2022 IEEE Wireless Communications and Networking Conference (WCNC) | 978-1-6654-4266-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/WCNC51071.2022.9771647

Proactive Dynamic Spectrum Sharing for URLLC Services Under Uncertain Environment via Deep Reinforcement Learning

Xingyun Chen[†], Liang Shan[‡], Xuanheng Li[†], Na Deng[†], Nan Zhao[†]
[†]School of Information and Communication Engineering, Dalian University of Technology, Dalian, China
[‡]Big Data Reconnaissance and Analysis Center, Shenyang Municipal Public Security Bureau, Shenyang, China
Email: XingyunC8196@mail.dlut.edu.cn; 740176227@qq.com; {xhli, dengna, zhaonan}@dlut.edu.cn

***Abstract*—To support the emerging applications with the coming of the Beyond-5G (B5G) era, e.g., Ultra Reliable Low Latency Communications (URLLC) services, our telecommunications networks have witnessed a serious spectrum shortage problem. According to our spectrum measurement campaign, we note that many bands are actually extremely under-utilized, even for the operators' ones, e.g., LTE spectrums. Thus, it is expected to share the idle spectrums for the B5G services. Nevertheless, how to determine an effective sharing strategy is non-trivial. It is necessary to jointly consider the spectrum requirement of primary networks and the traffic demand of secondary networks when making the sharing decision, which, however, are both uncertain and hardly known precisely in advance. In this paper, taking the uncertain network environment into account, we propose a Proactive Dynamic Spectrum Sharing (PDSS) scheme to employ the under-utilized LTE spectrums for URLLC service provisioning. We take the long-term overall utility as the objective to achieve a trade-off between two networks to avoid the performance degradation of primary networks, while fulfilling as many URLLC services as possible with quality of service (QoS) guarantee. To deal with the environment uncertainty, we develop a model-free deep reinforcement learning (DRL) based solution, which can proactively capture the feature of the uncertain environment and achieve the best sharing decision autonomously. Based on the real spectrum data, simulation results have shown the effectiveness of the proposed DRL based PDSS scheme.**

***Index Terms*—dynamic spectrum sharing, URLLC services, uncertain network environment, deep reinforcement learning.**

I. Introduction

Recently, with massive devices accessing to the network, our telecommunications networks have witnessed a huge demand on radio resource. According to the forecast report by Cisco, the wireless data traffic in 2022 will increase to six times as compared with that in 2017 [1]. Such traffic explosion aggravates the spectrum shortage problem and poses a great challenge to mobile network operators (MNOs) to provide Beyond-5G (B5G) services. However, under the current static spectrum allocation policy, the spectrum efficiency is very low in many regions, even under 30% as reported by the Shared Spectrum Company [2]–[4]. To verify it, we have also implemented a spectrum measurement campaign on China Mobile LTE spectrums ranging from 2615 MHz to 2635 MHz on the campus of Dalian University of Technology as presented in Fig. 1. From the result, it can be observed that the spectrum is under-utilized in most of the time. Facing the contradiction between the high spectrum demand and the low spectrum utilization, dynamic spectrum sharing (DSS) has emerged and regarded as a promising solution to address the dilemma [5], [6]. Based on the DSS, MNOs can dynamically employ the under-utilized spectrums from the legacy networks, e.g., LTE networks, to support the emerging spectrum-consuming B5G services, e.g., the delay-sensitive Ultra Reliable Low Latency Communications (URLLC) ones.

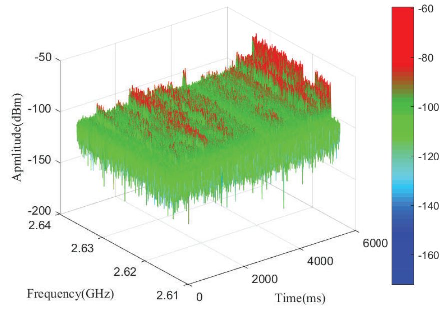


Fig. 1. Spectrum measurement on China mobile LTE spectrums.

