Reinforcement Learning for Efficient and Fair Coexistence Between LTE-LAA and Wi-Fi

Mengqi Han, Sami Khairy, *Student Member, IEEE*, Lin X. Cai, Yu Cheng, *Senior Member, IEEE*, and Ran Zhang, *Member, IEEE*

***Abstract*— Long-Term Evolution (LTE) over unlicensed spectrum extends LTE technology to the spacious unlicensed spectrum with readily available bandwidth. The provided capacity surge makes it one of the most high-profile technologies to meet the explosive growth of mobile traffic demand. Among its different variants, Licensed Assisted Access (LAA) is considered as a promising global solution attributed to its mandatory listen-before-talk (LBT) procedure. Nevertheless, although LBT effectively maintains transmission fairness between LTE and other unlicensed systems (e.g., Wi-Fi), the current LAA protocol specified in 3GPP Release 13 is far from perfect to achieve harmony coexistence. To this end, in this paper, we first develop an analytical model to evaluate the throughput performance of Category 4 (Cat 4) algorithm agreed in 3GPP release 13. Subject to the system fairness constraint, the aggregate throughput of LTE-LAA and Wi-Fi networks is maximized based on a semi branch and bound algorithm. To make the complex optimization tractable, reinforcement learning techniques are introduced to intelligently tune the contention window size for both LTE-LAA and Wi-Fi nodes. Specifically, a cooperative learning algorithm is developed assuming that the information between different systems is exchangeable. A non-cooperative version is subsequently developed to remove the previous assumption for better practicability. Extensive simulations are conducted to demonstrate the performance of the proposed learning algorithms in contrast to the analytical upper bound under various conditions. It is shown that both proposed learning algorithms can significantly improve the total throughput performance while satisfying the fairness constraints. Particularly, the proposed cooperative learning algorithm can closely approach the analytical bound.**

***Index Terms*—Licensed-assisted-access, unlicensed band, performance analysis, reinforcement learning.**

I. Introduction

IN THE last decade, the wireless communication industry have been experiencing explosive mobile data increase due to the prevalence of smart devices and data-intensive mobile applications. According to Cisco traffic white paper [1], the number of mobile network devices will reach 12 billion by 2020 and the mobile data traffic will reach 4.8 ZB per year by 2022. The wireless operators are significantly challenged to excavate more network capacity to meet the ever-growing mobile traffic demand. As a hard core for current cellular communications, the Long-Term Evolution (LTE) technology was originally designed to operate over the licensed band due to the scheduling based nature. Yet the scarcity of the licensed spectrum and the high cost of a licensed frequency have motivated the LTE operators to step into the unlicensed spectrum for vast and cost-effective bandwidth. The featured technology of LTE in the unlicensed band provides larger network capacity and better user experience at much lower cost [2].

However, due to the shared nature of unlicensed spectrum, the major challenge of LTE in the unlicensed band is to ensure harmony coexistence between LTE and other incumbent unlicensed systems with inherently different radio access technologies, particularly Wi-Fi. Wi-Fi systems adopt a contention-based scheme, i.e., carrier sensing multiple access (CSMA/CA) [3], to access the channel; while the LTE transmissions are based on slotted centralized scheduling. To prevent the sever performance degradation of Wi-Fi users, an efficient and fair coexistence scheme is indispensable. The current coexistence algorithms are categorized into two types: Listen-Before-Talk (LBT) based and non-LBT based algorithms. Specifically, Carrier Sensing Adaptive Transmission (CSAT), proposed by Qualcomm, is the major non-LBT based coexistence scheme [4]. Nevertheless, since CSAT schedules LTE transmissions in specified periods without sensing the channel ahead, the scheduled LTE transmissions may considerably affect the performance of other LTE nodes or unlicensed systems in an uncoordinated way. To this end, LAA, which mandates the LBT feature thus being more fitting globally, is specified in 3GPP release 13 [5]. Under LAA, LTE users are required to sense the channel before access to avoid collisions with other nodes. In this paper, we focus on the Cat 4 algorithm which is more general and adaptive compared to other existing protocols.

In the meantime, the ever-growing diversity and complexity of wireless network has made monitoring and managing the multitude of network elements intractable.Thanks to the remarkable advances in machine learning, embedding machine learning into future mobile networks is drawing unparalleled interest [6]–[10]. This trend is reflected in the machine learning based solution for problems ranging from power estimation [11] to wireless resource management [12]–[14], as well as wireless communication system design [15], [16]. For example, in [11], a

Manuscript received February 6, 2020; revised March 28, 2020; accepted April 23, 2020. Date of publication May 14, 2020; date of current version August 13, 2020. This work was supported by the National Science Foundation under Grants ECCS-1554576, CNS-1816908, and ECCS-1610874. The review of this article was coordinated by Dr. F. Tang. *(Corresponding author: Mengqi Han.)*

Mengqi Han, Sami Khairy, Lin X. Cai, and Yu Cheng are with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: mhan9@hawk.iit.edu; skhairy@hawk.iit.edu; lincai@iit.edu; cheng@iit.edu).

Ran Zhang is with the Department of Electrical and Computer Engineering, Miami University, Oxford, OH 45056 USA (e-mail: zhangr43@miamioh.edu).

Digital Object Identifier 10.1109/TVT.2020.2994525

machine learning-based model is proposed to estimate the power consumption of electric vehicle by extracting the knowledge of historical trips. In [12], a neural network-based reinforcement learning algorithm is used to solve the cloud resource allocation problem. Similarly, reinforcement learning algorithm is applied to find the optimal virtual machine allocation policy in [13]. Within the scope of machine learning, deep learning and reinforcement learning are two methods of most interest in wireless networks. In deep learning, neural network layers are used to achieve a brain-like feature extraction and accurate inference from a huge volume of pre-collected examples, whereas the goal of reinforcement learning is to train the agent to achieve an optimal policy from its on-the-fly interactions with the environment. Due to the random nature of the wireless environment and the the dynamic activities of network entities, pre-collecting sufficient examples for all the possible system states may not be feasible, thus impacting the adaptability of deep learning-based mechanisms in many practical occasions. Therefore, reinforcement learning-based methods becomes increasingly popular as the agent requires no prior knowledge of the system and can learn while interacting with and observing the system. To this end, we resort to reinforcement learning in this work to optimize the configurations of both LTE-LAA and Wi-Fi users to achieve their harmony coexistence. However, to the best of our knowledge, the existing coexistence schemes based on learning algorithms either rely only on simulations without investigating the gap to the analytical bound, or only focus on a single performance metric (i.e., fairness or throughput). In addition, there is no related papers considering joint optimization of the parameters from both LTE and Wi-Fi networks in a distributed manner. In this paper, a distributed reinforcement learning based coexistence scheme is proposed to jointly select the optimal window size for both networks while considering the trade off between throughput and fairness. An analytical upper bound is also devised for performance comparison.

Our contributions can be summarized in four-fold. First, an analytical model to evaluate the throughput performance of Wi-Fi and LTE-LAA networks is developed. Second, subject to the system fairness constraint, the upper bound of the total system throughput is calculated employing a branch-and-bound based algorithm. Third, depending on whether the information of the two systems is exchangeable or not, we propose both cooperative and non-cooperative learning algorithms to intelligently tune the contention window size for both networks. Multi-armed bandit learning technique is adopted to obtain the optimal coexisting performance due to its stateless property. More details of the justification on the multi-armed bandit learning can be found in Section V. Finally, extensive simulations are conducted to demonstrate that the proposed learning algorithms can significantly improve the coexistence performance. Particularly, the cooperative version can closely approach the analytical bound.

The remainder of this paper is organized as follows. Section II summarizes the related works. Section III describes the access protocol of Cat 4. The analytical framework is developed in Section IV to investigate the coexisting performance of LTE-LAA and Wi-Fi systems. In Section V, the proposed learning algorithms are elaborated. Numerical results on performance of the proposed algorithms are presented in Section VI. Section VII concludes the paper and discusses the possible future work.

