

GTDM-CSAT: an LTE-U self Coexistence Solution based on Game Theory and Reinforcement Learning

Pedro M. de Santana, José M. de C. Neto, Fuad M. Abinader Jr. and Vicente A. de Sousa Jr.

Abstract—There is substantial literature covering both problems and solutions related to the operation of Long Term Evolution (LTE) networks in unlicensed spectrum (LTE-U) while in coexistence with other technologies, such as Wi-Fi. However, a seldom explored scenario is the coexistence between multiple LTE-U networks. Within this scenario, a big issue is establishing optimal configurations that take into account fairness among different operators coexisting in the same unlicensed spectrum coverage area. Solutions to this problem could, for instance, react to changes in the environment by "tuning" different system configurations. We propose a game theoretical reinforcement learning algorithm, called GTDM-CSAT, aiming to maximize the LTE-U aggregated throughput while keeping channel access fairness among different access points. GTDM-CSAT uses the relative data rate offered by the system to adapt the LTE-U ON-OFF time. For this, we formulate the problem as a Markovian game, where the LTE-U operators coexist on a two-zero-sum game. The solution for the best ON-OFF time ratio is defined by applying a modified Minimax Q-learning algorithm for finding the game equilibrium. We perform simulations following 3GPP specifications using the ns-3 simulator for evaluating GTDM-CSAT under different traffic load scenarios. Results indicate gains in the system aggregated throughput by improving the individual data rate of each operator.

Index Terms—Markovian Games, LTE-U, Minimax Q-learning.

I. INTRODUCTION

The demand for solutions enabling the coexistence between different broadband wireless access systems in the same unlicensed spectrum bands has crystallized into multiple existing standardized solutions. For instance, 3rd Generation Partnership Project (3GPP) Release 13 introduced the Licensed-Assisted Access (LAA), enabling LTE to coexist with legacy systems in 5 GHz ISM bands by means of a Wi-Fi-like channel access mechanism named Listen Before Talking (LTE-LBT) [1]. Also, LTE-U (LTE-Unlicensed) is an industry standard [2], [3] based on the continuous interleaved switch of LTE ON-OFF operation within a duty-cycle period. There are also other LTE/Wi-Fi coexistence standards, such as 3GPP's LTE Wi-Fi Aggregation (LTE-LWA) [4] and MuLTEFire [5]. As government initiatives, LTE over Citizen

Broadband Radio Service (CBRS) [6] allows operation of LTE access networks over the 3.5 GHz spectrum band in the United States (US), while the so-called 'whitespaces spectrum' [7] is being regulated in some countries for utilization by unlicensed devices within a determined geographical region.

Most of recent research related to LTE-U/Wi-Fi coexistence [8]–[11] is based on Wi-Fi Alliance's draft on test procedures for Wi-Fi/LTE-U coexistence [12]. These works provide different means for controlling the ON-OFF ratio time of the duty-cycle of LTE-U, aiming to evaluate the achieved LTE-U/Wi-Fi combined throughput. A common assumption in these studies is the constant interference profile, an unrealistic condition given that offered traffic variation is not only a common situation in contemporary systems, but is determinant for the performance of any coexistence strategy. There are solutions that take such dynamic traffic conditions under consideration, such as Carrier-Sensing Adaptive Transmission (CSAT) [3], a Qualcomm solution for LTE-U that senses the medium for a long period for coexisting network activity, and adapts LTE transmission duration proportionally. In our previous works [13]–[15], we present and discuss the problems and solutions related to the operation of LTE networks in the unlicensed spectrum, especially the coexistence with Wi-Fi. However, there is scarce literature on the coexistence between multiple LTE-U networks, an expected scenario given the spectrum scarcity faced by some operators, and one which we believe needs to be properly addressed.

In this work we propose a novel strategy for the LTE-U/LTE-U coexistence, called Game Theory Decision Making for Carrier-Sensing Adaptive Transmission (GTDM-CSAT). GTDM-CSAT adaptation to varying traffic conditions is applied over the duty-cycle period, varying when each coexisting LTE network transmits or shuts down for a defined period. This problem is formulated by a Markovian Game where these two operators are in coexistence, and a two-zero-sum-game model is taken into account. The main goal is to find an optimal ON-OFF time ratio leading the modeled game to an equilibrium, using the aggregated LTE-U/Wi-Fi throughput as the target parameter.

This paper is organized as follows. Section II reviews the literature on Wi-Fi/LTE coexistence domain, as well as works applying game theory to this problem. We then introduce the concept of stochastic games and define the Minimax Q-Learning method on Section III, while the LTE-U self-coexistence problem is presented as a stochastic game in Section IV. The proposed GTDM-CSAT algorithm, as well as its input parameters, is presented in Section V, with the system modeling described in Section VI and the evaluation

Pedro M. de Santana is with the Samsung R&D Institute (SIDIA), Manaus, AM, Brazil (e-mail: p.santana@samsung.com). Fuad M. Abinader Jr. is currently with Nokia Bell Labs (e-mail: fuad.abinader@nokia-bell-labs.com), Paris, France.

José M. de C. Neto and Vicente A. Sousa Jr. are with the Federal University of Rio Grande do Norte, Natal, RN, Brazil (e-mails: {martinsee, vicente.sousa}@edu.ufrn.br).

The proof of concept simulations provided by this paper was supported by High Performance Computing Center at UFRN (NPAD/UFRN). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Digital Object Identifier: 10.14209/jcis.2019.17

scenario presented in Section VII, where some simulations evidence the need for adaptive adjust of the OFF time of LTE-U. Finally, GTDM-CSAT evaluation results are presented and discussed in the Section VIII, followed by final remarks and future investigations in the Section IX.

II. RELATED WORKS

This section reviews literature related to optimization of parameters for coexistence in the unlicensed spectrum, highlighting the contributions from our proposed solution.

A. LTE/WiFi Coexistence performance

Wi-Fi/LTE-LAA coexistence was modeled and analyzed using Markov chain in [16], specifically addressing the time domain interference profile. In [17], the authors propose an analytical model for LTE-LAA and its coexistence related features, as well as valuable recommendations about LTE-LAA and Wi-Fi performance analysis. Technical details on LTE-LAA physical layer, channel design and multicarrier operation in unlicensed spectrum are presented in [18].

