A Robustness Analysis of Learning-based Coexistence Mechanisms for LTE-U Operation in Non-Stationary Conditions

J. Pérez-Romero, O. Sallent, R. Ferrús, R. Agustí
Dept. of Signal Theory and Communications,
Universitat Politècnica de Catalunya (UPC),
Barcelona, Spain
E-mail: [jorperez, sallent, ferrus, ramon]@tsc.upc.edu

Abstract— **The use of Long Term Evolution (LTE) in the unlicensed 5 GHz band, referred to as LTE-U, is a promising enhancement to increase the capacity of LTE networks and meet the requirements of future systems. This paper considers a Q-learning based Channel Selection strategy to decide the most appropriate channel to use for downlink traffic offloading in the unlicensed band as a mechanism to greatly facilitate the coexistence among several LTE-U and/or Wi-Fi systems in the same band. The focus is placed on analyzing the robustness of the proposed approach in front of non-stationary conditions in the wireless environment. Simulation results allow assessing quantitatively the capability of the proposed strategy to relearn proper solutions when changes in the environment occur. Furthermore, the analysis evaluates quantitatively how fast the learning process has to be compared to the variations in the environment in order to retain an LTE-U throughput performance very close to the optimum one.**

Keywords - LTE-U; Unlicensed bands; Channel Selection, Q-learning; Coexistence; Non-stationarity.

I. INTRODUCTION

Long Term Evolution Unlicensed (LTE-U), also known as Licensed Assisted Access (LAA), is a promising enhancement in the 3GPP ecosystem that enables LTE to operate and coexist with other technologies in unlicensed bands [1][2]. Although licensed spectrum remains 3GPP operators' top priority to deliver advanced services and better user experience because the benefits of licensed spectrum such as controlled Quality of Service (QoS) cannot be matched by unlicensed spectrum, the use of unlicensed spectrum will be an important complement to meet the ultra-high capacity needs foreseen for 4G and beyond.

In contrast to the use of Wi-Fi in unlicensed spectrum, LTE-U offers several aspects that are attractive to operators, such as better spectrum efficiency thanks to more advanced radio features, simplified network management and tracking of Key Performance Indicators through a single Radio Access Network, improved network management and load balancing through tighter integration, etc.

The introduction and adoption of LTE-U brings a number of challenges to be addressed. Due to the use of unlicensed spectrum, LTE-U must support fair access of multiple LTE-U and Wi-Fi networks. This will require LTE-U to adapt to the presence of other LTE-U and Wi-Fi networks, while Wi-Fi uses its current mechanisms. Therefore, issues such as coexistence with Wi-Fi systems operating in unlicensed spectrum, unpredictable interference to LTE-U from other technologies and coexistence among cells of the same or different operators need to be resolved.

Although early discussions among players agreed that the core technology should be as much frequency agnostic as possible, a clear focus is placed on unlicensed operation in the 5 GHz band. Furthermore, given that typical traffic today is asymmetric, the first focus for LTE-U is on leveraging supplemental downlink capabilities over unlicensed spectrum. In this way, licensed band LTE provides reliable connection for mobility, signaling, voice and data in uplink and downlink, while LTE-U boosts data rates and capacity in downlink.

Under the above framework, the authors have proposed in [3] the exploitation of learning mechanisms to support the channel selection functionality, which is in charge to decide the most appropriate channel in the unlicensed band to set-up an LTE-U carrier. The proposal was motivated by previous investigations in connection to Cognitive Radio Networks [4]-[7], Wi-Fi networks [8] and, more in general, some evidences of the benefits that such artificial intelligence-based machine learning mechanisms can bring into the management and operation of wireless networks (see e.g., [9][10]).

