Wireless Personal Communications
https://doi.org/10.1007/s11277-021-08432-0

Cellular Licensed Band Sharing Technology Among Mobile Operators: A Reinforcement Learning Perspective

Minsu Shin[1] · Danish Mehmood Mughal[1] · Seungil Park[2] · Sang-Hyo Kim[1] · Min Young Chung[1]

Accepted: 22 March 2021

Abstract

Next-generation wireless networks will need to support of very high data rates and low–latency communications, which will require a new wireless radio technology paradigm. The growing number of mobile users is causing spectrum scarcity; and hence, an efficient spectrum utilization method is required. Conventional scheduling-based resource allocation scheme in wireless networks under limited resources is a challenging due to the complex network situations, dynamic network environment, and diverse needs for future networks. To overcome resource scarcity in mobile networks, spectrum sharing among multiple operators may be an efficient solution. Traditional methods of dynamic spectrum sharing are model-dependent, and they are not robust to the changing wireless environments. To enable low-latency communications for complex future wireless networks, efficient machine learning algorithms can be used across the wireless network infrastructure. Integrating machine learning for resource allocation can leverage intelligent and efficient mechanisms for dynamic wireless networks. To efficiently and intelligently utilize the scarce resources of dynamic networks, this paper proposes an efficient machine learning-based spectrum sharing method among multiple mobile network operators (MNOs). A mobile network operator uses the idle slots of the another operator and transmits the information efficiently. Using the neural network model, each MNO learns the spectrum utilization of other MNOs and selects the idle slots of other MNOs. Simulation results have been computed and compared with the conventional scheme where resources are not shared. These simulation results show that the proposed neural network model can efficiently learn the network quickly, and spectrum sharing can lead to improved network performance in terms of the delay, user-perceived throughput, resource usage, packet drop, and sum throughput of the network.

Keywords Spectrum sharing · Neural network · Deep learning · Resource selection · Mobile network operators · Reinforcement learning

✉ Min Young Chung
mychung@skku.edu

Extended author information available on the last page of the article

Published online: 02 April 2021

1 Introduction

Recently, wireless communication networks have observed a huge demand from new wireless technologies, such as virtual reality, augmented reality, holographic communications, etc. [1]. These technologies will be a key feature of 6G communications. To enable these technologies, high-speed networks with minimum latency and high data rates are highly desirable. In conventional mobile networks, each network operator is provided with a specific band and the operator uses that band to provide the necessary services to its users. However, the wireless spectrum is a scarce resource and the available spectrum bands will not be enough to cope with the new technologies or the capacity required to enable those technologies [2]. Due to the static resource allocation in conventional wireless networks, the wireless band remains underutilized. This reduces effective resource utilization; and as a result, lots of wireless resources are wasted. Thus, efficient spectrum management schemes are required in order to address the underutilization of the spectrum resources and to meet the requirements of increasing demands of spectrum for new wireless communication services. Since the licensed spectrum is underutilized, its optimal utilization can be enabled by letting other network operators to use the idle resources. Thus, an efficient spectrum sharing method between multiple network operators can enhance spectrum utilization and reduce the delay of the network.

Recently, shared spectrum access has become an important component for next-generation communication systems. Shared spectrum access can provide opportunities for dynamic spectrum assignment and improve the spectrum utilization [3, 4]. Using a cognitive radio networks is a type of spectrum sharing method where unlicensed users can share and access the licensed spectrum of the primary users. In [5], the authors have proposed a dynamic spectrum access method for cognitive radio networks where secondary devices, working in unlicensed bands, can occupy the licensed band of primary users opportunistically. With recent developments in the applications of machine learning to wireless networks, methods for intelligent resource sharing in wireless networks (specifically in cognitive radio networks) have been studied, as in [6] and [7]. In [8], a cooperative spectrum sensing scheme based on machine learning techniques for cognitive radio networks was established. Additionally. the authors in [9] have proposed learning algorithms for efficient spectrum sensing in cognitive radio networks. In [10] and [11], the authors have analyzed the spectrum occupancy and resource allocation in cognitive radio networks is analyzed via machine learning tools.

With the recent advancements in technology and fast computation capabilities, neural networks (NNs) have been applied for wireless communications. NNs have benefits over the other schemes because an NN model can be well-trained in fast changing wireless networks. Some efforts have been made regarding the application of NNs for efficient resource sharing specifically in cognitive radio networks. In [12], the authors proposed a channel sensing mechanism using an artificial NN for cognitive radio networks to sense the primary radio signals. In [13] and [14], the authors have briefly surveyed NN models for their application in wireless networks. A joint resource allocation and network access problem was considered in [15] for maximizing the normalized throughput.

A wireless network is dynamic, and getting the data for such an unpredictable network is very challenging. Reinforcement learning is a type of machine learning algorithm where a dataset is not required. In reinforcement learning, an agent learns which action needs to be taken in an environment in order to maximize the reward [16]. There is a continuous interaction between the agent and the environment, and the reward is calculated based on action taken by the agent. Reinforcement learning techniques can be used by the network entities for decision making in dynamically changing network conditions. By continuously learning the network dynamics, efficient resource management, mobility management, and networking can be ensured [17]. In [18], the authors proposed a reinforcement learning-based mode selection and sub-channel allocation in D2D networks. The authors in [19] utilized reinforcement learning to model a resource management algorithm for both licensed and unlicensed frequency bands in LTE-U systems. Q–learning is a famous algorithm of reinforcement learning, which seeks the best result in a given environment. The authors in [20], proposed a distributed Q–learning algorithm for spectrum management in ultra-dense networks. Additionally in [21], the authors studied a spectrum access problem to manage the interference in a heterogeneous network.

To maximize the normalized throughput of the unlicensed band while guaranteeing the quality of service (QoS) of users, the authors considered the joint resource allocation and network access problem using a learning mechanism [15]. The authors in [22] proposed a reinforcement learning mechanism for cognitive radio networks. In [23] and [24], the authors proposed a model for resource sharing as a multi-agent reinforcement learning problem for the spectrum sharing problem in vehicular networks.

In [25], the authors investigated a distributed Q-learning mechanism for channel selection in an unlicensed 5 GHz band in LTE-U systems. Based on prior experience, the proposed mechanism learns to select the appropriate channel for downlink transmission.

With recent improvements in computational power, deep learning (DL) has been studied in various wireless communication systems [14]. Deep learning can provide intelligent services for wireless networks, especially for resource management and optimizing network operations (e.g., power control, interference control, throughput etc.), in real time [26, 27].

In [28], a neural network is applied for the bandwidth allocation decision-making process. The proposed neural network model efficiently reduces the network latency. However, the implication of machine learning in wireless networks is still new, and a lot of work still has to be done in this field. To increase the spectrum efficiency, efficient machine learning–based algorithms with low complexity are needed. Moreover, most of the work related to spectrum sharing has been done in the field of cognitive radio networks where one primary and one secondary network are considered. However, not much research work has been done for multi-operator networks where multiple operators are sharing the spectrum bands with each other and using the idle slots of other operators. Conventionally, all operators have their own bands, and they manage their users with their own resources. Intelligent spectrum sharing among multiple MNOs can lead to more effective utilization of network resources and improve the network performance.

However, it is possible that the resources of one mobile network operator are underutilized while another network operator has lots of users. It can be possible in different scenarios like in an event, where users of one mobile network has lots of users in one cell, while other network operator has less users. In such a case, it is possible that the resources are not sufficient to manage lots of users, thus, users experience delays and a low quality of experience. In such a scenario, a possible solution would be to allow the network operator to use the resources of the other operators, which have less users. This can reduce the delay and improve the user's quality of experience. However, in dynamic wireless networks, an efficient and robust mechanism is required for effective sharing of resources. Motivated by this, we have proposed a deep reinforcement learning based spectrum sharing technique in a multi–operator network, where multiple network operators share the band to provide services to users. Using the deep Q–network model, an efficient and robust mechanism for resource sharing is proposed. The results have been plotted in terms of the delay, throughput, and user–perceived throughput (UPT). Simulation results show the network performance improvement as compared to the conventional resource selection mechanism where resources are not shared with other network operators.

