Proactive Resource Management in LTE-U Systems: A Deep Learning Perspective

Ursula Challita*, Li Dong*, and Walid Saad†

*School of Informatics, The University of Edinburgh, Edinburgh, UK,

Emails: {ursula.challita, li.dong}@ed.ac.uk.

†Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA. Email: walids@vt.edu.

Abstract

LTE in unlicensed spectrum (LTE-U) is a promising approach to overcome the wireless spectrum scarcity. However, to reap the benefits of LTE-U, a fair coexistence mechanism with other incumbent WiFi deployments is required. In this paper, a novel deep learning approach is proposed for modeling the resource allocation problem of LTE-U small base stations (SBSs). The proposed approach enables multiple SBSs to proactively perform dynamic channel selection, carrier aggregation, and fractional spectrum access while guaranteeing fairness with existing WiFi networks and other LTE-U operators. Adopting a proactive coexistence mechanism enables future delay-intolerant LTE-U data demands to be served within a given prediction window ahead of their actual arrival time thus avoiding the underutilization of the unlicensed spectrum during off-peak hours while maximizing the total served LTE-U traffic load. To this end, a noncooperative game model is formulated in which SBSs are modeled as Homo Egualis agents that aim at predicting a sequence of future actions and thus achieving long-term equal weighted fairness with WLAN and other LTE-U operators over a given time horizon. The proposed deep learning algorithm is then shown to reach a mixed-strategy Nash equilibrium (NE), when it converges. Simulation results using real data traces show that the proposed scheme can yield up to 28% and 11% gains over a conventional reactive approach and a proportional fair coexistence mechanism, respectively. The results also show that the proposed framework prevents WiFi performance degradation for a densely deployed LTE-U network.

Index Terms

LTE-unlicensed (LTE-U); small cell; unlicensed band; long short term memory (LSTM); game theory; proactive resource allocation

I. Introduction

LTE in unlicensed bands (LTE-U) has emerged as an effective solution to overcome the scarcity of the radio spectrum [1]. Using LTE-U, a cellular small base station (SBS) can improve its

performance by simultaneously accessing licensed and unlicensed bands. However, to achieve the promised quality-of-service (QoS) improvements from LTE-U, many challenges must be addressed ranging from effective co-existence with existing WiFi networks to resource allocation, multiple access, and inter-operator spectrum sharing [1].

If not properly deployed, LTE-U can significantly degrade the performance of WiFi [1]. There has been a number of recent works [2]–[9] that study the problem of enhanced LTE-U and WiFi coexistence. This existing body of works can be categorized into two groups: channel access [2]–[5] and channel selection [7]–[9]. The authors in [2]–[4] propose different channel access mechanisms based on listen-before-talk (LBT) that rely on either an exponential backoff [2], a fixed/random contention window (CW) size [3], or an adaptive CW size [4]. Nevertheless, an exponential backoff approach leads to unnecessary retransmissions while a fixed CW size cannot handle time-varying traffic loads thus yielding unfair outcomes. The authors in [5] develop a holistic approach for both traffic offloading and resource sharing across the licensed and unlicensed bands but considering one SBS. In [6], the authors study the problem of resource allocation with uplink-downlink decoupling for LTE-U. The authors in [10] propose an inter-network coordination scheme with a centralized radio resource management for the coexistence of LTE and WiFi. However, this prior art is limited to one unlicensed channel and does not jointly account for channel selection and channel access. In other words, these works do not analyze the potential gains that can be obtained upon aggregating or switching between different unlicensed channels. Operating on a fixed unlicensed channel limits the amount of cellular data traffic that can be offloaded to the unlicensed band and leads to an increase in the interference level caused to neighboring WiFi access points (WAPs) operating on that same channel.

In terms of LTE-U channel selection, the authors in [7] propose a distributed approach based on Q-learning. A matching-based solution approach is proposed in [8], which is both distributed and cooperative. Moreover, the work in [9] combines channel selection along with channel access. Despite the promising results, all of these works [7]–[9] consider a reactive approach in which data requests are first initiated and, then, resources are allocated based on their corresponding delay tolerance value. Nevertheless, this sense-and-avoid approach can cause an underutilization of the spectrum due to the impulsive reconfiguration of the spectrum usage that does not account for the future dynamics of the network. Despite the predominance of the reactive LTE-WiFi coexistence solutions, cellular data traffic networks are known to exhibit statistically fluctuating

and periodic demand patterns, especially applications such as file transfer, video streaming and browsing [11], therefore providing an opportunity for the network to exploit the predictable behavior of the users to smooth out the traffic over time and reduce the difference between the peak and the average load. Therefore, in a *proactive* approach, rather than reactively responding to incoming demands and serving them when requested, an SBS can predict traffic patterns and determine future off-peak times so that incoming traffic demand can be properly allocated over a given time window.

Therefore, the main motivation for adopting a proactive LTE-WiFi coexistence scheme is to avoid the underutilization of the unlicensed spectrum during off-peak hours. This is mainly accomplished by either serving a fraction of the LTE-U traffic when requested or shifting part of it to the future, over a given time window, so as to balance the occupancy of the unlicensed spectrum usage across time and, consequently, improve its degree of utilization. From the LTE-U network perspective, this will increase its transmission opportunities on the unlicensed spectrum, reduce the collision probability with WAPs and other SBSs and, hence, provide a boost for its throughput. Moreover, a proactive resource allocation scheme can exploit the inherent predictability of the future channel availability status so as to allocate resources in a window of time slots based on the predicted requests. This, in turn, can lead to a decrease in the probability of occurrence of a congestion event while ensuring a degree of fairness to the wireless local area network (WLAN).

The main contribution of this paper is a novel deep reinforcement learning algorithm based on long short-term memory (RL-LSTM) cells for proactively allocating LTE-U resources over the unlicensed spectrum. The LTE-U resource allocation problem is formulated as a noncooperative game in which the players are the SBSs. To solve this game, we propose an RL-LSTM framework using which the SBSs can autonomously learn which unlicensed channels to use along with the corresponding channel access probability on each channel taking into account future environmental changes, in terms of WLAN activity on the unlicensed channels and LTE-U traffic loads. Unlike previous studies which are either centralized [9] or rely on the coordination among SBSs [4], our approach is based on a self-organizing proactive resource allocation scheme in which the SBSs utilize past observations of the network state to build predictive models on spectrum availability and to intelligently plan channel usage over a finite time window. The use of long short term memory (LSTM) cells enables the SBSs to predict a sequence of interdependent actions over a long-term time horizon thus achieving long-term fairness among different underlying technologies. We show that, upon convergence, the proposed algorithm

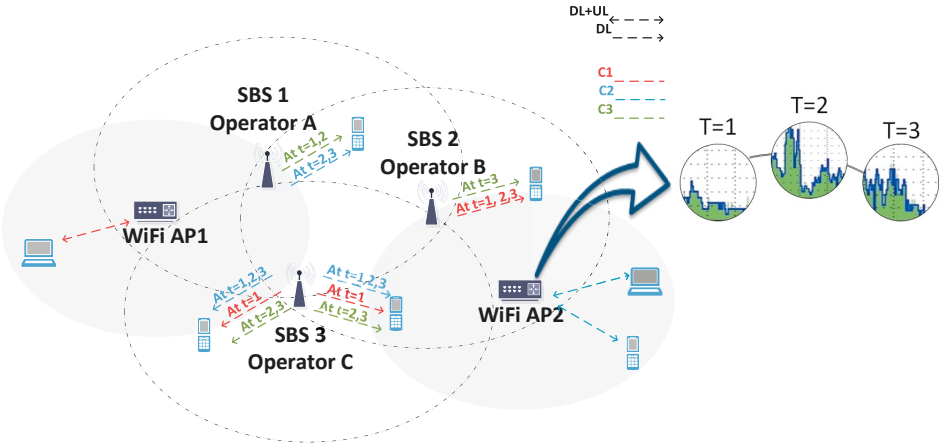

Fig. 1: Illustration of the system model. In the above example, 3 SBSs belonging to different operators and 3 unlicensed channels are only shown for simplicity. The channel selection vector over a time window of 3 epochs is also shown.

reaches to a mixed-strategy distribution which constitutes a mixed-strategy Nash equilibrium (NE) for the studied game. We also show that the gain of the proposed proactive resource allocation scheme and the optimal size of the prediction time window is a function of the traffic pattern of the dataset under study. To the best of our knowledge, *this is the first work that exploits the framework of LSTMs for proactive resource allocation in LTE-U networks*. Simulation results show that the proposed approach yields significant rate improvements compared to conventional reactive solutions such as instantaneous equal weighted fairness, proportional fairness and total network throughput maximization. The results also show that the proposed scheme prevents disruption to WLAN operation in the case large number of LTE operators selfishly deploy LTE-U in the unlicensed spectrum. In terms of priority fairness, results show that an efficient utilization of the unlicensed spectrum is guaranteed when both technologies, LTE-U and WLAN, are given equal weighted priorities for transmission on the unlicensed spectrum.

The rest of this paper is organized as follows. In Section II, we present the system model. Section III describes the proposed coexistence game model. The LSTM-based algorithm is proposed in Section IV. In Section V, simulation results are analyzed. Finally, conclusions are drawn in Section VI.

II. System Model

Consider the downlink of an LTE-U network composed of a set $\mathcal{J}$ of J LTE-U SBSs belonging to different LTE operators, a set $\mathcal{W}$ of W WAPs, and a set $\mathcal{C}$ of C unlicensed channels as shown in Fig. 1. Each SBS $j \in \mathcal{J}$ has a set $\mathcal{K}_j$ of K_j LTE-U UEs associated with it. We focus on

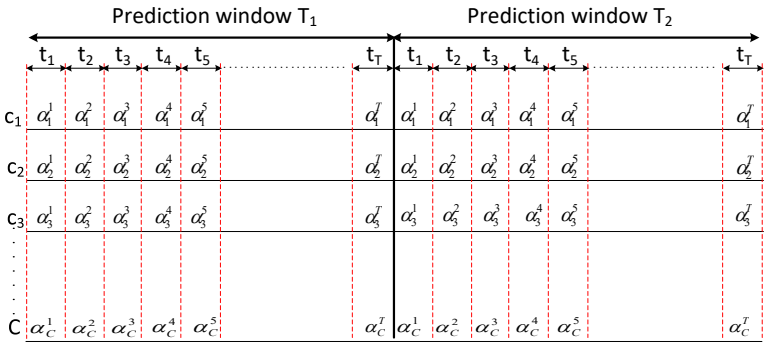

Fig. 2: The division of the time domain into multiple time windows T, each of which consists of multiple time epochs t.

the operation of the SBSs over the unlicensed band, while the licensed spectrum resources are assumed to be allocated in a conventional way [12]. Both SBSs and WAPs adopt the LBT access scheme and, thus, at a particular time, a given unlicensed channel is occupied by either an SBS or a WAP. We consider the LTE carrier aggregation feature using which the SBSs can aggregate up to five component carriers belonging to the same or different operating frequency bands [13]. This, in turn, would enable the SBSs to operate on multiple unlicensed channels simultaneously thus maximizing their data rate during a particular transmission opportunity.