The results obtained in [3] for a fully decentralized Q-learning mechanism that exploits prior experience are very promising. The evaluations presented in an indoor scenario with small cells belonging to different Mobile Network Operators (MNOs) revealed that the proposed approach is able to achieve a performance between 96% and 99% of the optimum ideal achievable throughput. Nevertheless, in order to build a fully consistent and reliable framework providing high confidence to MNOs to implement such advanced mechanisms in their evolved 4G and future 5G networks, a more comprehensive analysis is needed. In this respect, this paper intends to deepen into the intrinsic performance of the Q-learning solution under more challenging conditions. In particular, this paper analyses the robustness of the proposed approach against non-stationary conditions in the wireless environment (i.e., changes in the spectrum usage of sharing users) that tend to negatively affect the learning process.

The rest of the paper is organized as follows. Section II discusses the considered learning-based channel selection approach for LTE-U coexistence. Then, Section III presents the associated system model and the simulation-based framework for performance assessment. Section IV presents

This work is supported by the Spanish Research Council and FEDER funds under RAMSES grant (ref. TEC2013-41698-R).

the performance results under different conditions and section V summarizes the main conclusions.

II. LEARNING-BASED CHANNEL SELECTION IN LTE-U

Channel Selection (also denoted 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. Therefore, it can be used as a frequency-domain coexistence mechanism to safeguard that LTE-U is a "good neighbor" in unlicensed bands without requiring modifications in LTE PHY/MAC standards, e.g., just by enabling small cells to properly choose the cleanest channel based on received power measurements. If interference is found in the operating channel and there is another cleaner channel available, the transmission can be switched to the new channel using LTE Rel. 10/11 procedures [11]. This ensures that the interference is avoided between the small cell and its neighboring Wi-Fi devices and/or other LTE-U small cells, provided that there are clean frequencies available. Clearly, the design of a proper Channel Selection functionality can greatly improve the overall efficiency of the LTE-U operation and it should avoid, whenever possible, that multiple small cells have to share the same channel.

From an architectural point of view, different approaches for Channel Selection can be envisaged: (a) fully distributed case, where each small cell makes decisions on its own, (b) intra-operator coordination, where decisions for a given small cell take into consideration knowledge about other small cells' configurations belonging to the same operator, (c) inter-operator coordination, where also information about small cells from other operators in the area is available and (d) coordination also with managed Wi-Fis in the area. Notice that unmanaged legacy Wi-Fis unable to explicitly provide information about their configuration may also be present in the scenario. Clearly, higher coordination levels will ease the Channel Selection decision-making. However, higher coordination levels involve more demanding network coordination architectures, information exchange protocols and procedures, etc. In this respect, while the purpose of this work is not to deepen into suitable architectures and comparative trade-off analysis, this paper shifts the focus towards a fully decentralized approach and explores to what extent the inclusion of a smart Channel Selection logic can overcome the intrinsic disadvantages associated with the fact that no explicit knowledge about the other small cells and/or Wi-Fis operating in the area is available.

From a decision-making logic point of view, exploiting learning from past experience seems a pertinent principle in the LTE-U context [3]. Each small cell may autonomously learn what channels are usually not being used by its neighbors and then tend to select such free channels. Thanks to this learning capability, general scanning procedures over the 5 GHz band conducted systematically to look for the cleanest channel can be avoided or reduced to a minimum. Besides, learning from the own experience in using a channel can help in overcoming situations like the hidden node problem where a small cell can detect a channel as not used but some of its served terminals can experience severe interference conditions from other small cells and/or Wi-Fis. 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.

A. Formulation of the proposed Q-learning solution

With all the above, and following the prior work of [3], this paper considers the use of a Q-learning solution as an efficient means to carry out a distributed Channel Selection in a practical while at the same time efficient way. Q-learning is a type of Reinforcement Learning (RL) technique [12] where learning is achieved through the interaction with the environment, so that the learner discovers which actions yield the most reward by trying them. In this way, each small cell progressively learns and selects the channels that provide the best performance based on the previous experience.