In [19], using a Monte Carlo approach, the authors establish a novel metric for numerically demonstrating the Wi-Fi/LTE-U medium access sharing fairness, based on Wi-Fi performance loss ratio and LTE-U duty-cycle rate. In [20], the authors propose an analytical model to compute the probability of collision faced by Wi-Fi, and estimate its downlink throughput performance while coexisting with LTE-U. They also run ns-3 network simulations to validate their model accuracy estimation.

In [21], the authors address the LTE-U/Wi-Fi coexistence problem as an optimization of the LTE-U channel access probability. They propose a new proportional fair allocation scheme, based on equal channel times to every competing entity including idle periods, successful transmissions and collisions for the WiFi network. In [22], the authors use stochastic geometry to model and analyze the coexistence performance of Wi-Fi/LTE-U in a multi-RAT network. The coverage probability and spatial throughput of Wi-Fi and LTE-U are derived, as well as a simple expression for the Wi-Fi retention probability, revealing how the density of Wi-Fi/LTE-U affect the retention probability of Wi-Fi.

Practical measurements of LTE-U and Wi-Fi coexistence in different indoor scenarios are presented in [23], showing the LTE-U impact over Wi-Fi throughput and latency performance considering a full-buffer traffic scenario with different LTE-U duty-cycles. Results showed that, depending on the duty-cycle value, LTE-U disproportionately damages Wi-Fi throughput and degrades real-time Wi-Fi communications in a way that would likely be irritating to Wi-Fi users. The authors also highlight the need for an open and collaborative R&D for finding a more fair LTE-U approach just like 3GPP LAA solution.

Performance evaluation for LTE-LAA/Wi-Fi coexistence in dense deployment scenarios is presented in [24]–[28]. The results for low/moderate traffic loads evidence that LTE-LAA is a better neighbor for Wi-Fi even when compared to the scenario with two coexisting Wi-Fi systems. In one hand, Wi-Fi experiences meaningful performance degradation when

coexists with LTE-LAA in dense deployment and heavy traffic load scenarios. However, comparing the coexistence of identical technology in [28], the LTE-LAA/LTE-LAA combination is more robust to interference than Wi-Fi/Wi-Fi one.

An adaptive LTE-LAA scheme presented in [29] claims a harmonious coexistence and fair spectrum sharing among LTE-LAA and Wi-Fi. This scheme uses a variable LTE transmission opportunity (TXOP) followed by a variable muting period, similar to CSAT. This way, besides LBT feature, co-located Wi-Fi networks can also exploit the muting period (like LTE-U) to gain access to the wireless medium. The results also show a strong dependency of parameters configuration, but the proposed scheme can significantly improve the coexistence among LTE-LBT and Wi-Fi in a fair manner. Once again, the authors of [27] affirm that Listen-Before-Talking mechanism itself is not enough to provide a fair coexistence, and it is necessary a constant parameter adjustment, suggesting that artificial intelligence techniques could be analyzed and deployed.

B. Game Theory

Several works have applied game theory to solving telecommunication problems related to resource allocation in standalone and coexistence scenarios.

The authors in [30] surveyed game theory-based solutions for LTE-A radio resource allocation, discussing the fair spectral sharing. They propose to fill some gaps in the literature by presenting detailed techniques that can improve efficiency in the spectrum utilization. In [31], the authors propose a two-level scheme for resource allocation, based on cooperative game theory. In the first level, finite resources are allocated in classes, and a Nucleolus game approach is proposed for seeking fairness for an inter-class distribution. By the second level, the authors suggest a proportional fair algorithm aiming an intra-class resource distribution.

When it comes Wi-Fi spectrum sharing, the work presented in [32] proposes a game theory model based on interactions among Wi-Fi devices and channel usage. Their objective is to improve efficiency on the channel access control for coexisting Wi-Fi APs in unlicensed spectrum bands.

In [33], a coalition game is proposed for access point (AP) selection in a LTE-U and Wi-Fi coexistence scenario. The users with diverse traffic demand are capable of accessing either LTE-U or Wi-Fi, and they operate based on the proposed coalition formation game. Simulations were performed using ns-3 network simulator, and the results demonstrate a considerable increase at the overall throughput of LTE-U and Wi-Fi that outperforms the standalone cases.

In [34], another coalition game-based solution is proposed for LTE-LAA and Wi-Fi coexistence. The use of time resources, via almost blank subframe (ABS), and the traffic offloading are taken into account for a Shapley fairness game. Simulations show that resource allocation using this approach can sure present better fairness.

Multi-game approach for LTE-U and Wi-Fi coexistence is proposed in [35], [36]. In [35] the authors propose

a framework where dynamic channel selection and inter-operator spectrum sharing are modeled as distinct games for resource allocation. Taking this method into account, the paper show, by simulation results, that sum-throughput for LTE-U is higher when compared with LBT.

In [36] the multi-game framework proposed uses Nash bargaining game (NBG) for sharing time resources in unlicensed bands for LTE-U and Wi-Fi. Internally, the LTE-U uses a bankrupt game for its resource allocation. The major conclusion of this work states that using this method in a dense deployment scenario the Wi-Fi is less damaged by LTE-U when compared with LTE-LBT.

Finally, in [37] the game theory approach for spectrum sharing problem in LTE-U and Wi-fi coexistence is combined with decision tree method. The results show improvements in throughput and spectrum efficiency keeping the traffic load balanced and guarantying fairness.

C. Q-learning

In [38], a technique named double Q-learning for coexistence among LAA and Wi-Fi is devised. This solution is based on DTX (Discontinuous Transmissions), and TPC (Transmit Power Control) applied for channel selection, which leads to an improved LAA and Wi-Fi coexistence. In [39], Q-learning is also used for LAA channel selection in a 3GPP indoor coexistence scenario. The author also proposes a game theoretical solution and compare its performance with Q-learning. As for DC-based solutions, a Q-learning solution is proposed in [40] for dynamic duty-cycle selection in co-channel LTE-U/Wi-Fi situations, with results indicating an enhancement on maximum capacity considering FTP as the offered traffic.