With regard to the DSS between two networks, how to find an optimal spectrum sharing policy to effectively utilize the idle spectrum resources of primary network for the service provisioning in the secondary network is the key, which, however, is a very challenging problem, considering the uncertain spectrum environment of primary network and the uncertain traffic demand of secondary network. On the one hand, for the primary network who opens the spectrums for sharing, it needs to reserve sufficient spectrums to ensure its own network performance. Nevertheless, the requirement on spectrums in the future is usually dynamic and uncertain, making such decision difficult to determine [7], [8]. Reserving less spectrums will influence the throughput of its own network, while more reservation will lead to low sharing efficiency. Hence, it's necessary to predict the future spectrum requirement for the primary network and determine the amount of spectrums for

The work of X. Li was supported in part by the National Natural Science Foundation of China under Grant 61801080, in part by the Project funded by China Postdoctoral Science Foundation 2020M680038, in part by the Dalian Talents Innovation Support Program under Grant 2021RQ053, and in part by the Doctoral Research Initiation Fund of Liaoning Province under Grant 2019-BS-049. *Corresponding author: Xuanheng Li.*

reservation accordingly. On the other hand, for the secondary network who attempts to use the sharing spectrums for service provisioning, how much spectrum it needs depends on its traffic demands. Intuitively, if the demand of the secondary network could be known precisely, the primary network only needs to share the spectrums that can just fulfill the demands. Nevertheless, such decision is hard to make. It needs to consider the quality of service (QoS) guarantee issue to determine the needed spectrums, especially for the delay-sensitive URLLC services, and also the uncertainty on the traffic demand. As a result, for the DSS, it is necessary to proactively evaluate the requirement on both primary and secondary networks, and make an effective sharing decision by jointly considering their performance to achieve a trade-off between them.

Many recent works have studied DSS with the consideration of the environmental uncertainty. Most of them employed certain specific models by regarding the uncertain factors as priori knowledge. In [9], Deng *et al.* proposed a hybrid spectrum access scheme in OFDMA-based cognitive networks to improve both spectrum and energy efficiency under the assumption that the information on both channel state and traffic demand are available. Asaduzzaman *et al.* designed a stochastic optimization framework in [10] to facilitate DSS among different operators to increase the spectrum utilization, where the available bandwidth of primary networks is assumed to be known. In [11], Lertsinsrubtavee *et al.* developed an adaptive spectrum handoff scheme for DSS to increase the achievable rate, where the short-term future primary users' activities are assumed to be available and the expectation of the unavailable period is employed. In [12], Akshay *et al.* proposed a layered dynamic spectrum access approach for DSS to increase the throughput of the network, which relies on the average traffic demand of users. Koudouridis *et al.* designed a distributed radio resource management scheme for ultra-dense networks in [13] to minimize the spectrum sharing cost by assuming that the traffic demand follows a poisson point process. In [14], Li *et al.* developed a novel deep sensing paradigm for future DSS applications, where a Bernoulli filter algorithm is designed based on a stochastic discrete-state Markov chain modelling to evaluate the occupancy state of primary spectrum bands. Although the model-based approaches could obtain an optimal solution from different perspectives, the adopted models might not be accurate, or even not exist, and the explicit models are usually hardly obtainable in practice. Furthermore, most of them consider primary users and secondary users separately during the sharing decision making without considering the overall network performance.

In this paper, considering the uncertain environment on both spectrum requirement of primary network and traffic demand of secondary network, we propose a Proactive Dynamic Spectrum Sharing (PDSS) scheme. Without loss of generality, we consider the LTE network as the primary one and the Beyond-5G (B5G) network as the secondary one. We take the long-term overall utility of both networks as the objective to achieve a trade-off between spectrum reservation and spectrum sharing to avoid the performance degradation in the primary network, while fulfilling as many URLLC services in the secondary network as possible with QoS guarantee. To deal with the environment uncertainty issue, we develop a model-free solution based on deep reinforcement learning (DRL) to make the scheme proactively capture the feature of both the uncertain spectrum requirement of primary network and the uncertain traffic demand of secondary network. In this way, the proposed PDSS scheme can adapt to the dynamic environment and achieve the best spectrum sharing decision autonomously. Our main contributions are summarized as follows:

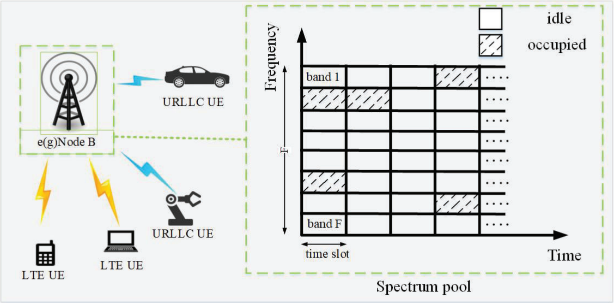


Fig. 2. Network model.