II. Related Work

In the literature, most prior works studying the performance of coexistence between LTE-LAA and Wi-Fi users focus on the access protocol of Category 4 (Cat 4) algorithm which is proposed in the 3GPP Release 13. There are three other categories defined in the 3GPP Release 13, yet with less generality. In Cat 1, LTE-LAA users access the channel in an on/off pattern, which is adverse to the ongoing Wi-Fi transmissions. Cat 2 and 3 adopt the feature of LBT, a similar channel access scheme to CSMA/CA in Wi-Fi, but the backoff either has a deterministic time or a fixed window size. In Cat 4 algorithm, LTE-LAA nodes adopt a similar channel access scheme with Wi-Fi nodes which is CSMA/CA. But, Cat 4 and CSMA/CA may have different backoff mechanisms, sensing time and slot time. For example, Wi-Fi nodes use the binary exponential backoff (BEB), whereas the backoff window size in Cat 4 is q which is chosen in a dynamic range $[X, Y]$. In [17], an analytical framework to investigate the downlink coexistence performance between Wi-Fi and LTE-LAA is proposed when LTE-LAA adopt the fixed window size. The main reason that Cat 4 algorithm attracts the most attention is that Cat 4 can perform similarly to CSMA/CA protocol which is adopted by Wi-Fi users. In [18], authors propose a Markov model to study the downlink LBT Cat 4 mechanism considering that the network parameters, i.e., sensing time and slot time, are the same as Wi-Fi. Similarly, in [19], the backoff processes of Wi-Fi and LTE-LAA are both modeled as bi-dimensional Markov chain. To the best of our knowledge, most previous works consider that the LTE-LAA users use the same network parameters as Wi-Fi. The impacts when LTE-LAA users adopt different parameters setting are seldom investigated. In [20], authors point out that the collisions between LTE-LAA users and Wi-Fi users can be significantly mitigated if the sensing time and slot duration of LTE-LAA users are different from Wi-Fi users. Therefore, to alleviate the collisions and improve the spectrum efficiency, the network parameters should be separately and appropriately selected.

Besides, some existing works focus on the fairness issue. Since LTE over unlicensed spectrum was first proposed, there have been serious concerns about the fairness preservation of other existing unlicensed technologies, such as Wi-Fi and Bluetooth technologies. In [21], the impact of the key channel access parameters on the fairness and the throughput of the LAA/Wi-Fi coexistence system was investigated. According to the numerical results in [22], the performance of Wi-Fi could be starved if LTE adopts the scheduling based access without sensing the channel before transmitting in the unlicensed band. As Wi-Fi nodes can only transmit when the channel is not occupied, LTE users would keep transmitting while Wi-Fi users would be forced to stay in the backoff stage in the coexisting situation. Therefore, it is desired to design a fair and efficient coexistence scheme to protect Wi-Fi performance by regulating LTE transmissions. Several

coexistence schemes have been proposed to improve the performance of coexistence between LTE and Wi-Fi [23]–[29]. Most proposed algorithms dynamically adapt the protocol parameters for either LTE or Wi-Fi. Also, most of the related works are LTE performance-oriented. In [23], a Maximum Contention Window Timer Mechanism (MCWTM) is proposed for LTE-LAA user which requires LTE-LAA users to adopt the similar feature as Wi-Fi, i.e., maximum retry limit to improve the performance of LTE-LAA users. Authors in [24] propose an enhanced LBT which jointly considers the total system throughput and fairness. In the proposed algorithm, the optimum idle period is selected for WLAN. Similarly in [25], blank sub-frame is introduced for LTE users to improve fairness. During the blank sub-frame, LTE users are not allowed to transmit, so Wi-Fi users could get more opportunities to transmit. Other than selecting the idle period, some algorithms consider the adaption of the window size. In [26], the proposed algorithm update the window size in regard to the slot utilization ratio which is calculated using busy time observed in an observation window. In this paper, the window size of LTE users is chosen in the range $[X, Y]$. If the utilization ratio is lower than a predefined threshold, the value of X is decremented by a step size. Meanwhile, the value of Y is incremented by one step size when the utilization ratio is larger than the threshold. The proposed algorithm can improve the LAA throughput performance 5% comparing to fixed window size. Note that, adapting only the range of window size could only improve the performance slightly. Following the similar concept of step size, authors in [27] jointly consider the adaptation of CCA period and window size. The value of CCA is updated based on the fairness constraint, while the window size is updated based on the delay constraint.

With the capability of machine learning in solving problem in a dynamic but statistical environment, some related works focus on the algorithm design based on machine learning. In [28], a Q-learning based algorithm is proposed to autonomously select the combination of transmission time and muting period to improve only the fairness between the LTE-U and Wi-Fi networks. In [29], the authors propose a Neural Network based scheme which can select the optimal window size for LTE-U based on the predicted number of Negative Acknowledgements (NACKs). According to the simulation results, the proposed scheme can improve the performance in terms of throughput and latency. Authors in [30] propose a Q learning algorithm to dynamically select the duty cycle for LAA users to improve the coexistence performance. But the definitions of the state and cost function are heavily dependent on some predefined the target value which is not flexible in practice. In [31], a Q learning based algorithm is put forward to dynamically select the ratio of blank frame during one sub-frame and the size of sub-frame for only LTE users. Yet the proposed algorithm only consider the performance of LTE users.

To the best of our knowledge, there is no learning based algorithm which can select the key parameters for both LTE-LAA and Wi-Fi networks in a distributed manner. Motivated by this, we first propose a cooperative learning algorithm in which the throughput information of both networks can be exchanged. For better practicality, a non-cooperative learning algorithm is proposed when the information of both networks is not exchangeable. Moreover, we compare the performance of both algorithms with some existing algorithms and the developed analytical upper bound.

III. Description of LTE-LAA Protocol

As specified in 3GPP Release 13, LTE-LAA nodes are required to perform an Initial Clear Channel Assessment (ICCA) which is at least 20 us prior to transmission. LTE-LAA nodes can transmit on the channel immediately if the channel is sensed idle during ICCA. And LTE-LAA nodes can occupy the channel for at most $13/32q$ ms, where the value of q is selected in a predefined range $[X, Y]$ [32]. If the channel is not idle during ICCA, the node must perform an Extended CCA, where the channel should be observed idle for N ECCA slots. We can notice that N slots do not need to be consecutive. The value of N is randomly selected in the range $[0, q-1]$. In other words, N is similar to the backoff counter for Wi-Fi nodes. The value of N would be reduced by one when the channel is idle for one slot. When LTE-LAA node finds the channel busy during ECCA, the value of N would freeze and resume when the channel is sensed idle again for one slot. The values of both X and Y are configurable. There are several candidates for the contention window size, e.g., fixed, exponential. The Cat 4 algorithm is illustrated in Fig. 1. As shown in Fig. 1, the LTE-LAA node enters ECAA when the channel is busy during ICCA. During ECCA, the LTE-LAA node keeps sensing the channel until it is idle for N ECCA slots after a failed ICCA. In Cat 4 algorithm, LTE-LAA nodes behave similarly as Wi-Fi nodes which adopt CSMA/CA as the channel access scheme. But, Cat 4 and CSMA/CA may have different backoff mechanisms, sensing time and slot time. For example, Wi-Fi nodes use the binary exponential backoff (BEB), whereas the backoff window size in Cat 4 is q which is chosen in a dynamic range $[X, Y]$.

IV. Analytical Model

In this section, we develop the analytical model to evaluate the performance of LTE-LAA and Wi-Fi nodes. Specifically, we consider that N_l LTE-LAA nodes (can be either base station or user equipment) and N_w Wi-Fi nodes (can be either access point or station) are coexisting in the same unlicensed frequency.

We consider a more general scenario where both uplink and downlink traffic can be transmitted in the unlicensed spectrum to leverage the full benefits of LTE operation in unlicensed spectrum.

Both LTE-LAA and Wi-Fi nodes carry persistent traffic. And the channel is ideal where all the nodes can sense the transmissions of others. In this work, all nodes adopt the feature of LBT to access the channel. Based on our previous work [20], when the sensing time and slot time of LTE-LAA users are carefully selected, the probability that LTE-LAA nodes collide with Wi-Fi nodes is negligible. Thus, in the analytical model, we consider that the sensing time and slot time of LTE-LAA are co-prime with Wi-Fi nodes. The collisions between Wi-Fi

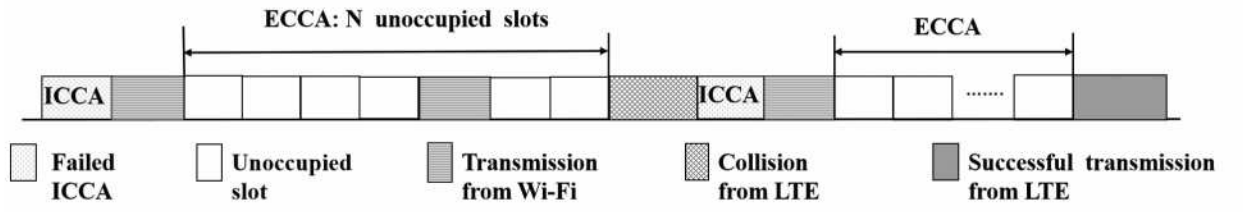


Fig. 1. Cat 4 algorithm.

