On the Implementation of Channel Selection for LTE in Unlicensed Bands using Q-learning and Game Theory algorithms

A. Castañé, J. Pérez-Romero, O. Sallent
Department of Signal Theory and Communications,
Universitat Politècnica de Catalunya (UPC),
Barcelona, Spain
E-mails: annacastane.92@gmail.com, jorperez@tsc.upc.edu, sallent@tsc.upc.edu

Abstract—**The use of Long Term Evolution (LTE) in the unlicensed 5 GHz band, referred to as LTE-U or Licensed Assisted Access (LAA), is a promising enhancement to increase the capacity of LTE networks and meet the requirements of future systems. This paper analyses the use of fully distributed channel selection mechanisms for facilitating the coexistence among different LTE-U and/or Wi-Fi systems operating in the same band. Specifically, the paper focuses on a Q-learning and a Game Theory based approach. The implementation considerations of both approaches are discussed in relation to current 3GPP specifications and a comparison in terms of performance is presented to analyze the convergence time, the signaling requirements and the impact of errors in the throughput estimation.**

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

I. INTRODUCTION

Long Term Evolution - Unlicensed (LTE-U), which has been standardized in the context of 3rd Generation Partnership Project (3GPP) through the Licensed Assisted Access (LAA) feature of Release 13, is a promising enhancement that enables Long Term Evolution (LTE) to operate and coexist with other technologies in unlicensed bands [1][2][3], with a clear focus on the 5 GHz band. Although licensed spectrum remains cellular 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 Fourth Generation (4G) and beyond systems. 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 using LTE in the licensed band provides reliable connection for mobility, signaling, voice and data in both uplink and downlink.

Compared to the usage of Wi-Fi in unlicensed spectrum, LTE-U offers several features that are attractive to operators. Nevertheless, the introduction and adoption of LTE-U brings a number of challenges to be addressed as LTE-U must support fair access of multiple LTE-U and Wi-Fi networks. So as to allow that multiple LTE-U small cells and Wi-Fi access points share the same operating channel different channel access mechanisms can be used, such as Listen Before Talk (LBT).

Channel selection functionality is the mechanism used to decide the operating channel where a LTE small cell sets up an unlicensed carrier. This mechanism is fundamental in order to facilitate the coexistence of LTE with other systems sharing the same unlicensed band, such as other LTE small cells or Wi-Fi access points.

As discussed in [4] and [5], a fully distributed channel selection 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 has been modeled in terms of the learning parameters performance in prior works using Q-learning [4][5] and Game Theory [6] principles. Taking these prior works as starting point, this paper focuses on the implementation considerations regarding these two approaches. For this purpose, the paper discusses the practical implementation of both algorithms based on current 3GPP specifications and presents an implementation-based performance comparison between the two techniques in terms of the convergence time and the signaling requirements associated to the required channel selections. In addition, in Game Theory, the impact of errors in the estimation of the throughput is also assessed.

The rest of the paper is organized as follows. Section II discusses the operation of the channel selection process in LTE-U in relation to the involved associated signaling procedures. It also summarizes the operation of both the Q-learning and Game Theory algorithms and discusses the implementation considerations. Section III presents some implementation-based performance results to compare both approaches, while section IV summarizes the main conclusions.

II. LEARNING-BASED CHANNEL SELECTION IN LTE-U

A. Channel selection in LTE-U

In the context of LTE, the use of unlicensed bands has been standardized in the LAA feature by combining the use of

licensed and unlicensed spectrum using carrier aggregation technology. For a User Equipment (UE) in connected mode, the configured set of serving cells includes a Primary Cell (PCell) that uses a licensed carrier and at least one Secondary Cell (SCell) operating in the unlicensed spectrum [7].

The channel selection (also denoted as carrier selection) refers to the mechanism used to decide the operating channel (i.e., center frequency and associated bandwidth) of the unlicensed carrier(s) in a cell. Whenever the channel selection functionality decides a channel change or the activation of a new channel, UEs and SCells need to be properly configured. This includes an UE Radio Resource Control (RRC) Connection Reconfiguration procedure used to add a SCell with the selected channel to the set of component carriers of each UE. Furthermore, in both configurations a Medium Access Control (MAC) activation/deactivation process is needed for each UE in order to activate the configured SCell and be able to receive both the Physical Downlink Control Channel (PDCCH) and the Physical Downlink Shared Channel (PDSCH).