Our goal is to jointly determine the dynamic channel selection, carrier aggregation, and fractional spectrum access for each SBS, while guaranteeing long-term airtime fairness with WLAN and other LTE-U operators. The main motivation for adopting a long-term fairness approach is to avoid the underutilization of the unlicensed spectrum by either serving part of the LTE-U traffic when requested or shifting part of it in the future over a given time window in a way that would balance the occupancy of the unlicensed spectrum usage across time and, consequently, improve its degree of utilization. This will subsequently result in an increase in the transmission opportunities for LTE-U as well as a decrease in the collision probability for the WLAN. To realize this, we need to dynamically analyze the usage of various unlicensed channels over a particular time window. To this end, we divide our time domain into multiple time windows of duration T, each of which consists of multiple time epochs t, as shown in Fig. 2. Our objective is to proactively determine the spectrum allocation vector for each SBS at $t = 0$ over T while guaranteeing long-term equal weighted airtime share with WLAN. To guarantee a fair spectrum allocation among SBSs belonging to different operators, we consider inter-operator interference along with inter-technology interference. In fact, inter-operator interference is the

consequence of the selfish behavior of different operators and could result in a degradation in the spectral efficiency if not managed. Next, we define $x_{j,c,t} = 1$ if channel c is selected by SBS j during time epoch t, and 0, otherwise, and $\alpha_{j,c,t} \in [0, 1]$. $x_{j,c,t}$ determines the channel c that is used by SBS j during time t and $\alpha_{j,c,t}$ is the channel access probability of SBS j on the unlicensed channel c at time t.

A contention-based protocol is used for channel access over the unlicensed band. In this protocol, prior to transmission, the SBS applies clear channel assessment to detect the state of the channel (idle or busy) based on the detected energy level. If the channel is idle, the SBS gets a transmit opportunity for up to 10 LTE sub-frames; otherwise, it keeps monitoring the channel until it becomes idle. We consider an exponential backoff scheme for WiFi while the SBSs adjust their CW size (and thus the channel access probability) on each of the selected channels in a way that would guarantee a long-term equal weighted fairness with WLAN and other SBSs. In fact, small CW sizes lead to an increase in the collision probability while large CW sizes result in too much time spent waiting in idle slots. Therefore, an efficient access method should adapt the value of the CW to the current traffic conditions.

To derive the throughput achieved by an LTE-U user equipment (UE) and a WAP, we first define the stationary probability of each WAP w and each SBS j, τ_w and $\tau_{j,c,t}$ respectively. The stationary probability is the probability with which a given base station attempts to transmit in a randomly chosen slot. Considering an exponential backoff scheme for WiFi, the stationary probability with which WAPs transmit a packet, τ_w, will be given by [14]:

$$\tau_w = \frac{2(1-2q_w)}{(1-2q_w)(\mathrm{CW}_{\min}+1) + q_w \mathrm{CW}_{\min}(1-(2q_w)^m)}, \tag{1}$$

where q_w is the collision probability of a WAP, m is the maximum backoff stage where $\mathrm{CW}_{\max} = 2^m \mathrm{CW}_{\min}$. $\mathrm{CW}_{\min}$ and $\mathrm{CW}_{\max}$ are the minimum and maximum contention window size, respectively. For LTE-U, m=0 since no exponential backoff is considered, and, thus, the stationary probability of an SBS on a given unlicensed channel c during time epoch t will be $\tau_{j,c,t} = \frac{2}{\mathrm{CW}_{j,c,t}+1}$, where $\mathrm{CW}_{j,c,t}$ is the contention window size of SBS j on channel c during time epoch t. Therefore, we do not consider a contention stage for LTE-U. Instead, the SBSs adjust their CW size adaptively to control their channel access probability over the unlicensed band. The collision probability of a WAP is defined as $q_w = 1 - \prod_{v=1, v\neq w}^{W}(1-\tau_v)\prod_{j=1}^{J}(1-\tau_{j,c,t})$, where c is the channel used by WAP w. The throughput R_w of a WAP w will be:

$$R_w = \frac{P_{w,\text{succ}} \cdot E[D_w]}{P_{w,\text{idle}} \cdot \theta + P_{w,\text{busy}} \cdot T_b}, \tag{2}$$

where $E[D_w]$ is the expected payload size for WAP w, $P_{w,\text{succ}} = \tau_w \prod_{v=1,v\neq w}^{W}(1-\tau_v)\prod_{j=1}^{J}(1-\tau_{j,c,t})$ is the probability of a successful transmission, $P_{w,\text{idle}} = \prod_{j=1}^{J}(1-\tau_{j,c,t})\prod_{w=1}^{W}(1-\tau_w)$ is the probability of an idle slot, and $P_{w,\text{busy}} = 1-\prod_{j=1}^{J}(1-\tau_{j,c,t})\prod_{w=1}^{W}(1-\tau_w)$ is the probability of a busy slot, regardless of whether it corresponds to a collision or a successful transmission. θ and T_b are, respectively, the average durations of an idle and a busy slot and, thus, the denominator in (2) corresponds to the mean duration of a WiFi medium access control (MAC) slot.

The achievable airtime fraction for an LTE-U SBS j on channel c at time t is:

$$\alpha_{j,c,t} = \tau_{j,c,t} \prod_{i=1,i\neq j}^{J}(1-\tau_{i,c,t}) \prod_{w=1}^{W}(1-\tau_w). \tag{3}$$

The airtime fraction represents the time allocated for an SBS on channel c during time t. Thus, the total throughput of all $K_{j,t}$ UEs that are served by SBS j during time epoch t is:

$$R_{j,t} = \sum_{c=1}^{C} \alpha_{j,c,t} r_{j,c,t}, \tag{4}$$

where

$$r_{j,c,t} = \sum_{k=1}^{K_{j,t}} B_c \log\Big(1+\frac{P_{j,c,t}h_{j,k,c,t}}{I_{j,c,t}+B_cN_0}\Big). \tag{5}$$

Here, $I_{j,c,t} = \sum_{w=1}^{W} P_{w,c,t}h_{w,k,c,t} + \sum_{i=1,i\neq j}^{J} P_{i,c,t}h_{i,k,c,t}$ is the interference level on SBS j when operating on channel c during time t and B_c is the bandwidth of channel c. $P_{j,c,t}$ is the transmit power of SBS j on channel c during time t. $h_{j,k,c,t}$ is the channel gain between SBS j and UE k on channel c during time t. N_0 is the power spectral density of additive white Gaussian noise. Since SBSs and WAPs both adopt LBT, then one cell may occupy the entire channel at a given time thus transmitting *exclusively* on a given channel c. However, hidden and exposed terminals could be present on a given channel which can result in interference and thus a degradation in the throughput.

Given this system model, next, we develop an effective spectrum allocation scheme that can allocate the appropriate unlicensed channels along with the corresponding channel access probabilities to each SBS simultaneously over T, at $t = 0$.

III. Proactive Resource Allocation Scheme for Unlicensed LTE

A. Proactive Resource Allocation Game

We formulate the resource allocation problem as a noncooperative game $\mathcal{G}$=$(\mathcal{J}, \mathcal{A}_j, u_j)$ with the SBSs in $\mathcal{J}$ being the players, each of which must choose a channel selection and channel

access pair $a_{j,c,t} = (x_{j,c,t}, \alpha_{j,c,t}) \in \mathcal{A}_j$ at $t = 0$ for each t of the next time window T. The objective of each SBS j is to maximize its total throughput over the selected channels:

$$u_j(\boldsymbol{a}_j, \boldsymbol{a}_{-j}) = \sum_{t=1}^{T} \sum_{c=1}^{C} \alpha_{j,c,t} r_{j,c,t}, \tag{6}$$

where $\boldsymbol{a}_j = [(a_{j,1,1}, \cdots, a_{j,1,T}), \cdots, (a_{j,C,1}, \cdots, a_{j,C,T})]$ and $\boldsymbol{a}_{-j}$ correspond, respectively to the action vector of SBS j and all other SBSs, over all the channels $\mathcal{C}$ during T. Note that the utility function (6) of SBS j depends on its actions as well as those of other SBSs which makes the formulation of a game model suitable for this problem. This is mainly due to the interference from other SBSs transmitting on the same channel as SBS j as it was shown previously in the definition of the rate expression in (5). The goal of each SBS j is to maximize its own utility:

$$\max_{a_j \in \mathcal{A}_j} u_j(\boldsymbol{a}_j, \boldsymbol{a}_{-j}) \quad \forall j \in \mathcal{J}, \tag{7}$$

$$\text{s.t.} \qquad \alpha_{j,c,t} \leq x_{j,c,t} \quad \forall c, t, \tag{8}$$

$$\sum_{c=1}^{C} x_{j,c,t} \leq \min(M_c, C) \quad \forall t, \tag{9}$$

$$\sum_{t_T=1}^{t} \sum_{c=1}^{C} \alpha_{j,c,t_T} B_c \leq \sum_{t_T=1}^{t} f(L_{j,t_T}) \quad \forall t, \tag{10}$$

$$\alpha_{w,c,t} + \alpha_{j,c,t} + \sum_{i=1, i \neq j}^{J} \alpha_{i,c,t} \leq t_{\max} \quad \forall c, t, \tag{11}$$

$$x_{j,c,t} \in \{0, 1\}, \quad \alpha_{j,c,t} \in [0, 1] \quad \forall c, t, \tag{12}$$

where M_c is the total number of unlicensed channels which an SBS can aggregate. (8) allows the allocation of a channel access proportion for SBS j on channel c during t only if SBS j transmits on channel c at time t. (9) guarantees that each SBS can aggregate a maximum of M_c channels at a given time t. (10) limits the amount of allocated bandwidth to the required demand where $f(L_{j,t})$ captures the relationship between bandwidth requirement and offered load. (11) captures coupling constraints which limit the proportion of time used by SBSs and WLAN on a given unlicensed band to the maximum fraction of time an unlicensed channel can be used, $t_{\max}$[1]. (12) represents the feasibility constraints.