According to our state-of-the-art findings, we point out that although there are several papers exploring game theory applied in the coexistence domain, few are the ones that take into account the LTE-U/LTE-U coexistence case. In fact, the LTE-U Forum itself, in [41], assumed that for multi-operator LTE-U deployment scenario, the LTE-U APs would perform channel selection. However, since there are a limited number of channels in the ISM spectrum, the probability of two LTE-U small cells sharing the same unlicensed channel is very high, especially in a dense deployment scenario. Hence, our work proposes a novel solution for the co-channel LTE-U/LTE-U coexistence problem, filling a gap we believe exists for this scenario. In particular, we apply game theory (suitable to model the competition of limited resources), and reinforcement learning techniques (convenient to dynamic decision-making problems), as meticulously described in this section.

III. STOCHASTIC GAMES

In reinforcement learning, a Markov Decision Process (MDP) formalization [42] is a method for modeling a *single* adaptive agent that interacts with an environment characterized by a probabilistic transition function. The MDP is defined by (a) a set of S states, (b) a set of A actions, (c) a transition probability function, $T : S \times A \rightarrow [0, 1]$, and (d) a reward function, $R : S \times A \rightarrow \mathbb{R}$, that represents the effects of

each action on the state environment. The goal is to maximize the agent’s reward by finding a policy function $\pi : S \rightarrow A$ providing the optimal action decision for each state. As for *secondary agents* according to MDP perspective, they are modeled as part of the environment, with some fixed behavior.

By extending MDP via the use of stochastic games [43], also called Markov games, it is possible to model dynamic interactions that change the environment in response to each player action. In other words, stochastic games expand the MDP approach to include multiple adaptive agents with interacting or competing goals.

In its general form, an N -player stochastic game is defined by a set of states (games), S , and a collection of action sets, $\{A_1, A_2, \dots, A_N\}$, one for each agent of the environment. The state transition probability function and reward function for each agent are defined, respectively, by $T : S \times A_1 \times A_2 \times \dots \times A_N \rightarrow [0, 1]$ and $R : S \times A_1 \times A_2 \times \dots \times A_N \rightarrow \mathbb{R}$. The overall goal is to find an optimal strategy that maximizes the reward of each and all agents.

This paper considers a special case of a stochastic game, where there are two agents in the environment competing for resources such that one agent gain necessary implies a same amount of loss to its opponent. This diametrical perspective, called two-player zero-sum stochastic game, allows us to use a single reward function that each agent tries to maximize itself.

Following this approach, each player state transitions are controlled by their current state, action, and the corresponding opponent action. For that, we can define $T_i^j(s, a, \sigma)$ as the probability function of transition to state s given that the player i takes the action a while his opponent takes the action σ .

A. Solving Two-Player Zero-Sum Stochastic Game

Solving an MDP means finding an optimal policy that maximizes the sum of rewards received from the environment. A naive approach for that solution consists of enumerating and evaluating all possibles policies and return the best one, however, the total number of policies is an exponential function of the number of states and the number of actions [44].

An alternative approach to finding the best policy is through Bellman’s equation [45]. According to Richard Bellman’s principle of optimality [46], an optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision. Hence, we can break the MDP problem into sub-problems, reducing the complexity to a polynomial function of the number of states and the number of actions [45]. Bellman’s dynamic programming is elegant. However, its development assumes the availability of an explicit model that encompasses the transition probability from one state to another. Unfortunately, in practical situations, such a model is not available.

We can cope with this issue by applying the Q-learning [47], a model-free reinforcement learning algorithm. The Q-Learning formulation is based on a Q function, updated whenever it receives a reward of r from a state transition after

the agent takes a certain action a from a finite set of actions A . Hence, denoting the agent's decision time steps as t , the function Q_t is updated at the next decision time, $t + 1$, when a reward, r , is observed for a state transition from s_t to s_{t+1} after taking an action a_t , as shown by equation 1:

$$Q_{t+1}^i(s_t; a_t^o) = \alpha Q_t^i(s_t; a_t^o) + (1 - \alpha) [r + \max_{a \in A} Q_t^i(s_{t+1}; a^o)] \quad (1)$$

where α and γ are the learning rate and the discount factor, respectively. Indicating the effect of the future reward to the current state [45]. The goal is to decide the optimal action at the current state s_t as:

$$a = \text{arg max}_{a \in A} Q_t^i(s_t; a^o) \quad (2)$$

According to the Q-learning convergence proof [48], Q-learning converges to the optimum action-values with 100% probability, so long as all actions are repeatedly sampled in all states and the action-values are represented discretely under reasonable conditions on the Markovian environment. So, through experience (i.e., trial-and-error interactions with the dynamic environment), the agent is capable to find a policy indicating what action maximizes rewards under any given circumstances. For this reason, Q-Learning is highly suited for solving Markov decision problems without explicit knowledge of the transition probabilities [45].

However, Q-Learning itself is not able to solve stochastic games, since secondary opponents behavior is accounted for just as part of the environment. In fact, two or more interacting Q-Learning agents, in a certain competitive situation, could provide either low or unfair performance between them [49].

In [50], Michael L. Littman proposes a new algorithm that widens Q-Learning technique for solving stochastic games, specifically two-player zero-sum stochastic games, called Minimax Q-Learning. This solution is a combination of the Q-Learning and linear programming to a single matrix game, providing an optimal solution for any fully competitive game, where two players have opposite goals and reward functions.