The overall process involves the exchange of six signaling messages (Table I) between the eNodeB and each UE. Considering [8], the RRC time needed in order to configure a SCell is 31.7 ms per UE and the time needed in order to activate a SCell is 8 ms per UE.

TABLE I. SIGNALING MESSAGES [8][9]

Signaling messages
RRC ConnectionReconfiguration
HARQ (Hybrid Automatic Repeat reQuest) ACK
Scheduling Request
DCI (Downlink Control Information) 0
RRC ConnectionReconfiguration Complete
MAC Control Element (CE) subheader

B. *Q-learning algorithm and implementation*

Q-learning is a type of Reinforcement Learning (RL) technique [10] 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 (SC) progressively learns and selects the channels that provide the best performance based on the previous experience. In the considered algorithm, described in details in [4][5], 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:

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

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. The reward is given by the average normalized throughput that has been obtained by the small cell in the channel. Based on the $Q(i,k)$ value functions, the channel selection decision-making for the small cell i follows the softmax policy in which channel k is chosen with probability:

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

where $\tau(i)$ is a positive parameter called *temperature*. Therefore, channels with high $Q(i,k)$ values are selected with higher probability.

In order to implement this learning technique, there is the need to gather some information to perform the throughput computation (i.e. the reward). The throughput is computed by dividing the number of successfully transmitted bits over a certain observed period when there is data to transmit in the Packet Data Convergence Protocol (PDCP) buffer when the SCell is activated. The number of acknowledged data bits can be easily counted from the HARQ buffers at the MAC layer. Instead, notice that counting the data bits in the IP level or PDCP Service Data Unit (SDU) level would imply a complex implementation, because at these levels the system does not know if a packet will be sent using a licensed or an unlicensed carrier.

C. *Game Theory algorithm and implementation*

In this case the channel selection problem is modelled as a game in which each small cell is a player and the actions made by each player are the selected channels. Specifically, this paper considers the Iterative Trial and Error Learning - Best Action (ITEL-BA) algorithm described in [6] which was proved to converge to a Nash Equilibrium (NE) [11].

In ITEL-BA, each SC retains a benchmark action (i.e. a benchmark channel to select) and the corresponding benchmark reward as a reference to evolve the action selection strategy. At a certain time, a channel 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 can be *content*, *discontent*, *hopeful* or *watchful*. The general idea is that a *content* player will be selecting the benchmark action most of the time, and will occasionally experiment with new actions according to a probability $\varepsilon \ll 1$ called exploration rate. Instead, a *discontent* player will try out new actions frequently, eventually becoming *content*. The *hopeful* and *watchful* moods correspond to transitional situations, triggered by changes in the behavior of other players (or in the environment), and they will facilitate updates in the values of the benchmark action and reward to cope with these changes. The reader is referred to [6] for a detailed specification of the ITEL-BA algorithm.

Like in the Q-learning approach, the considered reward is the obtained normalized throughput when using a channel. However, an essential differential aspect with respect to Q-learning approach is that the channel selection of a SC in ITEL-BA is based on both the actual reward obtained with the current channel of this SC (which can be easily measured through the same procedure described for Q-learning) and the hypothetical reward that would be obtained if using a different channel (which needs to be estimated in some way or another).

Assuming that the SC i is using channel k, the hypothetical reward (i.e. throughput) that it would get in another channel $k' \neq k$ can be estimated as follows:

$$\hat{r}(i,k') = f\left(SINR(i,k')\right)\frac{1}{M(k')} \quad (3)$$

where $f(SINR(i,k'))$ is a function that maps the Signal to Interference and Noise Ratio (SINR) experienced by the UEs of SC i in the channel k' with the throughput and $M(k')$ is the number of SCs that would be sharing the channel k' in the time domain by means of the LBT strategy (e.g. if there are 2 SCs sharing the same channel this means that each one would get one half of the throughput).

