International Journal of Mobile Computing and Multimedia Communications
Volume 14 • Issue 1

Multi-Agent Reinforcement Learning-Based Resource Management for V2X Communication

Nan Zhao, Hubei University of Technology, China
Jiaye Wang, Hubei University of Technology, China
Bo Jin, Hubei University of Technology, China
Ru Wang, Hubei University of Technology, China
Minghu Wu, Hubei University of Technology, China*
Yu Liu, The First Construction and Installation Co., Ltd. of China Construction Third Engineering Bureau, China
Lufeng Zheng, The First Construction and Installation Co., Ltd. of China Construction Third Engineering Bureau, China

ABSTRACT

Cellular vehicle-to-everything (V2X) communication is essential to support future diverse vehicular applications. However, due to the dynamic characteristics of vehicles, resource management faces huge challenges in V2X communication. In this paper, the optimization problem of the comprehensive efficiency for V2X communication network is established. Considering the non-convexity of the optimization problem, this paper ulitizes the markov decision process (MDP) to solve the optimization problem. The MDP is formulated with the design of state, action, and reward function for vehicle-to-vehicle links. Then, a multiagent deep Q network (MADQN) method is proposed to improve the comprehensive efficiency of V2X communication network. Simulation results show that the MADQN method outperforms other methods on performance with the higher comprehensive efficiency of V2X communication network.

KEYWORDS

Comprehensive Efficiency, *Multi-Agent Deep Q Network*, *V2X Communication*

INTRODUCTION

With the development of an intelligent transportation system, vehicle-to-everything (V2X) communication can improve traffic efficiency, road safety, and vehicle entertainment experience through wireless connection between road infrastructure and vehicles (Prathiba et al., 2021; Haapola et al., 2021). In V2X communication, vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) links are used to support the various vehicle applications (Chen et al., 2017; Thunberg et al., 2021).

DOI: 10.4018/IJMCMC.320190 *Corresponding Author

The different access requirements of vehicles and the special spectrum of V2X communication, make meeting the demands of massive data transmission difficult. Therefore, resource management is beneficial to improve the performance of a V2X communication network. Lee et al. (2017) used an efficient cluster-based resource management scheme to improve cellular user sum rate, average packet, and throughput. Bahonar et al. (2021) proposed a low-complexity resource allocation method for dense cellular V2X networks, and Bischoff et al. (2021) proposed a decentralized V2X resource allocation that can be used for cooperative driving. Zhou et al. (2021) described a multichannel access management approach for software defined in a cellular V2X network, and Pang et al. (2021) proposed an intelligent network resource management system to overcome the high-mobility edge computing problem of a 5G vehicle network. Although many documents describe the use of conventional optimization methods to solve the problem of V2X resource management, it is actually difficult to find the optimal solution by obtaining global channel status information.

To address the challenges caused by obtaining the global channel status information, we adopt the reinforcement learning (RL) method (Zhao et al., 2021; Zhao et al., 2020) in this paper. Currently one of the most powerful machine learning tools, RL is usually applied to time-varying dynamic systems (Wu et al., 2018; Yan et al., 2018) and the wireless network (Simsek et al., 2018; Zhao et al., 2018). Liang et al. (2019) proposed a multi-agent reinforcement learning (MARL) approach for spectrum sharing in a vehicle network. Zhang et al. (2019) proposed a deep reinforcement learning method to solve the problem of the resource allocation and the model selection in a cellular V2X network. Liu et al. (2020), described the use of a deep reinforcement learning method to optimize the spectrum efficiency and the energy efficiency of a V2X network. Choi et al. (2021) described a distributed congestion control method based on deep reinforcement learning to improve the traffic efficiency of a cellular V2X network.

However, most of the aforementioned literature lacks comprehensive consideration of the randomness of V2V link transmission data and vehicle dynamics in V2X communication networks. Therefore, this paper proposes a multi-agent depth Q network (MADQN) method to meet the aforementioned challenges.