TABLE I
LIST OF NOTATIONS

Description	Notation
Number of Wi-Fi nodes	N_w
Number of LTE BS	N_l
Service time of a single Wi-Fi node	D_w
Service time of a single LTE-LAA node	D_l
Backoff window size of LTE-LAA BS	C_l
Minimum window size of Wi-Fi node	C_w
Average number of transmissions of LTE-LAA node during D_l	A_l
Average number of transmissions of Wi-Fi node during D_w	A_w
Collision time of a Wi-Fi node	T_{wc}
Transmission time of a LTE BS	T_l
A slot time of LTE-LAA node	σ_l
Sensing time for LTE-LAA node	CCA_l
Collision probability of LTE-LAA node	P_l
Collision probability of Wi-Fi node	P_w

users and LAA happen when they transmit in the same slot, i.e., $CCA_l + \sigma_l n_l = DIFS + \sigma_w n_w$, where n_l, n_w are the minimum backoff counters of LAA and Wi-Fi users after one busy transmission, and σ_l, σ_w are the slot times of LAA and Wi-Fi, respectively. In this paper, we adopt the IEEE 802.11a/ac parameters, namely, DIFS $= 34\ \mu$s and $\sigma_w = 9\ \mu$s. On the other hand, LAA uses a CCA time of $20\ \mu$s and $\sigma_l = 20\ \mu$s. Unless the value of CCA_l is chosen to be $34 + 9r$, $r = \{-1, 0, 1, 2, \ldots\}$, and $\sigma_l = 9\,k$, $\{k = 1, 2, ..\}$, the collision probability between LAA and Wi-Fi users cannot be ignored; Otherwise, the collisions between LAA users and Wi-Fi users are negligible and can be assumed as zero. Notice that, our model can be ready to modify to consider the collision between LTE-LAA and Wi-Fi nodes. But, we choose to ignore due to the negligible impact on the throughput. Unlike the exponential backoff window adopted by Wi-Fi, LTE-LAA nodes are assumed to adopt a fixed window size which is chosen in the range $[X, Y]$. And the window size of all LTE-LAA nodes is identical which is denoted as C_l. Next, we derive the throughput of both LTE-LAA and Wi-Fi network. The main parameters used in the analytical part is listed in Table I. The channel access of both Wi-Fi and LTE-LAA nodes can be modeled using renewal theorem [33]. Denote D_w as the average duration between two successive successful transmissions of one tagged Wi-Fi node or the service time. Similarly, D_l denotes the average service time of one tagged LTE-LAA node. We analyze all the possible events could happen during D_l and D_w. Since one Wi-Fi node can successfully transmit a frame on every D_w slots, thus on average it can successfully transmit D_l/D_w frames during D_l. Therefore, the total successful transmissions including Wi-Fi and LTE-LAA nodes during D_l can be written as,

For each LTE-LAA node, C_l is fixed. Thus, the average backoff time before one transmission is given by,

$$\overline{W_l} = \frac{\frac{X+Y}{2} - 1}{2} = \frac{C_l - 1}{2}. \tag{1}$$

Define A_w and A_l as the average number of transmissions of one Wi-Fi node during D_w and the average number of transmissions of one LTE-LAA node during D_l, respectively. For each Wi-Fi node, multiple transmissions can be made before either the transmission is successful or the retry limit is reached. Denote τ_w as the transmission probability of Wi-Fi node, the conditional collision probability P_w can be calculated as $1 - (1 - \tau_w)^{(N_w - 1)}$. Given the conditional collision probability P_w and the retry limit m, the average number of transmissions during D_w is given by,

$$A_w = \frac{1 - {P_w}^{m+1}}{1 - P_w}. \tag{2}$$

Unlike Wi-Fi node, LTE-LAA node keeps transmitting until the frame is successfully transmitted. Given the probability that LTE-LAA node collides with other nodes as P_l, the average number of transmissions during D_l is,

$$A_l = \frac{1}{1 - P_l}. \tag{3}$$

Among A_w transmissions, only the last one transmission is successful. Given that each Wi-Fi node experiences $(A_w - 1)$ unsuccessful transmissions during D_w on average, the total unsuccessful transmissions from the aspect of Wi-Fi users during D_l can be calculated as $\frac{Dl/D_w(A_w - 1)N_w}{2}$. We assume that the collision occurs when two users are transmitting concurrently. Because, the probability that three or more than three users transmit in the same slot is extremely low. Summing up all the possible events, D_l can be expressed as,

$$D_l = T_{st} + \frac{(A_l - 1)N_l(T_l + CCA_l)}{2} + \frac{D_l/D_w(A_w - 1)N_w T_{wc}}{2} + A_l\overline{W_l}\sigma_l. \tag{4}$$

Note that the T_{wc} includes the duration of CCA_w and can be written as,

$$T_{wc} = T_{data} + ACK_{timeout} + CCA_w \tag{5}$$

The service time D_w of Wi-Fi nodes can be derived similarly as LTE-LAA nodes and given by

$$D_w = N_{ws}T_w + \frac{1}{2}(A_w - 1)N_w T_c + \overline{W_w} + \frac{D_w/D_l(A_l - 1)N_l T_l}{2} + N_l \frac{D_w}{D_l} T_l, \quad (6)$$

where $\overline{W_w}$ denotes the average backoff time of a Wi-Fi node during D_w. T_{ws} denotes the duration of a successful transmission from Wi-Fi user and can be written as,

$$T_{ws} = T_{data} + T_{ACK} + CCA_w \quad (7)$$

The first line in (6) represents the average transmissions from Wi-Fi nodes and average waiting time due to the backoff. Meanwhile, the second line represents the transmissions from LTE-LAA nodes.

For Wi-Fi nodes, before the retry limit is reached, the backoff window size will be doubled when a collision occurs. Given the minimum window size is C_w and the retry limit is m, the average backoff time during D_w is written as,

$$\overline{W_w} = \sum_{i=0}^{m-1} P_w^i (1 - P_w) \sum_{j=0}^{i} \frac{2^j C_w}{2} + P_w^m \sum_{j=0}^{m} \frac{2^j C_w}{2}. \quad (8)$$

After obtaining the service time of both LTE-LAA and Wi-Fi nodes, we can derive the throughput of two networks.

During D_l, each LTE-LAA node transmits at rate R for T_l slots, therefore, the total throughput of N_l LTE-LAA nodes in the unlicensed band is given by:

$$Th_l = \frac{N_l T_l R}{D_l}. \quad (9)$$

During D_w, each Wi-Fi node transmit a packet of PL bits. Therefore, the total throughput of all Wi-Fi nodes is given by,

$$Th_w = \frac{N_w PL}{D_w}. \quad (10)$$

A. Optimization Problem

In this section, we formulate the optimization problem. The backoff window size C_l of LTE-LAA nodes and the minimum window size C_w of Wi-Fi nodes can be selected to maximize the total system throughput, while to ensure the fair coexistence. Although, the window size of LTE-LAA nodes can be selected between $[X, Y]$ in Cat 4 algorithm. But the performance can be slightly improved if the value of X and Y are both adaptive. In addition, when the average window size is given, there are different possible $[X, Y]$ ranges which can achieve the same average window size. Therefore C_l is the average window size. A tolerance parameter ζ is introduced to guarantee fairness between LTE-LAA and Wi-Fi network. The optimization problem can be formulated as

$$\begin{aligned} &\underset{C_l, C_w}{\text{maximize}}\ Th_l + Th_w \\ &\text{subject to } C_{l_{\min}} \le C_l \le C_{l_{\max}}, \\ &\qquad C_{w_{\min}} \le C_w \le C_{w_{\max}}, \\ &\qquad \left| \frac{Th_l}{Th_w} - 1 \right| \le \zeta, \end{aligned} \quad (11)$$

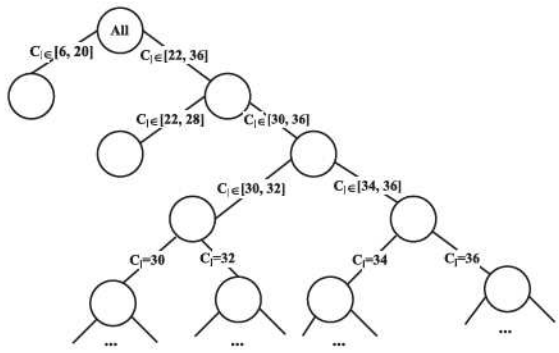


Fig. 2. SBnB algorithm.

where the first two constraints represent the constraints on C_l and C_w and the last constraint specifies the fairness requirement.

For non-convex problems, the optimal solution can be obtained by Branch and Bound (BnB) optimization method [34]. Thus, we propose a semi-BnB(SBnB) algorithm to find the optimal solution. In the classic BnB algorithm, the searching space is divided into different subsets which is called branching or splitting. The algorithm keeps track of both lower bound and upper bound of each branch. These bounds are used to prune the search space. Because the bound on the branch may prove that candidate solutions in this branch do not contain an optimal solution. Thus, many branches can be terminated. The SBnB algorithm to find the optimal window size is shown in Fig. 2. In this example, the window size of both Wi-Fi and LTE-LAA nodes can be chosen from $\{6, 8, 10 \ldots 36\}$. Thus, we first split the search space based on the window size of LTE-LAA users and calculate only the lower bound of the ratio $|\frac{Th_l}{Th_w} - 1|$ on each branch. If the lower bound on this branch cannot satisfy the fairness requirement, this branch will be pruned.