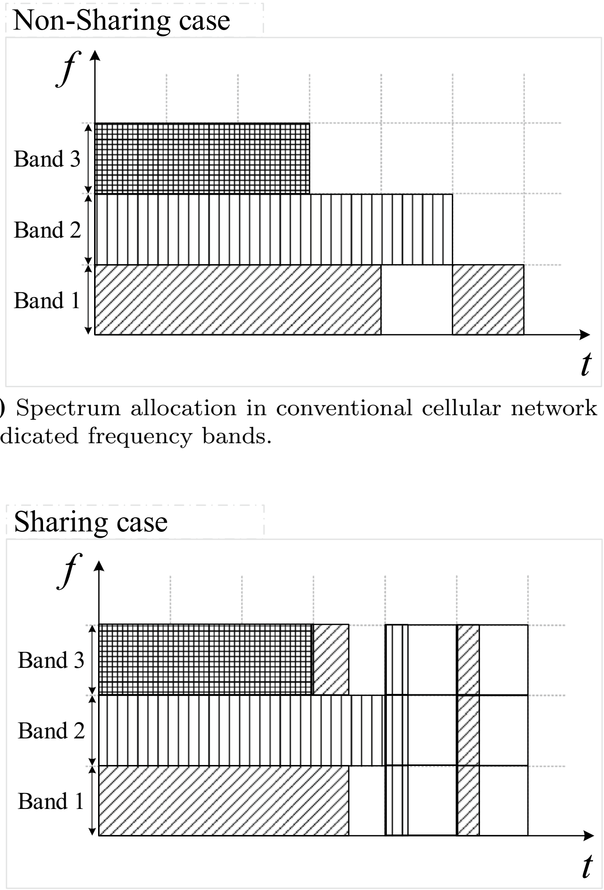
The rest of the paper is organized as follows. Section 2 briefly discusses the proposed scheme and explain the preliminaries required to understand the proposed model. In Section 3, we evaluate the performance and explain the simulation environment, performance metrics, and simulation results. Lastly, we conclude the paper in Sect. 4.

2 Proposed Scheme

In this section we will explain the proposed scheme and provide information about the required preliminaries. We consider a multi-operator network environment consisting of three MNOs that offer wireless network services to their users. We assume that all three MNOs are sharing the same coverage area; in other words, they are co-located. It is assumed that each MNO has a given wireless band through which it can communicate to its users. Considering the downlink transmission, data arrives at each MNO with arrival rate λ. For simplicity, we consider a special case where all users of an MNO have the same λ. Moreover, all MNOs have a queue for data storage where newly arrived data packets are stored and later processed. All MNOs share information about the amount of data packets with each other, and they also share the information related to number of serving UEs and their arrival rates. Using the received information from other MNOs, an MNO estimates the resource band utilization of other MNOs and utilizes their bands during their idle periods.

We explain the proposed scheme using an example as shown in Fig. 1. When a data packet arrives at MNO 1 it stays in the queue if MNO 1 is already transmitting data to a user; otherwise it is processed based on the available spectrum band. Considering the limited resources, if MNO 1 is serving lots of users, then users may experience considerable delay. In this situation, even if other MNOs (MNO 2 and MNO 3) are not transmitting data (no packet in their queues), MNO 1 cannot use those idle resources to transmit the packets and those resources are wasted as shown in Fig. 1a. One possible solution is to increase the resources but this is an expensive solution. To efficiently utilize the resources and reduce the delay of its users, spectrum sharing environment is proposed where multiple operators share their resource bands with each other, shown in Fig. 1b. Thus, the proposed spectrum sharing environment allows MNO 1 to try to use the idle resources of other MNOs and transmit its data packets . Considering the time slotted operation of the network, an MNO can occupy one or more frequency bands for one or multiple time slots.

The proposed scheme can effectively reduce the delay and improve network performance. However, since MNOs do not share information about resource selection before transmission, if there are multiple MNO transmitters and multiple MNOs select the same resource at a time, a collision can occur. This can lead to a reduced network performance; hence, a proper resource selection mechanism for a spectrum sharing environment is required. To overcome the collision problem and efficiently utilize the network resources, we used a reinforcement learning based mechanism. The proposed reinforcement learning algorithm is based on a deep NN, and it can efficiently learn the networks and select the idle resources. Reinforcement learning has benefits over other learning schemes because it can tackle changing environments quite efficiently.



(a) Spectrum allocation in conventional cellular network with dedicated frequency bands.

(b) Spectrum allocation in cellular network with shared frequency bands.

Fig. 1 Resource allocation procedure

At the beginning of the network simulation, each MNO exchanges the information about the amount of data packets in its queue to other MNOs through the Backhaul link. Based on the received information about the number of data packets in the queue of other MNOs, each MNO decides when to use the resources of other MNOs. Depending on the action taken by the MNO (i.e. transmitting the data packet using the idle resources of other MNOs), a reward function is calculated. Based on the reward function, the deep NN of the MNO learns the wireless network environment and once deep NN is trained, efficient resource utilization can be achieved. Each MNO has a deep NN and it selects idle wireless bands based on output of the designed deep NN. The deep NN is trained by a modified Q–learning based deep neural network using a specially designed reward function. This reward function is implemented at each node of the output layer. During the training phase, the reward is calculated and the weights of the NN are updated based on the reward function. The backpropagation algorithm is used to update the weights. If collision occur, a negative reward is calculated; thus the NN readjust its weights, and tries again. After some timeslots NN can learn the environment, enabling efficient spectrum sharing among multiple MNOs. In later sub-sections, we will explain the Q–learning algorithm and the proposed learning mechanism.

2.1 Q–Learning Model

There are several types of reinforcement learning such as Markov decision process and Q–learning etc. In Q–learning a sequence of actions are taken that result in the maximum reward. The obtained reward value is also called the Q–value. In a given states s, an action a is taken and the Q–value is calculated as

$$Q'(s, a, \theta_k) = r(s_t, a_k) + \gamma \max_{a' \in A} Q(s'_t, a'_k; \theta_k). \tag{1}$$

In Eq. 1, $r(s_t, a_k)$ is the immediate reward based on the action taken at the k^{th} iteration and at tth state. $Q(s', a')$ is the set of output values referred to as the Q–values. Maximum value is chosen from these Q–values with parameter θ_k at k^{th} iteration where θ_k are some parameters that define the Q-values. However, the Q–learning algorithm requires more memory and time especially when there are many actions per state, and hence it is less effective in a complex scenario. To overcome the limitations of Q–learning algorithm, neural fitted Q–learning (NFQ) is introduced. In NFQ, a state is provided as an input to the Q–network, and different outputs are computed for each possible action. Thus, the computation of $max_{a' \in A} Q(s', a'; \theta_k)$ can be obtained in a single forward pass in the neural network for a given state s'.

In NFQ, the values of θ_k are updated by stochastic gradient descent (or a variant) in $Q(s, a; \theta_k)$. The loss value is calculated using the the equation as given by [29]:

$$L_{DQN} = (Q(s, a; \theta_k) - Q'(s, a; \theta_k))^2. \tag{2}$$

Mnih et al. introduced the deep Q–network (DQN) algorithm which shows strong performance by directly learning from the images in ATARI games [30]. In their scheme, after every $C \in \mathbb{N}$ interactions, θ_k^- is updated in $Q(s'_t, a_k, \theta_k^-)$. Because of the fixed value C, the proposed scheme results in preventing the instability to propagate in the network quickly and reduces the risk of divergence.

