Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9985972, 15 pages
https://doi.org/10.1155/2021/9985972

WILEY | Hindawi

Research Article

Reinforcement Learning for Joint Channel/Subframe Selection of LTE in the Unlicensed Spectrum

Yuki Kishimoto, Xiaoyan Wang, and Masahiro Umehira

Graduate School of Science and Engineering, Ibaraki University, 4-12-1 Nakanarusawa, Hitachi, Ibaraki 316-8511, Japan

Correspondence should be addressed to Xiaoyan Wang; xiaoyan.wang.shawn@vc.ibaraki.ac.jp

Received 25 March 2021; Accepted 18 May 2021; Published 2 June 2021

Academic Editor: Keping Yu

In recent years, to cope with the rapid growth in mobile data traffic, increasing the capacity of cellular networks is receiving more and more attention. To this end, offloading the current LTE-advanced or 5G system's data traffic from licensed spectrum to the unlicensed spectrum that is used by WiFi systems, i.e., LTE-Licensed-Assisted-Access (LTE-LAA), has been extensively investigated. In the current LTE-LAA system, a Listen-Before-Talk (LBT) approach is implemented, which requires the LTE user also perform carrier sense before the transmission. However, fair LTE-WiFi coexistence is still hard to guarantee due to their unbalanced frame sizes and traffic loads. In the LTE-LAA system, the optimal channel selection and subframe number adjustment are the keys to realize efficient spectrum utilization and fair system coexistence. To this end, in this paper, we propose a reinforcement learning-based joint channel/subframe selection scheme for LTE-LAA. The proposed approach is implemented at the LTE access points with zero knowledge of the WiFi systems. The results of extensive simulations verify that the proposed approach can significantly improve the fairness and packet loss rate compared with baseline schemes.

1. Introduction

In the past few years, we have witnessed a phenomenal growth in mobile data traffic. This growth is accelerated by the increasing number of mobile and Internet of Things (IoT) devices and the popularity of spectrum-hungry wireless applications such as online games and high definition videos. Since most of the mobile data traffic is carried by cellular networks, both the academia and industry have made many attempts to increase the capacity of LTE networks to accommodate this surging growth and exploit the possible enhancement on the future 5G networks [1–3]. Originally, the mobile communication system LTE is capable of providing 150 (Mbps) data rate with a maximum bandwidth of 20 (MHz). As the demand for high-speed communication is further increasing, currently used LTE-advanced system utilizes Carrier Aggregation (CA) technology to speed up the communication by bundling multiple 20 (MHz) LTE carriers.

The highest licensed spectrum used for downlink LTE communication in Japan is 3.5 (GHz), and the unlicensed spectrum is 5 (GHz). Since the communication capacity is proportional to the frequency bandwidth, aggregating multiple noncontiguous channels in the unlicensed 5 (GHz) band enables higher capacity of LTE networks. However, the unlicensed band is already used by other wireless systems such as WiFi networks. Based on the fact that LTE is a schedule-based technology, which would severely degrade the performance of WiFi by letting its transmission backoffs continuously, it is necessary to modify LTE to enable the fair coexistence between different wireless systems. To this end, Licensed-Assisted-Access (LAA) technology [4] has been proposed in 2013, which uses Listen-Before-Talk (LBT) approach to let LTE system assess the channel state before transmitting. Additionally, in March 2020, 3GPP (Third-Generation Partnership Project) committed to 5G NR (New Radio) in unlicensed spectrum in Release 16, which extends the LAA from LTE to 5G [5, 6].

However, even with LBT, the fair coexistence issue between LTE and WiFi systems on unlicensed spectrum is still nontrivial, due to their unbalanced frame size and traffic volume. For instance, the transmission duration for LTE varies from 2 to 10 milliseconds, but a typical WiFi transmission only lasts for a few hundreds of microseconds [7]. Considering a dense scenario that multiple LTE and WiFi systems

colocate to share multiple channels, the optimal channel and subframe number (In this article, we use the term "subframe number" to indicate "the number of subframes"). Selections are the keys for efficient and fair spectrum utilization. Furthermore, in a dynamic network, it is very important and challenging to dynamically adjust the optimal channel and subframe number according to the varying environment.

The channel selection and subframe number adjustment problems have been widely investigated in recent years. The most common channel selection method is to select a channel with minimum received power by using channel assessment. This method, unfortunately, suffers from low channel utilization efficiency and fails to well support the dynamic network environment. A learning-based channel selection mechanism for LTE operation in unlicensed bands was proposed in [8]. However, it needs global information of the coexistence system and does not take fairness into consideration. In [9], Challita et al. proposed a proactive channel selection scheme for LTE-U system by exploiting deep learning. However, it requires a WiFi traffic load distribution dataset as input and thus is hard to be applied in a dynamic environment. In [10], an online learning distributed channel selection scheme for 5G NR-U has been proposed, which focuses on optimal channel selection for uplink traffic offloading by formulating it to a noncooperative game. Regarding the work related on subframe number adjustment, the idea of blank LTE subframe was first proposed by Almeida et al. [11]. The goal is to achieve fair medium access by giving more transmission opportunities to WiFi systems. Based on [11] and recent advances in learning techniques [12–14], a Q-learning-based muting period selection scheme was proposed for fair LTE-WiFi coexistence in [15]. However, it focuses on maximizing the LTE throughput and can only work in single channel scenario.

Recently, joint channel and subframe number selection problem has been investigated. A joint user association and resource allocation approach for LTE-WiFi coexistence was proposed in [16], which aims at maximizing the number of users supported by LTE. However, they do not consider the traffic balance problem between LTE and WiFi systems. In [17], a double Q-learning-based scheme was proposed to achieve efficient LTE-WiFi coexistence by jointly considering channel selection, discontinuous transmission, and transmit power control. However, the goal of this research is efficiently utilizing the idle time as much as possible, instead of achieving fair medium access. In [18], the authors proposed a duty cycle optimization scheme by considering both the fairness and the throughput. However, the proposed scheme assumed that the throughput information needs to be exchanged between LTE and WiFi systems to perform Q-learning, which is hard to realize in reality. In [19], a deep reinforcement learning-based dynamic resource allocation algorithm to reduce the latency of devices has been proposed, which focuses on reducing the latency of mission critical devices in accessing uplink resources of the small cell network. In [20], a novel framework that uses flying UAV-enabled networks to provide service for VR users in an LTE-U system has been proposed, which does not take into account the fairness between LTE and WiFi systems. In [21], the coexistence of LTE and ZigBee networks at the unlicensed frequency band of 2.4 GHz is studied, which focuses on performance evaluations. To summarize, the aforementioned works cannot be applied on an LTE-LAA system in a distributed and dynamic way, with the purpose of achieving fairness and efficiency simultaneously.

To this end, in this paper, we propose a joint channel/-subframe number selection scheme for LTE-LAA system by exploiting reinforcement learning technique. The proposed scheme is distributedly implemented at LTE Access Points (APs) and requires zero knowledge from the WiFi systems. In the proposed scheme, the LTE AP monitors the traffic volume on the currently used channel and dynamically learns the optimal channel and subframe number selections to efficiently utilize the spectrum resource and maintain each system's throughput to its target value as close as possible. To minimize the frequent channel switching created by traffic variation, we further propose an enhanced scheme with channel switch penalty. The effectiveness of the proposed scheme is verified by extensive simulation results. We compare the proposed scheme with baseline schemes to show its superiority in terms of fairness and packet loss rate. Although the proposed joint channel/subframe number selection scheme focuses on LTE-LAA system, it could be easily extended to 5G NR-U system.

The rest of the paper is organized as follows. Section 2 briefly introduces the preliminaries of LTE-LAA system. Section 3 presents the system model and the proposed reinforcement learning-based scheme. Section 4 provides the simulation scenario and evaluation results. Section 5 presents the issue on undesired channel switches in a complicated dense scenario and presents an enhanced scheme with evaluation results. Finally, Section 6 concludes this paper.

2. Preliminaries of LTE-LAA System

2.1. Channel Access Policy. In LTE-LAA systems, efficient and fair channel access methods are required. LBT mechanism [4] has been proposed, with the purpose that LTE communications in the unlicensed spectrum do not significantly deteriorate the performance of nearby WiFi systems. With LBT, LTE system also performs carrier sense before the transmission, and the transmission is initiated only when the channel is idle. Figure 1 shows the concept of LBT.

2.2. Channel Selection Scheme. Appropriate channel selection is important to fully utilize the spectrum resource, especially when there are multiple wireless systems colocate nearby. In a common channel selection scheme [22], APs periodically monitor all the channels and calculate the average received power at a constant interval. The channel with the minimum received power in the current interval will be selected to access at the next interval. Sensing-based channel selection scheme has the following drawbacks. Firstly, since channel sensing and communications cannot be performed simultaneously, the channel utilization efficiency is low. Secondly, since the sensing period is randomly assigned, the sensing result may not accurately reflect the real channel utilization

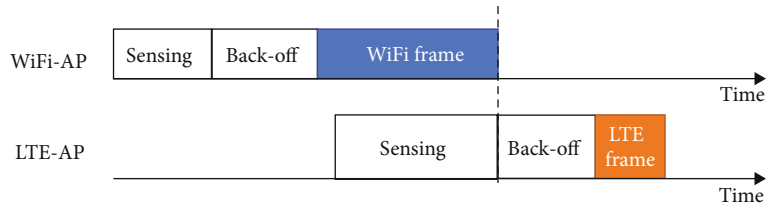


Figure 1: Conceptual diagram of LBT.