Given the fact that different operators and technologies have equal priorities on the unlicensed spectrum, we incorporate the Homo Egualis (HE) anthropological model, an inequity-averse based fairness model, into the strategy design of the agents [15].

[1] $t_{\max}$ depends on the channel access method in the unlicensed band and should be strictly less than 1 in the case of LBT, otherwise, the channel will always be sensed busy and devices would not be able to access it.

Definition 1. *Inequity aversion* is the preference for fairness and resistance to incidental inequalities. In other words, it refers to the willingness of giving up some material payoff in order to move in the direction of more equitable outcomes.

In an HE society, agents focus not only on maximizing their own payoffs, but also become aware of how their payoffs are compared to other agents' payoffs [15], [16]. Therefore, their utility function is influenced not only by their own reward, but also by envy and altruism. An agent is altruistic to others if its payoff is above an equitable benchmark and is envious of the others if its payoff exceeds that benchmark and therefore, an unfair distribution of resources among agents results in disutility for inequity-averse agents. The HE concept comes from the anthropological literature in which Homo sapiens evolved in small hunter-gatherer groups without a centralized governance [15].

In fact, we incorporate the notion of airtime fairness in the modeling of our HE agents. The average airtime per radio system is considered as one of the most important fairness metrics in the unlicensed band and is the focus of this work [17]. Our motivation for considering a time-fair channel allocation scheme is to overcome the rate anomaly problem that arises when different nodes use distinct data rates, which leads to the slowest link limiting the system performance [4], [17], and [18]. Therefore, to model our players as HE agents, we consider the following two coupling constraints for the allocated airtime fraction on each channel c for each SBS j:

$$\frac{1}{w_{j,c}}\frac{1}{T}\frac{\sum_{t=1}^{T}\alpha_{j,c,t}}{\sum_{t=1}^{T}\bar{L}_{j,t}}=\frac{1}{w_{i,c}}\frac{1}{T}\frac{\sum_{t=1}^{T}\alpha_{i,c,t}}{\sum_{t=1}^{T}\bar{L}_{i,t}}\ \forall c\in\widehat{\mathcal{C}}_j, i\in\widehat{\mathcal{S}}_{j,c}(i\neq j), \tag{13}$$

$$\frac{1}{T}\frac{\sum_{t=1}^{T}\sum_{n\in\mathcal{S}_{c,t}}\alpha_{n,c,t}}{P_{\text{LTE}}\sum_{t=1}^{T}\sum_{n\in\mathcal{S}_{c,t}}\bar{L}_{n,t}}=\frac{1}{T}\frac{\sum_{t=1}^{T}\alpha_{w,c,t}}{P_{\text{WiFi}}\sum_{t=1}^{T}L_{w,c,t}}\ \ \forall c\in\widehat{\mathcal{C}}_j, \tag{14}$$

where $\widehat{\mathcal{C}}_j$ is the subset of channels used by SBS j during T. $\mathcal{S}_{c,t}$ is the subset of SBSs that are transmitting over channel c, $c\in\widehat{\mathcal{C}}_j$, during time t and $\widehat{\mathcal{S}}_{j,c}$ is the subset of other neighboring SBSs, $i\neq j$, that are using the same channel $c\in\widehat{\mathcal{C}}_j$ as SBS j during T. $\bar{L}_{j,t}=L_{j,t}-\sum_{c'}f(\alpha_{j,c',t})$ corresponds to the remaining traffic that needs to be served by SBSs j with $L_{j,t}$ being the *total* aggregate traffic demand of SBS j on channel c during time epoch t. $f(.)$ corresponds to the served traffic load as a function of the airtime allocation. c' represents all the set of channels except channel c. $\alpha_{w,c,t}=\min(f(L_{w,c,t}),\, t_{\max}-\alpha_{j,c,t}-\sum_{i\in\mathcal{S}_{j,c,t}}\alpha_{i,c,t})$ is the airtime allocated for WLAN transmissions over channel c during time t. P_{WiFi} and P_{LTE} correspond to the priority metric defined for each technology when operating on the unlicensed band. These parameters allow adaptation of the level of fairness between LTE-U and WLAN.

Constraint (13) represents inter-operator fairness which guarantees an equal weighted airtime allocation among SBSs belonging to different operators on a given channel c. The adopted notion of fairness is based on a long-term weighted equality over T, as opposed to instantaneous weighted equality. $w_{j,c} = \sum_{t=1}^{T} x_{j,c,t}$ is the weight of SBS j on channel c during T and thus different SBSs are assigned different weights on each channel c based on the number of time epochs t a given SBS j uses that particular channel. (14) defines an inter-technology fairness metric to guarantee a long-term equal weighted airtime allocation over T for both LTE-U and WiFi. Therefore, (13) and (14) reflect the inequity aversion property of the SBSs.

In fact, the optimal value of T, which corresponds to the time window size that allows the maximum achievable throughput for LTE-U as compared to the reactive approach, is dataset dependent. Next, we characterize the optimal value of T under a uniform traffic distribution.

Proposition 1. For a uniform traffic distribution, the optimal value of T is equal to 1.

Proof. Under a uniform demand model, the traffic load for each of SBS j and WAP w is an independent and identically distributed (i.i.d.) sequence of random variables which implies that all requests of the same user are statistically indistinguishable over time. In our model, WAPs are passive in that their channel selection action is fixed and, thus, the activity on a given channel is characterized by the level of activity of WAPs operating on that channel. In that case, the WLAN traffic load on each channel also follows a uniform distribution. At the convergence point, (8)-(14) are satisfied and, hence, the average airtime allocated to the LTE-U network on channel c over the time window T will be:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{j\in\mathcal{S}_{c,t}}\alpha_{j,c,t} = \frac{P_{\text{LTE}}}{P_{\text{WiFi}}}\frac{\sum_{t=1}^{T}\sum_{j\in\mathcal{S}_{c,t}}\bar{L}_{j,t}}{\sum_{t=1}^{T}L_{w,c,t}}\frac{1}{T}\sum_{t=1}^{T}\alpha_{w,c,t} \quad \forall c\in\mathcal{C}, \tag{15}$$

However, for the case of uniform traffic demand, the channel selection vector over T is the same for each SBS because the network state is the same for every t in T. Moreover, if an SBS aggregates multiple channels, then its load on each channel is the same for each t in T. This implies that $\bar{L}_{j,t}$ for each SBS j is uniform over T and thus $\frac{\sum_{t=1}^{T}\sum_{j\in\mathcal{S}_{c,t}}\bar{L}_{j,t}}{\sum_{t=1}^{T}L_{w,c,t}} = \frac{\sum_{j\in\mathcal{S}_{c,t}}\bar{L}_{j,t}}{L_{w,c,t}}$. Consequently, (15) can be written as:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{j\in\mathcal{S}_{c,t}}\alpha_{j,c,t} = \frac{P_{\text{LTE}}}{P_{\text{WiFi}}}\frac{\sum_{j\in\mathcal{S}_{c,t}}\bar{L}_{j,t}}{L_{w,c,t}}\frac{1}{T}\sum_{t=1}^{T}\alpha_{w,c,t} \quad \forall c\in\mathcal{C}, \tag{16}$$

When $T = 1$, the airtime allocated to the LTE-U network on channel c will be:

$$\sum_{j\in\mathcal{S}_{c,t}}\alpha_{j,c,t} = \frac{P_{\text{LTE}}}{P_{\text{WiFi}}}\frac{\sum_{j\in\mathcal{S}_{c,t}}\bar{L}_{j,t}}{L_{w,c,t}}\alpha_{w,c,t} \quad \forall t, c\in\mathcal{C}, \tag{17}$$

Over a fixed time window T, the average airtime allocated to the LTE-U network on channel c can be written as:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{j\in\mathcal{S}_{c,t}}\alpha_{j,c,t}=\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{P_{\text{LTE}}}{P_{\text{WiFi}}}\frac{\sum_{j\in\mathcal{S}_{c,t}}\bar{L}_{j,t}}{L_{w,c,t}}\alpha_{w,c,t}\Big) \tag{18}$$

$$=\frac{P_{\text{LTE}}}{P_{\text{WiFi}}}\frac{\sum_{j\in\mathcal{S}_{c,t}}\bar{L}_{j,t}}{L_{w,c,t}}\frac{1}{T}\sum_{t=1}^{T}\alpha_{w,c,t}\ \ \forall c\in\mathcal{C}. \tag{19}$$

(19) is equivalent to (16) and, hence, our proposed framework does not offer any gain for the LTE-U network when considering a time window T larger than 1 in the case of a uniform traffic pattern. This completes the proof. ■

From Proposition 1, we conclude that the gain of our proposed long-term fairness notion is evident in the case of traffic fluctuations. Under a uniform traffic distribution, the SBSs cannot make use of future off-peak times to shift part of their traffic forward in time and, hence, the gain is limited to predicting the network state for the next time epoch only. It is also worth noting that the gain of the proactive scheduling approach decreases in the case of a highly congested WLAN network. This is mainly due to the fact that the system becomes more congested with incoming requests, thereby restricting the opportunities of shifting part of the LTE-U load in the future.