In particular, in the proposed approach each small cell i stores a value function $Q(i,k)$ that measures the expected reward that can be achieved by using each channel k according to the past experience. Whenever a channel k has been used by the small cell i, $Q(i,k)$ is updated following a single state Q-learning approach with null discount rate given by [12]:

$$Q(i,k) \leftarrow (1-\alpha_L) Q(i,k) + \alpha_L \cdot r(i,k) \quad (1)$$

where $\alpha_L \in (0,1)$ is the learning rate and $r(i,k)$ is the reward that has been obtained as a result of the current use of the channel k. Assuming that the target of the channel selection is to find a channel that maximizes the throughput, the reward function considered in this paper is given by $r(i,k) = \overline{R(i,k)} / R_{\max}$, where $\overline{R(i,k)}$ is the average throughput that has been obtained by the i-th small cell in channel k as a result of the last selection of this channel. R_{max} is a normalization factor. At initialization, i.e., when channel k has never been used in the past by small cell i, $Q(i,k)$ is set to an arbitrary value Q_{ini}.

Based on the $Q(i,k)$ value functions, the proposed Channel Selection decision-making for the small cell i follows the softmax policy [12] in which channel k is chosen with probability:

$$\Pr(i,k) = \frac{e^{\frac{Q(i,k)}{\tau(i)}}}{\sum_{k'=1}^{K} e^{\frac{Q(i,k')}{\tau(i)}}} \quad (2)$$

where $\tau(i)$ is a positive parameter called *temperature*.

B. Challenges in non-stationary conditions

Most RL methods are designed to work assuming stationary environments [13]. Nevertheless, many practical environments are non-stationary as their dynamics might change due to some unknown or not directly perceivable causes. In the context of LTE-U operation, non-stationarity is an inherent characteristic: other LTE-U small cells or legacy Wi-Fi equipment deployed in the scenario may follow unknown channel selection patterns, newly installed small cells may appear in the scenario, etc. As a result, changing interference conditions on the air interface will be observed.

An interesting property of Q-learning is that it does not require a pre-specified model of the environment and,

therefore, it can be used in dynamic and non-stationary environments [14]. Certainly, non-stationary environments affect learning methods in a way that forces them to relearn the selection policy whenever changes in the scenario occur, since the policy which was calculated for a given situation might be no longer valid after a change. As a result, the time for relearning how to behave could lead to performance drops during the readjustment phase [13]. The robustness of the proposed Q-learning approach to deal with non-stationary situations as well as the achievable performance can be anticipated to be dependent on how often the changes occur and how long it takes to learn a new solution.

III. System Model and 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 [2]. 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. 1. Small cells SC1 to SC4 are owned by operator 1 (OP1), while SC5 to SC8 are owned by operator 2 (OP2). Small cells are deployed at height 6m while the antenna height of the mobile terminals is 1.5m. A total of 10 terminals (users) per operator are randomly distributed inside the building. Each user is associated to the SC of its own operator that provides the highest received power. The SC-to-terminal and SC-to-SC path loss and shadowing are computed using the ITU InH model in [15].

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.

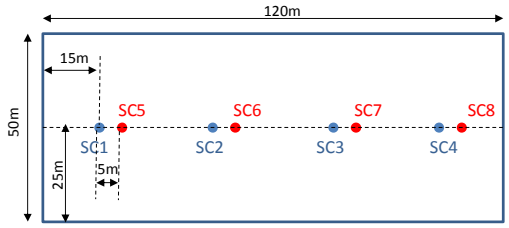


Fig. 1. Layout of the floor building

As required by the regulation of some markets like Europe and Japan for operating in unlicensed bands we assume the use of a Listen-Before-Talk (LBT) scheme that operates at milliseconds scale and regulates the transmissions of multiple SCs working in the same channel. For that purpose, following the strategy explained in [16] a SC using an LTE-U carrier will only transmit if it senses the channel as free during the Clear Channel Assessment (CCA) time (whose duration should be at least 18μs), meaning that the received power in this channel is below a given threshold TL. Then, transmission will be done during a maximum time of 10 ms followed by an idle period θ_{idle} of at least 5% of the transmission time, after which the CCA will be executed again. According to the formula in [16], TL is set to -70 dBm/MHz.