In this paper, we propose the MADQN method to find the optimal solution of the optimization problem; that is, the optimal spectrum allocation and transmission power selection strategy of V2V link to maximize the comprehensive efficiency of the V2X communication network. The main contributions of this work are as follows:

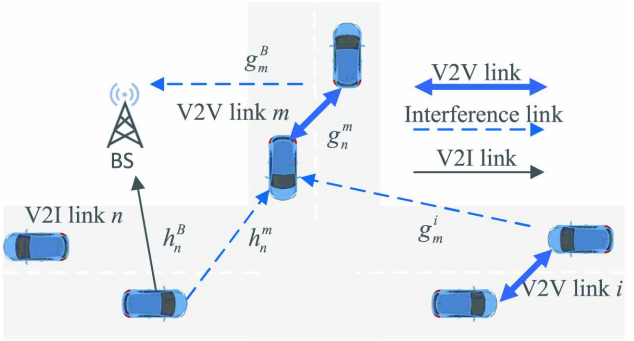
- Combining spectrum efficiency and energy efficiency, we defined the ratio of spectrum efficiency to the total power consumption as the efficiency of a V2I link and a V2V link. Then, we formulated an optimization problem to maximize the comprehensive efficiency of a V2X communication network.
- To obtain better V2X communication network performance, we modeled the optimization problem as Markov decision process (MDP). Then, we defined the state, action and reward function of V2V links. Considering the non-convexity of the optimization problem, we proposed a new MARL strategy. We adopted the MADQN method to achieve the optimal strategy.
- Experimental results show that the MADQN method can improve the performances of V2I and V2V links, which can achieve the higher comprehensive efficiency of a V2X communication network.

In the rest of this article we introduce a system model, explain the problem formulation, describe the establishment of the MDP model to solve a series of problems caused by V2V link reliability, and illustrate the performance of the MADQN approach in V2X communications.

SYSTEM MODEL

As shown in Figure 1, the cellular V2X communication network consists of a base station (BS), MM V2V links, and NN V2I links. Assume that the spectrum of V2I links is divided into NN spectrum sub-bands with the bandwidth WW .

Figure 1.
V2X Communication in the Vehicular Networkv



If we consider that the V2V links may have the different spectrum sub-band options (Han et al., 2018), the binary spectrum-allocation vector is defined as $\Gamma_m^n \Gamma_m^n$. When the V2V link mm reuses spectrum sub-band nn occupied by V2I link nn, $\Gamma_m^n = 1 \Gamma_m^n = 1$. Otherwise, $\Gamma_m^n = 0 \Gamma_m^n = 0$. We assume that each V2V link can occupy at the most only one spectrum sub-band, as shown in equation (1).

$$\Sigma_{n=1}^{N} \Gamma_m^n \le 1, m \in \{1,2,\ldots,M\} \Sigma_{n=1}^{N} \Gamma_m^n \le 1, m \in \{1,2,\ldots,M\}. \tag{1}$$

When multiple V2V links reuse the same frequency spectrum sub-band, interference management is required to reduce the interference between V2V links. The channel gain from V2I link nn to the BS on the spectrum sub-band nn is denoted as $h_n^B h_n^B$. Let $p_n p_n$ be denoted as the transmission power of V2I link nn on spectrum sub-band nn. Then, the signal-to-interference-plus-noise ratio (SINR) of V2I link nn on the spectrum sub-band nn is calculated using the formula shown in equation (2).

$$\tau_n = \frac{p_n h_n^B}{\sigma^2 + \Sigma_m \Gamma_m^n p_m^n g_m^B} \tau_n = \frac{p_n h_n^B}{\sigma^2 + \Sigma_m \Gamma_m^n p_m^n g_m^B}, \tag{2}$$

In equation (2), $\sigma^2 \sigma^2$ is the noise power, $p_m^n p_m^n$ denotes the transmission power of the V2V link mm on the spectrum sub-band nn, and $g_m^B g_m^B$ is the interference channel gain of the V2V link mm to the BS on the spectrum sub-band nn. Then, the transmission rate of V2I link nn on the spectrum is calculated as shown in equation (3).

$$r_n = Wlog_2\left(1+\tau_n\right) r_n = Wlog_2\left(1+\tau_n\right). \tag{3}$$

Moreover, let $p_i^n p_i^n$ be denoted as the transmission power of the V2V link ii on the spectrum sub-band nn, then the SINR of V2V link mm on the spectrum sub-band nn can be expressed as shown in equation (4).

$$\tau_m^n = \frac{\Sigma_m \Gamma_m^n p_m^n g_m^n}{\sigma^2 + p_n h_n^m + \Sigma_{i \ne m} \Sigma_m \Gamma_i^n p_i^n g_m^i} \tau_m^n = \frac{\Sigma_m \Gamma_m^n p_m^n g_m^n}{\sigma^2 + p_n h_n^m + \Sigma_{i \ne m} \Sigma_m \Gamma_i^n p_i^n g_m^i}, \tag{4}$$