Hence, considering a finite set of predefined states S and actions A , the goal is to define for each player i a function $V^i(s^o)$ to be the expected reward for an optimal deterministic policy starting from a state $s \in S$, and $Q^i(s; a; o^o)$ as the expected reward for taking the action a when the opponent chooses the action o from the state s , and continuing optimally thereafter, as shown by Equations (3) and (4).

$$V^i(s^o) = \max_{a \in A} \min_{o \in A} Q^i(s; a; o^o) \quad (3)$$

$$Q^i(s; a; o^o) = R^i(s; a; o^o) + \gamma \sum_{s' \in S} T^i(s; a; o^o; s') V^i(s'^o) \quad (4)$$

Then, for every state transition from s to s' , we can treat the $Q^i(s; a; o^o)$ values as immediate payoffs in an unrelated sequence of matrix games (one for each state, s), each of which can be solved optimally using a similar Q-Learning technique, as shown by equation (5).

$$Q^i(s; a; o^o) = r + \max_{a \in A} \min_{o \in A} Q^i(s'; a'; o'^o) \quad (5)$$

In [51], the authors demonstrate that this Minimax Q-Learning holds essential properties of the standard Q-Learning, including:

- Policies can be computed independently;
- The Q function is enough to specify the policy function;
- Value iteration methods can be applied, and they guarantee convergence;
- There's a unique optimal solution;
- It converges to the Nash equilibrium under certain conditions.

The strength of the *minimax* criterion is that it allows the agent to converge to a fixed strategy that is guaranteed to be safe in the sense it does as well as possible against the worst possible opponent. A convergence proof for that algorithm was provided subsequently by Littman and Szepesvari in [52].

IV. LTE-U SELF-COEEXISTENCE PROBLEM MODELING

In a situation of two coexisting LTE-U operators (herein named Operator 1 and Operator 2), their channel access profiles will behave similarly as illustrated in Figure 1. Basically, for each fixed time window defined by the *Almost-Blank-Subframe* (T_{ABS}), both operators determine a certain time for starting transmitting (T_{start}) and stopping transmitting (T_{stop}), thus providing a total duration time of continuous operation (T_{on}) and inactivity (T_{off}). As a consequence, both systems are susceptible to either a certain collision period (T_c) or IDLE channel period (T_{IDLE}), causing network efficiency loss.

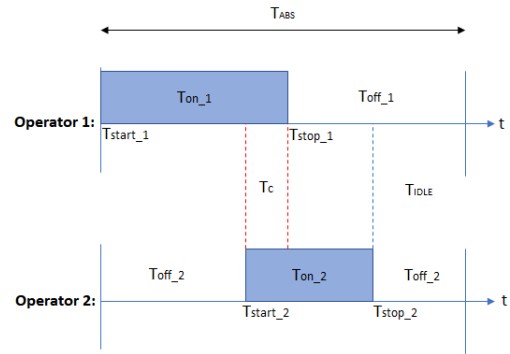


Fig. 1. Channel access profile of two competing LTE-U operators.

Assuming $\forall k \neq k^0$, and $k, k^0 \in \{1, 2\}$, according to Figure 1, we can also analytically infer that:

$$T_{on_k} + T_{off_k} = T_{ABS} \quad (6)$$

$$T_{on_{k^0}} + T_{off_{k^0}} = T_{ABS} \quad (7)$$

$$T_{on_k} = T_{stop_k} - T_{start_k} \quad (8)$$

$$T_{on_k^0} = T_{stop_k^0} - T_{start_k^0} \quad (9)$$

$$T_c = T_{stop_k} - T_{start_k} \quad (10)$$

For an efficient collision-free transmission, where $T_c = T_{IDLE} = 0$, we have:

$$T_{stop_k} = T_{start_k} \quad (11)$$

$$T_{ABS} = T_{stop_k^0} - T_{start_k} \quad (12)$$

Hence, replacing Equations (8) and (11) in (6), we have:

$$T_{off_k} = T_{ABS} - T_{on_k} \quad (13)$$

$$= T_{ABS} - T_{stop_k^0} - T_{start_k^0} \quad (14)$$

$$= T_{ABS} - T_{start_k^0} - T_{start_k^0} \quad (15)$$

Replacing Equation (12) in (15):

$$T_{off_k} = T_{stop_k^0} - T_{start_k} - T_{start_k^0} + T_{start_k^0} \quad (16)$$

$$= T_{stop_k^0} - T_{start_k^0} \quad (17)$$

Therefore, according to Equations (9) and (17), the following identity can be obtained:

$$T_{off_k} = T_{on_k^0} \quad (18)$$

In other words, when it comes in time resource management for two competing LTE-U, the gain of an operator implies the same amount of loss for its opponent, thus as stated by two-person zero-sum game modeling.

V. GTDM-CSAT SOLUTION PROPOSAL

This section presents the configuration of Minimax Q-learning algorithm for GTDM-CSAT. First of all, we need to clearly define what constitutes the Agent. After that, we define the Q-learning process, with three main variables:

- The system's states;
- The actions that the agent and its opponent can perform;
- The reward from the environment.

In order to provide a simple approach suitable in practical situations, we propose the following Minimax-Q-learning framework:

- 1) The agents represent two LTE-U APs (Operators A and B), subject to mutual interference. GTDM-CSAT algorithm is embedded in them, making both capable of updating their OFF period within each 40 ms duty-cycle;
- 2) The actions that the agents can take are established on a set of duty-cycle patterns $A = \{20\%; 30\%; 50\%; 60\%; 80\%\}$. Herein, a duty-cycle pattern of 30% means that LTE-U gates off 70% of the duty-cycle time and transmits in the rest of 40 ms;

- 3) The reward is given by the aggregated of each operator transmitted data rate over its respective demanded data rate: $\frac{TX_{LTE_A}}{TX_{LTE_Atotal}} + \frac{TX_{LTE_B}}{TX_{LTE_Btotal}}$, where
 - a) TX_{LTE_A} and TX_{LTE_B} are the transmitted link data rates of the MCS scheme mapped from the current Channel Quality Indicator (CQI) of each operator LTE-U AP in accordance with the Adaptive Modulation and Coding (AMC) described in [53], [54].
 - b) TX_{LTE_Atotal} and TX_{LTE_Btotal} are the average traffic data rates offered for each operator;
- 4) Therefore, the system's states can be defined based on the following thresholds:

$$S = \begin{cases} 0; & \text{if } 0 \leq \frac{TX_{LTE_A}}{TX_{LTE_Atotal}} + \frac{TX_{LTE_B}}{TX_{LTE_Btotal}} < 1 \\ 1; & \text{if } 1 \leq \frac{TX_{LTE_A}}{TX_{LTE_Atotal}} + \frac{TX_{LTE_B}}{TX_{LTE_Btotal}} < 2 \end{cases} \quad (19)$$

Within above-mentioned definitions, we proposed the GTDM-CSAT based on the relative transmitted data rate of each operator. In addition to the ease of obtaining it from the PHY and MAC layers of LTE-U APs, this metric could also provide an indication of channel usage fairness for each operator. We claim this is a differential the proposed GTDM-CSAT because, in the case of two LTE-U APs having the same transmission data rate capabilities, this will not always indicate the same channel access priority, as their demanded data rate can vary.