- To the best of our knowledge, this is the first work to design the spectrum sharing strategy for URLLC services by jointly considering the uncertain environment on both spectrum requirement of LTE networks and traffic demand of URLLC services. Taking the performance of both networks into account, we propose a PDSS scheme, which can make sharing decisions based on the predicted environment to meet the spectrum requirement of LTE networks, and meanwhile, fulfilling as many URLLC service demands as possible with QoS guarantee.
- Unlike most existing works relying on explicit models of uncertain environment, considering the fact that such models might be hardly obtainable in practice, we develop a model-free DRL based solution, which can achieve the best spectrum sharing strategy proactively and autonomously by interacting with the uncertain environment. Real spectrum data is collected by SAM-60BX to verify the effectiveness of the proposed PDSS scheme.

II. System Model And Problem Formulation

A. Network Model

We consider a downlink 5G non-stand alone (NSA) network with LTE user equipments (UEs) and N URLLC UEs as shown in Fig. 2 [15]. A spectrum pool is established and managed by the LTE eNodeB (eNB) to share some under-utilized spectrums with the 5G gNodeB (gNB) for URLLC service provisioning. The sharing decision is made periodically in a time-slotted way. As presented in Fig. 2, the scheduled resource includes both time and frequency domains, also known as resource block (RB) as defined in the LTE standard [16].

The time domain is partitioned into multiple transmission time intervals (TTIs) with the duration of L, called time slots. The available frequency bandwidth is divided into F sub-bands, each of which is B KHz. The RB is the minimum resource unit that can be allocated to serve an UE. For any RB f, it has two possible states at each time slot t, i.e., idle or occupied, expressed as $c_{f,t} = 0$ and $c_{f,t} = 1$, respectively. For each time slot t, we denote the arrival rate of data packets of URLLC UE n as $\lambda_{n,t}$, where each one has the size of s_n. Then, the data volume brought by URLLC UE n at time slot t can be described as $z_{n,t} = \lambda_{n,t} \cdot s_n$. Since each URLLC UE needs to accomplish the data transmission during each time slot within the duration L, the delay requirement actually corresponds to a rate requirement denoted as $\tilde{R}_n^t = \frac{z_{n,t}}{L}$. For each time slot, the eNB will proactively decide which RBs among the total F ones to share, and we use $\mathbf{a}_t = [a_{1,t}, a_{2,t}, \ldots, a_{F,t}], a_{f,t} \in \{0, 1\}$ to represent the sharing strategy made by it, where $a_{f,t} = 1$ indicates RB f is shared at time slot t, otherwise, $a_{f,t} = 0$. The goal for the sharing decision making is to find the idle RBs to fulfill the URLLC services satisfying their rate requirements.

B. Communications Model

For each time slot, we employ a widely-used simplified model to calculate the power propagation gain between the BS and URLLC UE n [17], which is expressed as

$$g_n = \beta d_n^{-\xi}, \tag{1}$$

where β is a parameter related to the antenna characteristics and the average channel attenuation. d_n is the distance between URLLLC UE n and the BS, and ξ is the path loss factor. The transmission power of the BS on each RB is assumed to be the same as P. Then, the achievable data transmission rate on each RB can be calculated as

$$R_n = B \log\left(1 + \frac{P g_n}{\sigma^2}\right), \tag{2}$$

where σ^2 represents the noise power.