In equation (4), $g_m^n g_m^n$ is the channel gain of V2V link mm on the spectrum sub-band nn, $h_n^m h_n^m$ is the interference channel gain of V2I link nn to V2V link mm, and $g_m^i g_m^i$ is the interference channel gain of V2V link ii to V2V link mm. Then, the transmission rate of V2V link mm on spectrum sub-band nn can be obtained by the formula in equation (5).

$$r_m^n = Wlog_2\left(1+\tau_m^n\right)r_m^n = Wlog_2\left(1+\tau_m^n\right). \tag{5}$$

Considering the spectrum efficiency and energy efficiency of the V2X communication network (Ye et al., 2019), we define the efficiency of the V2I link as the ratio of spectrum efficiency to the total transmission power consumption, which can be expressed as shown in equation (6).

$$\eta_I = \frac{\sum_{n=1}^{N} r_n \,/\, w}{\left(\sum_{n=1}^{N} p_n + Np_c\right)}, \tag{6}$$

In equation (6), $p_c p_c$ is the circuit power. Similarly, the efficiency of the V2V link is defined as shown in equation (7).

$$\eta_v = \frac{\Sigma_{n=1}^{N}\Sigma_{m=1}^{M}\Gamma_m^n r_m^n \,/\, \sum_{n=1}^{N} w}{\left(\sum_{m=1}^{M} p_m^n + Mp_c\right)}. \tag{7}$$

Thus, the comprehensive efficiency of the V2X communication network is defined as shown in equation (8).

$$\eta_x = \lambda_\theta \eta_I + \lambda_\omega \eta_{IV} \eta_x = \lambda_\theta \eta_I + \lambda_\omega \eta_{IV}, \tag{8}$$

In equation (8), $\lambda_\theta \lambda_\theta$ and $\lambda_\omega \lambda_\omega$ are the weighting coefficients in the two V2X communication networks, respectively.

PROBLEM FORMULATION

In this paper, the global goal is to maximize the comprehensive efficiency of the V2X communication network, which is formulated as shown in equation (9).

$$\begin{aligned} &\max_{\Gamma_m^n, p_m^n}\left(\eta_X\right) \\ s.t.\quad &C1: \sum_{n=1}^{N}\Gamma_m^n \le 1, \forall m \in \left\{1,2,...,M\right\}, \\ &C2: \Gamma_m^n \in \left\{0,1\right\}, \\ &C3: 0 \le p_m^n \le P_{\max}, \end{aligned} \tag{9}$$

In this equation C_1C_1 and C_2C_2 are the constraints of spectrum allocation $\Gamma_m^n\Gamma_m^n$. Constraint C_3C_3 is the maximum transmission power $p_{max}p_{max}$ of the V2V link. Because of the non-convexity of the optimization issue shown in equation (9), obtaining the optimal strategy with the traditional optimization methods is difficult. In the next section, the MADQN method will be proposed to solve the above problem.

MADQN-BASED RESOURCE ALLOCATION SCHEME

Because the reliability of the V2V link brings great complexity in interference management, large state space, and action space, the above optimization problem is modeled as MDP $\left(S,A,R,P\right)\left(S,A,R,P\right)$ (Xiang et al., 2021), where SS is the state space, AA represents the action space, RR is the reward function, and PP represents the state transition probability. The V2V links act as agents to guide their own spectrum-allocation $\Gamma_m^n\Gamma_m^n$ and transmission power $p_m^np_m^n$ selections.

State Space

Assume that the transmission load ψ of V2V link mm is transmitted within the time step tt. Then, according to the transmission rate of V2V link mm, the remaining transmission load $z_m^tz_m^t$ can be obtained by the formula in equation (10).

$$z_m^t = \psi - \Sigma_n^N\Gamma_m^nr_m^n\left(t\right)z_m^t = \psi - \Sigma_n^N\Gamma_m^nr_m^n\left(t\right). \tag{10}$$

For each V2V link, the state space mainly contains the channel state information of the V2I link and the V2V link. In addition, the remaining transmission load $z_m^tz_m^t$ and remaining transmission time $T-tT-t$ are considered to better capture the queuing state of the V2V link. That is, the state space $s_m^ts_m^t$ of V2V link mm can be defined as shown in equation (11).