Hence, the algorithm can work to maximize the system's aggregated data rate simultaneously keeping a fair channel access priority between operators. Pseudo-code in Algorithm 1 illustrates the proposed Minimax-Q-Learning framework:

VI. SYSTEM MODELING

We model and evaluate the GTDM-CSAT in ns-3 simulator [55], an open-source network simulator based on discrete events, developed especially for research and educational purposes. The ns-3 is built upon C++, and it complies to technical norms of standards organizations of emerging networks like 3GPP, IEEE, and Wi-Fi Alliance. This is the main reason for using ns-3 as a prototyping tool for the performance analysis presented here.

LTE-U, LTE-LAA and Wi-Fi systems modeling, and their corresponding coexistence are kept in a repository called *laa/ns-3-lbt* [56], apart from the ns-3 main branch. The code is part of a project funded by the Wi-Fi Alliance and developed by the *Centre Tecnològic de Telecomunicacions de Catalunya* (CTTC) and by the University of Washington [57]. The main improvements implemented by this project are: (i) Wi-Fi's CCA procedure; (ii) frame synchronization modeling; (iii) AP selection modeling based on RRSI; (iv) propagation and interference modeling; (v) MIMO modeling. Two new LTE transmission models are available: (i) LTE-Duty cycle (LTE-DC), the LTE-U; and (ii) LTE Listen Before Talking (LTE-LBT), the 3GPP LTE-LAA. Additionally, some evaluation scenarios defined in [58] are ready to use (e.g., indoor and outdoor). Then, we generate the results of this

Algorithm 1: Minimax-Q-Learning algorithm for dynamic duty-cycle selection of a LTE-U AP during self coexistence.

```

1 Initialize
2 for  $s \in S; a \in A$  do
3   | Start Q-Table with random Q-Values  $Q^1 s; a; \sigma^0$ .
4 end
5 Estimate initial states  $s$ .
6 Learning:
7 Loop - for each agent:
8 Generate a random value  $r \in U[0; 1]^0$ 
9 if  $r < \theta$  then
10  | Select an action  $a \in A$  randomly;
11 else
12  | Select an action  $a \in A$  according to
    |  $\max_a \min_o Q^1 s; a; \sigma^0$ ;
13 end
14 Run  $a$ ;
15 Take the immediate reward  $r$ , which
    | is  $\frac{TX_{LTEA}}{TX_{LTEA}^{total}} + \frac{TX_{LTEB}}{TX_{LTEB}^{total}}$ ;
16 Observe the next state  $s^0 \in S$ ;
17 Update Q-table according to the expression:
18  $Q^1 s; a; \sigma^0 = r + \max_{a^1 \in A} \min_{o^1 \in A} Q^1 s^0; a^1; \sigma^0$ 
19  $s = s^0$ 
20 end loop

```

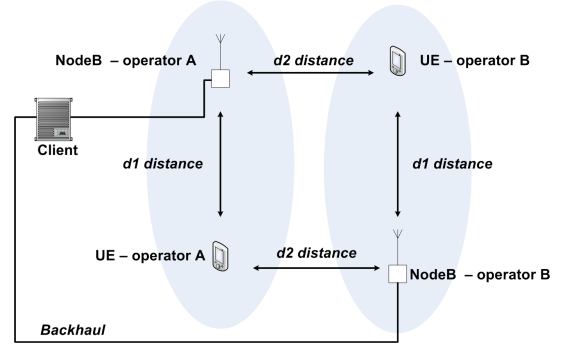


Fig. 2. 3GPP Simple Coexistence Scenario [58].

TABLE I
SYSTEM MODELING PARAMETERS.

LTE Parameters (Release 8 - FDD)	
Bandwidth	20 MHz - 100 PRBs
Carrier frequency	5 GHz
Link-to-system mapping	Vienna LTE Simulator (MIESM)
Packet Scheduler	Proportional Fair
CQI report	Wideband CQI (based on SINR) - 15 CQIs
ABS Pattern Duration	40 ms
Path loss and Shadow fading	ITU InH
Multi-antenna Technology	2x2 MIMO (5dBi - Omnidirectional)
Tx Power	18 dBm
Traffic Model (DL only)	UDP full-buffer

paper by using a modified version of changeset 26 of repository *lans-3-lbt* [56] (ns-3-lbt-056f37aea269.tar.bz2).

The implemented LTE-U transmission model follows the approach proposed in [2]. It takes advantage of the LTE Almost Blank Subframe (ABS) functionality to disable LTE transmissions in certain subframes. This component works by specifying an ABS pattern within a duty-cycle period of 40 ms. The configuration is implemented with a bit-mask of 40 values (one per LTE subframe), thus conforming to the format specified by 3GPP in [59]. Our contribution to the ns-3 code is two-fold: (i) Specific codes in order to set the bit-mask as well as to calculate the LTE-U AP transmitted data rate on runtime basis; and (ii) The implementation of a Q-learning algorithm to decide the suitable bit-mask.

VII. EVALUATION SCENARIO AND SYSTEM PARAMETERS

We evaluate the performance of LTE-U self-coexistence in the first coexistence scenario defined by 3GPP in [58]. This scenario is composed of two access points (LTE-U NodeBs) of different operators, and two LTE-U User Equipments (UEs), one of each operator, as shown in Figure 2.