For an online setup, the $N_{replay} \in \mathcal{N}$ values are stored in a memory called the replay memory for N_{replay} time steps [3]. The set of tuples $< s, a, r, s' >$ is selected randomly from the replay memory. This set of information is called a mini–batch. When updates are made using mini–batch, each mini–batch has less variance compare to a single tuple update. This allows for efficient parallelization of the algorithm.

2.2 Proposed Learning Mechanism

Applying Q-learning to the proposed multi-operator coexistence scenario requires some consideration. First, the proposed scenario has a complex priority system with different priorities among shared users depending on the band. In addition, it is necessary to define a method to determine whether to use different bands simultaneously. In other words, the MNO designs a rewards function that allows it to learn shared spectrum bands fairly and avoid interference. Lastly, in the proposed distributed scenario, a method of determining the structure of the NN and determining the compensation function is required to learn the frequency sharing method of the eNB of MNO.

Figure 2 shows the neural network model for the proposed scheme. The input layer consists of 12 nodes, where 4-nodes for each MNO. Among these four nodes, the first node shows the number of packets in the queue of that MNO. The remaining three nodes shows the percentage of resource usage in the first, second, and third bands respectively. In the NN model, there are two hidden layers, each consisting of 50 nodes. The output layer consists of three nodes, each node gives some real valued output indicating which band has to be used by the MNO.

More specifically, deep Q-network (DQN) algorithm is used. In DQN, a neural network model is used to approximate the Q–value function. Considering the fully connected neural network model, all the nodes are connected with the nodes of the hidden layer. The inputs are given at the input layer and information is propagated through the hidden layers. This information propagates to the output layer. The output layer selects sub–channels and the loss is estimated at the output layer. The loss function is then used to update the weight matrix using back propagation process. An optimizer function is used to minimize the loss. ReLU activation is used in the output layer and the hidden layers. Each node at the output layer consists of a reward function which is used to select a the resource band.

The input layer of the deep NN model is divided into three group; each group representing an MNO. As its input, an MNO considers two things. The first is the amount of traffic that needs to be transmitted from the eNB to the users. That is the determination of whether or not the transmission of the base station is necessary; is used as input data by looking at the number of packets in the eNB's queue, and this information is provided to the NN using first input of each group of the input layer. At this time, MNOs receive requests for ftp files from UEs at a Poisson arrival rate of λ. The MNOs estimate channel state information of each band. The distance of UEs from the eNB are calculated by the eNB. The affects of path loss, shadowing, and fading are added to estimate the signal–to–inference (SIR) information. Each MNO uses this information with the results of channel selection to determine how many resource will be used in each resource band. The information about the amount of resources used from each band is used to determine whether or not to use

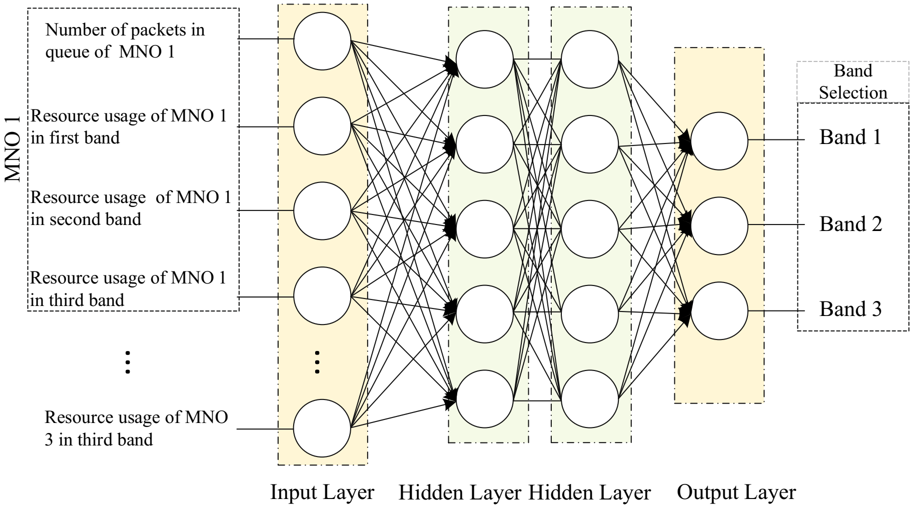

Number of packets in queue of MNO 1
Resource usage of MNO 1 in first band
Resource usage of MNO 1 in second band
Resource usage of MNO 1 in third band
MNO 1
Resource usage of MNO 3 in third band
Band Selection
Band 1
Band 2
Band 3
Input Layer
Hidden Layer
Hidden Layer
Output Layer

Fig. 2 Deep neural network design for the proposed scheme with 12 nodes in input layer, multiple hidden layers, and 3 on output layer

the frequency of the previous time and the frequency usage change of other operators. The resource usage amount for each band is input to the NN along with the number of packets in the base station queue. Therefore, in the frequency sharing scenario of three base stations, one base station requires four inputs.

The output of the neural network determines the operation. In frequency sharing, the operation is the same as the selection of frequency used. Unlike the DQN algorithm, the output of all output nodes determines the operation for selecting multiple bands. Each output node is designed to select each band, and whether a band is selected is determined according to the positive or negative division of the output value. For example, in order for the eNB to select bands 1 and 3 among three bands at t, the output of the neural network must have positive, negative, and positive values in that order.

Due to this unique output structure, the proposed model is different from DQN algorithm. In the DQN algorithm, the operation with the highest Q–value is selected based on one–hot coding, and only the corresponding output is updated according to the compensation value using ε–greedy technique. One operation of ε–greedy technique requires one reward value, the reflection of future rewards, and use of the target network for system stabilization. Our proposed technique works without these auxiliary actions because it uses all outputs and reflects different rewards for all outputs simultaneously. The frequency selection behavior and neural network learning structure considered by the proposed technique are shown in Fig. 3.

For the compensation value applied to each output value, a positive compensation value ($r = 1$) or a negative compensation value ($r = -1$) is used. Positive compensation induces the use of the corresponding band by increasing the output value, while negative compensation induces the use of the band by reducing the output value. The compensation

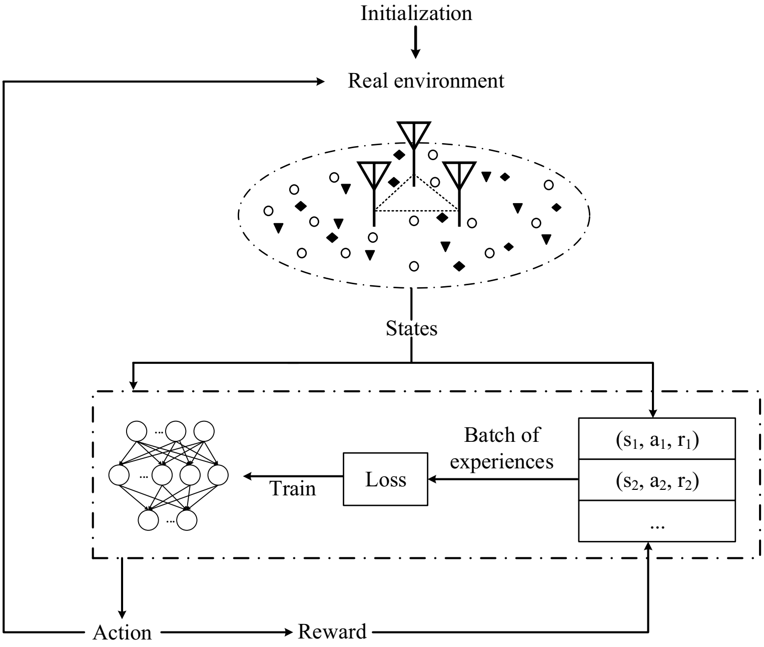


Fig. 3 Learning procedure for the proposed network

value is determined according to the input of each base station and the result of using the frequency band. Compensation value determination is based on several criteria. Each band reflects the characteristics of different usage priorities for each operator. Since one operator can access multiple bands at the same time, characteristics to avoid performance degradation due to collisions in multiple bands are considered when selecting an unused band of another operator. Moreover, in order to a prevent under utilization of resource bands (i.e. the resource band remains unused), it is also considered that if a MNO has data to transmit and a resource band is idle, it must use the idle resource band. Based on these criteria, positive or negative compensation is applied to each output. The reward is calculated and by using gradient decent method and weights are recalculated and backpropageted to the network. The proposed learning algorithm is summarized in Algorithm 1.