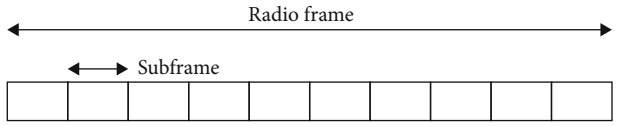


Figure 2: Radio frame structure for TDD downlink.

states. Last but not least, it works poor in a dynamic network environment due to the fixed long sensing cycle.

2.3. Radio Frame Structure for LTE. Similar as most of the previous researches [8, 9, 11, 15, 16], only downlink communication is considered in this work. Additionally, we assume that the LTE communications use Time Division Duplex (TDD), in which the radio frame structure is illustrated in Figure 2. The maximum number of subframes in one LTE frame is 10, and the length of one subframes is 1 (ms). Some of the subframes could be muted to give more channel access opportunities to potentially colocated systems such as WiFi [11, 23]. By dynamically varying subframe number, colocated LTE and WiFi systems could share the medium in a fair way.

2.4. LTE-U and LAA. An attempt to provide LTE communication in unlicensed frequency bands, i.e., LTE-Unlicensed (LTE-U) [24], has been originally proposed in 3GPP release 10. LTE-U mainly performs channel selection and duty cycle dynamic adjustment based on power measurements of the surrounding environment. However, the introduction of LTE-U may significantly degrade the performance of WiFi systems. To solve this problem, LTE-LAA was standardized in 3GPP release 13, which uses LBT to perform carrier sense before transmission. In this paper, the proposed scheme is based on LTE-LAA, in which the LBT is used.

3. Proposed Reinforcement Learning-Based Joint Channel/Subframe Selection Scheme

3.1. Reinforcement Learning. The proposed scheme is based on a typical reinforcement learning algorithm, i.e., Q-learning [25, 26]. Reinforcement learning is one of the three basic machine learning paradigms, by which the agent learns the optimal behavior through repeated interactions with the environment in discrete time steps. The learning is performed in an online fashion, and it does not need large amount of labeled data with correct input/output pairs. Figure 3 shows a conceptual diagram of Q-learning. In time step t, the agent observes the environmental state s_t and receives a reward r_t. The agent chooses an action a_t, and the environment evolves to the next state s_{t+1} and feeds reward r_{t+1} back to the agent. The action could be either

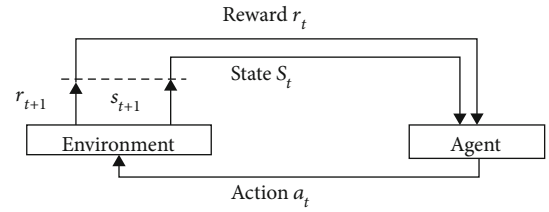


Figure 3: Conceptual diagram of the Q-learning.

Table 1: Definition of the proposed reinforcement learning-based scheme.

Parameter	Definition
State s	Low/high traffic volume
Action a	(channel, subframes number) selection pair
Reward R	$\begin{cases} (X_{\text{Own}}/L_{\text{frame}}) & \text{if (State : Low traffic volume)} \\ (X_{\text{Own}}/X_{\text{Own}} + X_{\text{Other}}) - \beta \cdot \rho_{\text{fair}} & \text{else} \end{cases}$

exploring the action space or exploiting the optimal action that gives the most cumulative reward as a result of a series of continuous actions. The optimal action is chosen based on a Q-table, in which the cell's value $Q(s_k, a_k)$, i.e., Q value, represents the value of the (state, action) pair. At the end of time slot t, the Q value is updated by

$$Q(s_t, a_t) \longleftarrow Q(s_t, a_t) + \alpha\{r_{t+1} + \gamma \max Q(s_{t+1}, a_t) - Q(s_t, a_t)\}, \quad (1)$$

where α is a learning rate and γ is a discount factor.

3.2. Proposed Scheme

3.2.1. Definition of the Proposed Scheme. In this paper, we propose a reinforcement learning-based joint channel/subframe number selection scheme, which is performed at individual LTE-APs. The state, action, and reward in the proposed scheme are defined as in Table 1. Specifically, we consider two states depending on the traffic volume of agent AP's currently used channel, i.e., low traffic volume state and high traffic volume state. The agent is in low traffic volume state if $X_{\text{Own}} + X_{\text{Other}} \leq L_{\text{frame}}$, i.e., the channel used by the AP, is unsaturated, or in high traffic volume state otherwise. Here, X_{Own} represents the agent AP's average subframe number in one learning cycle, and X_{Other} represents the average subframe number of all the other APs who share the same channel and are within the interference range with the agent

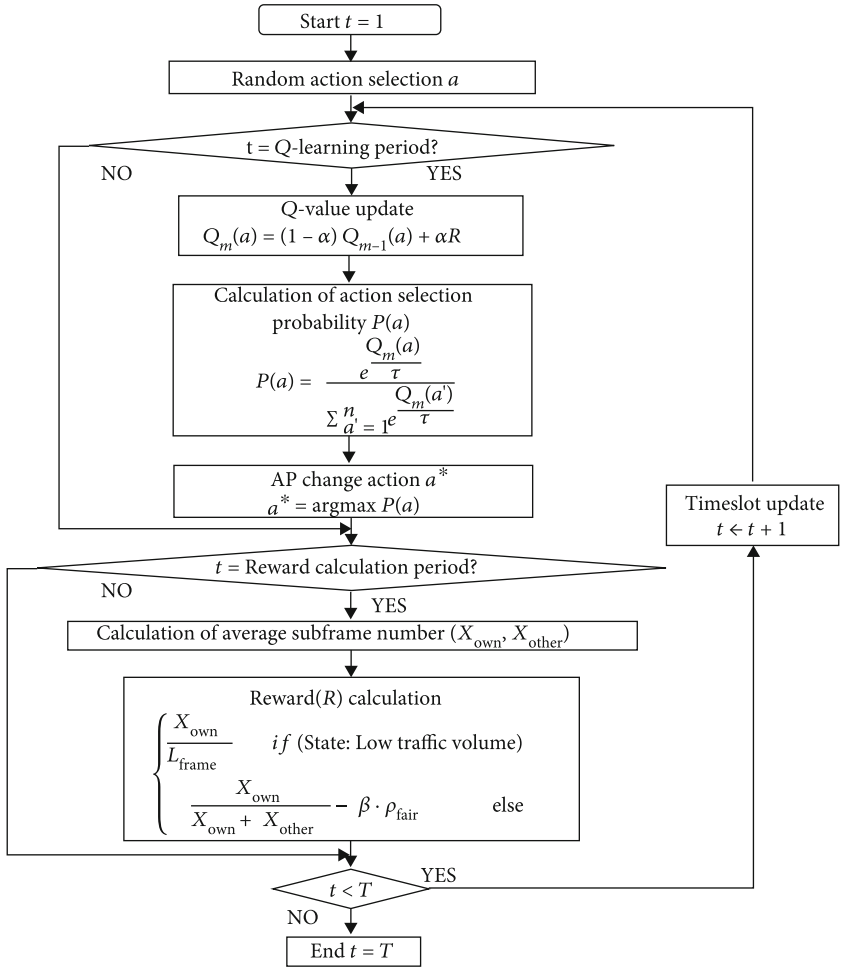


Figure 4: Flowchart of the proposed scheme.

AP. And L_{frame} is the subframe number in one frame. The actions are the (channel, subframe number) selection pairs. Notice that similar as the previous work [15, 17, 18], we focus on the selection of subframe number in this paper and leave the issue of selecting which subframe to use as the future work (In the simulations, the subframes are selected continuously in ascending order. More practical implementations that based on 3GPP standardization [4] will be considered in our future work). The reward functions are separately defined based on agent AP's current states as shown in Table 1. In low traffic state, the reward only focuses on its achieved throughput. But in high traffic state, both the throughput and fairness are taken into consideration. Here, β is a weight factor, ρ_{fair} denotes the fairness penalty which is calculated by Equation (2).

$$\rho_{\text{fair}} = \left| \frac{N_{\text{other}}}{1 + N_{\text{other}}} - \frac{X_{\text{Other}}}{X_{\text{Own}} + X_{\text{Other}}} \right|. \quad (2)$$

Here, $N_{\text{other}}/(1 + N_{\text{other}})$ represents the target fairness factor, where N_{other} is the number of APs that share the same channel and are within the interference range of the agent AP. Additionally, $X_{\text{Other}}/(X_{\text{Own}} + X_{\text{Other}})$ is the achieved fairness factor. The difference between them is defined as the fairness penalty, the lower it is, the fairer coexistence among different systems. The basic idea of the designed reward function is that in the low traffic volume state, all the APs could ideally send all the packets, but in the high traffic volume state, a fairness penalty is introduced to assure that all the APs using this channel could have equal opportunity to send the packets. Note that X_{Other} could be easily obtained by the agent AP thru channel sensing when it is not sending packets. N_{other} could be obtained by decoding the header the of the WiFi frames, which could be performed periodically when the scenario of the network is not highly dynamic. The interaction between LTE system and WiFi system is not required.