Herein, although it is possible to configure the distance $d2$ (between operators) as well as the distance $d1$ (between the LTE-U AP and the UE), both remain set to 10 m to model a situation of high coexistence interference. The *backhaul* generates UDP full buffer traffic in the *downlink* to the STAs. We present the main simulation parameters in Table I.

Despite its simplicity, this scenario allows us to easily control the overall system’s interference profile, thus enabling

the exploration of the proposed solution under a set of different network conditions.

VIII. RESULTS

As a means to evaluate the performance gains under different Q-learning GTDM-CSAT learning parameters (θ and γ), the following investigation scenario is proposed:

- Both operators (LTE-U A and LTE-U B) start the simulation with $udpRate = 75$ Mbps for a total simulation time of 40 s;
- At 5 s, LTE-U B decreases its offered data rate ($udpRate$) to 37.5 Mbps;
- At 10 s, LTE-U B decreases its $udpRate$ to 18.75 Mbps;
- At 15 s, LTE-U B returns its $udpRate$ to 75 Mbps, so that both LTE-U A and LTE-U B operators reach the same $udpRate$ again;
- At 20 s, LTE-U A decreases its $udpRate$ to 37.5 Mbps;
- At 25 s, LTE-U A decreases its $udpRate$ to 18.75 Mbps;
- At 30 s, LTE-U A returns its $udpRate$ to 75 Mbps, so that both LTE-U A and LTE-U B operators reach the same $udpRate$ once again.

The goal is to analyze a static preliminary scenario where one LTE-U operator deterministically changes its traffic rate, thus requiring the other LTE-U operator adaptation to a new interference profile. In this scenario, these changes encompass the main possible traffic rate combinations during a simulation time that allows up to 1000 state transitions. Besides, for every traffic rate changing, a maximum aggregated throughput is expected. Hence, based on this prior information, we linearly

evaluate the impact of the learning rate, α , and the discount factor, γ , in our simulation results, aiming to minimize the total number of iterations the algorithm takes to reach the expected aggregated throughput.

Following this perspective, we found that when $\alpha = 0.3$ and $\gamma = 0.9$ the algorithm reaches the expected aggregated throughput with an average of 15 iterations (40 ms each) in the worst case, showing the behavior of the transmitted data rate presented in Figure 3. This performance corroborates with [39], which confirmed the rapid convergence of a solution based on the stochastic game in comparison with traditional Q-Learning.

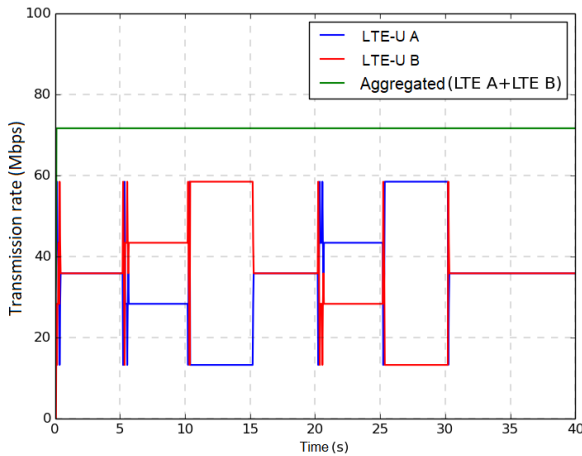


Fig. 3. Investigative result of two LTE-U coexisting operators.

Other straight conclusions can be taken from Figure 3.

We can note that during the time interval from 5 s to 15 s, when LTE-U B decreased its *udpRate*, the LTE-U A sequentially changed its *duty-cycle* pattern to increase its transmission rate, then keeping the aggregated transmitted data rate constant. The GTDM-CSAT transferred transmission time from LTE-U B to LTE-U A, because of the lower *udpRate* of LTE-U B. The same behavior is also observed during the time interval from 20 s to 30 s, when LTE-U A decreased its *udpRate* while LTE-U B sequentially increases its transmission time, and consequently, its transmission rate.

Based on these promising results, we evaluate the effectiveness of the GTDM-CSAT in the following scenario:

Simulations of 100-second long with LTE-DC APs randomly changing their *udpRate*;

Initially, an LTE-U AP is randomly selected (A or B) to chance its *udpRate*;

After the operator is selected, each *udpRate* variation follows a second uniform random variable to select one of three possible *udpRate* values: 75 Mbps, 37.5 Mbps, and 18.75 Mbps;

A third uniform random variable defines for how long the *udpRate* selected will remain, considering an interval from 4 s to 7 s;

The throughput (calculated in L3) is evaluated at the end of the simulation.

Now, the goal is to analyze the GTDM-CSAT performance compared to a fixed *duty-cycle* LTE-U solution in the same condition. Figure 4 presents our performance results by bar plots of received data throughput for some static *duty-cycle* patterns and the GTDM-CSAT Minimax-Q-learning. Plots from two operators are overlapped so that the aggregated throughput could be visualized.

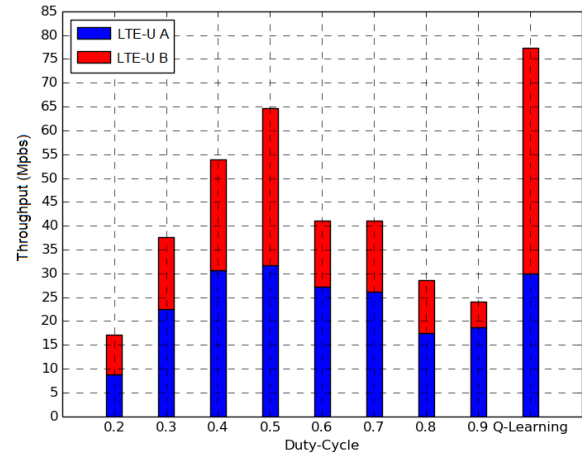


Fig. 4. Minimax-Q-Learning performance compared to a fixed *duty-cycle* LTE-U in a random *udpRate* scenario.