B. Equilibrium Analysis

Our game $\mathcal{G}$ belongs to the family of generalized Nash equilibrium problems (GNEPs) in which both the objective functions and the action spaces are coupled. To solve the GNEP, we incorporate the Lagrangian penalty method into the utility functions thus reducing it to a simpler Nash equilibrium problem (NEP). The resulting penalized utility function will be given by, $\forall(j\in\mathcal{J})$:

$$\begin{aligned}\widehat{u}_j(\boldsymbol{a}_j,\boldsymbol{a}_{-j})=\sum_{t=1}^{T}\sum_{c=1}^{C}\alpha_{j,c,t}r_{j,c,t}&-\rho_{1,j}\sum_{c=1}^{C}\sum_{t=1}^{T}\Big(\min(0,t_{\max}-\alpha_{w,c,t}-\alpha_{j,c,t}-\sum_{i=1,i\neq j}^{J}\alpha_{i,c,t})\Big)^2\\&-\rho_{2,j}\sum_{c\in\widehat{\mathcal{C}}_j}\sum_{i\in\widehat{\mathcal{S}}_{j,c}(i\neq j)}\frac{1}{T^2}\left(\frac{1}{w_{j,c}}\frac{\sum_{t=1}^{T}\alpha_{j,c,t}}{\sum_{t=1}^{T}\bar{L}_{j,t}}-\frac{1}{w_{i,c}}\frac{\sum_{t=1}^{T}\alpha_{i,c,t}}{\sum_{t=1}^{T}\bar{L}_{i,t}}\right)^2\\&-\rho_{3,j}\sum_{c\in\widehat{\mathcal{C}}_j}\frac{1}{T^2}\left(\frac{\sum_{t=1}^{T}\sum_{n\in\mathcal{S}_{c,t}}\alpha_{n,c,t}}{P_{\text{LTE}}\sum_{t=1}^{T}\sum_{n\in\mathcal{S}_{c,t}}\bar{L}_{n,t}}-\frac{\sum_{t=1}^{T}\alpha_{w,c,t}}{P_{\text{WiFi}}\sum_{t=1}^{T}L_{w,c,t}}\right)^2,\end{aligned} \tag{20}$$

where $\rho_{1,j}$, $\rho_{2,j}$ and $\rho_{3,j}$ are positive penalty coefficients corresponding to constraints (11), (13), and (14), respectively. For our reformulation, we consider equal penalty coefficients for all players for each coupled constraint, $\rho_{1,j} = \rho_1$, $\rho_{2,j} = \rho_2$ and $\rho_{3,j} = \rho_3$. This allows all SBSs to have equal incentives to give up some payoff in order to satisfy the coupled constraints. To determine the values of ρ_1, ρ_2 and ρ_3, we adopt the incremental penalty algorithm in [19] where it has been shown that there exists some penalty parameters $\boldsymbol{\rho}_l^* = [\rho_1^*, \rho_2^*, \rho_3^*]$ at which the coupled constraints can be satisfied.

In our game $\mathcal{G}$, $\alpha_{j,c,t}$ is a continuous variable bounded between 0 and 1, however, for a particular network state, we are interested only in a certain region of the continuous space where the optimal actions are expected to be. Therefore, we will propose a sampling-based approach to discretize $\alpha_{j,c,t}$ in Section IV. Under such a discretization of the action space, we turn our attention to mixed strategies in which players choose their strategies probabilistically. Such a mixed-strategy approach enables us to analyze the frequency with which players choose different channels and channel access combinations. In fact, the optimal policy is often stochastic and therefore requires the selection of different actions with specific probabilities [20]. This, in turn, validates our choice of adopting a mixed strategy approach as opposed to a pure strategy one that is oriented towards finding deterministic policies. A player can possibly choose different possible actions with different probabilities which enables it to play a combination of strategies over time. Moreover, unlike pure strategies that might not exist for a particular game, there always exists at least one equilibrium in mixed strategies [21].

Let $\Delta(\mathcal{A})$ be the set of all probability distributions over the action space $\mathcal{A}$ and $\boldsymbol{p}_j = [p_{j,\boldsymbol{a}_1} \cdots, p_{j,\boldsymbol{a}_{|\mathcal{A}_j|}}]$ be a probability distribution with which SBS j selects a particular action from $\mathcal{A}_j$. Therefore, our objective is to maximize the expected value of the utility function, $\overline{u}_j(\boldsymbol{p}_j, \boldsymbol{p}_{-j}) = \mathbb{E}_{\boldsymbol{p}_j}\left[\widehat{u}_j\left(\boldsymbol{a}_j, \boldsymbol{a}_{-j}\right)\right] = \sum_{\boldsymbol{a}\in\mathcal{A}} \widehat{u}_j(\boldsymbol{a}_j, \boldsymbol{a}_{-j}) \prod_{j=1}^{J} p_{j,\boldsymbol{a}_j}$.

Definition 2. A mixed strategy $\boldsymbol{p}^*{=}(\boldsymbol{p}_1^*, \cdots, \boldsymbol{p}_J^*){=}(\boldsymbol{p}_j^*, \boldsymbol{p}_{-j}^*)$ constitutes a *mixed-strategy Nash equilibrium* if, $\forall j \in \mathcal{J}$ and $\forall \boldsymbol{p}_j \in \Delta(\mathcal{A}_j)$, $\overline{u}_j(\boldsymbol{p}_j^*, \boldsymbol{p}_{-j}^*) \geq \overline{u}_j(\boldsymbol{p}_j, \boldsymbol{p}_{-j}^*)$.

Here, we note that any finite noncooperative game will admit at least one mixed-strategy Nash equilibrium [21]. To solve for the mixed-strategy NE of our game $\mathcal{G}$, we first consider the simpler scenario in which the number of SBSs is less than the number of unlicensed channels. Then, we develop a learning algorithm to handle the more realistic scenario in which the number of SBSs is much larger than the number of unlicensed channels.

Remark 1. If the number of SBSs is less than the number of available unlicensed channels (i.e., $J \leq C$), then the mixed-strategy NE solution will simply reduce to a pure strategy that is reached when all SBSs occupy disjoint channels during each time epoch of the time window T.

To show this, we consider two cases depending on whether or not carrier aggregation is enabled. Let $M_c = 1$. Consider the state in which each SBS is operating on a different unlicensed channel. If SBS j changes its channel from c to c' on which SBS i is transmitting, then it would have to share channel c' with SBS i in an equal weighted manner (based on the inter-operator fairness constraint). This leads to a decrease in the reward function of SBS i on channel c' (and potentially for SBS j), which makes SBS i deviate to another channel that is less occupied (e.g., c). Therefore, a given strategy cannot be a best response (BR) strategy for SBS i in case it results in its transmission on the same channel as SBS j. Therefore, all strategies that result in more than one SBS occupying the same channel are dominated by the alternative where different SBSs transmit on disjoint channels and hence cannot correspond to BR strategies. Consequently, at the NE point, all SBSs play their BR strategies that would result in each SBS occupying a disjoint channel. Similarly for $M_c > 1$. If SBS j transmits on multiple channels, then aggregating a channel that is already occupied by SBS i would make SBS i change its operating channel to a less congested one. This implies that an SBS would not aggregate more channels unless they are not occupied by other SBSs.

Therefore, we can conclude that our proposed scheme results in having less number of SBSs on each of the unlicensed bands. This leads to a lower collision probability on each channel and a better coexistence with WLAN. Moreover, enabling carrier aggregation does not necessarily allow LTE to offload more traffic to the unlicensed band. On the other hand, our proposed scheme can avoid causing performance degradation to WLAN in case a large number of LTE operators deploy LTE-U in the unlicensed bands.

Now, when $J > C$, multiple SBSs will then potentially have to share the same channel. In this case, the mixed-strategy NE is challenging to characterize, and therefore, next, we propose a learning-based approach for solving our game $\mathcal{G}$. Given the fact that each SBS needs to learn a *sequence* of actions over the time window T at $t = 0$ based on a *sequence* of previous network states, the proposed learning algorithm must be capable of generating data that is sequential in nature. This necessitates the knowledge of historical traffic values as well as future network states for all the time epochs of the following time window T. Moreover, in order to satisfy the

long-term fairness constraints (13) and (14), future actions cannot be assumed to be independent due to the long-term temporal dependence among these actions. Conventional reinforcement learning algorithms such as Q-learning and multi-armed bandit take as an input the current state of the network and enable the prediction of the next state only and therefore do not account for the interdependence of future actions [22]. To learn several steps ahead in time, recursive learning can be adopted. However, such an approach uses values already predicted, instead of measured past values which produces an accumulation of errors that may grow very fast. In contrast, deep learning techniques, such as time series prediction algorithms, are capable of learning long-term temporal dependence sequences based on input sequences [23]. This is viable due to their adaptive memory that allows them to store necessary previous state information to predict future events. Therefore, next, we develop a novel time series prediction algorithm based on deep learning techniques for solving the mixed-strategy NE of our game.

IV. RL-LSTM FOR SELF-ORGANIZING RESOURCE ALLOCATION

The proposed game requires each SBS to learn a sequence of actions over the prediction time window T, at $t = 0$, without any knowledge of future network states. This necessitates a learning approach with memory for storing previous states whenever needed while being able to learn a sequence of future network states. Employing LSTMs is therefore an obvious choice for learning as they are capable of generating data that is sequential in nature [23]. Consequently, we propose a novel sequence level training algorithm based on RL-LSTM that allows SBSs to learn a sequence of future actions at operation time based on a sequence of historic traffic load thus maximizing the sum of their future rewards.

LSTMs are a special kind of "deep" recurrent neural networks (RNNs) that are capable of storing information for long periods of time and hence learning the long-term dependency within a given sequence [24]. In essence, LSTMs process a variable-length sequence $\boldsymbol{y} = (y_1, y_2, ..., y_m)$ by incrementally adding new content into a single memory slot, with gates controlling the extent to which new content should be memorized, old content should be erased, and current content should be exposed. Unlike conventional RL techniques (e.g., Q-learning) that can learn the action for the next state only, LSTM networks are capable of predicting a sequence of future actions [23]. Predictions at a given time step are influenced by the network activations at previous time steps thus making LSTMs suitable for our application in which an action at time t depends

on all previous and future actions within the current window T. The total number of parameters W in a standard LSTM network with one cell in each memory block is given by:

$$W = n_c \times n_c \times 4 + n_i \times n_c \times 4 + n_c \times n_o + n_c \times 3 \tag{21}$$

where n_c is the number of memory cells, n_i is the number of input units, and n_o is the number of output units. The computational complexity of learning LSTM models per weight and time step is linear i.e., $O(1)$. Therefore, the learning computational complexity per time step is $O(W)$ [25].

Consequently, we consider an end-to-end RL-LSTM based approach to train the network to find a mixed-strategy NE of the game $\mathcal{G}$. LSTMs have three types of layers, one input and one output layer as well as a varying number of hidden layers depending on the dataset under study. For our dataset, adding more hidden layers does not improve performance and thus one layer is sufficient. Moreover, in order to allow a sequence to sequence mapping, we consider an encoder-decoder model. The encoder network takes the input sequence and maps it to a vector of a fixed dimensionality. The encoded representation is then used by the decoder network to decode the target sequence from the vector. Fig. 3 summarizes the proposed approach. The traffic encoder takes as an input the historical traffic loads and learns a vector representation of the input time-series. The multi-layer perceptron (MLP) summarizes the input vectors into one vector. In our scheme, an MLP is required to encode all the vectors together since a particular action at time t depends on the values of all other input vectors (i.e., traffic values of all SBSs and WLAN on all the unlicensed channels). The action decoder takes as an input the summarized vector to reconstruct the predicted action sequence. All SBSs have the same input vector for the traffic encoders and thus they share the same traffic encoders. On the other hand, SBSs learn different action sequences and thus different SBSs use different action decoders.