Algorithm 1 Proposed Learning Mechanism for Wireless Band Selection

```
Initialize replay memory $D$ to capacity $N$
Initialize action-value function $Q$ with random weights $\theta$
Initialize sequence $s_1 = \{x_1\}$ and preprocessed sequence $\phi_1 = \phi(s_1)$
for $t = 1, T$ do
    select all actions $a_{t_k}$
    Execute action $a_{t_k}$ in the simulation environment and observe reward $r_{t_k}$
    Store transition $(s, r_{t_k})$ in $D$
    Sample random minibatch of transitions $(s_j, a_j, r_j\ )$ from $D$
    Set $y_j = r_j$
    Perform a gradient decent step on $(y_j - Q(s_j, a_j; \theta))^2$ with respect to the network
    performance parameters $\theta$
end for
```

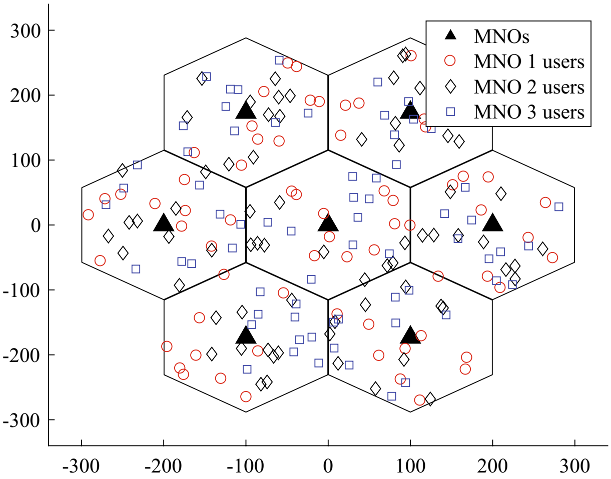
3 Performance Evaluation

This section presents the performance evaluation of the proposed scheme and compare it with the conventional resource scheduling mechanism. To evaluate the performance of the proposed scheme, simulations have been performed on a computer with a Windows–10 based operating system(core i7), 8 GB of RAM, and an RTX–2070 super GPU. Simulations have been carried in Python, and Tensorflow and Numpy libraries are used. The performance of the proposed scheme is evaluated in terms of the average network delay, resource usage, UPT, and network sum–throughput.

3.1 Simulation Environment

For simulation purposes, a seven cell network is used, where users are randomly deployed as shown in Fig. 4. Each cell consists of three MNOs, which are sharing their bands with each other. Moreover, the inter-site distance of each cell is 200 m. It is assumed that all of the MNOs are co-located at the center of each cell, as shown in Fig. 4. In a cell, each MNO has 10 users and the communication frequency is 4 GHz. For each MNO, a bandwidth of 10 MHz bandwidth to each MNO and the total bandwidth is 30MHz. Furthermore, for each MNO, 50 sub-channels with a bandwith of 180 KHz, each, are considered. For each simulation run, we consider a fixed number of users. The user mobility is also considered

Fig. 4 Simulation environment



and it is assumed that users are moving randomly with a speed of 3 km/h. In this experiment, we predict the resource band availability based on the shared information, and intelligent resource band allocation. Moreover, it is assumed that the antenna of each MNO has an omni-directional transmission pattern, and its height is 25 m. The antenna gain of each MNO's antenna is 17 dBi. Considering the downlink transmission, the transmitting power of an antenna is 46dBm. Specifically, we assess the delay performance, UPT, and sum–throughput of the cellular network using the proposed scheme for resource allocation in shares spectrum environment. These results are the compared with the conventional scheme.

For the neural network, the ftp file size is 1 MB and the discount value is 0.9. The scheduling interval is 1 ms and weights are updated after 100 ms. Moreover, the ReLU activation function is used in the output layer and the hidden layers to find the cost/loss function, and the Adam optimizer function is used. The batch size is 150 and the number of items in the replay memory is 1500. The packets arrive at each MNO with arrival rate λ, and it is assumed that the queue length of each MNO is infinite. However, a packet is dropped once its time reaches the value of τ_{drop}. For this simulation, we assume that τ_{drop} = 500 ms.

3.2 Performance Metrics

In order to evaluate the performance of the proposed scheme different performance metrics have been calculated. In this paper, the average delay, resource usage, UPT, packet drop ratio, and sum throughput have been calculated.

The duration of time between when a packet arrives and when it is transmitted to a user; is the delay that a packet experiences. A packet arrives in the network queue with the arrival rate (λ), and it is served and transmitted to a user based on the network load. If a collision occurs, the packet is retransmitted and delay is added. In the conventional resource allocation scheme, a packet arrives at an MNO and it is served using the allocated frequency band, as shown in Fig. 1a. If the frequency band is occupied, the arrived packet will face a long delay. The proposed scheme, on the other hand, can utilize idle resources of other MNOs, shown earlier in Fig. 1b. In order to decide whether or not to choose an idle band for packet transmission, the proposed learning mechanism is used; this is detailed in subsection 2.1. Using the proposed scheme, if packets can be transmitted quickly, the delay performance can be improved considerably.

The amount of network resources an MNO uses to transmit packets for a given amount of resources, is the resource usage percentage of that MNO. In the conventional scheme, if the resources are idle, others MNOs cannot use them; thus the resource usage percentage will be reduced. In comparison, the proposed scheme can efficiently utilize the resources by using the idle resources effectively.

UPT is an important performance metric. UPT (during the active time), can be defined as the size of a burst divided by the time between the arrival of the first packet of a burst and the reception of the last packet of a burst. More data can be transmitted in less time when the spectrum is efficiently shared among multiple MNOs. The proposed scheme can efficiently learn the network, and utilize idle resources efficiently. This can improve the UPT performance, as compared to the conventional resource scheduling mechanism where resources are not shared among multiple MNOs.

The packet drop ratio is the amount of packets dropped divided by the total number of packets arrived. When a packet arrives in the queue, it can stay there for some specific period (τ_{drop}). When the delay of the packet increases to the τ_{drop} value, the packet is dropped. In the proposed scheme, since an MNO can use the idle slots of other MNOs, the packet drop ratio can be reduced in comparison with the conventional resource allocation mechanism.