V. Proposed Learning Schemes

As mentioned in the previous section, the optimal window size of both Wi-Fi nodes and LTE-LAA nodes can be derived using analytical model. But the full network information including the number of nodes in the other network and the traffic pattern must be available for either network. In addition, the traffic is also dynamic in terms of the average packet size and the arrival rate. Therefore, in this section, we discuss how to apply reinforcement learning technique to select the optimal backoff window size for both LTE-LAA and Wi-Fi nodes in a distributed manner with no prior knowledge.

Unlike supervised learning or unsupervised learning, reinforcement learning can only make the optimal decisions based on the experience. The optimal action in reinforcement learning is the action that has the highest cumulative long term reward.Instead of using other learning techniques like Q learning, multi-armed bandit learning is adopted in this work due to its stateless property. The "stateless" means that there is no state transition and the reward is only dependent on the action. In classic multi-armed bandit problem, we are given a slot machine with n arms with each arm having its own rigged probability

distribution of success [35]. Since the true probability distributions of n arms are unknown, the ultimate goal is to find the optimal arm with the best value regardless of the state. The reinforcement learning algorithms can be divided into two categories: tabular solution based on estimating action value and function approximation-based solution. Note that multi-armed bandit learning focuses on learning/estimating the mean value of each arm which falls into the scope of tabular solution. Specifically, for our existence problem, either LAA or Wi-Fi users are only faced with different selections of window size which can be modelled as different arms (action). The reward of each action is random due to the unawareness of the actions taken by other users. Therefore, multi-armed bandit learning is suitable to our coexistence problem.

In reinforcement learning, no prior knowledge of environment is needed. There are two important matrices which are Q-matrix and Reward-matrix (R-matrix) in reinforcement learning. R-matrix is a matrix that has the number of states as rows and number of actions as column. R-matrix stores the instant reward obtained from each action. While Q-matrix stores the estimated value of each action. Note that Q-matrix has the same dimension as the R-matrix. In multi-armed bandit learning, there is no state information which means you are only faced repeatedly with different actions. Thus, both Q matrix and R-matrix have only one row.

Denote $Q_t(a)$ as the estimated value of action a on the t-th time step. One straightforward way to estimate the mean value of action a is to average all the rewards when action a was selected. Given that action a has been selected $N_t(a)$ times prior to time t, the estimated value can be written as,

$$Q_t(a) = \frac{\sum_{i=1}^{N_t(a)} R_i}{N_t(a)}, \tag{12}$$

where R_i is the instant reward when action is selected for the i-th time. Obviously, we can maintain a record of all previous rewards and then perform the calculation to get the estimated $Q(a)$ value. However, the problem is that the memory and computational requirements would grow over time as more rewards are obtained. For simplicity, we can devise incremental formulas for updating averages with small, constant computation required to process each new reward.

Denote t_k and $t_{(k-1)}$ as the time when action a is selected for the k-th time and the $(k-1)$-th time. Denote $Q_{t_k}(a)$ and $Q_{t_{(k-1)}}(a)$ as the estimated value of action a when action a is selected for k-th time and (k-1)-th time. Given the instant reward of action a at t_k as R_{t_k}, $Q_{t_k}(a)$ can be iteratively updated as,

$$Q_{t_k}(a) = Q_{t_{(k-1)}}(a) + \frac{1}{k}(R_{t_k} - Q_{t_{(k-1)}}(a)). \tag{13}$$

The value of $Q(a)$ is not changed between t_{k-1} and $t_{(k-1)}$. Because the value of $Q(a)$ is only updated when action is selected again. In reinforcement learning, there is always a dilemma between the exploration and exploitation. When the agent exploits the environment, it will choose the current optimal action based on previous trials. Meanwhile, when the agent explores, it will randomly choose an action hoping that it may yield a higher reward. In other words, an agent can investigate new actions by exploring, while by exploiting it selects the best action from the already investigated actions. ϵ-greedy and softmax are two common approaches to find a balance between exploration and exploitation. In this paper, we adopt the ϵ-greedy policy as the exploration strategy. For ϵ-greedy policy, the probability that an agent selects a random action is ϵ and probability that the action with the highest Q-value is chosen is $(1-\epsilon)$. Instead of using a constant ϵ, we adopt an adjustable ϵ in that the value of ϵ should be high in the beginning to ensure that more explorations are performed. But when the agent has already investigated some actions, more exploitation should be performed. Therefore, in this paper, we decay the value of ϵ after a number of N_ϵ by a p_ϵ (e.g.,0.1) and keep ϵ the same when ϵ reaches the minimum value $\epsilon_{\min}$.

Based on whether the information of both networks can be exchangeable or not, two learning based algorithms are proposed, i.e., cooperative learning algorithm and non-cooperative learning algorithm. When two networks are cooperative, the throughput information of both systems can be obtained by both networks. On the other hand, when two networks are not cooperative, each network can only obtain the information of its own. Both learning algorithms can be performed in a distributed manner which means LAA and Wi-Fi networks perform the learning algorithm independently and separately. Since the learning algorithm in this work focus on the coexistence between Wi-Fi network and LAA network, to ensure fairness between different users in the same network, we assume that all LTE-LAA nodes adopt the same window size and the minimum window size of all the Wi-Fi nodes are the same.

A. *Cooperative Learning Algorithm*

We define each element in cooperative learning as follows.

- Agent:
  Because, we assume that all Wi-Fi nodes use the same minimum window size and all LTE-LAA nodes adopt the same window size. Thus, the AP of the Wi-Fi network can be considered as an agent. Also, the BS of the LTE-LAA network is another agent.
- Action:
  The action of each agent is to select the window size that can maximize the system throughput while satisfying the fairness constraint. For Wi-Fi AP, the minimum window size C_w can be chosen from set $\{C_w, C_w+2 \ldots C_{w_{\max}}\}$. Similarly, the window size C_l can be selected in set $\{C_l, C_l+2 \ldots C_{l_{\max}}\}$.
- Reward:
  We define the reward function for each action as,

$$R(\text{action}) = \begin{cases} Th_l + Th_w + 100, & \text{if } \left|\frac{Th_l}{Th_w} - 1\right| \le \zeta \\ 100 - |\frac{Th_l}{Th_w} - 1|10, & \text{if } \left|\frac{Th_l}{Th_w} - 1\right| \ge \zeta \end{cases}$$

  When the fairness requirement is not satisfied, the larger reward means that the ratio between $\frac{Th_l}{Th_w}$ is closer to 1. Meanwhile, when the fairness requirement is met, larger reward means larger total system throughput. Adding 100 when the fairness constraint is satisfied can ensure the

Algorithm 1: Cooperative Learning Algorithm for Optimal Contention Window Selection.

Input:
$C_{w_{min}}$, the minimum value of C_w ; $C_{w_{max}}$, the maximum value of C_w;
$C_{l_{min}}$, the minimum value of C_l; $C_{l_{max}}$, the maximum value of C_l;
ϵ, the exploration policy parameter; N_ϵ, the number of iterations before reducing ϵ;
ϵ_{min}, the minimum value of of ϵ; ζ, the tolerance rate;
Output:
C_w^*, the optimal C_w; C_l^*, the optimal C_l;
For both Wi-Fi and LTE-LAA nodes ;
while *convergence is not reached* **do**
  if N_ϵ *has been reached* **then**
    $\epsilon = max(\epsilon_{min}, \epsilon - p_\epsilon)$
  Randomly choose $p_e \in [0, 1]$;
  if $p_e < \epsilon$ **then**
    Enter exploration ;
    Select the next action randomly
  else
    Enter exploitation;
    Select the next action with maximum Q value
  Receive an reward ;
  if $|\frac{Th_l}{Th_w} - 1| < \zeta$ **then**
    Receive the reward: $Th_l + Th_w + 100$
  else
    Receive the reward: $100 - |\frac{Th_l}{Th_w} - 1|10$
  Update the Q table;

reward when the fairness constraint is satisfied must be larger than the reward when the fairness constraint is not satisfied.

In Algorithm 1, we present the procedure of the cooperative learning. The proof of convergence of the cooperative learning algorithm can be found in [36]. In [36], authors prove that the convergence can be guaranteed for multi-agent system if one state Q learning algorithm is used and all agents have common interest. In our cooperative learning algorithm, the updating rule of Q matrix which is in (12) is same as one state Q learning in [36]. In addition, the agents in our proposed learning algorithm have same throughput-related reward, thus having common interest. Therefore, the convergence of the cooperative learning algorithm is guaranteed.