As we can analyze from Figure 4, the GTDM-CSAT Minimax-Q-Learning method provides the best aggregated throughput performance (77 Mbps), which is around 12 Mbps higher than the best fixed *duty-cycle* performance at $dc = 0:5$ (65 Mbps). Furthermore, analyzing each operator individually, we can note that, when operating with Minimax-Q-Learning, LTE-U B achieved its maximum throughput performance (43 Mbps), while LTE-U A reaches 30 Mbps, which is very close to its maximum throughput obtained at $dc = 0:5$ (32 Mbps). In other words, the GTDM-CSAT Minimax-Q-Learning solution not only maximized the aggregated throughput but also balanced channel access priority among operators. This is an evident consequence of maximizing the relative data rate as proposed in section V, making possible to detect data rate demand of each operator and effectively equalize the channel usage.

Enabling, however, multiple LTE-U APs coexistence, such as a dense LTE-U deployment scenario, requires extending the minimax criteria to a non-cooperative general-sum stochastic game. The authors in [60] propose an interesting framework for this, called Nash-Q-learning. This solution enables the agents to perform learning updates based on assuming Nash equilibrium behavior over their current Q-values, finding each player optimal strategy toward Nash equilibrium for each iteration. Nevertheless under the cost of exponential complexity in the number of agents. We address analyzing such a solution as a short-term future work perspective.

IX. CONCLUSION AND FUTURE INVESTIGATION

This work presented a game theory and reinforcement learning solution, named GTDM-CSAT, for co-channel LTE-U

self-coexistence problem in unlicensed spectrum. We first modeled the problem as a Markov decision process and then introduced a solution based on Minimax Q-Learning algorithm. We implemented this solution in ns-3, an open-source network simulator, and performed simulations based on an official 3GPP scenario according to specifications defined by Wi-Fi Alliance for coexistence in 5 GHz carrier frequency. Results showed that the proposed solution provides significant gains of aggregated throughput while still keeps a fair resource sharing between operators.

Our next steps will be extending the GTDM-CSAT solution for multiple operators and reproducing the simulations in a dense deployment scenario, e.g., the 3GPP indoor coexistence scenario defined in [58]. Furthermore, we also intend to evaluate the GTDM-CSAT for other LTE coexistence solutions, as LTE-LAA and LTE-LWA. The signaling for changing coexistence information among LTE and Wi-Fi APs is also our target for future works.