For the channel model between an MNO and a UE, the path loss model is considered. Moreover, time-block modeled Rayleigh fading and the log normal shadowing model are used. The received signal at a UE transmitted by an MNO can be given as

$$y_t = P_t - P_L - c[n] - \sum_{i=1}^{n-1} I_i, \tag{3}$$

where P_t is the transmitting power of an MNO and P_L represents the path loss model certified by the International Telecommunication Union [31] which also includes the effect of the Doppler shift due to mobility. This value is given by following formula

$$P_L = 22.0\log_{10}(d) + 28.0 + 20\log_{10}(f_c). \tag{4}$$

Here, $\log_{10}(d)$ represents the distortion in the signal due to the path loss, and the $20\log_{10}(f_c)$ term shows the effect of the Doppler shift. Moreover, $c[n]$ is the time-block Rayleigh fading channel gain, following the Rayleigh-Smith model given in [32]. I_i is the amount of interference faced by a UE from the transmitting MNO of the i^{th} cell. Considering n MNOs in an area, the total interference experienced by a UE will be the sum of all the interferences by $n-1$ transmitting MNOs. Considering the high interference network environment, the impact of noise will be negligible; thus, it is not considered in this paper. The signal–to–interference ratio (SIR) represented by (γ) in this paper, at a UE can be given as

$$\gamma = \frac{y_t}{\sum_{i=1}^{n-1} I_i}, \tag{5}$$

where n is the number of interfering MNOs. Finally, the sum throughput is defined for the performance metrics in this paper. The sum throughput is calculated using the Shannon capacity formula:

$$T = \log_2(1 + \gamma), \tag{6}$$

where γ is the SIR at a user, as given by Eq. 5.

3.3 Simulation Results

Fig. 5 plots the average delay of the network as the simulation time progresses from 0 to 200s. It can be seen that once the network is trained, the average delay of the proposed reinforcement learning scheme outperforms the conventional scheduling based resource allocation. Moreover, as the arrival rate (λ) increases, the delay of the network also increases. This is because that there will be more packets in the network's queue. It can also be seen that as the simulation time increases, the delay of the proposed network decreases considerably. Taking advantage of the learning procedure, MNOs learn the network and share the band efficiently. Once the network is well trained, there will be fewer collisions; thus, the delay of the network reduces. On the other hand, the delay of the conventional scheme remains constant.

In Fig. 6, the resource usage of the network is plotted versus the simulation time as simulation time progresses from 0 to 200s for different arrival rate (λ) values. Conventionally, even if the resource band of an MNO is idle, other MNOs cannot use it. Thus, the resources are not utilized efficiently. However, in the proposed scheme, an MNO can use the idle resources of other MNOs. This improves the network resource utilization. In the figure, it can be seen that as the simulation time increases, the resource usage of the proposed scheme improves relative to the conventional scheme. For lower values of λ, the resource usage of the proposed scheme is comparable with those of the conventional one. This is because there are not many packets in the network queue. However, as the value of λ increases, the proposed scheme outperforms the conventional scheme. The number of packets in the queue increases, and sharing the resource bands can lead to more effective utilization of resources. This shows the performance gain of the proposed scheme, where network learns about the resource usage and then efficiently utilizes the resources.

Fig. 5 Average delay performance of proposed scheme and conventional scheme for different values of λ as simulation time increases

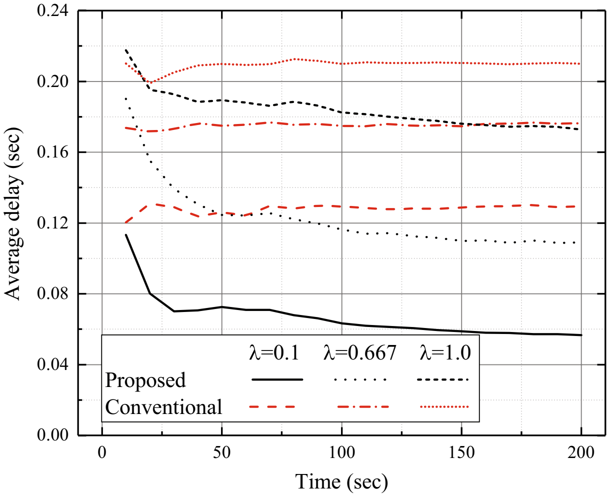


Fig. 6 Comparison of resource usage of the network of proposed and conventional schemes as the simulation time increases for different values of λ

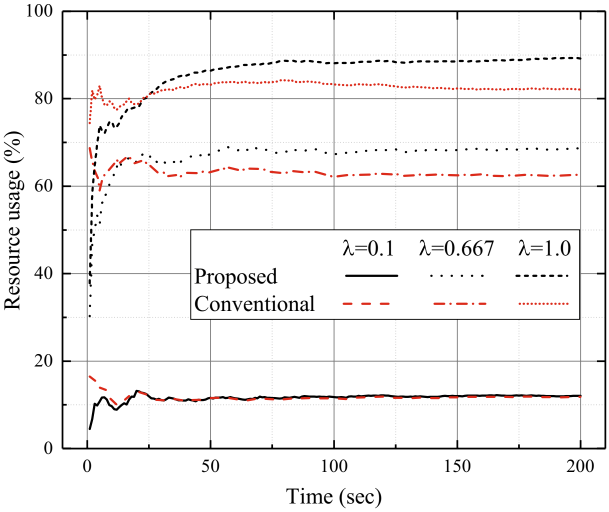


Fig. 7 plots the UPT against the simulation time, for different values of λ. The UPT is the amount of data a user can get during a time period. It can be observed that the proposed scheme shows possible performance gain in terms of the UPT. Sharing the band can reduce the amount of time a user has to wait to receive packets. Reducing the packet delay can lead to an improved quality of experience. Thus, by using the proposed scheme, resources can be scheduled intelligently. As a result, users can enjoy more throughput, as compared to the scheduling–based resource allocation.

The packet drop ratio is plotted in Fig. 8 for both the conventional and proposed schemes. For lower λ values, the packet drop ratio of the proposed scheme is slightly higher than the conventional scheme. This is because the learning mechanism requires some time resources; thus, the packet drop occurs. However, as λ increases, the performance of the proposed scheme improves. It can also be seen that the proposed neural network can learn the environment quickly, and the packet-drop reduces quickly for the proposed scheme.

Fig. 7 User perceived throughput comparison of the proposed scheme and conventional scheme as simulation time increases for different values of λ

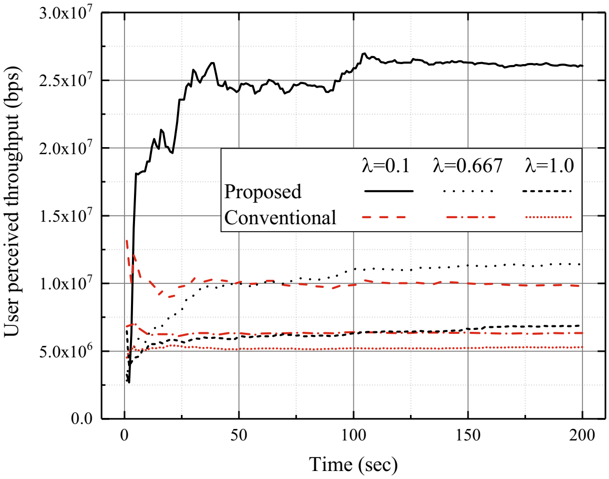
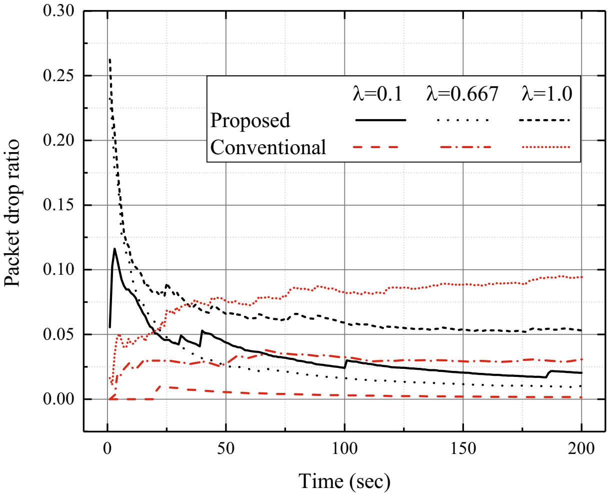


Fig. 8 The comparison of packet drop ratio for proposed scheme and conventional scheme as simulation time varies for different values of λ



The sum throughput of the network is plotted for the proposed and conventional schemes versus the simulation time in Fig. 9. When using the proposed learning mechanism, the network throughput remains almost the same or better than the conventional scheme. Once the network is well trained, the proposed scheme utilizes the resources efficiently; thus, delay of the network reduces. However, the amount of data transmitted over a timeslot is same for both proposed and conventional schemes. Thus, even by using the learning mechanism, the proposed scheme provides efficient results.