B. Non-Cooperative Learning Algorithm

We define each element in non-cooperative as follows.

- Agent:
  The Wi-Fi AP and the LTE-LAA BS are the agents. We assume that the Wi-Fi AP can choose the window size for

Algorithm 2: Non-cooperative Learning Algorithm for Optimal Contention Window Selection.

Input:
$C_{w_{min}}$, the minimum value of C_w; $C_{w_{max}}$, the maximum value of C_w;
$C_{l_{min}}$, the minimum value of C_l; $C_{l_{max}}$, the maximum value of C_l;
ϵ, the exploration policy parameter; N_ϵ, the number of iterations before reducing ϵ;
ϵ_{min}, the minimum value of of ϵ; ζ, the tolerance rate;
Output:
C_w^*, the optimal C_w; C_l^*, the optimal C_l;
For Wi-Fi AP ;
while *convergence is not reached* **do**
  if N_ϵ *has been reached* **then**
    $\epsilon = max(\epsilon_{min}, \epsilon - p_\epsilon)$
  Randomly choose $p_e \in [0, 1]$;
  if $p_e < \epsilon$ **then**
    Enter exploration ;
    Select the next action randomly
  else
    Enter exploitation;
    Select the next action with maximum Q value
  Receive an reward ;
  if $|\frac{On_l}{On_w} - 1| < \zeta$ **then**
    Receive the reward: $Th_w + On_l + 100$
  else
    Receive the reward: $100 - |\frac{On_l}{On_w} - 1|10$
  Update the Q table;
For LTE-LAA BS
while *convergence is not reached* **do**
  if N_ϵ *has been reached* **then**
    $\epsilon = max(\epsilon_{min}, \epsilon - p_\epsilon)$
  Randomly choose $p_e \in [0, 1]$;
  if $p_e < \epsilon$ **then**
    Enter exploration ;
    Select the next action randomly
  else
    Enter exploitation;
    Select the next action with maximum Q value
  Receive an reward ;
  if $|\frac{On_l}{On_W} - 1| < \zeta$ **then**
    Receive the reward: $Th_l + On_w + 100$
  else
    Receive the reward: $100 - |\frac{On_l}{On_w} - 1|10$
  Update the Q table;

TABLE II
PARAMETERS

MAC/PHY Header	34/16 Bytes
Wi-Fi Payload	1500 B
σ_w	9 ms
σ_l	20 ms
SIFS/PIFS/DIFS/ACK	16/25/34/44 μs
T_l	1 ms
Transmission rate	54 Mbps

all Wi-Fi nodes. Similarly, the window size of all LTE-LAA nodes can be controlled by one BS.

- Action:
  For each agent, the actions can be chosen is the selection of window size. For Wi-Fi AP, window size C_w can be chosen from set $\{C_w, C_w+2 \ldots C_{w_{\max}}\}$. Similarly, the window C_l can be selected in set $\{C_l, C_l+2 \ldots C_{l_{\max}}\}$.
- Reward:
  Since in non-cooperative learning, the throughput of other network is not obtainable, one indicator which can imply the throughput of other network should be adopted. Thus, instead of using the real ratio of $\frac{Th_l}{Th_w}$ in cooperative learning, in the non-operative algorithm, on time is introduced to represent the throughput of other network. On time is defined as the sum of both successful transmissions and collisions. When Wi-Fi AP performs the learning algorithm, the reward function is defined as,

$$R(\text{action}) = \begin{cases} Th_w + On_l + 100, & \text{if } \left|\frac{On_l}{On_w} - 1\right| \le \zeta \\ 100 - |\frac{On_l}{On_w} - 1|10, & \text{if } \left|\frac{On_l}{On_w} - 1\right| \ge \zeta, \end{cases}$$

  where On_l and On_w are the on time for LTE and Wi-Fi nodes.
  When LTE-LAA BS performs the learning algorithm, the reward is defined as,

$$R(\text{action}) = \begin{cases} Th_l + On_w + 100, & \text{if } |\frac{On_l}{On_w} - 1| \le \zeta \\ 100 - |\frac{On_l}{On_w} - 1|10, & \text{if } \left|\frac{On_l}{On_w} - 1\right| \ge \zeta \end{cases}$$

The non-cooperative learning algorithm is described in Algorithm 2.

VI. Performance Evaluation

To validate the proposed analytical model and evaluate the proposed learning algorithms, extensive simulations have been performed using MATLAB. The main parameters used in the simulations are listed in Table II.

To better illustrate the effects of parameter setting, we plot the collision probability and system throughput under different combinations of CCA and slot duration. For Wi-Fi nodes, the parameters are set based on the IEEE 802.11a/ac standard, i.e., $CCA_w = 34\ \mu$s and $\sigma_w = 9\ \mu$s [37]. The probability that Wi-Fi and LTE-LAA nodes transmit concurrently is shown in Fig. 3(a). The collision probability can not be ignored if the LTE-LAA

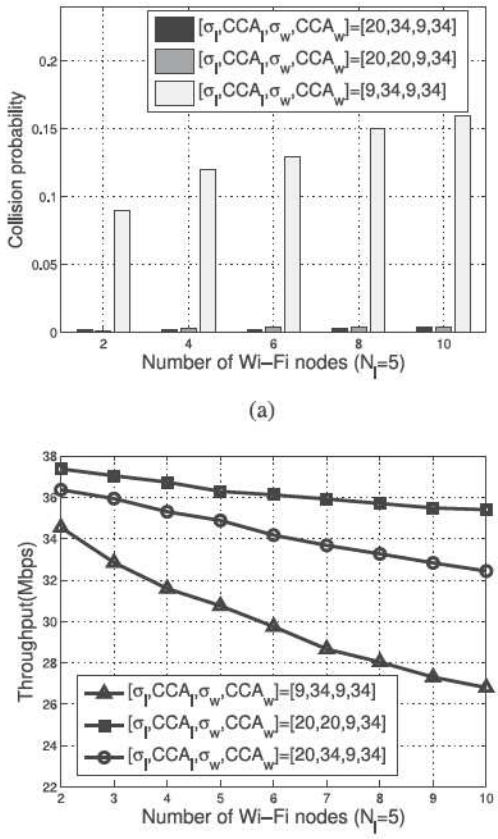


Fig. 3. Effects of parameter setting. (a) Collision probability. (b) System throughput.

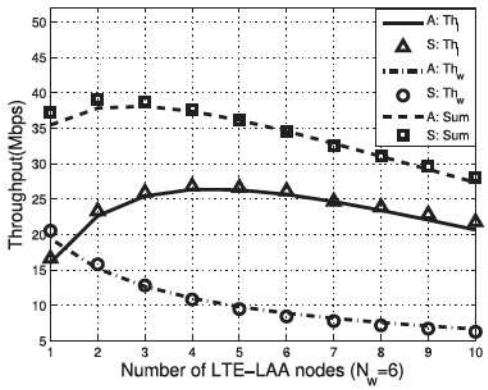


Fig. 4. Throughput vs N_l.

and Wi-Fi nodes adopt the same slot duration and CCA time. Also, the collision probability increases with the number of Wi-Fi nodes due to the increased contentions. It can be seen that the collision probability can be greatly mitigated when the slot duration σ_l of LTE-LAA nodes is co-prime with the slot duration σ_w of Wi-Fi nodes. Therefore, the system with $\sigma_l = 20\ \mu$s performs much better than the system with $\sigma_l = 9\ \mu$s which is shown in Fig. 3(b). In addition, the system throughput with $CCA_w = 20\ \mu$s and $\sigma_w = 20\ \mu$s is larger than the system where $CCA_w = 34\ \mu$s and $\sigma_w = 20\ \mu$s. Although both settings have

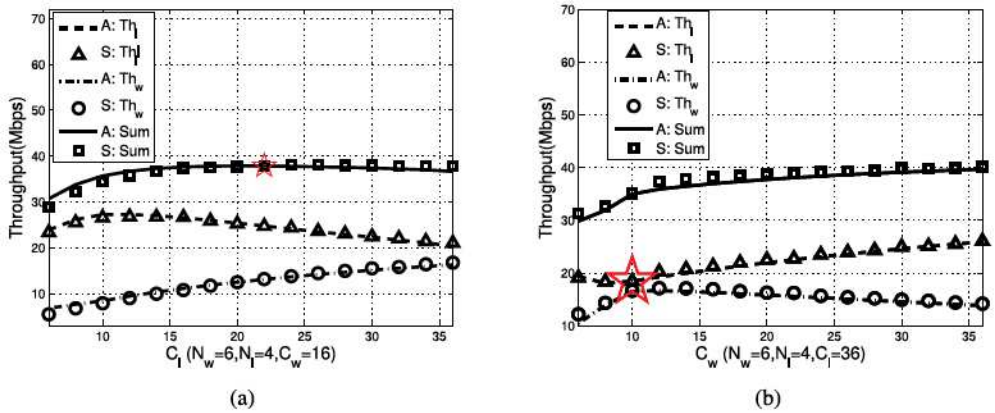


Fig. 5. Throughput when one system is adaptive. (a) Throughput vs C_l: (b) Throughput vs C_w.