Under the above considerations, assuming that the Channel Selection functionality of the i-th SC has chosen the k-th channel for carrying out LTE-U downlink transmissions, the total aggregated throughput served by this cell is modelled for simulation purposes as:

$$R(i,k)=\sum_{n=1}^{N(i)}\frac{B}{N(i)}S\left(SINR_n(i,k)\right)\frac{1-\theta_{idle}}{M(i,k)} \qquad (3)$$

$N(i)$ is the total number of users being served by the i-th SC. $SINR_n(i,k)$ is the signal to noise and interference ratio observed by the n-th user when downlink data is transmitted on the k-th channel. It depends on the propagation conditions between the n-th user and the i-th SC and on the interference generated by other SCs that use the k-th channel and that, when they transmit, they are detected at the i-th cell below threshold TL (i.e., they are not sharing the channel in the time domain based on the LBT). θ_{idle}=0.05 is the fraction of time associated with the idle periods imposed by the LBT strategy (the CCA time is already included in these idle periods). $M(i,k)$ is the number of SCs that are sharing in the time domain the k-th channel with the i-th small cell following the LBT strategy (i.e., those that, when they transmit, they are received above threshold TL at the i-th SC). Finally, following the model presented in Section A.1 of [17], $S(SINR_n(i,k))$ is a function ranging between 0 and S_{max}=4.4 b/s/Hz that provides the spectral efficiency in b/s/Hz as a function of $SINR_n(i,k)$. With all the above considerations, the normalization factor of the reward is $R_{\max}=B\cdot S_{\max}\cdot(1-\theta_{idle})$.

The Q-learning algorithm is configured with Q_{ini}=0.5 and different values of the learning rate α_L will be analyzed. The temperature is set initially to τ_0=0.15 and the value is decreased following a logarithmic cooling approach [3] during the simulation.

Simulation time is measured relative to generic units denoted as "time steps". All the SCs are transmitting data to their users during the whole simulation.

It is assumed that the SCs of OP1 execute the Q-learning approach. To characterize the rate at which the Q-learning mechanism interacts with the environment, we consider that the time between consecutive Channel Selection decisions made by an SC of OP1 is modelled as a geometrical random variable with average T time steps. In turn, as a source of non-stationarity in the scenario, SCs of OP2 carry out random channel selections, which introduce the variability in the environment. The time between two consecutive Channel Selections made by an SC of OP2 is modelled as a geometrical random variable with average Δ, which characterizes the rate at which the environment changes from OP1's perspective.

IV. Results

A. Intrinsic behavior of the Q-learning algorithm

To illustrate the behavior of the algorithm for different configurations of α_L and T, let consider first the case with K=4 channels and a stationary situation where SC5 to SC8 are using always channels k=1, 2, 3, 4, respectively. Fig. 2a, b, c