From the above results and discussion, it can be clearly observed that the proposed scheme performs better than the conventional resource allocation mechanism. However, it is not guarantee that the proposed scheme will always performs better in different scenarios. To observe this, we have computed the cumulative distribution function (CDF) of the delay for different simulations. Each simulation has random UE location. Fig. 10 shows the CDF plots of the average delay of the proposed scheme for each MNO, as computed by

Fig. 9 Sum throughput performance comparison of the proposed scheme and conventional scheme as simulation time increases for different values of λ

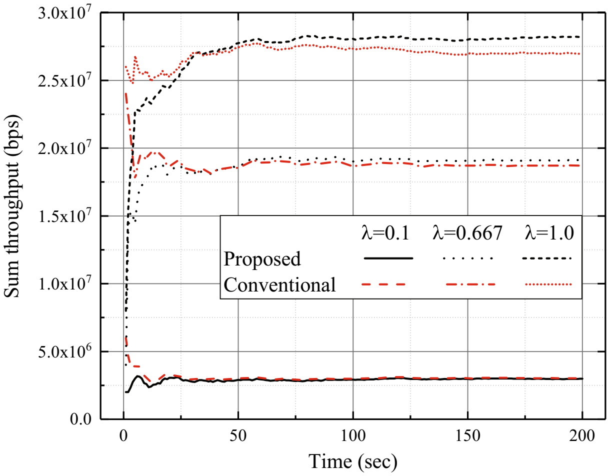
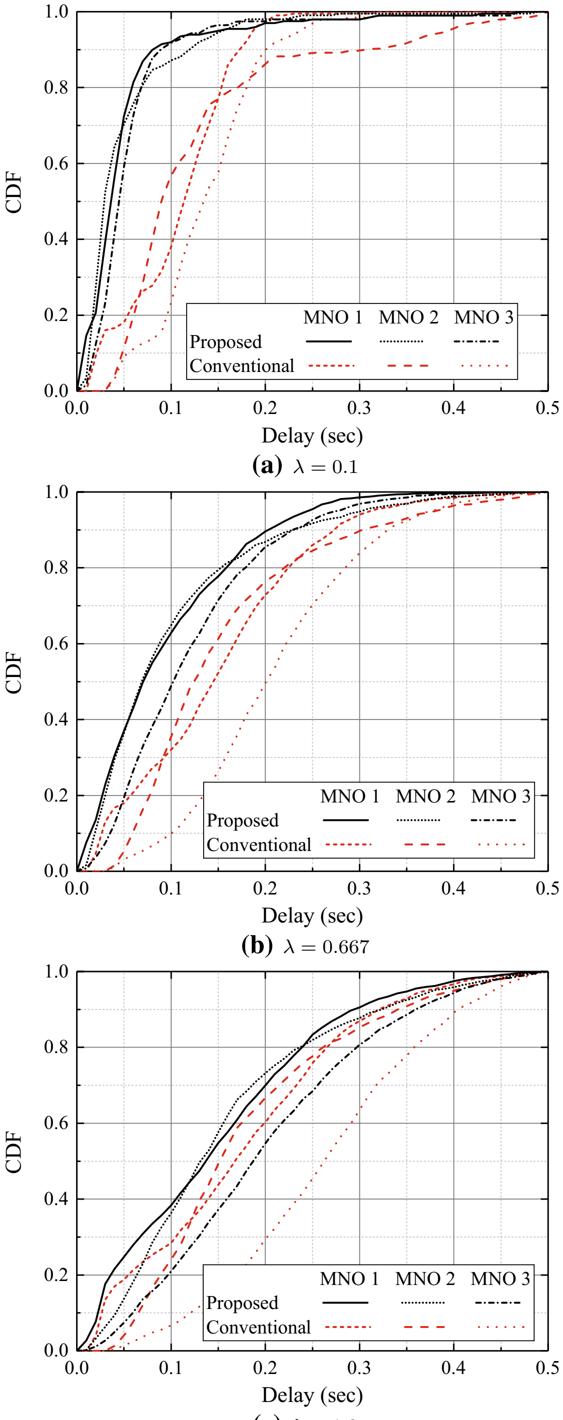


Fig. 10 CDF plots of the delay between packet transmission for conventional and proposed schemes



(a) $\lambda = 0.1$

(b) $\lambda = 0.667$

(c) $\lambda = 1.0$

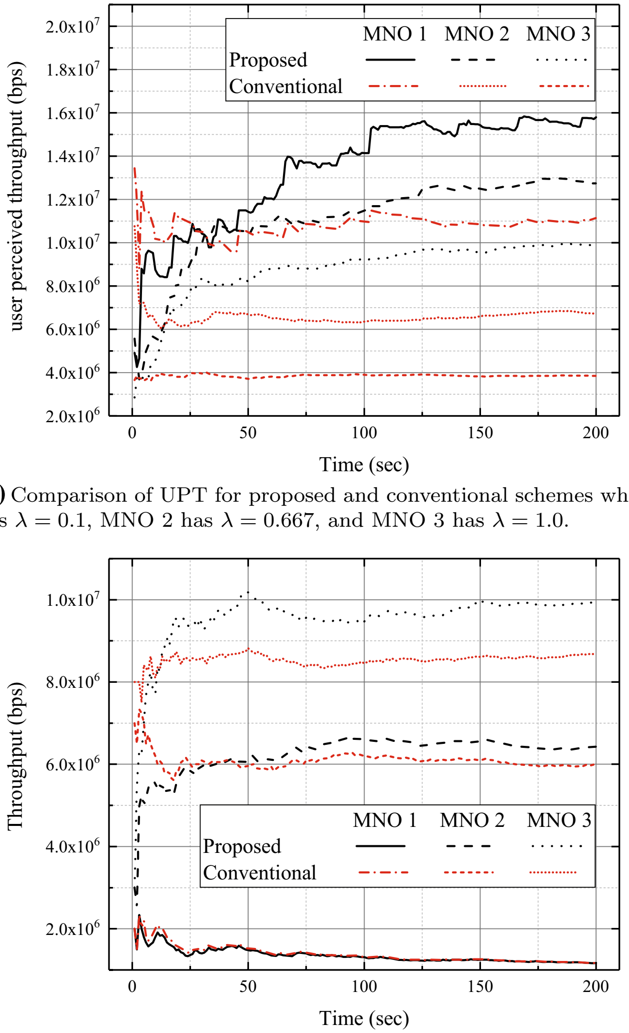
averaging the delay of 10 simulations. The CDF is then compared with the conventional resource allocation scheme for different values of λ. In Fig. 10a, the CDF of delay values for the proposed and conventional schemes when $\lambda = 0.1$ are plotted for each MNO. Sharing the resources among multiple MNOs can reduces the delay, and it can be seen in the figure that the proposed scheme reduces the delay. Similarly, Fig. 10b and c plot the CDF of delay values for $\lambda = 0.667$ and $\lambda = 1.0$, respectively. The delay performance improves by about 41%, 51%, and 13% for λ values of 0.1, 0.667, and 1.0, respectively, as compared to the conventional scheme. It can also be observed from the figures that, as the value of λ increases, the delay performance reduces; however, it is still better than the conventional scheme. This degradation in performance occurs because as λ increases, the number of packets in the queue will increase; thus, a packet has to wait for longer in the queue. However, since the proposed scheme can utilize the idle slots of other MNOs, it has better performance than the conventional scheme.