$$s_m^t = \left[g_m^n, g_m^i, h_n^m, g_m^B, z_m^t, T-t\right]s_m^t = \left[g_m^n, g_m^i, h_n^m, g_m^B, z_m^t, T-t\right]. \tag{11}$$

Action Space

Considering that the V2V links need to select spectrum allocation $\Gamma_m^n\Gamma_m^n$ and transmission power $p_m^np_m^n$, we define the action space $a_m^ta_m^t$ of the V2V link mm in each time slot tt using the formula shown in equation (12).

$$a_m^t = \left\{\Gamma_m^n, p_m^n\right\}a_m^t = \left\{\Gamma_m^n, p_m^n\right\}, \tag{12}$$

where $p_m^n \in \left\{P_1, P_2, ..., P_k\right\}p_m^n \in \left\{P_1, P_2, ..., P_k\right\}$.

Reward Function

Considering that there are two possibilities for the V2V link to be transmitted successfully or unsuccessfully, we design two different reward parameters. If the remaining transmission load is $z_m^tz_m^t$, the transmission rate $r_m^nr_m^n$ of the V2V link is taken as the reward coefficient. Otherwise, the

constant C serves as the reward parameter. Thus, the reward function $r_t r_t$ at each time step tt can be written as shown in equation (13).

$$r_t = \begin{cases} \Sigma_{n=1}^{N}\Sigma_{m=1}^{M}\Gamma_m^n r_m^n\left(t\right)\eta_X, & if z_m^t \geq 0, \\ C\eta_X, & otherwise. \end{cases} r_t = \begin{cases} \Sigma_{n=1}^{N}\Sigma_{m=1}^{M}\Gamma_m^n r_m^n\left(t\right)\eta_X, & if z_m^t \geq 0, \\ C\eta_X, & otherwise. \end{cases} \tag{13}$$

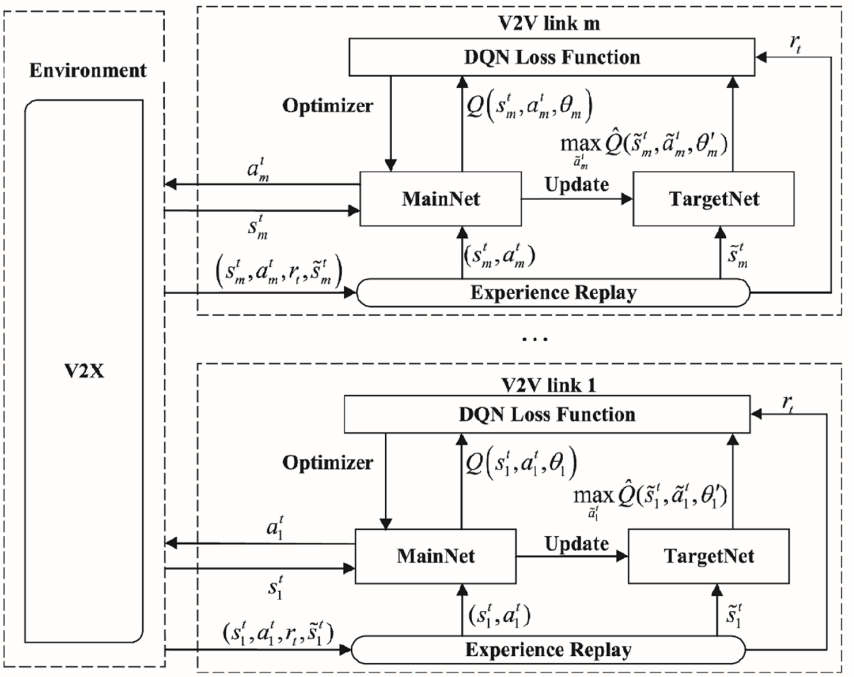
Learning Algorithm

To address the problems discussed in the previous sections, the Q learning method is usually used to learn an optimal policy and then choose the action that obtains the maximum reward from the environment. However, the larger state space $s_m^t s_m^t$ and action space $a_m^t a_m^t$ will result in a larger Q table that will require more time spent on space for exploration and storage. Thus, we propose the MADQN algorithm to deal with this problem.

As shown in Figure 2, the framework of DQN is composed of the vehicle environment and V2V links. MADQN uses the deep neural network model to realize state estimation of agents. To train effectively and update the Q network, the MADQN algorithm has two important strategies. On one hand, experience replay is applied to keep the historical experience, thus ensuring the relative independence of training data. On the other hand, the V2V link mm has the same structure of the MainNet (weight $\theta_m \theta_m$) and TargetNet (weight $\theta_m^{'} \theta_m^{'}$).