C. Problem Formulation

For the PDSS scheme, the eNB will periodically make the spectrum sharing strategy every time slot. For the primary network, the eNB expects to reserve sufficient RBs for LTE UEs while maximizing the utilization of idle spectrum resources. For each time slot t, there are three possible cases after the sharing strategy is made: 1) Right Decision (RD). When the eNB judges the RB state successfully, we call it a RD case. That is for each RB f, if it is in the idle state, it is chosen for sharing, i.e., $c_{f,t} = 0$, $a_{f,t} = 1$, otherwise, $c_{f,t} = 1$, $a_{f,t} = 0$. 2) Wrong Decision (WD). When the eNB decides to share a RB in the occupied state, we call it a WD case. 3) Conservative Decision (CD). For an idle RB, if the eNB does not decide to share, we call it a CD case. It would not influence the performance of the primary network as in WD cases, but might affect the spectrum supply for the secondary network. Considering the three cases, we define the utility of the primary network at time slot t as

$$U_1(t) = A_{\mathrm{R}}(t) + \mu_0 A_{\mathrm{C}}(t) - \mu_2 A_{\mathrm{W}}(t), \tag{3}$$

where μ_0 and μ_2 are bias parameters, and A_{R}, A_{C}, A_{W} represents the number of RBs corresponding to the RD, CD and WD case, respectively.

For the secondary network, the gNB expects to make the best use of sharing RBs to serve as many URLLC UEs as possible with the QoS guarantee. Thus, we define the number of served URLLC UEs as its utility, which can be expressed as

$$\textbf{P1:} \quad U_2(t) = \max_{\{\mathbf{o}, \boldsymbol{\varphi}\}} \sum_{n \in \mathcal{N}} o_n(t) \tag{4a}$$

$$s.t. \sum_{n \in \mathcal{N}} \varphi_n(t) \leq \sum_{f=1}^{F} a_{f,t}, \tag{4b}$$

$$o_n(t) \tilde{R}_n^t \leq \varphi_n(t) R_n, \forall n, \tag{4c}$$

where $o_n(t)$ is a binary variable, representing whether to provide service to URLLC UE n. $\varphi_n(t)$ is a variable indicating the number of RBs assigned to URLLC UE n by the gNB at time slot t. (4b) means at the time slot t, the number of RBs allocated to URLLC UEs by the gNB cannot exceed the number of the sharing RBs. (4c) is the QoS guarantee constraint for each admitted URLLC UE.

Taking the long-term overall utility as the objective, the proposed PDSS scheme can be formulated as

$$\textbf{P2:} \quad \max_{\{\mathbf{a}\}} \frac{1}{T} \sum_{t=1}^{T} \tau U_1(t) + U_2(t) \tag{5a}$$

$$s.t. \; a_{f,t} \in \{0, 1\}, \forall t, f, \tag{5b}$$

where τ is a weighting factor to achieve a trade-off between the primary network and the secondary network.

III. A Proactive Spectrum Sharing Scheme Based On Deep Reinforcement Learning

Deep reinforcement learning combines deep learning (DL) with reinforcement learning (RL) to enable the agent to make right decisions under complex and uncertain environments. To achieve the best decision formulated as in P2, considering the uncertain environment on both spectrum requirement of primary network and traffic demand of secondary network, we will develop a DRL solution in this section.

A. Reinforcement Learning Framework

In each time slot t, the eNB will observe current state $\mathbf{s}_t$ and execute action $\mathbf{a}_t$ according to a certain policy $\pi(a\,|s)$. Then reward r_t will be calculated according to the next state that observed by the eNB in the next time slot. The RL method can make the eNB find the optimal policy $\pi^*(a\,|s)$ in the dynamic environment, aiming at maximizing the expected accumulated discounted reward, which can be described as

$$Q(\mathbf{s}_t, \mathbf{a}_t) = \mathbb{E}_\pi(\sum_{t=0}^{T} \gamma^t r(\mathbf{s}_t, \mathbf{a}_t)\,|\mathbf{s}_t, \mathbf{a}_t), \tag{6}$$

where γ is a discounted factor on the future reward, and $r(\cdot)$ is the reward function. Next, we define the state $\mathbf{s}_t$, the action $\mathbf{a}_t$, and the reward function $r(\mathbf{s}_t, \mathbf{a}_t)$ for the proposed PDSS scheme.