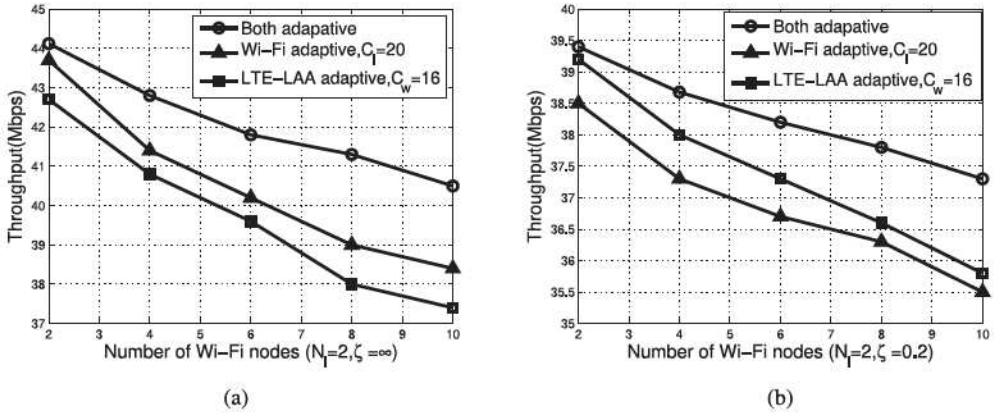


Fig. 6. Both adaptive vs Only one adaptive. (a) Without fairness constraint. (b) With fairness constraint.

negligible collision probability, the latter setting has lager CCA time which decreases the spectrum efficiency. The above results not only prove the importance of the parameter setting, but also justify the assumption on the collisions between LTE-LAA nodes and Wi-Fi nodes in the analytical parts.

The throughput of both LTE-LAA and Wi-Fi nodes are plotted in Fig. 4. The number of Wi-Fi nodes is six and C_l and C_w are both 16. As shown in Fig. 4, the total throughput of LTE-LAA network increases when the number of LTE-LAA nodes is smaller than four due to the increased chance of transmissions from LTE-LAA users. But when there are more than four LTE-LAA nodes, LTE-LAA throughput keeps dropping due to the increased number of collisions between two LTE-LAA nodes. The throughput of Wi-Fi nodes monotonically decreases when the number of LTE-LAA nodes increases. Since, with more LTE-LAA nodes, Wi-Fi nodes are less likely to win the competition and transmit in the unlicensed band. The fairness issue between LTE-LAA and Wi-Fi nodes becomes serious when the number of LTE-LAA nodes is larger than one. The fairness issue arises mainly due to the longer transmission time from LTE-LAA and the window size of LTE-LAA is not adaptive as Wi-Fi which adopts the exponential backoff window. To improve the coexistence performance, the window size of LTE-LAA nodes should be more adaptive. In Fig. 5, we plot the performance when the window size of only one network is adaptive. It can be observed in Fig. 5(a) that the throughput of Wi-Fi nodes increases as the window size of LTE-LAA nodes increases due to the longer waiting time of LTE-LAA nodes provides more opportunities for Wi-Fi nodes to transmit. In addition, when C_l equals 22, the maximum system throughput can be obtained. But, LTE-LAA and Wi-Fi nodes can not achieve similar performance when only the window size of LTE-LAA is adaptive. The results in Fig. 5(b) show that the unfairness issue can be improved if the window size of Wi-Fi users is adaptive. Also, when C_w equals 10 both networks have similar throughput. Simulation results validate the proposed analytical model.

According to the previous figures, the coexistence performance can be improved when the window size of only one network is appropriately selected. In Fig. 6, we plot the comparisons between two window size are adaptive and only the window size of either LTE-LAA or Wi-Fi is adaptive. When the window size of LTE and Wi-Fi nodes are both adaptive, the range of C_l and C_w are both [6,36]. In Fig. 6(a), we can clearly observe that the system where window size of both networks are adaptive performs much better than that only the window size of one network is adaptive when there is no fairness constraint. Especially, total system throughput increases to 44 Mbps when there are two Wi-Fi nodes. And when the number of Wi-Fi nodes is 10, the difference between double adaptive and only one is adaptive becomes almost 3M. The gap between both adaptive

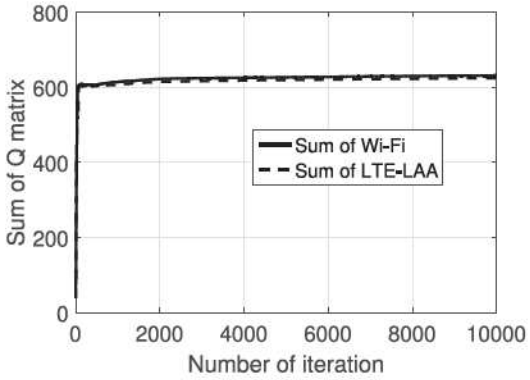


Fig. 7. Sum of Q matrix (Cooperative learning).

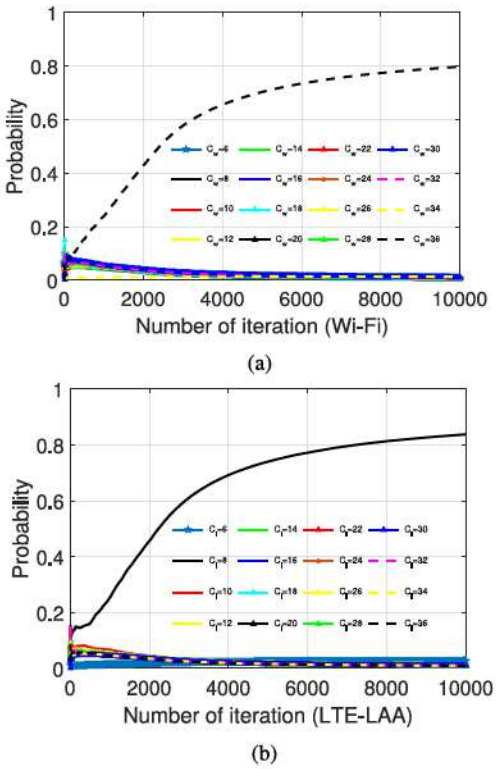


Fig. 8. Probability of action selection. (a) Wi-Fi AP. (b) LTE-LAA BS.

and only one is adaptive becomes larger when the number of Wi-Fi nodes increases. Because the appropriate selection window size can mitigate the collisions when there are more users. Thus, the selection of the window size becomes more critical when there are more competitions in the channel. To jointly consider the fairness and the system, the performance comparison between double adaptive and single adaptive given fairness constraint is plotted in Fig. 6(b). When the fairness tolerance rate is 0.2, the optimal total system throughput is smaller than the optimal total system throughput when there is no fairness constraint. Although the gap between double adaptive and only one is adaptive becomes smaller when there is fairness constraint, both adaptive system still provides gain compared to only one network is adaptive.

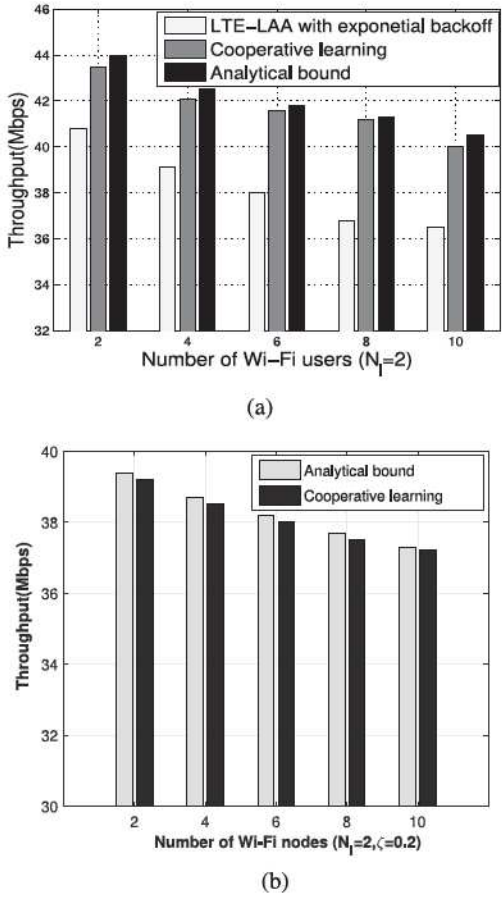
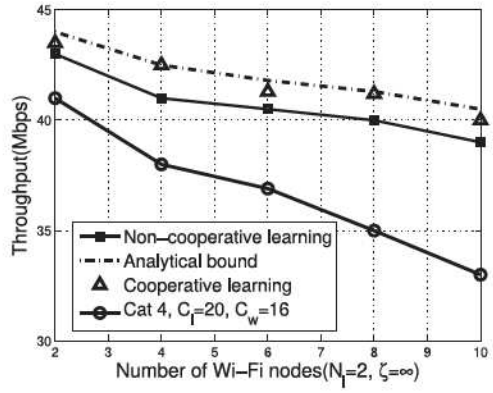


Fig. 9. Performance of cooperative learning. (a) Without fairness constraint. (b) With fairness constraint.