The estimation of $SINR(i,k')$ in (3) can be done as follows:

$$SINR(i,k') = \frac{P(i,k')}{RSSI(i,k')} \approx \frac{P(i,k)}{RSSI(i,k')} \quad (4)$$

where $P(i,k')$ is the useful received signal by the UEs of cell i if it was using channel k' and $RSSI(i,k')$ is the Received Signal Strength Indicator that measures the total received power by the UEs in channel k'. The value of $P(i,k')$, can be estimated as the useful received signal by the UE from the SC i at the current frequency k, $P(i,k)$, assuming that the link between the UE and the SC experiences similar path loss and shadowing conditions in both channels k and k'. The value of $P(i,k)$ can be obtained from the Reference Signal Received Power (RSRP) measurements made in channel k. Both the RSRP and the RSSI are measured by the UEs and reported to the SC i. For this purpose, the RRC protocol is used to configure the measurement reports of the UEs [12].

In turn, the function $f(\cdot)$ in (3) to map the estimated SINR with the throughput is implementation-dependent. One possible example for this function could be the one provided in Section A.1 of [13].

Finally, the number of SCs sharing the channel k', $M(k')$, can be estimated by the SC i using the RRC measurement reports provided by the UEs, which indicate the Physical Cell Identity (PCI) that UEs detect of each measured cell in channel k'.

An alternative option, in order to avoid measurement reports to gather the last parameters, is to use a Downlink (DL) receiver within the SC [14] running a utility that can configure RF card into sniff mode and PHY into Network Monitor Mode (NMM) state in order to listen to the downlink signals to measure signal level and detect presence of other cells in the vicinity. This option would allow a better estimation of $M(k')$ than the approach based on RRC measurement reports because it provides the measurements done at the SC where the LBT procedure is executed. However, the main drawback is that the SC is not able to schedule any users during the time period where NMM measurement is performed.

Based on all the above considerations, the estimation of the hypothetical reward for applying the ITEL-BA algorithm according to (3) is feasible in practice although it may be subject to estimation errors.

III. Performance Evaluation

A. Scenario

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 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 UE per operator are randomly distributed inside the building.

The 5 GHz unlicensed band is considered, organized in K channels of bandwidth B=20 MHz, numbered as k=1,..,K. The Q-learning algorithm is configured after the analysis performed in depth in [6] with α_L=0.1, temperature $\tau(i)$ adjusted based on a logarithmic cooling function with initial temperature τ_0=0.15 and the initial value of the $Q(i,k)$ for channels that have not been used is set to Q_{ini}=0.5. The rest of simulation parameters are taken from [6].

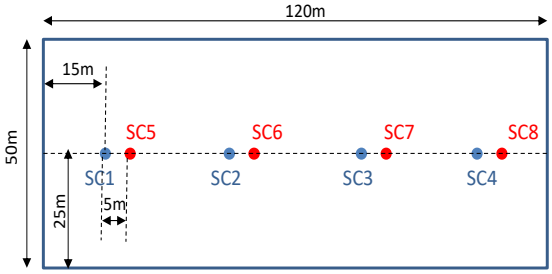

Fig. 1 Layout of the floor building

Simulation time is measured relative to generic units denoted as "time steps". It is assumed that, in each time step, every SC executes the channel selection algorithm (either Q-learning or ITEL-BA) to decide if the channel currently used by the SC has to be changed.

B. Convergence time

In the following we analyze the time needed by ITEL-BA and Q-learning to reach convergence. For the Q-learning, it is assumed that convergence occurs when every SC i in the scenario has identified a channel k^* with selection probability $Pr(i,k^*) \geq 99\%$. For ITEL-BA, convergence occurs when the system has reached the situation in which all the SCs are in *content* state and none of the SCs can improve its throughput by unilaterally changing the channel they are using (i.e. their benchmark action), meaning that a NE has been reached.

Fig. 2 plots the average convergence time for both techniques for the cases K=4 and K=8 channels. For each case, the presented results are the average of 10^5 random realizations.