3.2.2. Flowchart and Action Selection Probability. Figure 4 shows the flowchart of the proposed scheme which is performed distributedly in each LTE-AP. Initially, the agent AP selects a random channel and subframe number. After that, when the time slot t equals the reward calculation period, the average subframe number $X_{\text{Own}}, X_{\text{Other}}$ in one cycle is calculated, and the corresponding reward R is obtained. When the time slot t equals the Q-learning period, the Q value ($Q(a)$) of action a is updated by Equation (3).

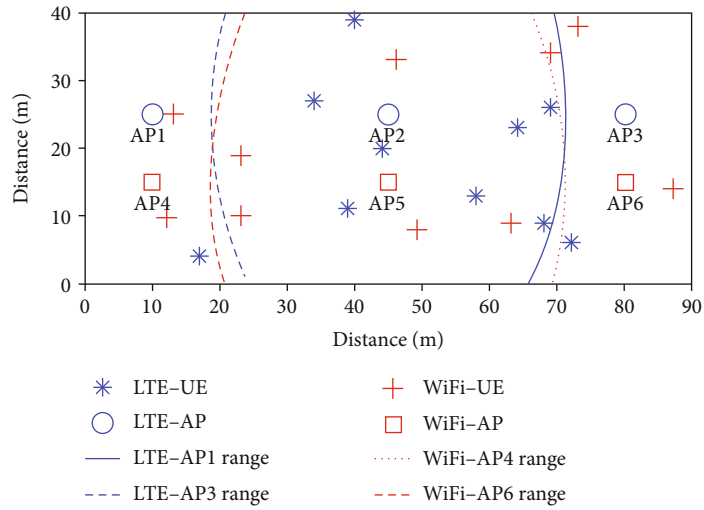


Figure 5: Simple LOS indoor scenario.

Table 2: Major parameters for simulations [15, 23].

Parameter	Value	Unit
Transmission power (P_{tx})	15	dBm
Antenna gain (G_{tx})	5	dB
Frequency band (f_c)	5	GHz
Bandwidth	20	MHz
Timeslot duration	9	μs
Timeslot number	40,000,000	
DIFS (T_{DIFS})	34	μs
SIFS (T_{SIFS})	16	μs
Backoff	[0,31]	Timeslot
Buffer size	10	Packets
Throughput calculation interval (T_{thr})	50,000	Timeslot
Radio frame length (L_{radio})	10	ms
Subframe length	1	ms

$$Q_m(a) = (1-\alpha)Q_{m-1}(a) + \alpha R, \tag{3}$$

where m is the number of times of Q-learning and α is the learning rate. It is considered that the larger the value of α, the easier it is to adapt to changing network conditions. In this paper, α evolves by Equation (4).

$$\alpha = 1 - \delta m, \tag{4}$$

where δ is a parameter that adjusts the changing speed of α. After the Q value is updated, the selection probability $P(a)$ for each action a is calculated by Equation (5).

$$P(a) = \frac{e^{(Q_m(a))/\tau}}{\sum_{a'=1}^{n} e^{(Q_m(a'))/\tau}}, \tag{5}$$

where n is the number of possible actions and τ is a design parameter representing the range of possible values of $P(a)$. The agent AP switches to the action $a^* = \text{argmax}\, P(a)$ to select the channel and subframe number.

To balance the exploration and exploitation, the parameter τ in Equation (5) gradually decreases by Equation (6)

$$\tau = \frac{\tau_0}{\log_2(1 + m/Z)}, \tag{6}$$

where τ_0 is an initial value of τ and Z is a parameter indicating the changing speed of τ. It is obvious that τ gradually decreases as m increases.

When τ is large, all actions are selected with almost the same probability regardless of the Q value (exploring for Q value). But when τ is small, the action with the largest Q value will be selected more easily (exploitation of Q value).

4. Simulation Results

4.1. Simulation Environment. We firstly consider a simple indoor LOS (Line-Of-Sight) environment to validate the performance of the proposed scheme. Similar to the previous work [8, 9, 11, 16, 17, 27], we only consider downlink communication in this work. Figure 5 shows the considered simple indoor LOS scenario with size $40(\text{m}) \times 90(\text{m})$, where three LTE-APs (APs 1-3) and three WiFi-APs (APs 4-6) are colocated to share two channels. The proposed joint channel/subframe number selection scheme is implemented at AP1 and AP3. The available channels are CH1 and CH2, and the selectable subframe numbers are [2, 4, 6, 8, 10]. To validate if the proposed scheme can adapt to the network's traffic load variations, we consider a dynamic network environment, in which the traffic volume at LTE-AP2 increases in the middle of simulations. To be specific, AP2 increases its subframe number from 1 to 10 at 180(s) and uses CH1 fixedly. Three WiFi APs use static channels, i.e., AP4, AP5, and AP6 assigned to CH1, CH1, and CH2, respectively. There are 10 LTE-UEs and 10 WiFi-UEs, each randomly placed in the network. Table 2 shows the major parameters

for simulations. WiFi system uses 802.11n standard protocol and is performed with frame aggregation.

Next, based on [28], we calculate the ranges of APs, beyond which the transmissions are unable to detect. The Threshold Level (TL) (/1 MHz) at which the carrier sensing of LBT is possible is calculated by $-73 + (23 - P_H)$, where P_H represents the transmission power P_{tx} when the antenna gain $G_{tx} = 0$ (dBi). According to the settings given in Table 2, P_H (dBm) of AP is calculated as $P_H = P_{tx} + (G_{tx} - 0) = 15 + 5 = 20$. Therefore, TL (/1 MHz) is calculated as $\{-73 + (23 - 20)\} = -70$. Since a 20 (MHz) bandwidth is assumed in the simulation, TL (dBm) is obtained as $10 \log_{10}(20/10^7)$. According to the settings given in Table 2, the received power P from the AP is calculated by Equation (7) according to [29].

$$P = -16.9 \log_{10} d - 12.8 - 20 \log_{10} f_c, \quad (7)$$

where d represents the distance (m) between the transmitting station and the receiving station and f_c represents the frequency band (GHz) used for communication. By considering the condition that P equals TL, we have the ranges as approximately 61.3 (m), beyond which the transmissions are unable to detect. The ranges of AP1, AP3, AP4, and AP6 are shown in Figure 5. Therefore, APs 3 and 6 are unable to detect the transmissions of APs 1 and 4 and vice versa.

The WiFi parameters used in this paper are shown in Table 3 based on [15, 30, 31]. Accordingly, we have the WiFi data packet length T_{data} and the ACK length T_{ACK} as 704 (μs) and 28 (μs), respectively. Table 4 shows the parameters for the proposed scheme and common channel selection method. Here, the sensing period and sensing time for the sensing-based scheme are set based on [22], which corresponds to a 1% sensing time. The learning period and reward calculation period used in the proposed scheme are set to 50000 timeslots, which indicates that the AP may change its action every 450 (ms). The other parameters in Table 4 are Q learning's parameters which are carefully adjusted to balance the performance, convergence, and adaptive capacity. The LTE and WiFi packet arrival intervals follow an exponential distribution $\lambda e^{-\lambda x}$ where x is the packet arrival interval and $\lambda = 1/\mu$. The average packet arrival interval (ms) of LTE-AP, μ_{LTE} is set as

$$\mu_{\text{LTE}} = T_{\text{DIFS}} + B_M + L_{\text{radio}} + T_{\text{ACK}} + T_{\text{SIFS}} = 10.2195, \quad (8)$$

where B_M is the average backoff. And the average packet arrival interval of WiFi-AP, μ_{WiFi}, for WiFi APs 4, 5, and 6 are set to 1.42 and 5.68 (ms), respectively. These average packet arrival intervals correspond to LTE's subframe numbers 1 and 4, respectively.

4.2. Evaluation Metrics. We compare the proposed scheme with two baseline channel selection schemes, i.e., sensing-based scheme and max-throughput scheme. Specifically, sensing-based scheme selects the channel with the minimum received power at a fixed interval. The sensing time is 1% of the whole cycle which is randomly assigned. Additionally, the max-throughput scheme is an ideal method, which chooses the channel that obtains the maximum system throughput.

Table 3: Major parameters for WiFi.

Parameter	Value	Unit
PLCP preamble + headers duration (T_{plcp})	20	μs
PLCP service field (L_S)	16	Bits
MAC header ($L_{\text{MAC\_h}}$)	224	Bits
Tail bits (L_t)	6	Bits
ACK length (L_{ack})	112	Bits
Payload (D)	12000	Bits
OFDM symbol duration (T_{sym})	4	μs
Number of bits per OFDM symbol (n_{sym})	72	Bits

Table 4: Major parameters for the proposed scheme and sensing-based scheme [22].