Table 1 shows the average performance improvement by comparing the proposed reinforcement learning-based model with the existing conventional scheme, which does not consider the scenario of sharing resource bands. The performance metrics include the UPT, delay, resource usage, packet drop and sum throughput. The values are observed and recorded as the average of 10 different user locations (i.e., for 10 simulations runs) when the network in the proposed scheme is well–trained. From this table, we can observe that our proposed scheme outperforms the conventional resource allocation scheme. The improvement is especially noticeable for higher values of λ. The proposed scheme is very efficient in predicting the spectrum availability, which shows that employing a deep reinforcement learning approach to perform spectrum availability prediction in the simulated communication network is successful. Moreover, as the MNOs learn the network over time, the allocation of the spectrum resources improves and number of collisions among MNOs reduces which improves the performance of the network. However, for the proposed scheme, some resources are utilized to train the network. This can reduce the packet transmission, especially for low values of λ, as compared to the conventional scheme.

Moreover, in order to check the effect of different λ values of MNOs on the UPT and throughput of the network, results have been plotted in Fig. 11. It can be clearly seen that the proposed scheme outperforms the conventional scheme. Fig. 11a shows the UPT performance of the proposed scheme when MNOs have different λ values. It can be seen that the UPT of the proposed scheme of each MNO is greater than the conventional scheme. Moreover, with the increasing simulation time, the UPT also increases. This is because,

Table 1 Summary of the performance metrics

Performance metrics	$\lambda = 0.1$	$\lambda = 0.667$	$\lambda = 1.0$
UPT (Mbps) (prop.)	18.1148	9.5337	6.6382
UPT (Mbps) (conv.)	8.5725	6.1339	5.5662
Delay (sec) (prop.)	0.0553	0.1050	0.1508
Delay (sec) (conv.)	0.1192	0.1692	0.1864
Resource usage(%) (prop.)	11.3351	69.3927	87.5376
Resource usage(%) (conv.)	11.0136	60.5050	78.5842
Packet drop (%) (prop.)	1.1601	1.1707	5.5976
Packet drop (%) (conv.)	0.00000	4.2251	9.9250
Sum throughput (Mbps) (prop.)	1.0425	6.8268	9.4592
Sum throughput (Mbps) (conv.)	0.9971	6.3978	9.2086



(a) Comparison of UPT for proposed and conventional schemes when MNO 1 has $\lambda = 0.1$, MNO 2 has $\lambda = 0.667$, and MNO 3 has $\lambda = 1.0$.

(b) Comparison of throughput for proposed and conventional schemes when MNO 1 has $\lambda = 0.1$, MNO 2 has $\lambda = 0.667$, and MNO 3 has $\lambda = 1.0$.

Fig. 11 Plots of different performance matrices when values of λ are different for each MNO

in the proposed scheme, MNOs can share the band and the delay of users is reduced significantly. In the conventional scheme, even if another band is idle, other MNOs cannot access it. This increases the delay and reduces the UPT. Since users can receive information quickly, the throughput also increases. It can be seen in Fig. 11b that, for a lower value of λ, the throughput of the proposed and conventional scheme are almost similar; however, as the value of λ increases, the throughput of the proposed scheme improves significantly. This is because, for lower values of λ there will be very few packets in the queue of the network; hence, the delay will be very low for both schemes. However, once the value of λ increases, the delay of the conventional scheme increases, and more packets have to wait in the queue (even if other bands are idle). Alternatively, the proposed scheme utilizes the idle bands efficiently, and the throughput of the proposed scheme improves.

4 Conclusions

Resource scarcity is likely to be a big problem for future wireless networks. Sharing the spectrum among multiple mobile network operators may be an efficient solution to overcome problems related to spectrum scarcity. In this paper, we have proposed a machine learning-based spectrum sharing scheme where an MNO can utilize the idle resources of other MNOs. A modified version of deep Q–learning network is proposed and trained for varying network environments. The main idea is to use the deep Q–learning approach that can efficiently utilize the resource bands in the shared spectrum scenario. Through simulation results, different performance metrics of the proposed scheme have been obtained and compared to the conventional resource scheduling scheme. Simulation results show the performance improvement of the proposed scheme as compared to the conventional scheme in terms of the delay, user-perceived throughput, resource usage, and network sum throughput.

Acknowledgements This work was supported in part by Samsung Research in Samsung Electronics.

References

1. Calvanese Strinati, E., Barbarossa, S., Gonzalez-Jimenez, J. L., & Ktenas, D. (2019). 6G: The Next Frontier: From Holographic Messaging to Artificial Intelligence Using Subterahertz and Visible Light Communication. *IEEE Vehicular Technology Magazine, 14*(3), 42–50.
2. Zhang, L., Xiao, M., Wu, G., Alam, M., Liang, Y.-C., & Li, S. (2017). A survey of advanced techniques for spectrum sharing in 5G networks. *IEEE Wireless Communications, 24*(5), 44–51.
3. Lin, Y. T., Tembine, H., & Chen, K. C. (2012). Inter-operator spectrum sharing in future cellular systems. *IEEE Global Communications Conference (GLOBECOM),* 2597–2602.
4. Luo, J., Eichinger, J., Zhao, Z., & Schulz, E. (2014). Multi-carrier waveform based flexible inter-operator spectrum sharing for 5G systems. *IEEE DySPAN,* 449–457.
5. Moon, B. (2017). Dynamic spectrum access for internet of things service in cognitive radio-enabled LPWANs. *Sensors (Switzerland), 17*.
6. Bkassiny, M., Li, Y., & Jayaweera, S. K. (2013). A survey on machine learning techniques in cognitive radios. *IEEE Communications Surveys and Tutorials, 15*(3), 1136–1159.
7. Wang, W., Kwasinski, A., Niyato, D., & Han, Z. (2016). A survey on applications of model-free strategy learning in cognitive wireless networks. *IEEE Communications Surveys and Tutorials, 18*(3), 1717–1757.
8. Thilina, K. M., Choi, K. W., Saquib, N., & Hossain, E. (2013). Machine learning techniques for cooperative spectrum sensing in cognitive radio networks. *IEEE Journal on Selected Areas in Communications, 31*(11), 2209–2221.
9. Zhang, J., Kountanis, D. I., & Al-Fuqaha, A. (2012). Two novel learning algorithms to solve the spectrum sharing problem in cognitive radio networks, In *2012 International Conference on Systems and Informatics (ICSAI2012), Yantai*, 1472–1476.
10. Azmat, F., Chen, Y., & Stocks, N. (2016). Analysis of spectrum occupancy using machine learning algorithms. *IEEE Transactions on Vehicular Technology, 65*(9), 6853–6860.
11. Shrestha, A. P., & Yoo, S.-J. (2018). Optimal resource allocation using support vector machine for wireless power transfer in cognitive radio networks. *IEEE Transactions on Vehicular Technology, 67*(9), 8525–8535.
12. Popoola, J.J., & Van Olst, R. (2011). Application of neural network for sensing primary radio signals in a cognitive radio environment. *IEEE Africon'11, Livingstone, Zambia*, 1–6.