It is observed in Fig. 2 that Q-learning requires a higher number of time steps for converging than ITEL-BA. The differences are larger for the case K=4, in which the convergence time of Q-learning is about 10 times higher than that of ITEL-BA. Instead, for the case K=8, the average

convergence time of Q-learning is approximately twice than that of ITEL-BA. Indeed, while Q-learning converges faster for higher number of channels K, the average convergence time for ITEL-BA slightly increases with the number of channels K. The rationality behind this result is that the overall solution search space contains in proportion a lower number of NE with K=8 than with K=4. This means that the system needs to explore and discard more combinations before reaching an NE with K=8, so the convergence time increases [6]. In contrast, with Q-learning the average convergence time decreases when passing from K=4 to K=8. This behavior is due to the probability function definition (2), which converges faster with high K because of the denominator increment.

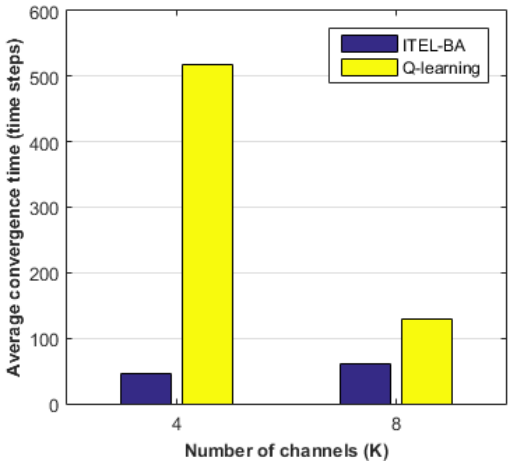

Fig. 2 Average convergence time comparison

To further assess the behavior of both algorithms, Fig. 3 depicts the average number of channel changes that have been performed by each SC before the system has reached convergence after performing 50 experiments corresponding to different spatial user distributions. It is observed that Q-learning tends to decrease the number of channels selections when K increases, whereas in ITEL-BA there are only very small variations, reflecting a similar behavior like in Fig. 2.

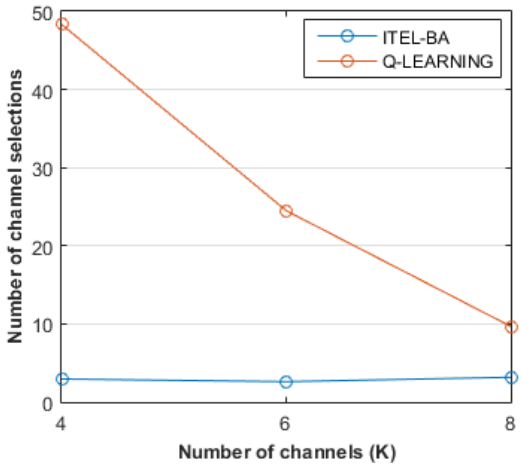

Fig. 3. Average number of channel selections performed by each SC before convergence

The number of channel selections shown in Fig. 3 can be directly translated into the signaling requirements for implementing each channel selection technique. Specifically, as discussed in Section II, each channel selection change involves a total of 6 signaling messages per UE (see Table I). Therefore, looking at Fig. 3 it can be concluded that for the case K=8, if Q-learning is used, the 10 channel selections that are required on average per UE will require a total of 60 signaling messages per UE. Instead, ITEL-BA will require approximately 24 signaling messages per UE.

Concerning the behavior of the algorithms after convergence is reached, in ITEL-BA this behavior is tightly related with the type of NE in which the system has converged to. Specifically, ITEL-BA can converge to either a strict or non-strict NE. A strict NE means that, for each SC, the channel associated to the benchmark action provides strictly the highest reward among all other options. Instead, with a non-strict NE, there may be, for some SCs, other channels that provide exactly the same reward than the benchmark action. When convergence to a non-strict NE occurs, an SC in *content* state can continue making additional channel changes as a result of the exploration stage. In particular, simulations have revealed that around 6 additional channel changes are performed after convergence in a period of 1000 time steps for K=8 when ITEL-BA converges to non-strict NE. On the contrary, this effect does not occur when ITEL-BA converges to a strict NE, so no additional channel changes are performed in this case.