Parameter	Value	Unit
Learning period	50,000	Timeslot
Initial Q value	0.5	
τ_0	0.4	
Z	35	
β	2	
δ	0.00025	
Reward calculation period	50,000	Timeslot
Sensing period	4,000,000	Timeslot
Sensing time	40,000	Timeslot/period

To evaluate the effectiveness of the proposed scheme, we consider three evaluation metrics, i.e., throughput, fairness, and packet loss rate.

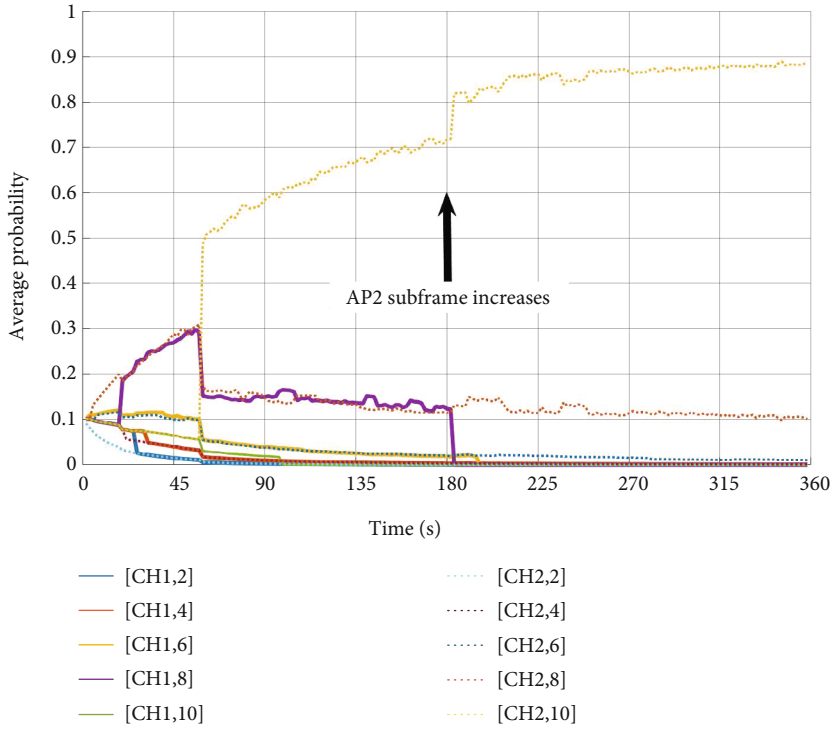
The throughput (Γ) is calculated by Equation (9).

$$\Gamma = r \times \frac{T_{\text{trans}}}{T_{\text{thr}}}, \quad (9)$$

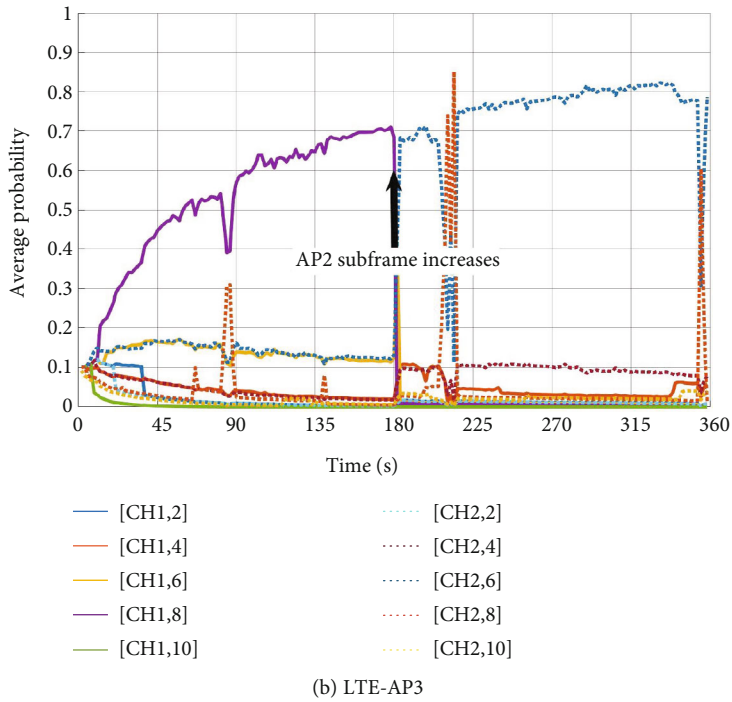
where r is the data rate and T_{trans} and T_{thr} indicate the AP's transmission time and the throughput calculation interval, respectively. QPSK (Quadrature Phase Shift Keying) modulation is adopted, and thus, the data rates for LTE and WiFi communications are 15.6 and 18 (Mbps), respectively [30]. In this work, the fairness (Ψ) needs to take into consideration the different traffic volume at different APs, which is defined by Equation (10).

$$\Psi = \frac{\Gamma}{\Gamma_{\text{ideal}}}, \quad (10)$$

where Γ_{ideal} is the throughput achieved when the AP operates in an isolated manner, i.e., access the channel without sharing it with any other systems. A low fairness value indicates that the AP's throughput is significantly affected by other systems that share the same channel.



(a) LTE-AP1



(b) LTE-AP3

Figure 6: Action selection probability for the proposed scheme.

The packet loss occurs when the queuing packets exceed the buffer size, and the lost packets plus the packets in the buffer are counted when we calculate the packet loss rate.

4.3. Simulation Results. First, we confirm the adaptive capacity of the proposed scheme in a dynamic network environment. Figure 6 shows the variance of action selection probability over time of LTE-AP1 and AP3. To make it clear, the illustrated probability is averaged by 4 times. The legend denotes (channel, subframe number). The arrow in the middle indicates the timing at which LTE-AP2 changes its subframe number proactively. We can observe that the action selection probability converges as time goes by. This indicates that the process of exploring for the Q value is gradually changing to the process of exploiting. From Figure 6(a), we can observe that AP1's highest action selection probability

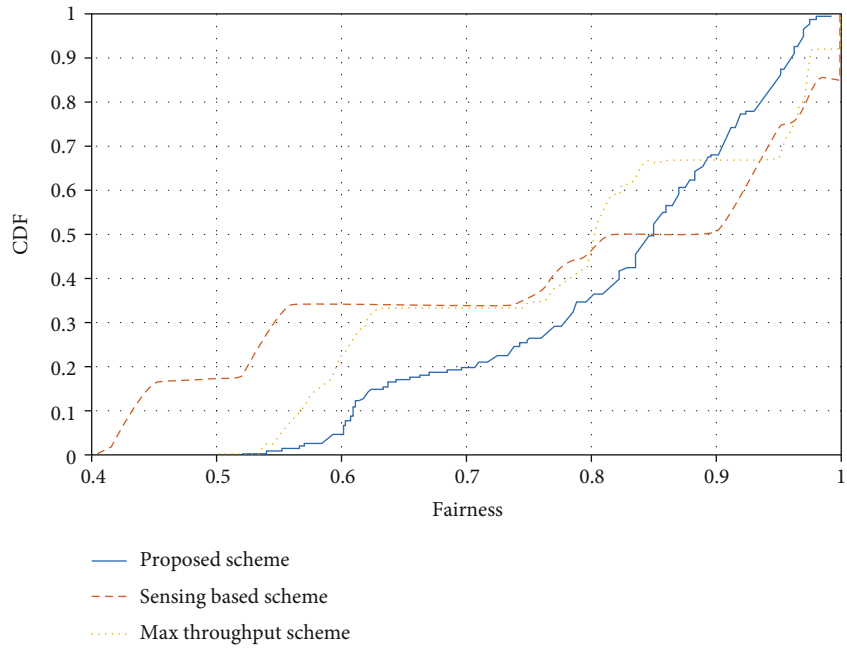


Figure 7: CDF of fairness for different schemes.