In the first step, we need to train the neural networks in order to learn the parameters of the algorithm that would maximize the proposed utility function. Therefore, the proposed algorithm is divided into *two phases, the training phase followed by the testing phase*. In the former, SBSs are trained offline before they become active in the network using the architecture given in Fig. 3. The input dataset represents the WiFi traffic load distribution on the unlicensed channels as well as the SBSs traffic load collected over several days. On the other hand, the testing phase corresponds to the actual execution of the algorithm after which the parameters have been optimized and is implemented on each SBS for execution during run time.

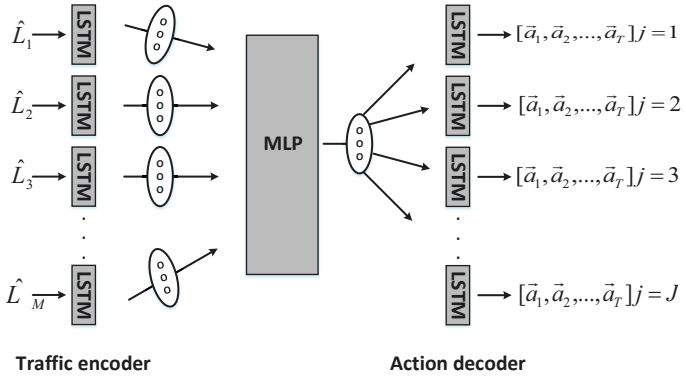

Fig. 3: Proposed framework.

For the training phase, we train the weights of our neural network using a policy gradient approach that aims at maximizing the expected return of a policy. This is achieved by representing the policy by its own function approximator and updating it according to the gradient of the expected reward with respect to the policy parameters [20]. Consider the set $\mathcal{M}$ of M history traffic sequences corresponding to either an SBS or WiFi on each unlicensed channel, where $M = J + C$. Let $\boldsymbol{h}_{m,t} \in \mathbb{R}^n$ and $\boldsymbol{h}_{j,t} \in \mathbb{R}^n$ be, respectively, the hidden vectors of the traffic encoder m and action decoder of SBS j at time t. $\boldsymbol{h}_{m,t}$ and $\boldsymbol{h}_{j,t}$ are then computed by:

$$\boldsymbol{h}_{m,t}=\phi\left(\boldsymbol{v}_{m,t}, \boldsymbol{h}_{m,t-1}\right), \quad \boldsymbol{h}_{j,t}=\phi\left(\boldsymbol{v}_{j,t}, \boldsymbol{h}_{j,t-1}\right), \tag{22}$$

where ϕ refers to the LSTM cell function [24] being used, and $\boldsymbol{v}_{m,t}$ is the input vector. For the encoder, $\boldsymbol{v}_{m,t} = \left[\widehat{L}_{m,t}\right]$ is the history traffic value. For the decoder, $\boldsymbol{v}_{j,t} = [\boldsymbol{W}_d \boldsymbol{e}(\boldsymbol{x}_{j,t-1}) || \alpha_{j,c,t-1}]$ is the vector of the previous predicted action where $\boldsymbol{e}()$ maps discrete value to a one-hot vector, $\boldsymbol{W}_d \in \mathbb{R}^{n \times N_x}$ is a matrix that is used to transform the discrete actions of each of the unlicensed channels into a vector, and N_x is the number of discrete actions. In our approach, we learn the channel selection vector for all the channels simultaneously and thus $\boldsymbol{x}_{j,t} = [x_{j,1,t}, \cdots, x_{j,C,t}]$.

To learn the mixed strategy of our proposed game, we need to initialize the action space with a subset of the continuous action space of $\alpha_{j,c,t}$. A naive approach for working with continuous action spaces is to discretize the action space; however, this approach would lead to combinatorial explosion and thus the well known problem of "curse of dimensionality" when highly discretizing our space and a loss in the accuracy of the predicted action when considering less discretized values. Therefore, we consider a sampling-based approach where we first define a probability distribution for the continuous variable $\alpha_{j,c,t}$ and for the discrete variable $x_{j,c,t}$ in order to deal with the large discrete action space as T increases. We use a softmax classifier to predict the distribution for the discrete variable $\boldsymbol{x}_{j,t}$ and a Gaussian policy for the distribution of

the continuous variable $\alpha_{j,c,t}$. For the Gaussian policy, the probability of an action is proportional to a Gaussian distribution with a parameterized mean and a fixed value for the variance in our implementation. The variance of the Gaussian distribution defines the area around the mean from which we explore the action space. For our implementation, the initial value of the variance is set to 0.06 in order to increase exploration and then is decreased linearly towards 0.02. Therefore, defining probability distributions for our variables allows the initialization of the action space $\mathcal{A}_j$ by sampling Z actions from the proposed distributions. This enables the SBSs to learn more accurate transmission probabilities for $\alpha_{j,c,t}$, as opposed to fixed discretization, thus satisfying the fairness constraints. The hidden vector $\boldsymbol{h}_{j,t}$ in the decoder is used to predict the t-th output actions $\boldsymbol{x}_{j,t}$ and $\alpha_{j,c,t}$. The probability vector over $\boldsymbol{x}_{j,t}$ and $\alpha_{j,c,t}$ can be defined, respectively, as:

$$\boldsymbol{x}_{j,t}|\boldsymbol{x}_{j,<t}, \alpha_{j,c,<t}, \widehat{\boldsymbol{L}}_t \sim \sigma\left(\boldsymbol{W}_x \boldsymbol{h}_{j,t}\right), \tag{23}$$

$$\mu_{j,c,t} = S\left(\boldsymbol{W}_\mu \boldsymbol{h}_{j,t}\right), \;\; \alpha_{j,c,t} \sim \mathcal{N}(\mu_{j,c,t}, \mathrm{Var}(\alpha_{j,c,t})), \tag{24}$$

where $\mu_{j,c,t}$ and $\mathrm{Var}(\alpha_{j,c,t})$ correspond to the mean value and variance of the Gaussian policy respectively, $\boldsymbol{W}_x \in \mathbb{R}^{|V_a| \times n}, \boldsymbol{W}_\mu \in \mathbb{R}^n$ are parameters, $\sigma(.)$ is the softmax function $\sigma(\boldsymbol{b})_q = \frac{e^{bq}}{\sum_{o=1}^{O} e^{bo}}$ for $q = 1, \cdots, O$, and $S(.)$ is the sigmoid function where $S(b) = \frac{1}{1+e^{-b}}$ and is used to normalize the value to $(0, 1)$. $\alpha_{j,c,t}$ is computed only when $x_{j,c,t} = 1$. The probability of the whole action sequence for SBS j, given a historic traffic sequence $\widehat{\boldsymbol{L}}$, $p_{j,\boldsymbol{a}_j|\widehat{\boldsymbol{L}}}$, is given by:

$$p_{j,\boldsymbol{a}_j|\widehat{\boldsymbol{L}}} = \prod_{t=1}^{T} p\left((\boldsymbol{x}_{j,t}, \alpha_{j,c,t})|\boldsymbol{x}_{j,<t}, \alpha_{j,c,<t}, \widehat{\boldsymbol{L}}_t\right), \tag{25}$$

where $\widehat{\boldsymbol{L}}_t$=$(\widehat{L}_{1,t}, \cdots, \widehat{L}_{M,t})$, $\boldsymbol{x}_{j,<t}$=$[\boldsymbol{x}_{j,1}, \cdots, \boldsymbol{x}_{j,t-1}]$, and $\mu_{j,c,<t}$=$[\mu_{j,c,1}, \cdots, \mu_{j,c,t-1}]$.

Our goal is to maximize the exact expectation of the reward $\widehat{u}_j(\boldsymbol{a}_j, \boldsymbol{a}_{-j})$ over the action space for the training dataset. Therefore, the objective function can be defined as:

$$\max_{a_j \in \mathcal{A}_j} \sum_{\mathcal{D}} \overline{u}_j(\boldsymbol{p}_j, \boldsymbol{p}_{-j}), \tag{26}$$

where $\mathcal{D}$ is the training dataset. For this objective function, the REINFORCE algorithm [26] can be used to compute the gradient of the expected reward with respect to the policy parameters, and then standard gradient descent optimization algorithms [20] can be adopted to allow the model to generate optimal action sequences for input history traffic values. Specifically, Monte Carlo sampling is adopted to compute the expectation.

In particular, we adopt the RMSprop gradient descent optimization algorithm for the update rule [27]. The learning rate of a particular weight is divided by a running average of the magnitudes of recent gradients for that weight. The RMSprop update rule is given by:

Algorithm 1 Training phase of the proposed approach.

Input: $\mathcal{J}; \mathcal{W}; \mathcal{C}; \widehat{L}_{j,t} \forall j \in \mathcal{J}, t; \widehat{L}_{w,c,t} \forall c \in \mathcal{C}$, t.

***Initialization*:** The weights of all LSTMs are initialized following a uniform distribution with arbitrarily small values.

***Training*:** Each SBS j is modeled as an LSTM network.

while Any of the coupled constraints is not satisfied **do**

  for Number of training epochs **do**

    for Size of the training dataset **do**

      Step 1. Run Algorithm 2 to compute the best actions for all SBSs.

      for j=1:J **do**

        Step 2. Sample actions for SBS j based on the best expected actions of other SBSs.

        Step 3. Use REINFORCE [26] to update rule and compute the gradient of the expected value of the reward function.

        Step 4. Update model parameters with back-propagation algorithm [28].

      end for

    end for

  end for

  Step 5. Using the incremental penalty algorithm, check the feasibility of the coupled constraints and update the values of $\boldsymbol{\rho}_l$ accordingly.

end while

Algorithm 2 Testing phase of the proposed approach.

Input: $\mathcal{J}; \mathcal{W}; \mathcal{C}; \widehat{L}_{j,t} \forall j \in \mathcal{J}, t; \widehat{L}_{w,c,t} \forall c \in \mathcal{C}$, t.

for For each SBS j **do**

  Step 1. ***Traffic history encoding*:** The history traffic of each SBS and WLAN activity on each channel is fed into each of the M LSTM traffic encoders.