1) State: For any time slot t, the state is defined as

$$\mathbf{s}_t = \{\mathbf{c}_{t-1}, \mathbf{z}_{t-1}\}, \tag{7}$$

$\mathbf{c}_{t-1}$ and $\mathbf{z}_{t-1}$ represent the states of F RBs and the data volume generated by N URLLC UEs in the previous time slot $t-1$, respectively.

2) Action: At each time slot t, the action taken by the eNB is $\mathbf{a}_t = [a_{1,t}, a_{2,t}, \ldots, a_{F,t}]$, which is an F-dimensional vector as described in Section II. The sharing RBs can be allocated to serve the URLLC UEs by the gNB.

3) Reward: The reward function is used to evaluate whether an action $\mathbf{a}_t$ executed by the eNB under state $\mathbf{s}_t$ is good. As formulated in P2, we take the overall utility as the reward function expressed as

$$R(t) = \tau U_1(t) + U_2(t), \tag{8}$$

where $U_1(t)$ and $U_2(t)$ respectively represent the utilitiy of LTE network and 5G network. τ is a weighting factor to achieve a trade-off between two networks.

B. DRL Based Proactive Dynamic Spectrum Sharing

We first build two deep neural networks (DNNs) with the same structure, namely, main network Q with parameter θ and target network $\hat{Q}$ with parameter $\hat{\theta}$. The main network Q maps the current state to a string of action values, called estimated Q values, representing the estimation of the expected accumulated discounted rewards of all actions. The estimated Q value for each state-action pair is denoted as $Q(\mathbf{s}_t, \mathbf{a}_t; \theta)$. The output of target network $\hat{Q}$ is employed to calculate the target Q value denoted as

$$y_t = r_t + \gamma \arg\max_{\mathbf{a}} \hat{Q}(\mathbf{s}_{t+1}, \mathbf{a}; \hat{\theta}), \tag{9}$$

The loss function is defined as the mean square error between the target Q value and the estimated Q value, i.e.,

$$L(\theta) = \mathbb{E}[(y_t - Q(\mathbf{s}_t, \mathbf{a}_t; \theta))^2]. \tag{10}$$

The whole algorithm process has been presented in **Algorithm 1**. At each time slot t, the eNB will get the current state $\mathbf{s}_t$, including both the state information of RBs $\mathbf{c}_{t-1}$ and the data volume of URLLC UEs $\mathbf{z}_{t-1}$, which will be fed into the main network Q as the input. The output is a 2^F dimensional vector, indicating the Q values for all the 2^F potential actions. Then the eNB will take an action $\mathbf{a}_t$ to determine which RBs to share based on ε-greedy policy, i.e., take the action with the maximum estimated Q value with possibility $1-\varepsilon$, and take a random action with possibility ε. At the beginning, a large value of ε will be set to explore the environment to avoid falling into a local optimum. As the algorithm goes on, ε will gradually decrease to make it exploit the best action. After taking an action, the network transfers to the next state $\mathbf{s}_{t+1}$ and the eNB will calculate the reward r_t according to (8). Then the experience tuple $\{\mathbf{s}_t, \mathbf{a}_t, \mathbf{s}_{t+1}, r_t\}$ will be stored in the memory unit for the DNN training.

As the eNB repeats the process above, the number of experience tuples in the memory unit keeps growing, and the oldest one will be replaced by the new one once the memory unit is overflowing. When the memory unit is full, the DNN training will be started. During the training process, a mini-batch of experience tuples will be sampled randomly from the memory unit to train the DNN. To be specific, for each tuple, $\mathbf{s}_t$ will be fed into the main network Q to calculate the estimated Q value corresponding to action $\mathbf{a}_t$, i.e., $Q(\mathbf{s}_t, \mathbf{a}_t; \theta)$. $\mathbf{s}_{t+1}$ will be fed into the target network $\hat{Q}$ to calculate the target Q value together with the reward r_t according to (9). Then the loss function can be obtained by (10), and the parameters θ of the main network will be updated based on the gradient descent method, which is expressed as