REFERENCES

- [1] 3GPP. Release 13. Accessed in 2019-05-06. [Online]. Available: http://www.3gpp.org/ftp/Information/WORK_PLAN/Description_Releases/
- [2] E. Almeida, A. M. Cavalcante, R. C. D. Paiva, F. S. Chaves, F. M. Abinader, R. D. Vieira, S. Choudhury, E. Tuomaala, and K. Doppler, "Enabling LTE/WiFi coexistence by LTE blank subframe allocation," in *IEEE ICC 2013*, 2013, doi: 10.1109/ICC.2013.6655388.
- [3] Qualcomm, "Qualcomm research LTE in unlicensed spectrum: Harmonious coexistence with Wi-Fi," Alcatel-Lucent, Ericsson, Qualcomm Technologies, Tech. Rep., 2014, accessed in 2019-05-06. [Online]. Available: <https://www.qualcomm.com/documents/lte-unlicensed-coexistence-whitepaper>
- [4] 3GPP. (2016) Collaboration on LTE - LWAN Integration. Accessed in 2019-05-06. [Online]. Available: http://www.3gpp.org/news-events/3gpp-news/1771-wlan_lte
- [5] Qualcomm. (2016) Multifire: LTE-like performance with Wi-Fi-like deployment simplicity. Accessed in 2019-05-06. [Online]. Available: <https://www.qualcomm.com/invention/technologies/lte/multifire>
- [6] C. Alliance. (2017) What is CBRS? - LTE in 3.5 GHz Shared Spectrum and What it Means for IoT. Accessed in 2019-05-06. [Online]. Available: <https://www.leverage.com/blogpost/what-is-cbrs-lte-3-5-ghz>
- [7] T. Baykas, M. Kasslin, M. Cummings, H. Kang, J. Kwak, R. Paine, A. Reznik, R. Saeed, and S. Shellhammer, "Developing a standard for TV white space coexistence: technical challenges and solution approaches," *Wireless Communications, IEEE*, vol. 19, no. 2, pp. 10–22, 2012, doi: 10.1109/MWC.2012.6155872.
- [8] P. M. de Santana, V. D. Melo, and V. A. de Sousa Jr., "Performance of license assisted access solutions using ns-3," *International Conference on Computational Science and Computational Intelligence*, 2016, doi: 10.1109/CSCI.2016.0181.
- [9] M. Mehrnosh, R. Patidar, S. Roy, and T. Henderson, "Modeling, simulation and fairness analysis of wi-fi and unlicensed LTE coexistence," 2018.
- [10] A. M. Baswade, T. A. Atif, B. R. Tamma, and A. Franklin, "A novel coexistence scheme for IEEE 802.11 for user fairness and efficient spectrum utilization in the presence of LTE-U," *Computer Networks*, July 2018, doi: 10.1016/j.comnet.2018.04.002.
- [11] V. Maglogiannis, D. Naudts, A. Shahid, and I. Moerman, "A Q-learning scheme for fair coexistence between LTE and Wi-Fi in unlicensed spectrum," *IEEE Access*, vol. 6, May 2018, doi: 10.1109/ACCESS.2018.2829492.
- [12] Wi-Fi Alliance, "Draft coexistence test plan - v0.8.4 - alpha," Tech. Rep., 2016, accessed in 2019-05-06. [Online]. Available: <https://www.wi-fi.org/file/draft-coexistence-test-plan>
- [13] F. M. Abinader, E. P. L. Almeida, F. S. Chaves, A. M. Cavalcante, R. D. Vieira, R. C. D. Paiva, A. M. Sobrinho, S. Choudhury, E. Tuomaala, K. Doppler, and V. A. Sousa, "Enabling the coexistence of LTE and Wi-Fi in unlicensed bands," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 54–61, Nov 2014, doi: 10.1109/MCOM.2014.6957143.
- [14] F. M. A. Jr., S. Choudhury, V. A. de Sousa Jr., F. de S. Chaves, A. M. Cavalcante, E. P. L. Almeida, R. D. Vieira, E. Tuomaala, and K. Doppler, "Distributed Wi-Fi Interference Coordination for Dense Deployments," *Wireless Personal Communications*, vol. 97, no. 1, pp. 1033–1058, 2017, doi: 10.1007/s11277-017-4550-3.
- [15] F. M. A. Jr., V. A. de Sousa Jr., S. Choudhury, F. de S. Chaves, A. M. Cavalcante, E. P. L. Almeida, R. D. Vieira, E. Tuomaala, and K. Doppler, "LTE/Wi-Fi Coexistence in 5 GHz ISM Spectrum: Issues, Solutions and Perspectives," *Wireless Personal Communications*, vol. 99, no. 1, pp. 403–430, 2018, doi: 10.1007/s11277-017-5114-2.
- [16] Y. Gao, X. Chu, and J. Zhang, "Performance analysis of LAA and WiFi coexistence in unlicensed spectrum based on markov chain," *IEEE GLOBECOM*, 2016, doi: 10.1109/GLOCOM.2016.7842129.
- [17] C. Cano, D. J. Leith, A. Garcia-Saavedra, and P. Serrano, "Fair coexistence of scheduled and random access wireless networks: Unlicensed LTE/WiFi," *IEEE/ACM Transactions on Networking*, 2016, doi: 10.1109/TNET.2017.2731377.
- [18] A. Mukherjee, J. F. Cheng, S. Falahati, H. Koorapaty, D. H. Kang, R. Karaki, L. Falconetti, and D. Larsson, "Licensed-assisted access LTE: Coexistence with IEEE 802.11 and the evolution toward 5G," *IEEE Communications Magazine*, 2016, doi: 10.1109/MCOM.2016.7497766.
- [19] Y. Pang, A. Babaei, J. Andreoli-Fang, and B. Hamzeh, "Wi-fi coexistence with duty cycled lte-u," *Wireless Communications and Mobile Computing*, 2017, doi: 10.1155/2017/6486380.
- [20] A. Abdelfattah and N. Malouch, "Studying the impact of lte-u on wi-fi downlink performance," *IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2016, doi: 10.1109/WiMOB.2016.7763261.
- [21] C. Cano and D. J. Leith, "Coexistence of wifi and lte in unlicensed bands: A proportional fair allocation scheme," *IEEE ICC - Workshop on LTE in Unlicensed Bands: Potentials and Challenges*, 2015, doi: 10.1109/ICCW.2015.7247522.
- [22] X. Wang, T. Q. S. Quek, M. Sheng, and J. Li, "Throughput and fairness analysis of wi-fi and lte-u in unlicensed band," *IEEE Journal on Selected Areas in Communications*, 2016, doi: 10.1109/JSAC.2016.2632629.
- [23] J. Andreoli-Fang, "Wi-fi vs. duty cycled lte-u: In-home testing reveals coexistence challenges," Cable Labs, Tech. Rep., 2015, accessed in 2019-05-06. [Online]. Available: <https://www.cablelabs.com/vran-over-docsis-cablelabs-making-reality>
- [24] O. El-Samadisy, M. Khedr, and A. El-Helw, "Performance evaluation of MAC for IEEE 802.11 and LAA LTE," *International Conference on Computational Science and Computational Intelligence*, 2016, doi: 10.1109/CSCI.2016.0178.
- [25] S. Dama, A. Kumar, and K. Kuchi, "Performance evaluation of LAA-LBT based LTE and WLANs co-existence in unlicensed spectrum," *Globecom Workshops*, 2015, doi: 10.1109/GLOCOMW.2015.7414071.
- [26] N. Rupasinghe and I. Guvenc, "Licensed-Assisted Access for WiFi-LTE coexistence in the unlicensed spectrum," *Globecom Workshops Emerging Technologies for 5G Wireless Cellular Networks*, 2014, doi: 10.1109/GLOCOMW.2014.7063546.
- [27] R. kwan, R. Pazhyannur, and V. Chandrasekhar, "Fair co-existence of Licensed Assisted Access LTE (LAA-LTE) and Wi-Fi in unlicensed spectrum," *7th CEEC*, 2015, doi: 10.1109/CEEC.2015.7332692.
- [28] P. M. de Santana, V. D. de Lima Melo, and V. A. D. S. Jr., "Performance of License Assisted Access solutions using ns-3," *CSCI*, 2016, doi: 10.1109/CSCI.2016.0181.
- [29] V. Maglogiannis, D. Naudts, A. Shahid, and I. Moerman, "An adaptive LTE listen-before-talk scheme towards a fair coexistence with wi-fi in unlicensed spectrum," *Telecommunication Systems*, Jan 2018, doi: 10.1007/s11235-017-0418-9.
- [30] S. Oulaouf, A. Haidine, and H. Ouahmane, "Review on using game theory in resource allocation for lte/lte-advanced," *IEEE Int. Conf. on Advanced Communication Systems and Information Security*, 2016, doi: 10.1109/ACOSIS.2016.7843946.
- [31] P. J. Ashidani and P. R. Guardieiro, "Two-level downlink resource allocation scheme based on cooperative game theory in lte networks," *IFIP Wireless Day*, 2014, doi: 10.1109/WD.2014.7020840.
- [32] T. Cuzanasukas and A. Anskaitis, "Game theory for future ieee 802.11 spectrum sharing," *IEEE Conference of Electrical, Electronic and Information Sciences*, 2015, doi: 10.1109/eStream.2015.7119481.
- [33] Y. Zhang, C. Jiang, J. Wang, Z. Han, J. Yuan, and J. Cao, "Coalition formation game based access point selection for lte-u and wi-fi coexistence," *IEEE Transactions on Industrial Informatics*, 2018, doi: 10.1109/TII.2018.2797100.
- [34] B. C. Chung and D. Cho, "Mobile data offloading with almost blank subframe in lte-laa and wi-fi coexisting networks based on coalition