  Step 2. ***Vector summarization*:** The encoded vectors are transformed to initialize action decoders.

  Step 3. ***Action decoding*:** Action sequence is decoded for each SBS j.

end for

$$\mathbb{E}[g^2]_t = \gamma \mathbb{E}[g^2]_{t-1} + (1-\gamma) g_t^2, \tag{27}$$

$$\theta_{t+1} = \theta_t - \frac{\lambda}{\sqrt{\mathbb{E}[g^2]_t + \epsilon}} g_t, \tag{28}$$

where θ_t corresponds to the model parameters at time t, g_t is the gradient of the objective function with respect to the parameter θ at time step t, $\mathbb{E}[g^2]_t$ is the expected value of the magnitudes of recent gradients, γ is the discount factor, λ is the learning rate and ϵ is a smoothing parameter.

On the other hand, the testing phase corresponds to the actual execution of the algorithm on each SBS. Based on historical traffic values, each SBS learns the future sequence of actions based on the learned parameters from the training phase. For practicality, we assume knowledge of historical measurements of the WiFi activity on each of the unlicensed channels through long-

term channel sensing[2] [5] and of other SBSs by exchanging past traffic information via the X2 interface. Consequently, the proposed algorithm offers a practical solution that is amenable to implementation. The training and the testing phases are given in Algorithms 1 and 2 respectively.

It is important to note that guaranteeing the convergence of the proposed deep learning algorithm is challenging as it is highly dependent on the hyperparameters used during the training phase. It has been shown in [30] that the learning rate and the hidden layer size are the two most important hyperparameters for the convergence of LSTMs. For instance, using too few neurons in the hidden layers results in underfitting which could make it hard for the neural network to detect the signals in a complicated data set. On the other hand, using too many neurons in the hidden layers can either result in overfitting [31] or an increase in the training time that could prevent the training of the neural network. Overfitting corresponds to the case when the model learns the random fluctuations and noise in the training data set to the extent that it negatively impacts the model's ability to generalize when fed with new data. Therefore, in this work, we limit our contribution to providing simulation results (see Section V) to show that, under a reasonable choice of the hyperparameters, convergence is observed for our proposed game. In such cases, it is important to characterize the convergence point of our proposed algorithm, which is given as follows.

Theorem 1. If Algorithm 1 converges, then the convergence strategy profile corresponds to a mixed-strategy NE of game $\mathcal{G}$.

Proof. In order to prove this theorem, we first need to show that the solution of the adopted multi-agent learning algorithm converges to an equilibrium point. In fact, every strict Nash equilibrium is a local optimum for a gradient descent learning approach but the reverse is not always true (Theorems 2 and 3 in [32]). Therefore, to show that a gradient-based learning method guarantees convergence of our proposed game to an equilibrium point, we define the following lemma.

Lemma 1. *The square of a linear function is convex. It follows that the payoff function of player j defined in (20) is an affine combination of convex functions, and hence is convex. Therefore,*

[2]We assume knowledge of WLAN historical traffic via long-term channel sensing for each of the C unlicensed channels using Simple Network Management Protocol (SNMP) statistics [29]. To realize this, an SNMP agent with multiple interfaces can run on each SBS to monitor the traffic values (i.e., byte counts) for periods of five minutes. This traffic is the total amount of bytes received and sent from all clients associated with each WAP during a particular time interval.

a gradient-based learning algorithm for our game $\mathcal{G}$ allows the convergence to an equilibrium point of that game.

Lemma 1 is the consequence of the convexity of the players' payoffs where it has been shown in [33] that under certain convexity assumptions about the shape of payoff functions, the gradient-descent process converges to an equilibrium point. However, convergence is only guaranteed under a decreasing step-size sequence [34]. Therefore, given the fact that we employ an adaptive learning rate method satisfying the Robbins-Monro conditions $(\lambda > 0, \sum_{t=0}^{\infty} \lambda(t) = +\infty, \sum_{t=0}^{\infty} \lambda^2(t) < +\infty)$, one can guarantee that under suitable initial conditions, our proposed algorithm converges to an equilibrium point.

Moreover, following the penalized reformulation of our game $\mathcal{G}$, one can easily show that a strategy that violates the coupled constraints cannot be a best response strategy. From [19], there exists $\boldsymbol{\rho}_l^*$ such that the incremental penalty algorithm terminates. Therefore, there exists a mixed strategy for which the coupled constraints are satisfied at $\boldsymbol{\rho}_l^*$. In that case, there is no incentive for an SBS to violate any of the coupled constraints, otherwise, its reward function would be penalized by the corresponding penalty function. Hence, all strategies that violate the coupled constraints are dominated by the alternative of complying with these constraints. Since in the proposed algorithm, the optimal strategy profile results in maximizing $\mathbb{E}_{\boldsymbol{p}_j}\left[\widehat{u}_j\left(\boldsymbol{a}_j, \boldsymbol{a}_{-j}\right)\right]$, we can conclude that the converged mixed-strategy NE is guaranteed not to violate the coupled constraints and hence it corresponds to a mixed-strategy NE for the game $\mathcal{G}$. Therefore, our proposed learning algorithm learns a mixed strategy of the game $\mathcal{G}$, by using a deep neural network function approximator to represent strategies, and by averaging those strategies via gradient descent machine learning techniques.

■

V. Simulation Results and Analysis

For our simulations, we consider a 300 m $\times$ 300 m square area in which we randomly deploy a number of SBSs and WAPs that share 7 unlicensed channels. We use real data for traffic loads from the dataset provided in [35] and divide it as 80% for training and 20% for testing. During the training phase, we randomly shuffle examples in the training dataset in order to prevent cycles when approximating the reward function. Table I summarizes the main simulation parameters. All statistical results are averaged over a large number of independent runs.

Table I:
SYSTEM PARAMETERS

Parameters	**Values**	**Parameters**	**Values**
Transmit power (P_t)	20 dBm	BW (channel)	20 MHz
CCA threshold	-80 dBm	Noise variance	92 dBm/Hz
Path loss	$15.3 + 50\log_{10}(m)$	SIFS	16 μs
Hidden size (encoder)	70	DIFS	34 μs
Hidden size (decoder)	70	$\mathrm{CW}_{\min}$	15 slots
time epoch (t)	5 min	$\mathrm{CW}_{\max}$	1023 slots
Action sampling (Z)	100 samples	ACK	256 bits
History traffic size	7 time epochs	P_{LTE}, P_{WiFi}	1, 1
Learning rate (λ)	0.01	LSTM layers	1
Learning rate decay (γ)	0.95	$t_{\max}$	0.9

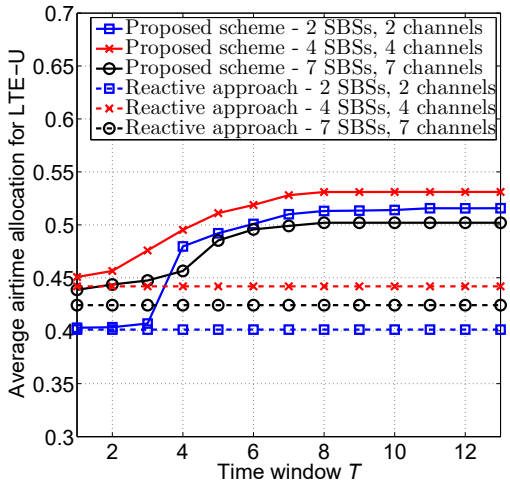

Fig. 4: The average throughput gain for LTE-U upon applying a proactive approach (with varying T) as compared to a reactive approach.

Fig. 4 shows the average throughput gain, compared to a reactive approach, achieved by the proposed approach for different values of T under three different network scenarios. Intuitively, a larger T provides the framework additional opportunities to benefit over the reactive approach, which does not account for future traffic loads. First, evidently, for very small time windows, the proactive approach does not yield any significant gains. However, as T increases, LTE-U network utilizes statistical predictions for allocating resources and thus the gains start to become more pronounced. For example, from Fig. 4, we can see that, for the case of 4 SBSs and 4 channels, the gain increases from 2% to 20% as T increases from 2 to 5, respectively. Eventually, as T grows, the gain of our proposed framework remains almost constant at the maximum achievable

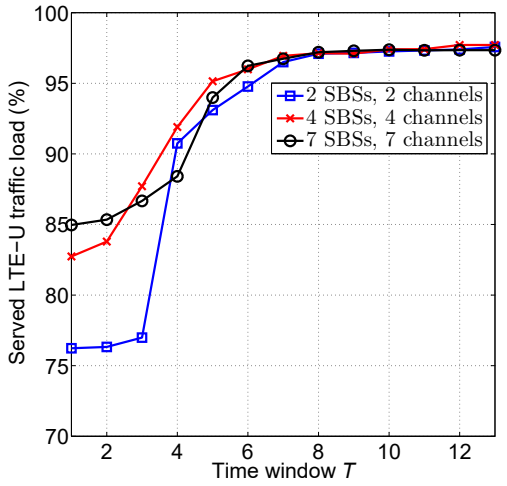

Fig. 5: The proportion of load served over LTE-U as a function of T.

value. This corresponds to the minimum value of T required to allow the LTE-U network smooth out its load over time and thus achieve maximum gain while guaranteeing fairness to WLAN.

In Fig. 5, we evaluate the proportion of LTE-U served load for different values of T. Fig. 5 shows that, as T increases, the proportion of LTE-U served traffic increases. For example, the proportion of served load increases from 82% to 97% for the case of 4 SBSs and 4 channels. Clearly, the gain of the LTE-U network stems from the flexibility of choosing actions over a large time horizon T. Unlike a reactive approach that allocates resources at time t based on the current network state only, our proposed proactive scheme takes into account future predictions of the network state along with the current state. Therefore, the optimal policy will balance the instantaneous reward and the available information for future use and thus maximizing the total load served over time. Based on the results given in Fig. 4 and Fig. 5, we can see that $T = 8$ is a suitable value of T for the studied dataset.