Algorithm 1 Proactive Spectrum Sharing Based on DRL

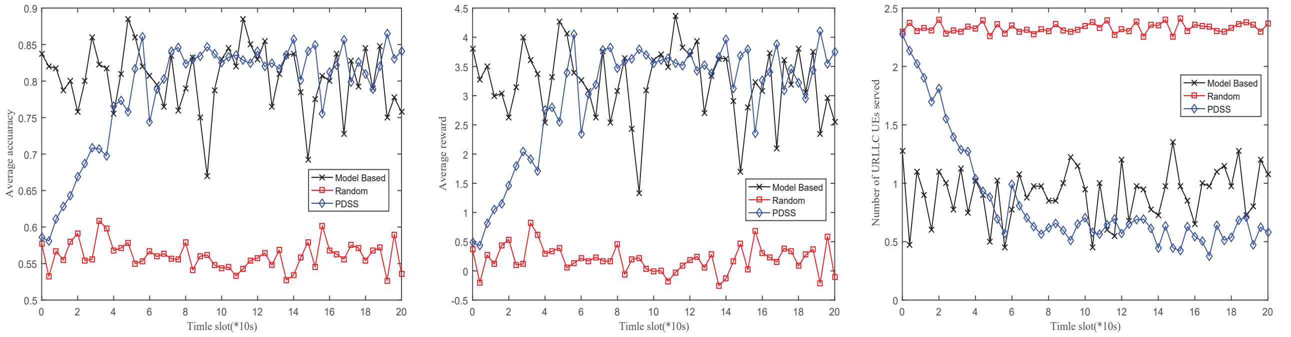
```
Initialize: main network $Q$ and target network $\hat{Q}$ with random parameter $\theta$ and $\hat{\theta}$, memory size $M$, mini-batch size $\rho$, learning rate $\alpha$, discount rate $\gamma$, target network update frequency $I$, explore rate $\varepsilon$, starting state $\mathbf{s}_0$, starting action $\mathbf{a}_0$, Train=true.
for $t$ = 1,2,...do
  If Train do
      Observe current state $\mathbf{s}_t = \{\mathbf{c}_{t-1}, \mathbf{z}_{t-1}\}$
      Make a spectrum sharing strategy $\mathbf{a}_t$ under current state $\mathbf{s}_t$ by $\varepsilon$-greedy policy, transfer to next state $\mathbf{s}_{t+1}$ and calculate $r_t$
      The eNB stores $\{\mathbf{s}_t, \mathbf{a}_t, \mathbf{s}_{t+1}, r_t\}$ in the memory unit.
        if $t > M$ do
            Replace the oldest experience tuple with the new one.
            Sample randomly a mini-batch tuples from memory unit to train the main network $Q$.
            Construct the loss function $L(\theta)$ with the evaluated $Q$ value and the target $Q$ value by formula (10).
            Perform a gradient descent step on the loss function by (11), and update parameter $\theta$.
            if $(t-M) \bmod I = 0$
              Copy the main network parameters to the target network as $\theta \to \hat{\theta}$.
            end if
        end if
  end If
end for
Algorithm end
```

$$\theta = \theta - \alpha \cdot \frac{dL(\theta)}{d\theta}, \tag{11}$$

where α is the learning rate. The parameter $\hat{\theta}$ of target network $\hat{Q}$ is copied from the parameter θ of the main network Q every I time slots.

IV. Simulation Results and Discussions

We consider a downlink 5G NSA network, where the eNB and the gNB are deployed in the same signal tower and located in the center of the cell coverage area with the radius of 500 m. 5 URLLC UEs are distributed randomly within the cell coverage area. We assume that the number of data packets generated by each URLLC UE follows the Poisson Point Process (PPP) with the arrival rate of λ (packets/sec) [18], which is randomly chosen from 20 to 100. Each URLLC packet size is fixed as 32 bytes [19]. The path loss exponent ξ is set to 3 and the antenna parameter β is set to 4 for all the communication links, and the transmission power of the BS on each RB is 20 dBm. As for the spectrum environment, we implement a spectrum measurement campaign on the campus of Dalian University of Technology. Specifically, it is carried



(a) The decision accuracy versus time slots.

(b) The average reward versus time slots.

(c) The number of served URLLC UEs versus time slots.