13. Chen, M., Challita, U., Saad, W., Yin, C., & Debbah, M. (2019). Artificial neural networks-based machine learning for wireless networks: A tutorial. *IEEE Communications Surveys and Tutorials, 21*(4), 3039–3071.
14. Zappone, A., Renzo, M. D., & Debbah, M. (2019). Wireless networks design in the era of deep learning: Model-based, AI-based, or both? *ArXiv preprint.*arXiv:1902.02647.
15. Tan, J., Xiao, S., Han, S. , & Liang, Y. C. (2018). A learning based coexistence mechanism for LAA-LTE based HetNets. In *Proc. 2018 IEEE International Conference on Communications*, 1–6.
16. Park, T., Abuzainab, N., & Saad, W. (2016). Learning how to communicate in the internet of things: Finite resources and heterogeneity. *IEEE Access, 4,* 7063–7073.
17. Sun, Y., Peng, M., Zhou, Y., Huang, Y., & Mao, S. (2019). Application of machine learning in wireless networks: Key techniques and open issues. *IEEE Communications Surveys and Tutorials, 21*(4), 3072–3108.
18. Sun, Y., Peng, M., & Poor, H. V. (2018). A distributed approach to improving spectral efficiency in uplink device-to-device-enabled cloud radio access networks. *IEEE Transactions on Wireless Communications, 66*(12), 6511–6526.
19. Chen, M., Saad, W., & Yin, C. (2017). Echo state networks for self-organizing resource allocation in LTE-U with uplink-downlink decoupling. *IEEE Transactions on Wireless Communications, 16*(1), 3–16.
20. Fan, C., Li, B., Zhao, C., Guo, W., & Liang, Y.-C. (2018). Learning-based spectrum sharing and spatial reuse in mm-Wave ultradense networks. *IEEE Transactions on Vehicular Technology, 67*(6), 4954–4968.
21. Alnwaimi, G., Vahid, S., & Moessner, K. (2015). Dynamic heterogeneous learning games for opportunistic access in LTE-based macro/femtocell deployments. *IEEE Transactions on Wireless Communications, 14*(4), 2294–2308.
22. Puspita, R. H., Shah, S. D. A., Lee, G., Roh, B., Oh, J., Kang, S., & (2019). Reinforcement learning based 5g enabled cognitive radio networks. In Proc. . (2019). *International Conference on Information and Communication Technology Convergence (ICTC)* (pp. 555–558). Korea (South): Jeju Island.
23. Liang, L., Ye, H., Li, G. Y., & (2019). Multi-agent reinforcement learning for spectrum sharing in vehicular networks. In: Proc. . (2019). *IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)* (pp. 1–5). France: Cannes.
24. Liang, L., Ye, H., & Li, G. Y. (2019). Spectrum sharing in vehicular networks based on multi-agent reinforcement learning. *IEEE Journal on Selected Areas in Communications, 37*(10), 2282–2292.
25. Sallent, O., Perez-Romero, J., Ferrus, R., & Agusti, R. (2015). Learning-based coexistence for LTE operation in unlicensed bands. *Proc. IEEE ICCW,* 2307–2313.
26. Wang, T., Wen, C. K., Wang, H., Gao, F., Jiang, T., & Jin, S. (2017). Deep learning for wireless physical layer: Opportunities and challenges. *China Communications, 14*(11), 92–111.
27. Mao, Q., Hu, F., & Hao, Q. (2018). Deep learning for intelligent wireless networks: A comprehensive survey. *IEEE Communications Surveys and Tutorials, 20*(4), 2595–2621.
28. Ruan, L., & Wong, E. (2018). Machine intelligence in allocating bandwidth to achieve low-latency performance. *ONDM,* 226–229.
29. Riedmiller, M. (2005). Neural fitted Q iteration-first experiences with a data efficient neural reinforcement learning method. *European Conference on Machine Learning* (pp. 317–328). Berlin, Heidelberg: Springer.
30. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., & Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529–533.
31. Series, M. (2009). Guidelines for evaluation of radio interface technologies for IMT-advanced. Technical report, ITU.
32. Silva, V., Abrao, T., & Jeszensky, P. J. (2004). Statistically correct simulation models for the generation of multiple uncorrelated Rayleigh fading waveforms. In *Eighth IEEE International Symposium on Spread Spectrum Techniques and Applications - Programme and Book of Abstracts (IEEE Cat. No.04TH8738) Sydney, NSW, Australia*, 472–476.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Minsu Shin received his B.S. and M.S., degree in electronic and electrical engineering from Sungkyunkwan University, Suwon, South Korea, in 2016 and 2018, respectively. He is currently pursuing his Ph.D. degree with the Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, South Korea. His research interests include 6G networks, small–cell network, Wireless Local Area Network, Software Defined Networking, and Network Function Virtualization

Danish Mehmood Mughal received his B.S. degree in electrical engineering from Mirpur University of Science and Technology, Mirpur, Pakistan, in 2015. He is currently enrolled in combined Masters and Ph.D. program at Department of Electrical and Computer Engineering, Sungkyunkwan University. His research interests include Vehicular communications, 5G and B5G networks, and wireless energy harvesting.

Seungil Park received the B.S. degree from the Department of Electronic and Electrical Engineering (EEE), POSTECH, Korea in 2011 and the M.S. and the Ph.D. degrees from the Department of Electrical Engineering and Computer Science, Seoul National University (SNU), Korea in 2013 and 2018, respectively. He is currently working with Samsung Research, Samsung Electronics, South Korea. His research interests include wireless networks, Device-to-Device (D2D) communication, Vehicle-to-Everything (V2X) communication, Spectrum Sharing, and Satellite Communications.

Sang-Hyo Kim received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Seoul National University, Seoul, South Korea, in 1998, 2000, and 2004, respectively. From 2004 to 2006, he was a senior engineer with Samsung Electronics. He was with the University of Southern California, Los Angeles, CA, USA, as a visiting scholar, from 2006 to 2007. In 2007, he joined the College of Information and Communication Engineering, Sungkyunkwan University, Suwon, South Korea, where he is currently a professor. In 2015, he had a one-year visit to the University of California at San Diego, San Diego, CA, USA, as a visiting scholar. His research interests include coding theory, wireless communications, and deep-learning-based communication systems. He has served as an editor for Transactions on Emerging Telecommunications Technologies and Journal of Communications and Networks in 2013.

Min Young Chung received his B.S., M.S., and Ph.D. degrees in electrical engineering from theKorea Advanced Institute of Science and Technology, Daejeon, Korea, in 1990, 1993, and 1999,respectively. From January 1999 to February 2002, he was a Senior Member of Technical Staffwith the Electronics and Telecommunications Research Institute, where he was engaged inresearch on the development of multiprotocol label switching systems. In March 2002, he joinedthe Faculty of Sungkyunkwan University, Suwon, Korea, where he is currently a Professor withthe Department of Electrical and Computer Engineering. His research interests include D2DCommunications, Software-Defined Networking (SDN), 5G wireless communication networks,and wireless energy harvesting. He worked as an editor on the Journal of Communications andNetworks from January 2005 to February 2011, and is a member of IEEE, IEICE, KICS, KIPS and KISS.

Authors and Affiliations

Minsu Shin[1] · **Danish Mehmood Mughal**[1] · **Seungil Park**[2] · **Sang-Hyo Kim**[1] · **Min Young Chung**[1]

Minsu Shin
qwerty1643@skku.edu

Danish Mehmood Mughal
danish@skku.edu

Seungil Park
si18.park@samsung.com

Sang-Hyo Kim
iamshkim@skku.edu

[1] College of Information and Communication Engineering, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon-si, Gyeonggi-do 16419, Republic of Korea

[2] Samsung Research, Samsung Seoul R&D Campus, 56, Seongchon–gil, Seocho–gu, Seoul, Republic of Korea