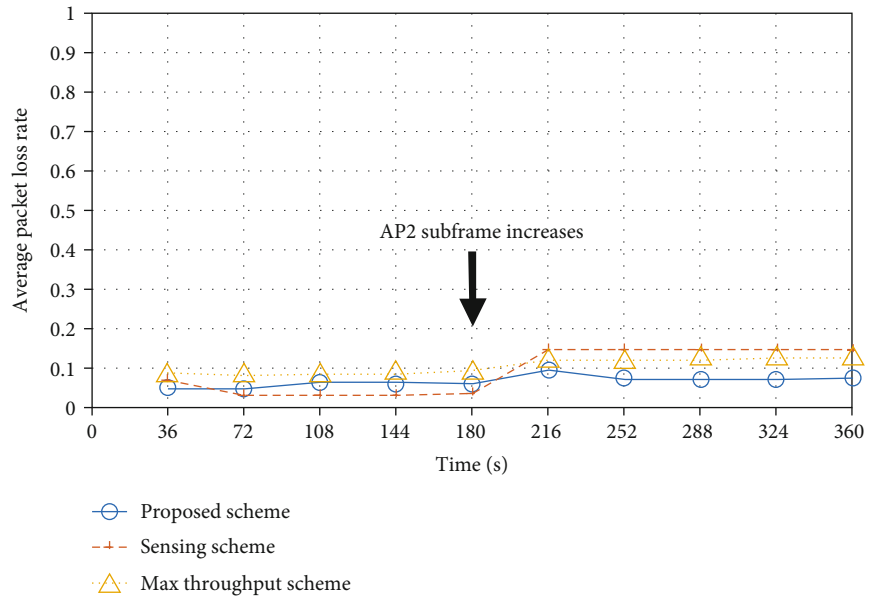


Figure 8: Average packet loss rate of all APs for different schemes.

is [CH2,10], which is the optimal result as expected. This is because there is no other AP in the range of AP1 that shares CH2, and thus, $X_{\text{Other}} = 0$. Therefore, large X_{Own} value can be obtained in the range of $X_{\text{Own}} + X_{\text{Other}} \leq L_{\text{frame}}$, which maximizes the reward. From Figure 6(b), for AP3, the probability of action [CH1,8] is the highest in most of the first half time. This is because subframe number of AP2 is 1, and thus, the value of X_{Other} is smaller in CH1 than that in CH2. Therefore, large reward could be obtained in the range of $X_{\text{Own}} + X_{\text{Other}} \leq L_{\text{frame}}$. However, when the subframe number of AP2 increases from 1 to 10 in the middle of the simulation, AP3's state changes from low traffic volume state into high traffic volume state. Accordingly, AP3's action with the highest probability changes to [CH2,6] immediately, which is as expected. This is because that using CH2 by sharing with AP6 could achieve higher reward. In addition, we can observe that the probability of [CH2,8] becomes high temporarily at around 210 (s). The reason is that the traffic volume of other APs within the transmission range is temporarily reduced at that time, due to the exponential distribution; thus, the reward of [CH2,8] becomes higher. Notice that the temporary changing of the action does not lead to

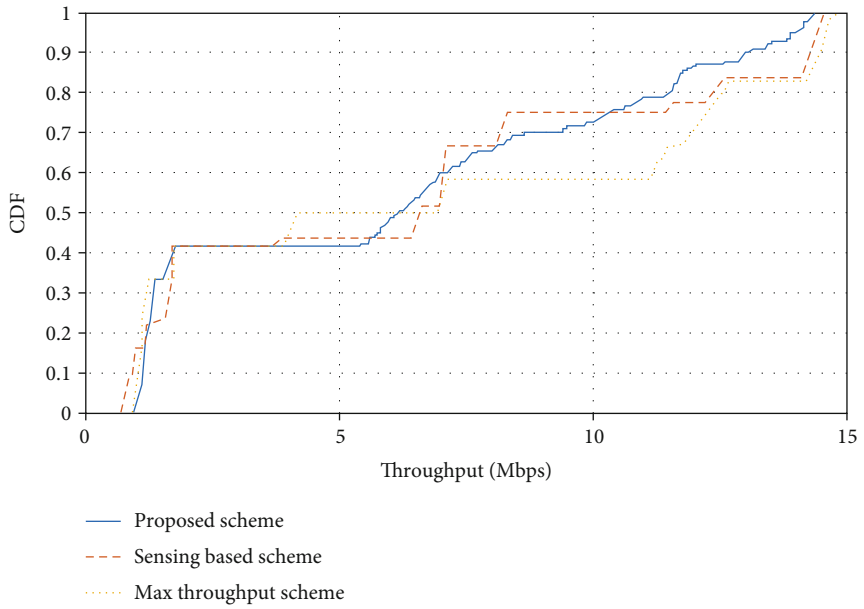


Figure 9: CDF of throughput for different schemes.

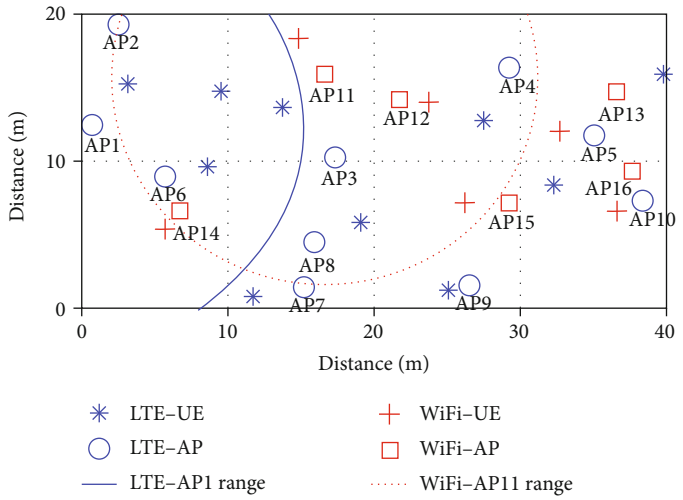


Figure 10: Dense indoor NLOS scenario.

channel switching, only the subframe number varies. Based on the previous analysis, we can conclude that the proposed scheme can adapt to the dynamic network environment automatically by selecting the ideal channel and subframe number in both low and high traffic volume states.

Next, in Figure 7, we show the Cumulative Distribution Function (CDF) of fairness for the proposed scheme, sensing-based scheme, and max-throughput scheme. All the results are averaged by three runs. We can observe that the proposed scheme significantly improves the fairness with low values, i.e., equal or lower than 0.84. This means that the proposed scheme can benefit the APs with poor performance. Additionally, the whole system's average fairness of the proposed scheme is 0.82, which is better than 0.77 for the sensing-based scheme and 0.79 for the max-throughput scheme.

Next, Figure 8 shows the comparison of the average packet loss rate of the proposed scheme, sensing-based scheme, and max-throughput scheme. We can observe that average packet loss rate of the proposed scheme is lower than that of two baseline schemes, when the system traffic volume is high.

Finally, Figure 9 shows the CDF of throughput of the proposed scheme, sensing-based scheme, and max-throughput scheme. As expected, the average throughput for all APs of the max-throughput scheme is the highest, i.e., approximately 6.8 (Mbps), compared with that of the proposed scheme and sensing-based scheme, i.e., 6.1 (Mbps) and 6.2 (Mbps), respectively. By the observations on the specific throughput results on different APs which have not been shown in this article due to limited space, we found that the low- and high-throughput values shown in Figure 9

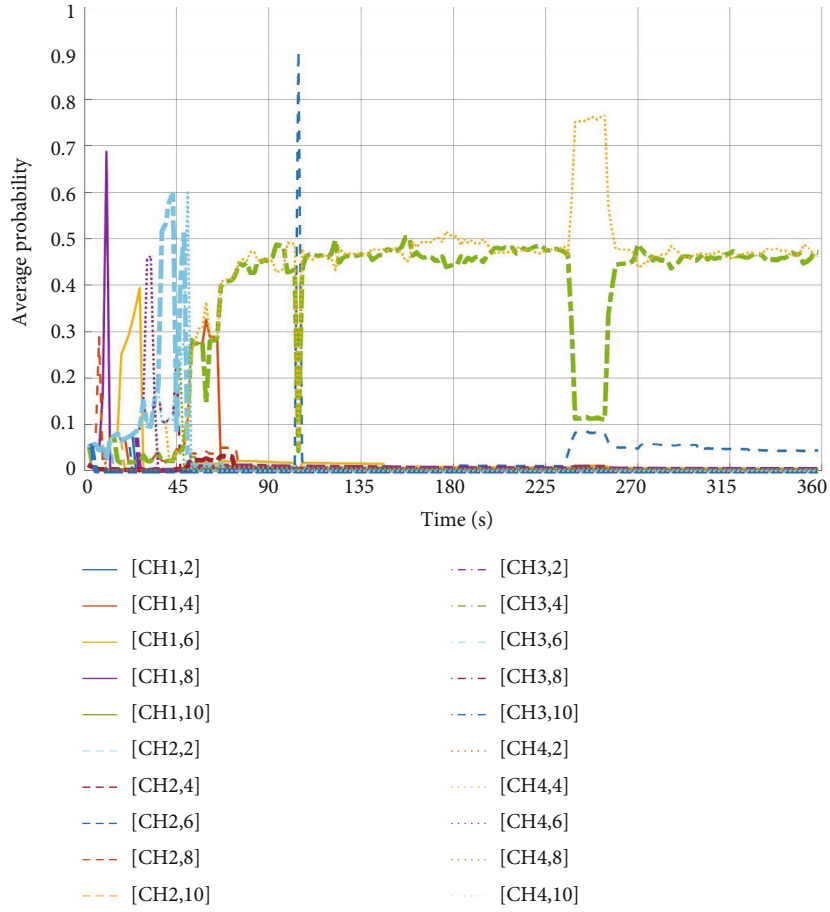


Figure 11: Action selection probability of AP5 for the original proposed scheme ($Z = 1$, $\delta = 0.001$).

correspond to WiFi and LTE systems' throughput, respectively. From Figure 9, we can observe that the throughput improvement of the max-throughput scheme mainly comes from the LTE-AP whose throughput is higher than 7 (Mbps). The reason is that in the max-throughput scheme, the LTE-AP uses the channel that leads to its maximal throughput with subframe number 10, without considering the negative impacts on WiFi systems. Notice that the max-throughput scheme is an ideal scheme which could only be realized in simulations.