Fig. 3. The performance metrics of three policies with $\tau = 4$.

out on the China Mobile LTE frequency ranging from 2615 to 2635 MHz. The bandwidth and the TTI of each RB is 180 KHz and 10 ms, respectively, and the guard bandwidth is 2 MHz. Then, we can obtain 100 signal power samples of 100 RBs accordingly at each time slot. We implement the measurement for 200 seconds, and thus get the samples for the 100 RBs in 20000 time slots. By setting a power threshold $P_{th} = -110\text{dBm}$, the samples are transformed to 1 or 0, representing the occupied or idle state. We randomly choose consecutive 5 RBs among 100 RBs for the simulation. We set the additive white Gaussian noise power to 10^{-10}W. The hyper parameters of DNNs are shown in Table I.

TABLE I
HYPER-PARAMETERS OF DNNS

Parameters	Values
Number of hidden layers	3
Number of neurons in hidden layers	256,256,128
Activation function	ReLu
Memory unit size M	2000
Mini-batch size ρ	512
Discount rate γ	0.8
Learning rate α	0.03
Update frequency of the target network I	100

We compare the proposed PDSS scheme with two other policies, i,e,. random policy and model based policy. For the random policy, the BS will randomly take actions at each time slot. As for the model based policy, we first discretize the state by setting two levels for λ and calculate the state transition probability according to the collected RB state samples. For each time slot, the BS will take an action to maximize the expectation of the immediate reward based on the state transition probability matrix. However, the actual state transition probability matrix is hardly obtained and the model based policy is used to evaluate the performance of the PDSS scheme as a baseline. We consider three performance metrics: 1) average reward. We observe the reward obtained every time slot and take average in 400 time slots as the result. This metric presents the overall utility of the scheme 2) decision accuracy (DA). We define the DA as

$$DA{=}\frac{\#RD + \omega \times \#CD}{\#RB}, \tag{12}$$

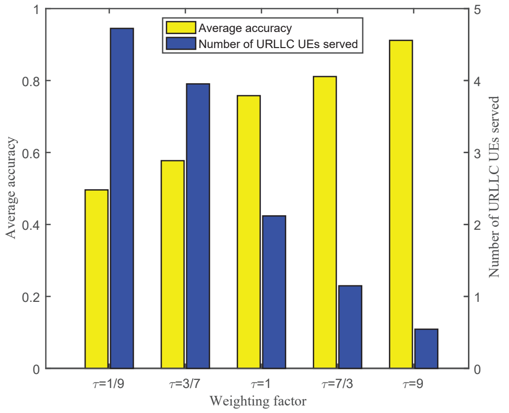


Fig. 4. The average accuracy and the number of served URLLC UEs under different weighting factors.

where $\#RD$ and $\#CD$ respectively represent the number of the right decision and the conservative decision, and $\#RB$ is the number of available RBs per time slot. ω is the measurement weight of such conservative choices, which is set as 0.5 in the simulation. DA presents the utility of the LTE network and we calculate the average of 400 consecutive samples for it. 3) number of URLLC UEs served. This metric can be obtained from (4a) every time slot and we average it every 400 time slots. This metric is used to evaluate utility of the 5G network.

In Fig 3, we can see that the DA and the average reward under the proposed PDSS scheme gradually increase and finally converge at a same level as the model based policy, which significantly outperforms the random policy. As a result, we can conclude that the PDSS scheme can well learn the feature of dynamic spectrum environment and proactively find idle RBs to share without relying on specific models. In addition, the proposed PDSS scheme can make the optimal spectrum sharing strategy to maximize the overall utility of both LTE and 5G networks.

Second, we adjust the weight factor τ to evaluate the trade-off on two networks by the PDSS scheme. Fig. 4 shows that the DA of the PDSS scheme gradually increases as the increasing of τ, but the number of served URLLC UEs is downward. It means that a larger value of τ will lead to a higher performance protection for the primary network with less spectrums sharing to the secondary network. Therefore, by adjusting the weighting factor, it can achieve a trade-off between primary and secondary networks from the perspective on the overall utility.