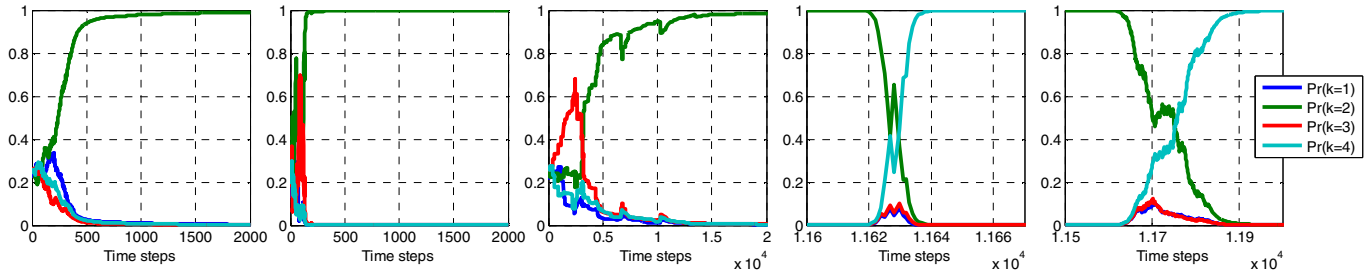
depict the evolution of the selection probabilities for the different channels in SC3. The configuration α_L=0.01 and T=1 time step is shown in Fig. 2a. It can be observed that, after an initial period when the different channels are selected with similar probabilities, SC3 learns to select channel k=2 with probability close to 1, which (together with the decisions made by SC1, SC2 and SC4 not shown here for the sake of brevity) can be proved to be an optimal decision. Fig. 2a reveals that the time needed to learn the solution is around 1000 time steps. Fig. 2b presents the case α_L=0.1 and T=1, meaning that the Q-learning update in (2) has less memory from the past. It can be observed how SC3 learns the same choice as in the previous case but the time needed by the learning process has been significantly reduced to around 150 time steps. Finally, Fig. 2c presents the case α_L=0.1 and T=100, representing that channel selections are done at a slower rate. The learning time has been increased up to around 1500 time steps. These results indicate that the time needed to learn a solution in a stationary situation where the SCs start from scratch and do not have any prior information of the environment exhibits proportionality to the ratio T/α_L (in the order of 10 to 15 times T/α_L).

Let consider now the case where the SCs of OP2 carry out random changes in the selected channel, so that the environment becomes non-stationary. Fig. 2d and Fig. 2e illustrate the evolution of the channel selection probabilities of SC3 in front of a channel change made by SC8 of OP2 which is located in the proximity of SC3. Fig. 2d corresponds to α_L=0.1 and T=1. Initially, SC3 has learnt to select channel k=2. After 11615 time steps, SC8 switches to this channel. This degrades the performance observed by SC3, so the selection probability of channel k=2 starts to decrease. After approximately 35 time steps, SC3 learns to select channel k=4 that offers a better performance. Fig. 2e depicts the same situation but with α_L=0.01 and T=1. The behavior is very similar but now the time to learn the new solution increases to around 350 time steps. Consequently, these results reveal that, when the SCs have already learnt a solution and they have to react in front of a change, the time to learn a new solution is in the order of 3.5 times the ratio T/α_L, which is about 4 times faster that when the solution needs to be learnt from scratch.

B. Performance analysis in varying environments

To assess the performance achieved by the Q-learning mechanism under non-stationary situations, let consider the case where the SCs of OP2 carry out random channel selections on average every Δ time steps. As a performance metric we consider the ratio between the total throughput achieved by the SCs of OP1 and the total throughput that would be achieved with an ideal optimum assignment (corresponding to the assignment of channels that maximizes the aggregated throughput of the SCs of OP1). Fig. 3 depicts this throughput indicator as a function of Δ, which reflects the variability of the environment or equivalently the period of time that the environment remains stationary (i.e., the lower the Δ the higher the non-stationarity). Results are presented for K=8 channels and for different combinations of α_L and T. Each point in the figure is the average of 50 experiments each one with a duration 1E6 time steps. It is observed that the achieved performance degrades when reducing Δ (i.e. when the environment varies more quickly). Fig. 3 also reflects that the performance achieved with T=1 and α_L=0.01 is very similar to that of T=10 and α_L=0.1 (i.e. the two combinations with T/α_L=100). The same occurs with the case T=10 and α_L=0.01 and the case T=100 and α_L=0.1 (i.e. the two combinations with T/α_L=1000). This reflects that the performance in front of a non-stationary environment mainly depends on the ratio T/α_L. As discussed in section IV.A, the time needed to learn solutions is proportional to this ratio. Therefore, as seen in Fig. 3 the performance when T/α_L=100 is better than when T/α_L=1000 because, whenever a change in the scenario occurs, SCs from OP1 are able to learn more rapidly a proper configuration. Besides, under more varying conditions (i.e., low Δ), the performance degrades more due to more frequent relearning processes. The performance drop is more noticeable when learning is performed more slowly (i.e., T/α_L=1000). Therefore, to ensure an adequate behavior of the learning, the time to learn new solutions (proportional to T/α_L) should be much smaller than the time Δ that the environment remains stationary.