Fig. 6 shows the average airtime allocated for the LTE-U network as a function of T for our proposed scheme as well as the centralized solution considering a proportional fairness (PF) utility function subject to constraints (8)-(12) with $T = 1$. The centralized solution is obtained using the branch-and-bound algorithm in [36]. From Fig. 6, we can see that for small values of T, the PF allocation offers higher airtime allocation for the LTE-U network. For example, for the scenario of 4 SBSs and 4 channels, PF offers airtime gains of 7% and 5% as compared to our proposed approach for $T = 1$ and 2 respectively. However, as T increases, our proposed scheme achieves more transmission opportunities for the LTE-U network as compared to the

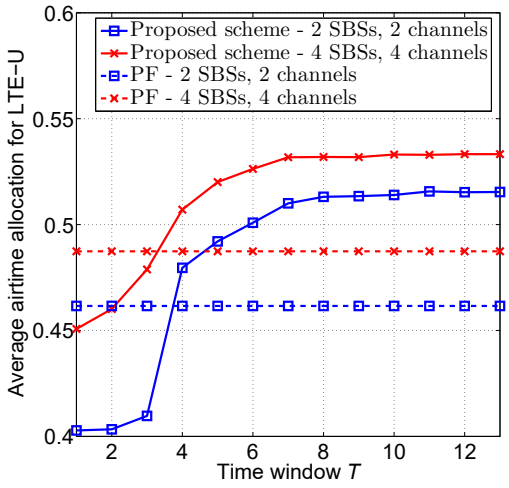

Fig. 6: The average airtime allocated for LTE-U (with varying T) for our proposed scheme as well as that of the centralized proportional fairness utility maximization.

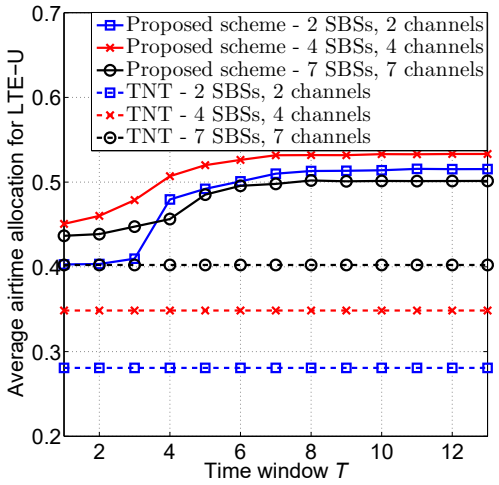

Fig. 7: The average airtime allocated for LTE-U (with varying T) for our proposed scheme as well as that of the centralized total network throughput utility maximization.

PF solution. For instance, for the scenario of 2 SBSs and 2 channels, our proposed scheme achieves an increase of 11% in the transmission opportunities for $T \geq 8$. This gain stems from the proactive resource allocation approach that allows more flexibility in spectrum allocation as T increases. Note that the resulting problem for the PF solution is a mixed integer nonlinear optimization problem (MINLP) and therefore, finding its solution becomes challenging for larger network scenarios due to the polynomial computational complexity.

Fig. 7 shows the average airtime allocated for the LTE-U network as a function of T for our proposed scheme as well as the centralized solution considering a total network throughput

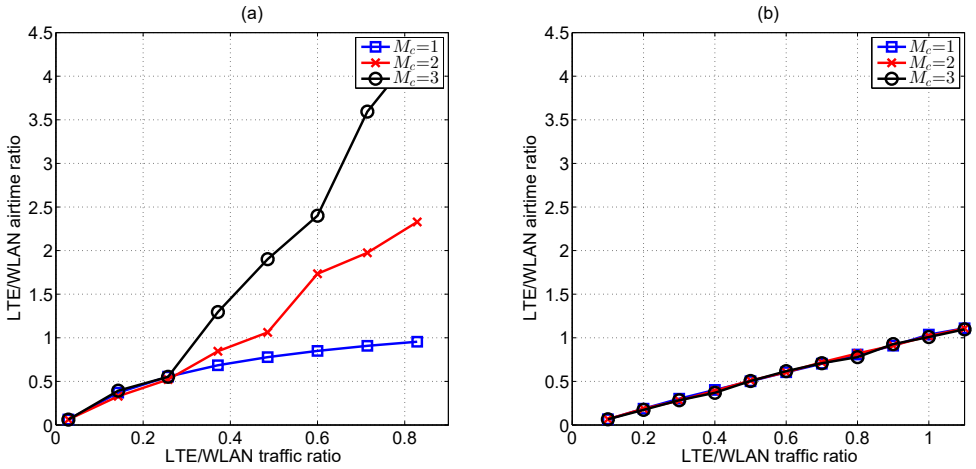

Fig. 8: LTE/WLAN airtime ratio as a function of the LTE/WLAN traffic ratio for 3 different values of M_c ($M_c = 1$, 2 and 3). The number of unlicensed channels is fixed to 7 and the number of SBSs is equal to 2 and 7 in (a) and (b) respectively.

(TNT) utility function subject to constraints (8)-(12) with $T = 1$. From Fig. 7, we can see that our proposed resource allocation scheme offers higher transmission opportunities for LTE-U for all values of T as compared to the centralized solution considering a TNT utility function. For example, for the case of 4 SBSs and 4 channels, the gain for our proposed approach can reach up to 52% for $T \geq 8$. This is due to the fact that the TNT utility function does not take fairness into account thus leading to a decrease in the LTE-U performance in the case of high WiFi offered load.

Fig. 8 shows the value of the LTE/WLAN airtime ratio under varying LTE/WLAN traffic ratio and for different values of M_c. We consider two different scenarios with varying number of SBSs (2 and 7 SBSs for scenarios (a) and (b) respectively), while the number of unlicensed channels is fixed to 7. Fig. 8 shows that inter-technology fairness is satisfied. This can be clearly seen in scenario (b) for the case of $M_c = 1$. For instance, when the traffic ratio is 1, LTE/WLAN airtime ratio is 1 and thus equal weighted airtime is allocated for each technology (given that $P_{\text{LTE}} = 1$ and $P_{\text{WiFi}} = 1$). From Fig. 8, we can also see that enabling carrier aggregation impacts the resource allocation outcome. In fact, we can see that a considerable gain in terms of spectrum access time can be achieved with carrier aggregation. For instance, in the case of 2 SBSs and 2 channels, the LTE/WLAN airtime ratio increases from 0.84 for $M_c = 1$ to 1.7 and 2.4 for $M_c = 2$ and 3 respectively for the value of 0.6 for LTE/WLAN traffic ratio. On the other hand, this gain decreases as more SBSs are deployed and for a densely deployed LTE-U network, there

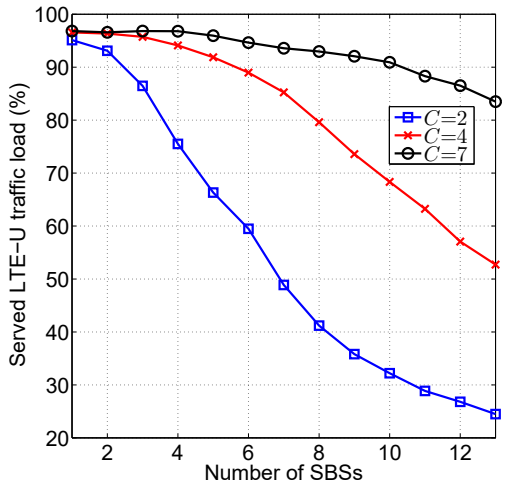

Fig. 9: The proportion of LTE-U served traffic load as a function of the number of SBSs and for different number of unlicensed channels ($C = 2$, 4, and 7).

is no need to aggregate more channels. This can be seen from (b) where the LTE-U network gets the same airtime share for $M_c = 1$, 2 and 3 (as also shown in Remark 1).

Moreover, Fig. 8 shows that deploying more SBSs does not necessarily allow more airtime for the LTE-U network. For example, LTE/WLAN airtime ratio of scenarios (a) and (b) corresponding to 0.6 LTE/WLAN traffic ratio is equal to 0.84 and 0.6 respectively for $M_c = 1$. Consequently, the proposed scheme can avoid causing performance degradation to WLAN in the case LTE operators selfishly deploy a high number of SBSs.

Fig. 9 investigates the proportion of served LTE-U traffic for different network parameters. From Fig. 9, we can see that, as the number of SBSs increases, the proportion of LTE-U served traffic, relative to its corresponding offered load decreases thus avoiding degradation in the WLAN performance in the case of a densely deployed LTE-U network. Moreover, reducing the number of unlicensed channels leads to a decrease in the proportion of LTE-U served traffic. Although the number of available unlicensed channels are not players in the game, they affect spectrum allocation action selection for each SBS. As the number of channels increases, the action space for the channel selection vector increases, thus giving more opportunities for an SBS to serve more of its offered load.

Fig. 10 shows the total network served traffic load as well as that of LTE-U and WiFi as a function of the priority fairness ratio on the unlicensed band $(P_{\text{LTE}}/P_{\text{WiFi}})$ for three different network scenarios considering $T = 6$. From Figs. 10 (b) and (c), we can see that more LTE-U and less WiFi traffic load is served as $P_{\text{LTE}}/P_{\text{WiFi}}$ increases and thus the priority fairness

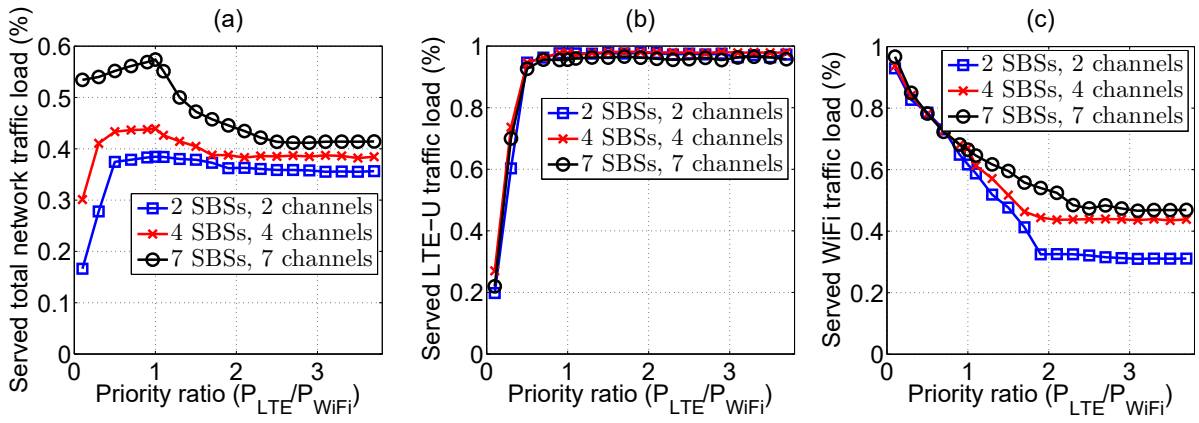

Fig. 10: The proportion of the (a) total network served traffic load (b) LTE-U served traffic load and (c) WiFi served traffic load as a function of the priority fairness ratio on the unlicensed band, ($\mathrm{P_{LTE}/P_{WiFi}}$).

parameters P_{LTE} and P_{WiFi} can be regarded as network design parameters that can be adjusted in a way that would avoid LTE-U from aggressively offloading traffic to the unlicensed bands in the case of high LTE-U traffic load. Moreover, from Fig. 10 (a), we can see that the optimal value of the priority fairness ratio $P_{\mathrm{LTE}}/P_{\mathrm{WiFi}}$ is 1. At $P_{\mathrm{LTE}}/P_{\mathrm{WiFi}} = 1$, the total network served traffic load is maximized while guaranteeing equal weighted airtime use of the unlicensed band for both technologies thus allowing an efficient utilization of the unlicensed spectrum. Therefore, our proposed spectrum sharing scheme achieves both efficiency and fairness.