5. Enhanced Joint Channel/Subframe Number Selection Scheme

5.1. Existing Problems and the Enhanced Proposed Scheme. To validate if the proposed scheme can work in a complicated scenario, we consider an extremely dense NLOS (Non-Line-Of-Sight) indoor environment in which the traffic is saturated in all channels [32]. As illustrated by Figure 10, there are 8 rooms (each of size 10 (m) × 10 (m)), and 10 LTE-APs (APs 1-10) and 6 WiFi-APs (APs 11-16) colocate to share 4 channels. There are 10 LTE-UEs and 6 WiFi-UEs in this scenario, which are associated to the closest APs. In this scenario, multiple APs are located within each other's communication range; thus, the traffic variance of one AP will affect the performance of the whole system significantly. The proposed scheme is implemented at all LTE-APs (APs 1-10). The available channels are CH1 to CH4, and the selectable subframe numbers are [2, 4, 6, 8, 10]. We still consider a dynamic network environment, in which the assigned channel at one WiFi-AP changes in the middle of the simulations. To be specific, AP14 proactively switches from CH3 to CH4 at 180 (s).

As an example, the ranges of LTE-AP1 and WiFi-AP11 are shown in Figure 10, beyond which the transmissions of AP1 and AP11 cannot be detected. These ranges at NLOS scenario are calculated as follows. TL (/1 MHz) at which the carrier sensing of LBT is possible is given by $-73 + (23 - P_H)$, so the receiving power is calculated by Equation (11) according to [29].

$$P = -43.3 \log_{10} d + 8.5 - 20 \log_{10} f_c. \quad (11)$$

By considering the condition that P equals TL, we have the ranges for NLOS scenario as approximately 14.4 (m), beyond which the transmissions cannot be detected.

μ_{WiFi} for WiFi AP 11, 13, 14, 16, 12, and 15 are set to 2.39 and 1.43 (ms), respectively. These average packet arrival intervals correspond to LTE subframe numbers 3 and 5, respectively.

In this extremely dense NLOS indoor scenario, we found that some of the APs' action selection probabilities do not

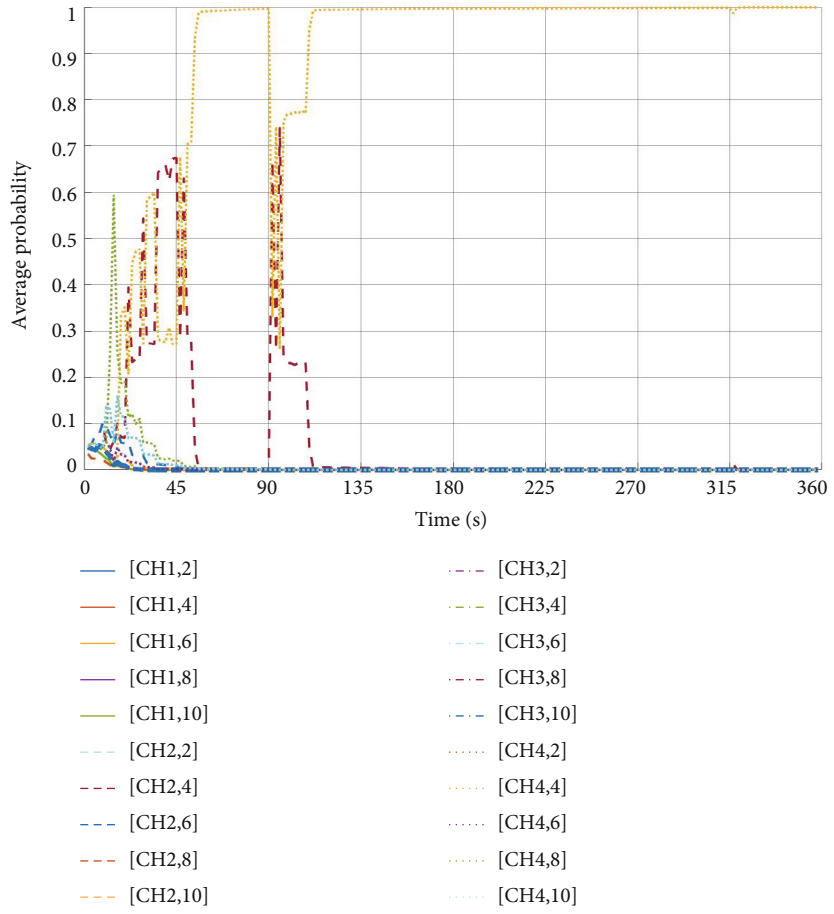


Figure 12: Action selection probability of AP5 for the enhanced proposed scheme.

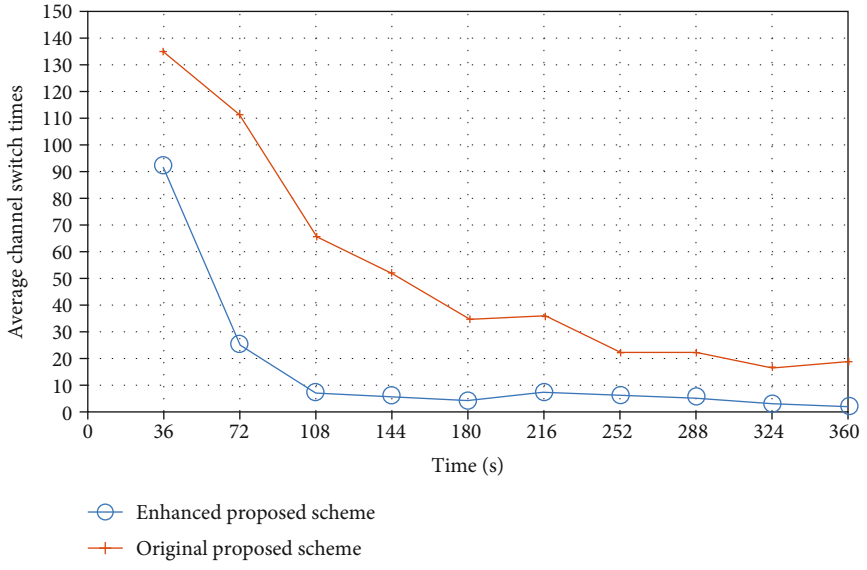


Figure 13: Average number of channel switches of all APs for different schemes.

converge by using the original proposed scheme. For instance, Figure 11 shows the variance of action selection probability over time for LTE-AP5. We can observe that CH3 and CH4 have the similar probabilities, and thus, the proposed scheme does not converge to one optimal action. The reason is that the traffic volume of CH1 and CH2 is almost saturated, and there is approximately the same amount of idle time in CH3 and CH4. This result means that using either CH3 or CH4 would lead to the same performance for AP5. However, in practice, it is not desirable to

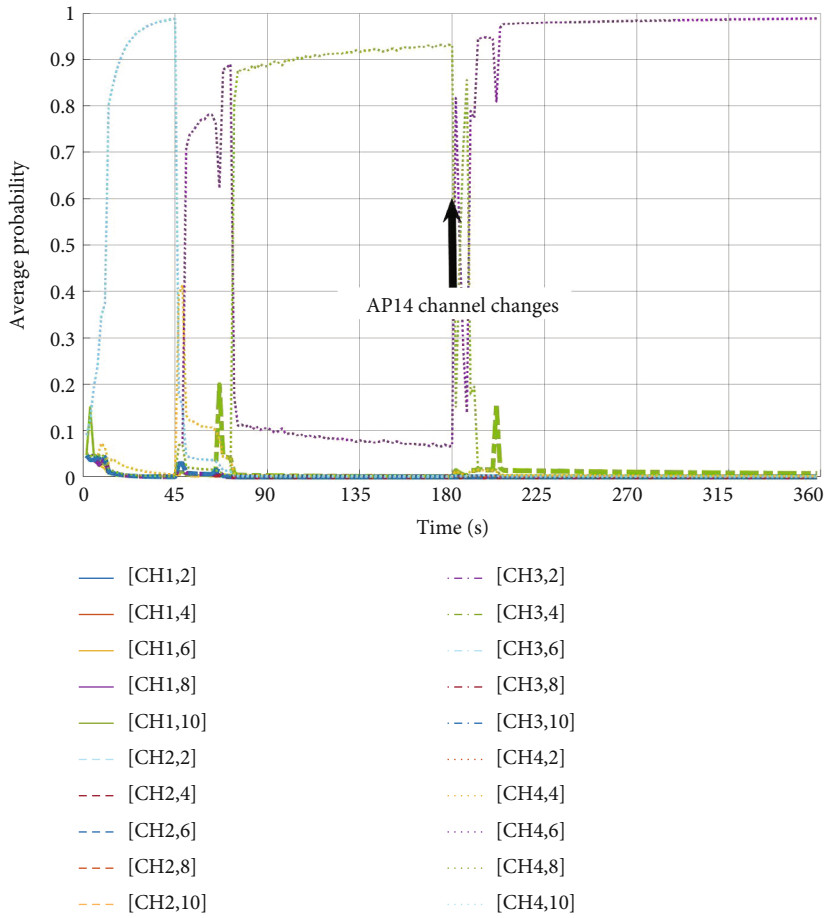


Figure 14: Action selection probability of AP1 for the enhanced proposed scheme.

switch channels frequently during communications. Noted that the sensing of channels also requires periodical channel switches. However, the channel switching during communications leads to not only the AP changes its channel but also all the connected end users.