As for the Q-learning case, although the probabilistic behavior of (2) might lead to some very sporadic channel changes after reaching convergence, the situations analyzed in the simulations of this paper have revealed that this effect is negligible.

C. *Study of the impact of errors in the throughput estimation in the Game Theory algorithm*

As discussed in Section II.C, the implementation of the ITEL-BA algorithm requires the estimation of the hypothetical reward (i.e. throughput) that a SC would obtain in all the other channels that this SC is not using. Since this estimation may be subject to errors, this sub-section analyzes the sensitivity of the ITEL-BA algorithm in front of these errors. For this purpose, an estimation error in the hypothetical reward is introduced in our simulation. It is modeled following a uniform distribution between $[-E_{max},E_{max}]$, where E_{max}(%) is the maximum relative error. So, the convergence time obtained in this chapter will be obtained simulating that the SC may take bad decisions due to the estimation errors which lead to wrong rewards in channels not being used.

Fig. 4 illustrates the average convergence time of ITEL-BA as a function of the maximum relative error E_{max} (%) for the cases K=4 and K=8 channels. The presented results are the average of 50000 random realizations. For errors lower than 20% only small differences in the average convergence time with respect to the case without errors (i.e. E_{max}=0%) are observed. Instead, when the maximum estimation error E_{max} is higher than 30%, which may be the case in practice due to the difficulties in the estimation procedure as described in Section II.C, the average convergence time suffers a substantial degradation for both K=4 and K=8, due to the high randomness in the estimated reward that leads to erratic

decisions made by the different SCs. However, when comparing the results of ITEL-BA in Fig. 4 with those of Q-learning shown in Fig. 2, it is worth mentioning that, even for the highest value of E_{max} (%) considered in the simulations, the convergence time of ITEL-BA is still better than that of Q-learning, in spite of the fact that Q-learning is not affected by estimation errors because it does not require any estimation of hypothetical rewards.

Fig. 5 depicts the impact of the estimation error in terms of the average number of channel selections made by each SC in ITEL-BA before the system has reached convergence for K=4. It is observed that the number of channel starts to increase for large values of E_{max}. However, when comparing the result with that of Q-learning shown in Fig. 3 it is observed that ITEL-BA still presents a lower number of channel selections even in the presence of estimation errors. The effects of non-strict NE leading to additional channel changes are also observed when errors in the estimation exist.

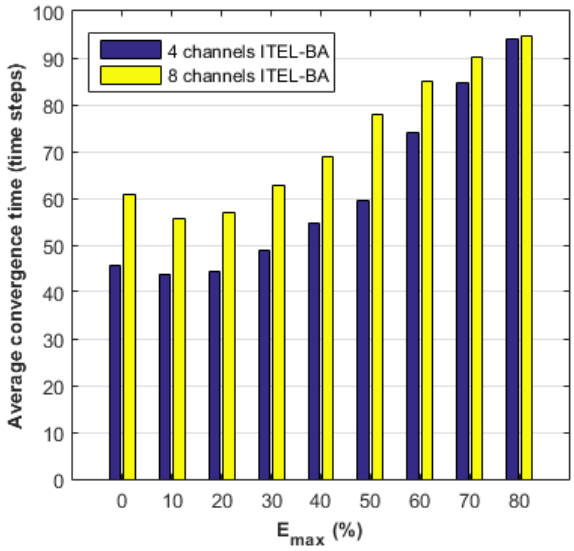

Fig. 4 Average convergence time of ITEL-BA in the presence of estimation errors

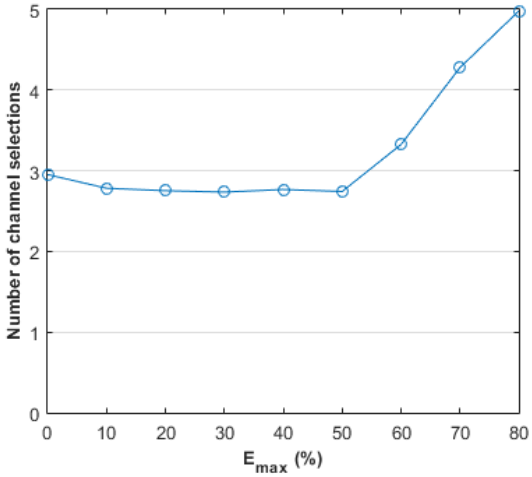

Fig. 5. Average number of channel selections per SC with K=4 as a function of the maximum estimation error.