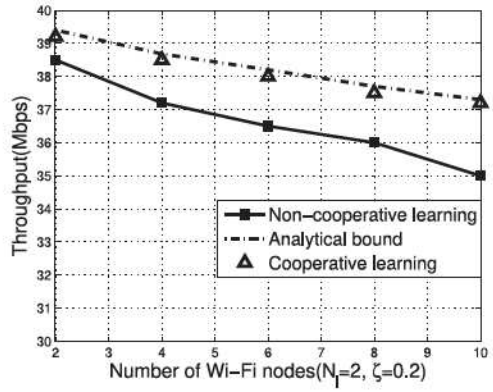
Fig. 7 illustrates the convergence of the cooperative learning algorithm. The horizontal axis is the number of iterations and the vertical axis is the sum of the values in the Q matrix. When the sum of the Q matrix converges, the agent has finished the learning procedure and can perform the optimal action in any state. As shown in the figure, the reward continues to grow in the beginning. After a sufficient amount of iterations, the agent has learned the optimal window size which can obtain the maximum throughput while satisfying the fairness requirement. Furthermore, the sum of Q matrix of both LAA and Wi-Fi tends to be the same.

In Fig. 8(a), the probability of action selection for LTE-LAA and Wi-Fi nodes are plotted when there are four LTE-LAA nodes and two Wi-Fi nodes. Based on the figures, we can see that the selection of the actions is uniformly distributed at the beginning of the learning. After some iterations, agents obtain some knowledge of the reward on different actions and select the actions based on the reward. For Wi-Fi AP, it tends to choose window size as 36 with 0.8 probability. And LTE-LAA BS tends to select the window size as 8.

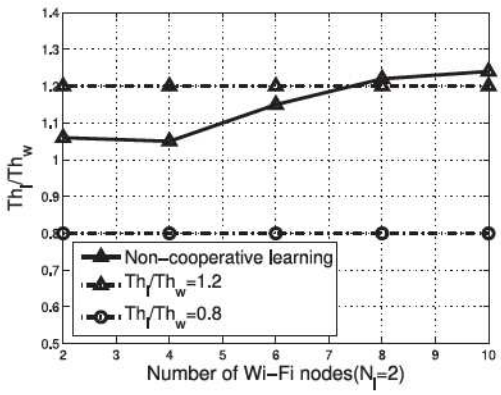
The performance comparisons between the analytical bound and cooperative learning algorithm with fairness constraint and without fairness constraint are plotted in Fig. 9. In Fig. 9(a), the tolerance rate is set to infinity which means the optimization



(a)



(b)



(c)

Fig. 10. Performance of non-cooperative learning. (a) Throughput without fairness constraint (b) Throughput with fairness constraint (c) $\frac{Th_l}{Th_w}$.

problem only considers the total system throughput. In addition, the average throughput when LTE-LAA nodes adopt the exponential backoff is also plotted in Fig. 9(a). The results show that the cooperative learning algorithm can closely reach the analytical bound and perform much better than exponential backoff. Similarly, in Fig. 9(b), the difference between the cooperative learning and analytical bound is sufficiently small.

The performance of the proposed non-cooperative learning algorithm is plotted in Fig. 10. In Fig. 10(a), we can observe that the non-cooperative learning algorithm can not perform as good as cooperative learning algorithm due to the less obtainable information. But the non-cooperative algorithm can still perform much better than the original Cat 4 algorithm. In Fig. 9(b), the throughput of the non-cooperative algorithm under fairness constraint is compared with analytical upper bound and cooperative algorithm. As shown in Fig. 10(b), with more Wi-Fi nodes, the gap between cooperative learning algorithm and analytical bound becomes larger. Because with more Wi-Fi nodes, the difference between $\frac{On_l}{On_w}$ and $\frac{Th_l}{Th_w}$ becomes larger due to the increased collisions from Wi-Fi nodes. Due to the similar reason, in Fig. 10(c), when the number of Wi-Fi nodes is small, the non-cooperative algorithm can satisfy the fairness requirement. But when there are more Wi-Fi nodes, the fairness constraint can not be satisfied.

VII. Conclusion

In this work, we have developed an analytical framework to evaluate the performance of LTE-LAA over unlicensed band using the protocol of Cat 4. The upper bound of total throughput has been calculated using the proposed analytical model. By considering that the full network information may not be obtainable and the network is dynamic, reinforcement learning has been introduced to adaptively tune the key parameters in both networks. Based on whether the information between different networks is exchangeable or not, two versions of distributed learning based algorithms which are cooperative and non-cooperative have been put forward. The simulation results have shown that both proposed learning algorithms can significantly improve the total throughput performance while satisfying the fairness constraints. Particularly, the cooperative learning can mostly reach the analytical upper bound. Due to the less obtainable information, non-cooperative algorithm cannot perform as well as cooperative learning algorithm, but it can still provide substantial gain compared to Cat 4 algorithm. In our future work, we will enhance the algorithms by considering the hidden node problem. Analysing LTE-LAA in multi-hop is also under investigation.

References

[1] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper," 2017. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[2] Z. M. Fadlullah, C. Wei, Z. Shi, and N. Kato, "GT-QoSec: A game-theoretic joint optimization of QoS and security for differentiated services in next generation heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 1037–1050, Feb. 2017.

[3] M. Han, S. Khairy, L. X. Cai, Y. Cheng, and F. Hou, "Capacity analysis of opportunistic channel bonding over multi-channel WLANs under unsaturated traffic," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1552–1566, Mar. 2020.

[4] R. Zhang, M. Wang, L. X. Cai, Z. Zheng, X. Shen, and L. Xie, "LTE-unlicensed: The future of spectrum aggregation for cellular networks," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 150–159, Jun. 2015.

[5] "3GPP tr 36.889 (V13.0.0) Study on Licensed-Assisted Access to unlicensed spectrum," 2015. [Online]. Available: http://www.3gpp.org/release-13

[6] C. Zhang, P. C. Zhang, P. Patras, H. Haddadi, Patras, and H. Haddadi. "Deep learning in mobile and wireless networking: A survey." *IEEE Commun. Surveys Tut.*, vol. 21, no. 3, pp. 2224–2287, 2019.

[7] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Commun. Surv. Tut.*, vol. 21, no. 4, pp. 3039–3071, Oct.–Dec. 2019.

[8] N. Kato *et al.*, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Trans. Wireless Commun.*, vol. 24, no. 3, pp. 146–153, Jun. 2017.

[9] F. Tang, Y. Kawamoto, N. Kato, and J. Liu, "Future intelligent and secure vehicular network toward 6G: Machine-learning approaches," *Proc. IEEE*, vol. 108, no. 2, pp. 292–307, Feb. 2020.

[10] Z. M. Fadlullah *et al.*, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Commun. Surv. Tut.*, vol. 19, no. 4, pp. 2432–2455, Oct.–Dec. 2017.

[11] B. Zheng, P. He, L. Zhao, and H. Li, "A hybrid machine learning model for range estimation of electric vehicles," in *Proc. IEEE Global Commun. Conf.*, Dec. 2016, pp. 1–6.

[12] H. Liang, X. Zhang, J. Zhang, Q. Li, S. Zhou, and L. Zhao, "A novel adaptive resource allocation model based on SMDP and reinforcement learning algorithm in vehicular cloud system," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 10018–10029, Oct. 2019.

[13] Q. Li, L. Zhao, J. Gao, H. Liang, L. Zhao, and X. Tang, "SMDP-based coordinated virtual machine allocations in cloud-fog computing systems," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1977–1988, Jun. 2018.

[14] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for interference management," *IEEE Trans. Signal Process.*, vol. 66, no. 20, pp. 5438–5453, Oct. 2018.

[15] A. Sadeghi, F. Sheikholeslami, A. G. Marques, and G. B. Giannakis, "Reinforcement learning for adaptive caching with dynamic storage pricing," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2267–2281, Oct. 2019.