To solve this problem, we further propose an enhanced joint channel/subframe number selection scheme to deal with this undesired frequent channel switches. In the enhanced proposed scheme, after the action selection probability is derived based on Equation (5), if the channel switch happens, the Q value is further updated by Equation (12).

$$Q_m(a^{\sim}) = Q_m(a^{\sim}) + \rho, \tag{12}$$

where $a^{\sim}$ represents the actions of channels other than the current channel. ρ represents a channel switch penalty which is set to -0.1 in the simulations. Since the channel switch penalty lowers the Q value for channels other than the channel currently used by the AP, the action of switching to another channel is suppressed. Additionally, in the enhanced proposed scheme, Z and δ are set to 10 and 0.001, respectively.

5.2. Simulation Results. Firstly, we verify that the enhanced proposed scheme could reduce the undesired frequent channel switches. In Figure 12, we show the variances of action selection probabilities over time of LTE-AP5. Compared with the result shown in Figure 11, we can observe that the action selection probability converges on [CH4,4] and approaches to 1 gradually. Thanks to the introduction of channel switch penalty, the frequent channel switch problem is solved. At around 90 (s), there is a temporary variance for action selection probability from [CH4,4] to [CH2,4]. The reason is that the traffic volume for all the APs is dynamically changing based on the exponential distribution. This temporary probability changing is undesired, since it leads to channel switches. The balance between stability and adaptability of the proposed scheme is our future work.

Next, Figure 13 shows the comparison of average number of channel switches between the enhanced proposed scheme and original proposed scheme. All the results are averaged by three runs. We can observe that average channel switch times of the enhanced proposed scheme are significantly lower than that of original proposed scheme. It could suppress the channel switch times to at most 10 times per 36 (s) after the scheme converges from 108 (s).

Next, we confirm that if the introduction of channel switch penalty would bring undesired negative effects on the adaptive capacity in a dynamic network environment. Figure 14 shows the variances of action selection probabilities over time of LTE-AP1 as an example. The arrow in the

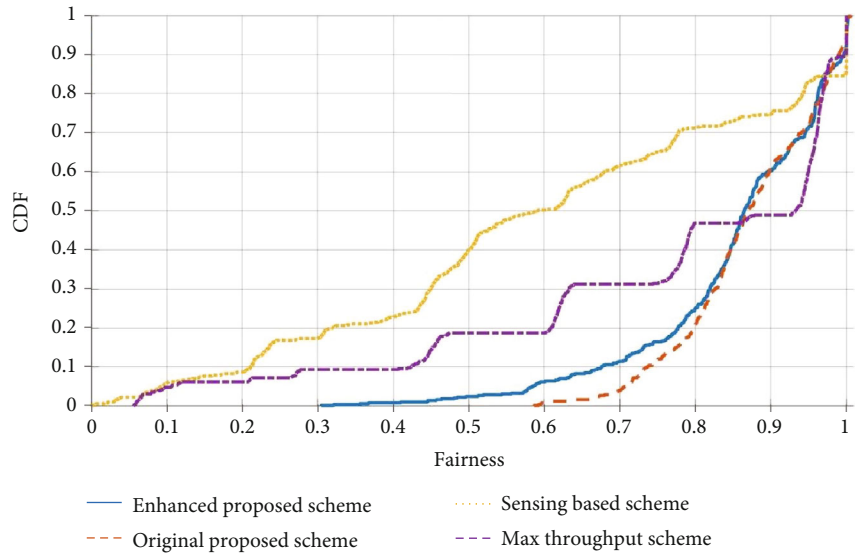


Figure 15: CDF of the fairness for different schemes.

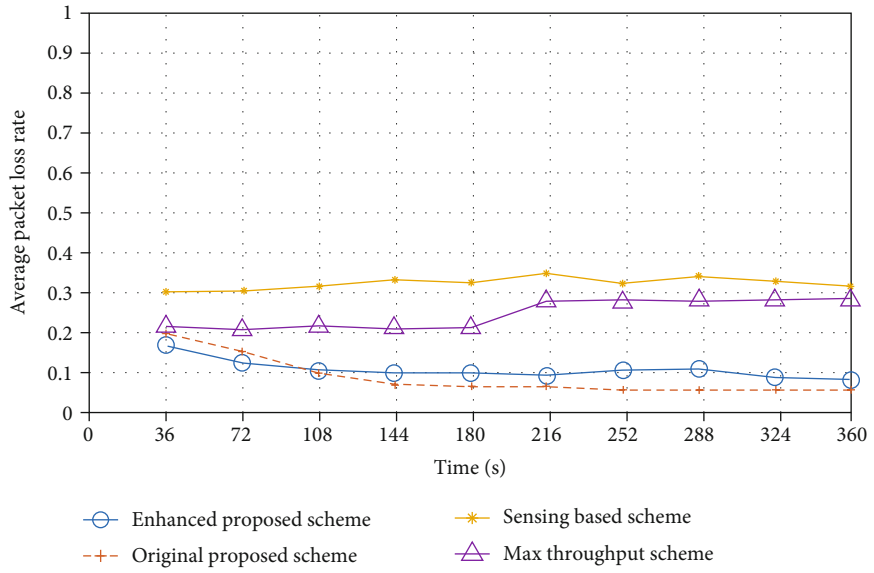


Figure 16: Average packet loss rate of all APs for different schemes.

middle indicates the timing that WiFi-AP14 proactively switches its channel. From Figure 14, we can observe that the probability of action [CH4,8] is the highest after the convergence in the first half time. This is because there is no other AP in the range of AP1 that shares CH4, and thus, $X_{\text{Other}} = 0$. Therefore, large X_{Own} value can be obtained in the range of $X_{\text{Own}} + X_{\text{Other}} \leq L_{\text{frame}}$, which maximizes the reward. However, when AP14 switches from CH3 to CH4 in the middle of the simulation, AP1's action with the highest probability changes to [CH4,6] accordingly, which is as expected. This is because that using CH4 by sharing with AP14 could achieve higher reward. Additionally, as for the selection results of other APs in this dynamic scenario, we confirm that they all make appropriate action adjustments according to this environment variance.

Next, we compare the performance of the enhanced proposed scheme, original proposed scheme, sensing-based scheme and max-throughput scheme in this dense NLOS scenario. In Figure 15, we show the CDF of fairness. We can observe that both the enhanced proposed scheme and original proposed scheme can significantly improve the fairness, especially for low fairness values, i.e., equal or lower than 0.96. Additionally, the whole system's average fairness of the enhanced proposed scheme is 0.85, which is better than 0.60 for the sensing-based scheme, 0.77 for the max-throughput scheme, and almost equal to 0.87 for the original proposed scheme.

Next, Figure 16 shows the comparison of the average packet loss rate. We can observe that the average packet loss rate of the enhanced proposed scheme is extremely lower

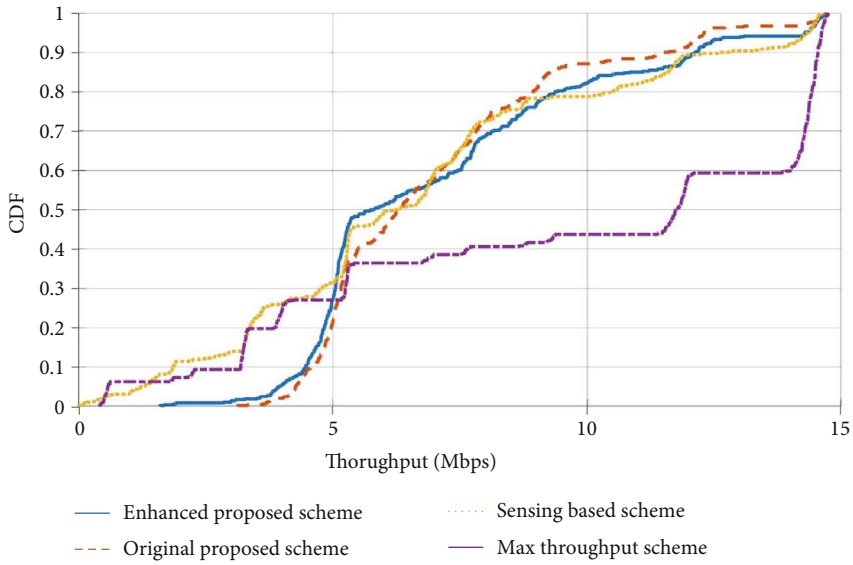


Figure 17: CDF of the throughput for different schemes.

than that of the sensing-based scheme and max-throughput scheme and slightly higher than that of the original proposed scheme. Moreover, for the max-throughput scheme, we can observe that the average packet loss rate increases by about 10% in the second half time. This indicates that the max-throughput scheme is unable to handle the channel switch of AP14. The average packet loss rates of the enhanced proposed scheme and the original proposed scheme are 0.11 and 0.090, respectively, which significantly outperform the sensing-based scheme and the max-throughput scheme, i.e., 0.33 and 0.25, respectively.