V. Conclusions

In this paper, we propose a proactive dynamic spectrum sharing scheme based on deep reinforcement learning algorithm to share LTE idle spectrum resources for providing URLLC services. It can enable the eNB to proactively make an optimal spectrum sharing strategy to maximize the overall network utility of both primary and secondary networks based on the historical observations. Simulation results on real spectrum data have shown that the proposed PDSS scheme can effectively learn the feature of environment uncertainty and make an optimal spectrum sharing strategy with maximum overall network utility.

References

[1] C. V. N., "Cisco annual internet report (2018–2023) white paper," *White Paper*, macrch, 2020.

[2] M. H. Islam, C. L. Koh, S. W. Oh, X. Qing, and W. Toh, "Spectrum survey in singapore: Occupancy measurements and analyses," in *Proc. Int. Conf. Cognitive Radio Oriented Wirel. Netw. Commun., CrownCom'08*, Singapore, 2008.

[3] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal," in *Proc. TAPAS'06*, Boston, MA, United States, 2006.

[4] S. Yin, D. Chen, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: A large-scale spectrum measurement study," *IEEE Trans. Mob. Comput.*, vol. 11, no. 6, pp. 1033–1046, 2012.

[5] G. Barb, M. Otesteanu, and M. Roman, "Dynamic spectrum sharing for LTE-NR downlink MIMO systems," in *Proc. ISETC'20*, pp. 1–4, Virtual, Timisoara, Romania, 2020.

[6] A. Roessler, "Impact of spectrum sharing on 4G and 5G standards a review of how coexistance and spectrum sharing is shaping 3GPP standards," in *Proc. IEEE EMCSI'17*, pp. 704–707, Washington, DC, United States, 2017.

[7] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, "Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1389–1401, 2019.

[8] Y. Liu, X. Wang, G. Boudreauand A. B. Sediq, and H. Abou-zeid, "Deep learning based hotspot prediction and beam management for adaptive virtual small cell in 5G networks," *IEEE Trans. Emerg. Topics Comput.*, vol. 4, no. 1, pp. 83–94, 2020.

[9] Q. Deng, Z. Li, J. Chen, F. Zeng, H. Wang, L. Zhou, and Y. Choi, "Dynamic spectrum sharing for hybrid access in OFDMA-based cognitive femtocell networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10830–10840, 2018.

[10] M. Asaduzzaman, R. Abozariba, and M. Patwary, "Dynamic spectrum sharing optimization and post-optimization analysis with multiple operators in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1589–1603, 2017.

[11] A. Lertsinsrubtavee and N. Malouch, "Hybrid spectrum sharing through adaptive spectrum handoff and selection," *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2781–2793, 2016.

[12] A. Kumar, A. Sengupta, R. Tandon, and T. C. Clancy, "Dynamic resource allocation for cooperative spectrum sharing in LTE networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5232–5245, 2014.

[13] G. P. Koudouridis and P. Soldati, "Joint network density and spectrum sharing in multi-operator collocated ultra-dense networks," in *Proc. Int. Conf. MOCAST'18*, pp. 1–4, Thessaloniki, Greece, 2018.

[14] B. Li, S. Li, A. Nallanathan, Y. Nan, C. Zhao, and Z. Zhou, "Deep sensing for next-generation dynamic spectrum sharing: More than detecting the occupancy state of primary spectrum," *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2442–2457, 2015.

[15] X. Zhao, J. Chen, P. Li, Z. Li, C. Hu, and W. Xie, "5G NSA radio access network sharing for mobile operators: Design, realization and field trial," in *Proc. ICCT'20*, pp. 454–461, Nanning, China, 2020.

[16] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation," *TS 36.211 (V10.3.0)*, 2011.

[17] Y. T. Hou, Y. Shi, and H. D. Sherali, "Spectrum sharing for multi-hop networking with cognitive radios," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 146–155, 2008.

[18] M. Series, "Minimum requirements related to technical performance for IMT-2020 radio interface(s)," *Report*, pp. 2410–0, 2017.

[19] P. Korrai, E. Lagunas, S. K. Sharma, S. Chatzinotas, A. Bandi, and B. Ottersten, "A RAN resource slicing mechanism for multiplexing of eMBB and URLLC services in OFDMA based 5G wireless networks," *IEEE Access*, vol. 8, pp. 45674–45688, 2020.