBECOM, 2013.

[10] L. Rose, S. Lasaulce, S. M. Perlaza, M. Debbah, "Learning Equilibria with Partial Information in Decentralized Wireless Networks", IEEE Communications Magazine, Vol. 49, No. 8, August, 2011, pp. 136-142.

[11] H. Young, "Learning by trial and error", Games and economic behavior, vol. 65, No.2, pp. 626-643, 2009.

[12] 3GPP TR 36.814 v9.0.0 "Further advancements for E-UTRA physical layer aspects", March, 2010.

[13] ETSI EN 301 893 v1.7.2 "Broadband Radio Access Networks (BRAN): 5 GHz high performance RLAN; Harmonized EN covering the essential requirements of article 3.2 of the R&TTE Directive", July, 2014.

[14] 3GPP TR 36.942 v12.0.0, "Radio Frequency (RF) system scenarios", September, 2014.