Finally, Figure 17 shows the CDF of average throughput. We can observe that both the enhanced proposed scheme and original proposed scheme could improve the throughput with low values, i.e., less than 5.0 (Mbps). Additionally, we can confirm that the average throughput of the max-throughput scheme is the highest, i.e., 9.5 (Mbps). And the average throughput of the enhanced proposed scheme is 7.2 (Mbps), which is better than 6.7 (Mbps) for the sensing-based scheme and 7.0 (Mbps) for the original proposed scheme.

To summarize, the enhanced proposed scheme can achieve almost the same fairness, packet loss rate, and throughput as the original proposed scheme, with significantly reduced channel switch times. Most importantly, the enhanced proposed scheme can realize system stability and adaptability simultaneously.

6. Conclusions

In this paper, we proposed a joint channel/subframe number selection scheme for the LTE-LAA system. It is able to achieve efficient channel utilization and fair system coexistence by exploiting reinforcement learning technique. We evaluated the effectiveness of the proposed scheme by computer simulations in two dynamic indoor environments and compared it with two baseline schemes. By using the proposed scheme, the optimal channel/subframe number can be selected even when the network conditions dynamically change. Compared with baseline schemes, both the fairness and packet loss rate are significantly improved. Especially in the extremely dense NLOS indoor scenario, by introducing the channel switch penalty, the system stability and adaptability are realized at the same time. In future work, we will consider using real WiFi traffic dataset in the simulations and optimize the learning parameters.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was supported by Grant-in-Aid for Scientific Research (C) (20K11764), the Telecommunications Advancement Foundation, and ROIS NII Open Collaborative Research (21FA01).

References

[1] W. S. H. M. W. Ahmad, N. A. M. Radzi, F. S. Samidi et al., "5G technology: towards dynamic spectrum sharing using cognitive radio networks," *IEEE Access*, vol. 8, pp. 14460–14488, 2020.

[2] K.-L. A. Yau, J. Qadir, C. Wu, M. A. Imran, and M. H. Ling, "Cognition-inspired 5G cellular networks: a review and the road ahead," *IEEE Access*, vol. 6, pp. 35072–35090, 2018.

[3] X. Wang, M. Umehira, B. Han, H. Zhou, P. Li, and C. Wu, "An efficient privacy preserving spectrum sharing framework for internet of things," *IEEE Access*, vol. 8, pp. 34675–34685, 2020.

[4] *Study on Licensed-Assisted Access Using LTE*, 3GPP Study Item RP-141397, Edinburgh, Scotland, 2014.

[5] J. Oh, Y. Kim, Y. Li, J. Bang, and J. Lee, "Expanding 5G new radio technology to unlicensed spectrum," in *2019 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, Waikoloa, HI, USA, December 2019.

[6] K. Yu, L. Lin, M. Alazab, L. Tan, and B. Gu, "Deep learning-based traffic safety solution for a mixture of autonomous and manual vehicles in a 5G-enabled intelligent transportation system," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2020.

[7] "IEEE Standard for information technology—telecommunications and information exchange between systems local and metropolitan area networks—specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," in *IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012)*, pp. 1–3534, IEEE, 2016.

[8] O. Sallent, J. Perez-Romero, R. Ferrus, and R. Agusti, "Learning-based coexistence for LTE operation in unlicensed bands," in *2015 IEEE International Conference on Communication Workshop (ICCW),*, London, UK, June 2015.

[9] U. Challita, L. Dong, and W. Saad, "Proactive resource management for LTE in unlicensed spectrum: a deep learning perspective," *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4674–4689, 2018.

[10] Y. Shi, Q. Cui, W. Ni, and Z. Fei, "Proactive dynamic channel selection based on multi-armed bandit learning for 5G NR-U," *IEEE Access*, vol. 8, pp. 196363–196374, 2020.

[11] E. Almeida, A. M. Cavalcante, R. C. D. Paiva et al., "Enabling LTE/WiFi coexistence by LTE blank subframe allocation," in *2013 IEEE International Conference on Communications (ICC),*, Budapest, Hungary, June 2013.

[12] Y. Li, J. Zhou, J. Tian, X. Zheng, and Y. Y. Tang, "Weighted error entropy based information theoretic learning for robust subspace representation," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021.

[13] Y. M. Li, J. T. Zhou, X. W. Zheng, J. Y. Tian, and Y. Y. Tang, "Robust subspace clustering with independent and piecewise identically distributed noise modeling," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, June 2019.

[14] J. Zhang, K. Yu, Z. Wen, X. Qi, and A. K. Paul, "3D reconstruction for motion blurred images using deep learning-based intelligent systems," *Computers, Materials & Continua*, vol. 66, no. 2, pp. 2087–2104, 2021.

[15] 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.

[16] J. Tan, S. Xiao, S. Han, Y. Liang, and V. C. M. Leung, "QoS-aware user association and resource allocation in LAA-LTE/-WiFi coexistence systems," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2415–2430, 2019.

[17] A. Glanopoulos, F. Foukalas, and T. A. Tsiftsis, "Efficient coexistence of LTE with WiFi in the licensed and unlicensed spectrum aggregation," *IEEE Transactions on Cognitive Communications and Networking*, vol. 2, no. 2, pp. 129–140, 2016.

[18] Y. Su, X. Du, L. Huang, Z. Gao, and M. Guizani, "LTE-U and Wi-Fi coexistence algorithm based on Q-learning in multi-channel," *IEEE Access*, vol. 6, pp. 13644–13652, 2018.

[19] M. Elsayed and M. Erol-Kantarci, "Deep reinforcement learning for reducing latency in mission-critical services," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab Emirates, December 2018.

[20] M. Chen, W. Saad, and C. Yin, "Echo state learning for wireless virtual reality resource allocation in UAV-enabled LTE-U networks," in *2018 IEEE International Conference on Communications (ICC)*, Kansas City, MO, USA, May 2018.

[21] I. Parvez, N. Islam, N. Rupasinghe, A. I. Sarwat, and I. Guvenc, "LAA-based LTE and Zig Bee coexistence for unlicensed-band smart grid communications," in *SoutheastCon 2016*, pp. 1–6, Norfolk, VA, USA, March-April 2016.

[22] S. Sengottuvelan, J. Ansari, P. Mähönen, T. G. Venkatesh, and M. Petrova, "Channel selection algorithm for cognitive radio networks with heavy-tailed idle times," *IEEE Transactions on Mobile Computing*, vol. 16, no. 5, pp. 1258–1271, 2017.

[23] A2A Research Inc, "LTE/LTE-A basics," 2020-7-2, http://www.a2a.jp/resources/lte_lte-A.pdf.

[24] S. Zinno, G. Di Stasi, S. Avallone, and G. Ventre, "On a fair coexistence of LTE and Wi-Fi in the unlicensed spectrum: a survey," *Computer Communications*, vol. 115, pp. 35–50, 2018.

[25] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[26] H. Zhou, X. Wang, M. Umehira, X. Chen, C. Wu, and Y. Ji, "Wireless access control in edge-aided disaster response: a deep reinforcement learning-based approach," *IEEE Access*, vol. 9, pp. 46600–46611, 2021.

[27] Y. Li, R. Liang, W. Wei, W. Wang, J. Zhou, and X. Li, "Temporal pyramid network with spatial-temporal attention for pedestrian trajectory prediction," *IEEE Transactions on Network Science and Engineering*, 2021.

[28] *ETSI, 2014a, EN 301 893, V1.7.2., Broadband Radio Access Networks (BRAN); 5 GHz High Performance RLAN; Harmonized EN Covering the Essential Requirements of Article 3.2 of the R & TTE Directive*, Sophia Antipolis: ETSI, 2014.

[29] *3GPP TR 36.814, V9.0.0, Further Advancements for E-UTRA Physical Layer Aspects*, 3GPP TR 36.814 V9.0.0 Release 9, 2010.

[30] M. Mehrnoush, V. Sathya, S. Roy, and M. Ghosh, "Analytical modeling of Wi-Fi and LTE-LAA coexistence: throughput and impact of energy detection threshold," *IEEE/ACM Transactions on Networking*, vol. 26, no. 4, pp. 1990–2003, 2018.

[31] S. Kubota and M. Morikura, *802.11 High-Speed Wireless LAN Textbook*, Impress, 2005.

[32] A. M. Voicu, L. Simić, J. P. de Vries, M. Petrova, and P. Mähönen, "Risk-informed interference assessment for shared spectrum bands: a Wi-Fi/LTE coexistence case study," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 3, pp. 505–519, 2017.