It is clear, that in general Game Theory has better performance than Q-learning when K is small. However, its implementation is a bit more complex. Therefore, there is a trade-off between performance and implementation depending on the number of available channels K in the scenario.

IV. Conclusions

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. Two different approaches for channel selection have been considered, one based on Q-learning and the other one based on Game Theory. The paper has discussed the implementation considerations in the context of 3GPP specifications and has presented a performance comparison between the two.

Results have shown that the Game Theory based approach has overall a better performance in terms of average convergence time. Differences are more significant when the number of channels is reduced. For example, for K=4 channels, the convergence time with the Game Theory approach is about 10 times lower than with the Q-learning approach. Furthermore, despite the fact that the Game Theory based approach can still perform some unnecessary channel changes due to the convergence to non-strict NE in some cases, it has been shown that the faster convergence of Game Theory approach also leads to a reduction in the required signaling for carrying out the channel selections.

The paper has also analyzed the sensitivity of the Game Theory solution to errors in the estimation of the hypothetical throughput that a small cell would obtain in an unlicensed channel that is not being used by this small cell. It has been found that the convergence time can almost double in the considered simulations, although the convergence in the presence of errors is still faster than that of the Q-learning approach. Nevertheless, the simple implementation of Q-learning may prevail as a choice criterion, particularly for large number of channels K, where the difference in convergence time between the Game Theory and Q-learning shrinks.

Acknowledgment

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

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] Qualcomm, “LTE in Unlicensed Spectrum: Harmonious Coexistence with Wi-Fi”, June 2014.

[4] O. Sallent, J. Pérez-Romero, R. Ferrús and R. Agustí, "Learning-based coexistence for LTE operation in unlicensed bands," In International Conference on Communication Workshop, IEEE ICCW 2015, 8-12 June 2015, London, United Kingdom. pp. 2307-2313. doi: 10.1109/ICCW.2015.7247525.

[5] J. Perez-Romero, O. Sallent, R. Ferrus and R. Agusti, "A Robustness Analysis of Learning-Based Coexistence Mechanisms for LTE-U Operation in Non-Stationary Conditions," In 82nd Vehicular Technology Conference, IEEE VTC 2015-Fall, 6-9 September 2015, Boston, MA. pp. 1-5. doi: 10.1109/VTCFall.2015.7390815.

[6] J. Pérez-Romero, O. Sallent, H. Ahmadi, I. Macaluso. "On Modeling Channel Selection in LTE-U as a Repeated Game". In Wireless Communications and Networking Conference, IEEE WCNC 2016, 3-6 April 2016, Doha, Qatar. pp. 1-6. doi:10.1109/WCNC.2016.7564743.

[7] 3GPP TS 36.300: " Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN)", January, 2016.

[8] S. Mohan, R. Kapoor, and B. Mohanty, "Latency in HSPA Data Networks," Qualcomm, Tech. Rep., 2011.

[9] Sharetechnote. LTE Advanced Carrier Aggregation. [Online] Available: http://www.sharetechnote.com/html/Lte_Advanced_CarrierAggregation.html. [Accessed: November 2016].

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

[11] 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.

[12] 3GPP TS 36.331, "Radio Resource Control (RRC) Protocol specification (Release13)", January, 2016.

[13] 3GPP TR 36.942 v12.0.0, "Radio Frequency (RF) system scenarios", September, 2014.

[14] 3GPP TR 36.922 v13.0.0, "TDD Home eNode B (HeNB) Radio Frequency (RF) requirements analysis", January, 2016.