Fig. 11 shows the average value of airtime allocated to the LTE-U network as a function of the number of epochs required for the network to converge while considering different values for the learning rate. The learning rate determines the step size the algorithm takes to reach the minimizer and thus has an impact on the convergence rate of our proposed framework. Moreover, an epoch, which consists of multiple iterations, is a single pass through the entire training set, followed by testing of the verification set. From Fig. 11, we can see that for $\lambda = 0.1$, our proposed algorithm requires more than 50 epochs to approximate the reward function, while, for $\lambda = 0.01$, it only needs 20 epochs. In fact, for $\lambda = 0.1$, we can see that our proposed algorithm fluctuates around a different region of the optimization space. Clearly, a learning rate that is too large can cause the algorithm to diverge from the optimal solution. This is because too large initial learning rates will decay the loss function faster and thus make the model get stuck at a particular region of the optimization space instead of better exploring it. On the other hand, a learning rate that is too small results in a low speed of convergence. For instance, for $\lambda = 0.0001$ and $\lambda = 0.00005$, our proposed algorithm requires ~ 40 epochs to converge. Therefore, although

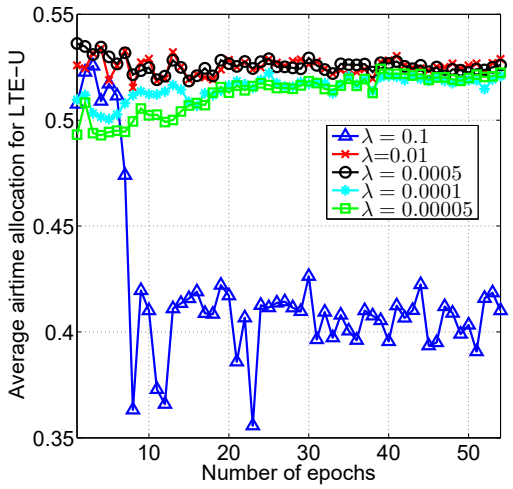

Fig. 11: The average airtime allocated for LTE-U as a function of the number of epochs for different values of the learning rate.

we use an adaptive learning rate approach, the optimization algorithm relies heavily on a good choice of an initial learning rate [37]. In other words, the initial value of the learning rate should be within a particular range in order to have good performance. Choosing a proper learning rate is an important key aspect that has an impact on the solution as well as the convergence speed. The optimal value of the initial learning rate is dependent on the dataset under study, where for each dataset, there exists an interval of good learning rates at which the performance does not vary much [30]. A typical range of the learning rate for the dataset under study falls approximately between 0.0005 and 0.01, requiring ~ 20 epochs.

VI. Conclusion

In this paper, we have proposed a novel resource allocation framework for the coexistence of LTE-U and WiFi in the unlicensed band. We have formulated a game model where each SBS seeks to maximize its rate over a given time horizon while achieving long-term equal weighted fairness with WLAN and other LTE-U operators transmitting on the same channel. To solve this problem, we have developed a novel deep learning algorithm based on LSTMs. The proposed algorithm enables each SBS to decide on its spectrum allocation scheme autonomously with limited information on the network state. Simulation results have shown that the proposed approach yields significant performance gains in terms of rate compared to conventional approaches that considers only instantaneous network parameters such as instantaneous equal weighted fairness, proportional fairness and total network throughput maximization. Results have also shown that

our proposed scheme prevents disruption to WLAN operation in the case large number of LTE operators selfishly deploy LTE-U in the unlicensed spectrum.

REFERENCES

[1] R. Zhang, M. Wang, L. Cai, Z. Zheng, X. Shen, and L. Xie, "LTE-unlicensed: the future of spectrum aggregation for cellular networks," *IEEE Wireless Communications*, vol. 22, no. 2, pp. 150–159, June 2015.

[2] B. Jia and M. Tao, "A channel sensing based design for LTE in unlicensed bands," in *Proc. of IEEE International Conference on Communication Workshop (ICCW)*. London, UK, June 2015.

[3] A. Mukherjee, J.-F. Cheng, S. Falahati, L. Falconetti, A. Furuskr, B. Godana, D. H. Kang, H. Koorapaty, D. Larsson, and Y. Yang, "System architecture and coexistence evaluation of licensed-assisted access LTE with IEEE 802.11," in *Proc. of IEEE International Conference on Communications (ICC)*. London, UK, June 2015.

[4] C. Hasan, M. K. Marina, and U. Challita, "On LTE-WiFi coexistence and inter-operator spectrum sharing in unlicensed bands: Altruism, cooperation and fairness," in *Proc. of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*. Paderborn, Germany, July 2016.

[5] U. Challita and M. K. Marina, "Holistic small cell traffic balancing across licensed and unlicensed bands," in *Proc. of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*. Malta, Nov. 2016.

[6] M. Chen, W. Saad, and C. Yin, "Echo state networks for self-organizing resource allocation in LTE-U with uplink-downlink decoupling," *IEEE Transactions on Wireless Communications*, To appear 2016.

[7] 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 *Proc. of the 82nd IEEE Vehicular Technology Conference (VTC Fall)*. Glasgow, Scotland, Sept. 2015.

[8] Y. Gu, Y. Zhang, L. X. Cai, M. Pan, L. Song, and Z. Han, "Exploiting student-project allocation matching for spectrum sharing in LTE-Unlicensed," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*. San Diego, CA, USA, Dec. 2015.

[9] Z. Guan and T. Melodia, "CU-LTE: Spectrally-efficient and fair coexistencebetween LTE and Wi-Fi in unlicensed bands," in *Proc. of IEEE Conference on Computer Communications (INFOCOM)*. San Francisco, CA, USA, Apr. 2016.

[10] S. Sagari, S. Baysting, D. Saha, I. Seskar, W. Trappe, and D. Raychaudhuri, "Coordinated dynamic spectrum management of LTE-U and Wi-Fi networks," in *Proc. of IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. Stockholm, Sweden, Sept. 2015.

[11] S. Ha, S. Sen, C. Joe-Wong, Y. Im, and M. Chiang, "TUBE: Time dependent pricing for mobile data," in *Proceedings of Special Interest Group on Data Communication (ACM SIGCOMM)*. Helsinki, Finland, Aug. 2012.

[12] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Q. S. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Wireless Communication Magazine*, vol. 18, no. 3, pp. 22–30, June 2011.

[13] 3GPPP TS 36.300 v10.2.0, "E-UTRA and E-UTRAN; Overall description; Stage 2 (Release 10)," 2011.

[14] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[15] H. Gintis, *Game theory evolving: A problem-centered introduction to modeling strategic behavior*. Princeton University Press, 2000.

[16] E. Fehr and K. Schmidt, "A theory of fairness, competition and cooperation," *Quarterly Journal of Economics (QJE)*, vol. 114, no. 3, pp. 817–868, 1999.

[17] Y. Xing, R. Chandramouli, S. Mangold, and S. Shankar, "Dynamic spectrum access in open spectrum wireless networks," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 24, no. 3, pp. 626–637, Mar. 2006.

[18] M. Heusse, F. Rousseau, R. Guillier, and A. Duda, "Idle sense: an optimal access method for high throughput and fairness in rate diverse wireless LANs," in *Proceedings of the annual conference of the Special Interest Group on Data Communication (SIGCOMM)*. Philadelphia, PA, Oct. 2005.

[19] M. Fukushima, "Restricted generalized Nash equilibria and controlled penalty algorithm," *Computational Management Science (CMS)*, vol. 8, no. 3, pp. 201–218, Aug. 2011.

[20] R. Sutton, D. McAllester, S. Singh, and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," *Advances in Neural Information Processing Systems*, vol. 12, pp. 1057–1063, 2000.

[21] J. Nash, "Non-cooperative games," *Annals of Mathematics*, vol. 54, pp. 286–295, 1951.

[22] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. The MIT Press, 1998.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," vol. 9, no. 8, pp. 1735–1780, 1997.

[24] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," in *Proceedings of International Conference on Learning Representations (ICLR)*. San Diego, CA, May 2015.

[25] H. Sak, A. Senior, and F. Beaufay, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," in *arXiv preprint arXiv:1402.1128*, 2014.

[26] R. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning*, vol. 8, no. 3, pp. 229–256, May 1992.

[27] T. Tieleman and G. Hinton, "Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude," Technical report, 2012.

[28] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

[29] E. Rozner, Y. Mehta, A. Akella, and L. Qiu, "Traffic-aware channel assignment in enterprise wireless LANs," in *Proc. of IEEE International Conference on Network Protocols (ICNP)*. Beijing, China, Oct. 2007.

[30] K. Greff, R. Srivastava, J. Koutnik, B. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Transactions on neural networks and learning systems*, 2016.

[31] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, June 2014.

[32] L. Peshkin, K. Kim, N. Meuleau, and L. P. Kaelbling, "Learning to cooperate via policy search," in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI-2000)*. San Francisco, CA, USA, June 2000.

[33] K. J. Arrow and L. Hurwicz, "Stability of the gradient process in n-person games." *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 280–295, June 1960.

[34] H. Robbins and S. Munro, "A stochastic approximation method," *Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, Sept. 1951.

[35] M. Balazinska and P. Castro, "IBM Watson Research Center," CRAWCAD, Feb. 2003.

[36] COINOR, "Basic open-source nonlinear mixed integer programming," http://www.coin-or.org/Bonmin/.

[37] C. Daniel, J. Taylor, and S. Nowozin, "Learning step size controllers for robust neural network training," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*. Phoenix, Arizona USA, Feb. 2016.