The previous results indicate that $\Delta/(T/\alpha_L)=\Delta\cdot\alpha_L/T$ is a proper metric to characterize the achieved throughput performance of the proposed Q-learning, since it compares the level of non-stationarity in the scenario to the learning capabilities. Fig. 4 plots the achieved throughput for K=4, 8 and 12 channels as a function of this metric. The presented figure is the result of multiple simulations done for different combinations of α_L=0.01 to 0.1, T=1 to 1500 time steps, and Δ=10000 to 50000 time steps. The vertical bars included in the figure represent the maximum variation obtained between the simulations of different combinations of parameters having the same value of $\Delta\cdot\alpha_L/T$. The fact that small differences (2% at most) are observed for different combinations of (Δ, α_L, T) leading to the same value of $\Delta\cdot\alpha_L/T$ confirms that this is the



(a) α_L=0.01, T=1 (b) α_L=0.1, T=1 (c) α_L=0.1, T=100 (d) α_L=0.1, T=1 (e) α_L=0.01, T=1

Fig. 2. Evolution of the channel selection probabilities for SC3: (a),(b),(c) correspond to a stationary environment; (d),(e) correspond to two changes in a non-stationary environment

relevant metric to consider.

It is observed in Fig. 4 that, whenever $\Delta \cdot \alpha_L / T$ >50 (approximately) the non-stationarity is not causing significant degradation. As a reference engineering criterion, this bound means that the time needed to learn (i.e. around $10 \cdot T/\alpha_L$) should be in the order of at least 5 times shorter than the period of time that the environment remains stationary (i.e. Δ). Indeed, for large values of $\Delta \cdot \alpha_L / T$, the algorithm is able to provide quite good performance in relation to the optimum ideal case (i.e. 98% of the optimum for K=12 channels, 96% for K=8 channels and 86% for K=4 channels). Better performance can be achieved with larger K because the sharing of the same channel between more than one SC can be avoided.

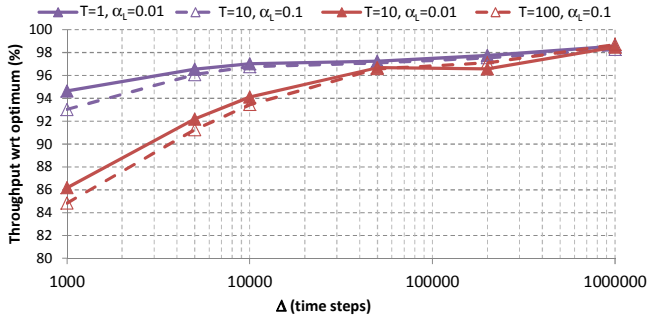


Fig. 3. Achieved throughput with respect to the optimum as a function of the variability of the environment for the case K=8 channels.

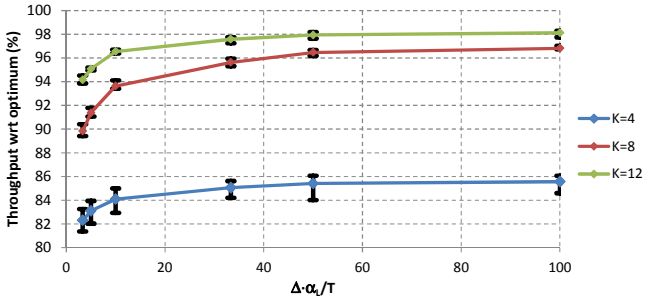


Fig. 4. Achieved throughput with respect to the optimum one for different configurations of α_L, T, Δ with K=4, 8 and 12 channels.