[16] F. Tang *et al.*, "On removing routing protocol from future wireless networks: A real-time deep learning approach for intelligent traffic control," *IEEE Trans. Wireless Commun.*, vol. 25, no. 1, pp. 154–160, Feb. 2018.

[17] C. Chen, R. Ratasuk, and A. Ghosh, "Downlink performance analysis of LTE and WiFi coexistence in unlicensed bands with a simple listen-before-talk scheme," in *Proc. IEEE Veh. Technol. Conf.*, May 2015, pp. 1–5.

[18] W. Wang, P. Xu, Y. Zhang, and H. Chu, "Performance analysis of LBT Cat4 based downlink LAA-WiFi coexistence in unlicensed spectrum," in *Proc. 9th Int. Conf. Wireless Commun. Signal Process.*, Oct. 2017, pp. 1–6.

[19] S. Khairy, L. X. Cai, Y. Cheng, Z. Han, and H. Shan, "A hybrid-LBT MAC with adaptive sleep for LTE LAA coexisting with Wi-Fi over unlicensed band," in *Proc. IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–6.

[20] M. Han, S. Khairy, Z. Chen, L. X. Cai, and Y. Cheng, "A performance comparison of LBE based coexistence protocols for LAA and Wi-Fi," in *Proc. IEEE Int. Conf. Commun.*, May 2018, pp. 1–6.

[21] M. Cierny *et al.*, "Fairness vs. performance in Rel-13 LTE licensed assisted access," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 133–139, Dec. 2017.

[22] C. Cano and D. J. Leith, "Coexistence of WiFi and LTE in unlicensed bands: A proportional fair allocation scheme," in *Proc. IEEE Int. Conf. Commun. Workshop*, Jun. 2015, pp. 2288–2293.

[23] X. Yan, H. Tian, and C. Qin, "A markov-based modelling with dynamic contention window adaptation for LAA and WiFi coexistence," in *Proc. IEEE Veh. Technol. Conf.*, Jun. 2017, pp. 1–6.

[24] H. Ko, J. Lee, and S. Pack, "A fair listen-before-talk algorithm for coexistence of LTE-U and WLAN," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10116–10120, Dec. 2016.

[25] E. Almeida *et al.*, "Enabling LTE/WiFi coexistence by LTE blank subframe allocation," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2013, pp. 5083–5088.

[26] F. Hao, Y. Chang, H. Li, J. Zhang, and W. Quan, "Contention window size adaptation algorithm for LAA-LTE in unlicensed band," in *Proc. Int. Symp. Wireless Commun. Syst.*, Sep. 2016, pp. 476–480.

[27] H. Yi, Y. Liu, F. Pingzhi, F. Sangsha, and M. Yongfu, "An adaptive access control mechanism for LAA and Wi-Fi coexistence in unlicensed band," in *Proc. 3rd IEEE Int. Conf. Comput. Commun.*, Dec. 2017, pp. 469–473.

[28] 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, pp. 27278–27293, 2018.

[29] Z. Ali, L. Giupponi, J. Mangues-Bafalluy, and B. Bojovic, "Machine learning based scheme for contention window size adaptation in LTE-LAA," in *Proc. IEEE 28th Annu. Int. Symp. Personal, Indoor, Mobile Radio Commun.*, Oct. 2017, pp. 1–7.

[30] N. Rupasinghe and I. Güvenç, "Reinforcement learning for licensed-assisted access of LTE in the unlicensed spectrum," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2015, pp. 1279–1284.

[31] Y. Liu and S. Yoo, "Dynamic resource allocation using reinforcement learning for LTE-U and WiFi in the unlicensed spectrum," in *Proc. 9th Int. Conf. Ubiquitous Future Netw.*, Jul. 2017, pp. 471–475.

[32] 3GPP TR 36.889 (V13.0.0), "Study on licensed-assisted access to unlicensed spectrum," Jun. 2015.

[33] L. X. Cai, X. Shen, J. W. Mark, L. Cai, and Y. Xiao, "Voice capacity analysis of WLAN with unbalanced traffic," *IEEE Trans. Veh. Technol.*, vol. 55, no. 3, pp. 752–761, May 2006.

[34] P. Belotti, J. Lee, L. Liberti, F. Margot, and A. Wächter, "Branching and bounds tightening techniques for non-convex MINLP," *Optim. Methods Softw.*, vol. 24, no. 4-5, pp. 597–634, 2009.

[35] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[36] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems," *AAAI/IAAI*, vol. 1998, no. 746-752, p. 2, 1998.

[37] *IEEE Standard for Information Technology– Telecommunications and Information Exchange Between Systemslocal and Metropolitan Area Networks– Specific Requirements–Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications–Amendment 4: Enhancements for Very High Throughput for Operation in Bands Below 6 GHz*, Standard 802.11ac-2013 (Amendment to IEEE Standard 802.11-2012, as amended by IEEE Standard 802.11ae-2012, IEEE Standard 802.11aa-2012, and IEEE Standard 802.11ad-2012), pp. 1–425, Dec. 2013.

Mengqi Han received the B.S degree from the Department of Electronic and information, Nanjing University of Science and Technology, Nanjing, China, in 2013, and the M.S. degree from the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, USA, in 2015. Now she is pursuing the Ph.D degree from the Department of Electrical and Computer Engineering in Illinois Institute of Technology. Her research interests include performance analysis of MAC protocol and protocol design for next generation wireless networks, wireless networks resource management, reinforcement learning.

Sami Khairy (Student Member, IEEE) received the B.S. degree in computer engineering from the University of Jordan, Amman, Jordan, in 2014 and the M.S. degree in electrical engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 2016. He is currently working towards a Ph.D. degree in electrical engineering at the Illinois Institute of Technology. His research interests span the broad areas of analysis and protocol design for next generation wireless networks, AI powered wireless networks resource management, reinforcement learning, statistical learning, and statistical signal processing. He received a Fulbright Predoctoral Scholarship from JACEE and the U.S. Department of State in 2015, and the Starr/Fieldhouse Research Fellowship from IIT in 2019. He is an IEEE Student Member and a Member of IEEE ComSoc and IEEE HKN.

Lin X. Cai received the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Waterloo, Canada, in 2005 and 2010, respectively. She was a Postdoctoral Research Fellow in Electrical Engineering Department at Princeton University in 2011 before she joined Huawei US Wireless R&D center as a Senior Engineer in 2012. She has been an Assistant Professor with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, Illinois, USA, since August 2014. Her research interests include green communication and networking, broadband multimedia services, and radio resource and mobility management. She received a Postdoctoral Fellowship Award from the Natural Sciences and Engineering Research Council of Canada (NSERC) in 2010, a Best Paper Award from the IEEE Globecom 2011, and an NSF Career Award in 2016. She is an Associated Editor of IEEE TRANSACTION ON WIRELESS COMMUNICATIONS, *IEEE Network Magazine*, and a Co-Chair for IEEE conferences.

Yu Cheng (Senior Member, IEEE) received B.E. and M.E. degrees in electronic engineering from Tsinghua University in 1995 and 1998, respectively, and a Ph.D. degree in electrical and computer engineering from the University of Waterloo, Canada, in 2003. He is now a Full Professor in the Department of Electrical and Computer Engineering, Illinois Institute of Technology. His research interests include wireless network performance analysis, network security, big data, cloud computing, and machine learning. He received a Best Paper Award at QShine 2007, IEEE ICC 2011, and a Runner-Up Best Paper Award at ACM MobiHoc 2014. He received the National Science Foundation (NSF) CAREER Award in 2011 and IIT Sigma Xi Research Award in the junior faculty division in 2013. He has served as several Symposium Co-Chairs for IEEE ICC and IEEE GLOBECOM, and Technical Program Committee (TPC) Co-Chair for WASA 2011 and ICNC 2015. He was a founding Vice Chair of the IEEE ComSoc Technical Subcommittee on Green Communications and Computing. He was an IEEE ComSoc distinguished Lecturer in 2016–2017. He is an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE INTERNET OF THINGS JOURNAL, and IEEE WIRELESS COMMUNICATIONS. He is an IEEE Senior Member.

Ran Zhang (Member, IEEE) received the B.E. degree (2010) in electronics information science from Tsinghua University, Beijing, China, and Ph.D. degree (2016) in electrical and computer engineering from the University of Waterloo, Canada, respectively. He then worked as a System Engineer at Huawei Ottawa Research Center from 2016 to 2018. He joined the Department of Electrical and Computer Engineering in Miami University, USA at 2018 as a Visiting Assistant Professor, and has been an Assistant Professor there since 2019. His research interests include artificial intelligence powered next-generation wireless communication and networking, channel coding, radio resource management in cellular networks, and Internet of Things.