V. Conclusions and Future Work

The use of LTE-U in the unlicensed 5 GHz band is a promising enhancement to meet the requirements foreseen for future systems. Coexistence between different systems operating in the same band is one of the key technical challenges to be resolved for a successful operation of LTE-U deployments. In this framework, this paper has addressed the Channel Selection functionality that decides the most appropriate channel to set-up an LTE-U carrier for supplemental downlink. In particular, the use of a distributed Q-learning mechanism has been considered, analyzing its robustness to operate in non-stationary environments.

The main conclusions that can be extracted from the presented analysis are: (i) The time needed by the learning process is in the order of 10 to 15 times the average time between two consecutive channel selection decisions divided by the learning rate. (ii) This factor can be reduced from 10-15 to 3-4 if the learning process does not start from scratch but uses previously learnt information to react in front of a change in the environment, (iii) The achievable LTE-U throughput performance mostly depends on the metric that relates the period of time that the environment remains stationary to the time needed for the learning process, (iv) Whenever this metric is above 50, the learning process is fast enough compared to the non-stationarity of the environment so the algorithm can achieve a throughput performance quite close to the optimum.

As part of future work, different intra and inter-operator coordination levels can be studied both in terms of architectural implications and Channel Selection strategy design in order to attain pros and cons versus the presented fully distributed solution. Finally, the combination of the proposed approach with other possible techniques based on Game Theory can also be developed.

References

[1] 3GPP workshop on LTE in unlicensed spectrum, Sophia Antipolis, France, June 13, 2014. http://www.3gpp.org/ftp/workshop/2014-06-13_LTE-U/

[2] 3GPP TR 36.889, "Study on Licensed-Assisted Access to Unlicensed Spectrum (Release 13)", November, 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] I. Macaluso, D. Finn, B. Ozgul, L. A. DaSilva, "Complexity of Spectrum Activity and Benefits of Reinforcement Learning in Dynamic Channel Selection", IEEE Journal on Selected Areas in Communications, Vol. 31, No. 11, November, 2013.

[5] J. Pérez-Romero, O. Sallent, R. Agustí, "Enhancing Cellular Coverage through Opportunistic Networks with Learning Mechanisms", GLOBECOM, December, 2013

[6] S. Chen, R. Vuyyuru, O. Altintas, A.M. Wyglinski, "Learning-based channel selection of VDSA Networks in Shared TV Whitespace", VTC Fall Conference, Quebec, Canada, September, 2012.

[7] Y. Li, H. Ji, X. Li, V.C.M. Leung, "Dynamic channel selection with reinforcement learning in cognitive WLAN over fiber", International Journal of Communication Systems, March, 2012.

[8] S. Chieochan, E. Hossain, J. Diamond, "Channel Assignment Schemes for Infrastructure-Based 802.11 WLANs: A Survey", IEEE Communications Surveys and Tutorials, Vol. 12, No. 1, 2010.

[9] 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.

[10] 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.

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

[12] R.S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, MIT Press, 1998.

[13] E.W. Basso, P.M. Engel, "Reinforcement learning in non-stationary continuous time and space scenarios", Anais do VII Brazilian Meeting on Artificial Intelligence, ENIA, Brazil, SBC Press, 2009.

[14] M. Abdoos, N. Mozayani, A.L.C. Bazzan, "Traffic Light Control in Non-stationary Environments based on Multi Agent Q-learning", 14th International IEEE Conference on Intelligent Transportation Systems, Washington DC, USA, October, 2011.

[15] 3GPP TR 36.814 v9.0.0 "Further advancements for E-UTRA physical layer aspects", March, 2010.

[16] 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.

[17] 3GPP TR 36.942 v12.0.0, "Radio Frequency (RF) system scenarios", September, 2014.