The collected environmental information is stored in the replay memory $D_m D_m$ by the experience replay. Then, mini-batches of samples from the replay memory $D_m D_m$ are randomly selected to

Figure 2.
The Framework of DQN-Based Vehicle Communication



update the Q network. In each step tt, based on the state $s_m^t s_m^t$, V2V link mm performs spectrum-allocation $\Gamma_m^n \Gamma_m^n$ and transmission power selection $p_m^n p_m^n$ according to the action value function $Q\left(s_m^t, a_m^t, \theta_m\right) Q\left(s_m^t, a_m^t, \theta_m\right)$, which is defined as shown in equation (14).

$$Q\left(s_m^t, a_m^t, \theta_m\right) = E\left[\sum_{t'=t}^{T} \gamma^{t'=t} r_{t'} \left| s_m^t, a_m^t, \theta_m \right.\right] Q\left(s_m^t, a_m^t, \theta_m\right) = E\left[\sum_{t'=t}^{T} \gamma^{t'=t} r_{t'} \left| s_m^t, a_m^t, \theta_m \right.\right], \tag{14}$$

In this equation, $0 < \gamma < 10 < \gamma < 1$ is the discount factor to represent the impact of the future reward. Then, based on the action $a_m^t a_m^t$, the environment transitions to a new state $s_m^t s_m^t$, and V2V link mm receives a reward $r_t r_t$ from the environment. With reward $r_t r_t$ and new state $\tilde{s}_m^t \tilde{s}_m^t$, the V2V link mm can update the weights of DQN by minimizing the loss function $L\left(\theta_m\right) L\left(\theta_m\right)$, which can be expressed using the formula shown in equation (15).

$$L\left(\theta_n\right) = E\left[\left(y_t^m - Q\left(s_m^t, a_m^t, \theta_m\right)\right)^2\right] L\left(\theta_n\right) = E\left[\left(y_t^m - Q\left(s_m^t, a_m^t, \theta_m\right)\right)^2\right], \tag{15}$$

In this equation, $y_t^m y_t^m$ is the target value to represent the optimization object of the TargetNet, which can be obtained using the formula in equation (16).

$$y_t^m = r_t + \gamma \max_{\tilde{a}_m^t} \hat{Q}\left(\tilde{s}_n^t, \tilde{a}_m^t, \theta_m^{'}\right) y_t^m = r_t + \gamma \max_{\tilde{a}_m^t} \hat{Q}\left(\tilde{s}_n^t, \tilde{a}_m^t, \theta_m^{'}\right), \tag{16}$$

In equation (16), $\hat{Q}\left(\tilde{s}_n^t, \tilde{a}_m^t, \theta_m^{'}\right) \hat{Q}\left(\tilde{s}_n^t, \tilde{a}_m^t, \theta_m^{'}\right)$ is the TargetNet updated every 100 steps. The proposed resource management policy based on the MADQN algorithm is summarized as Algorithm 1. First, the experience replay buffer is initialized with size $D_m D_m$, and the Q networks of the V2V links are randomly initialized. In each episode, the vehicle locations and large-scale fading need to be updated initially. The remaining transmission load $z_m^t z_m^t$ is reset. In each step of an episode, each V2V link estimates a local channel and compiles a sectional observation $s_m^t s_m^t$. The next environment state $\tilde{s}_m^t \tilde{s}_m^t$ and reward $r_t r_t$ can be obtained according to the V2V link's selection of an action $a_m^t a_m^t$. Next, the information $\left(s_m^t, a_m^t, r_t, \tilde{s}_m^t\right)\left(s_m^t, a_m^t, r_t, \tilde{s}_m^t\right)$ is stored in the replay memory $D_m D_m$. Then, a mini-batch of transitions is sampled randomly from the replay memory $D_m D_m$ to update the Q network. Moreover, equation (16) is used to optimize loss function $L\left(\theta_m\right) L\left(\theta_m\right)$ between the Q network and learning target. Finally, all V2V links start transmission according to the spectrum-allocation $\Gamma_m^n \Gamma_m^n$ and transmission power selected $p_m^n p_m^n$ by their own actions (table 1).

PERFORMANCE EVALUATION

Now we evaluate the performance of the MADQN method for resource management through simulations. We consider the intersection scene of vehicles based on the spatial Poisson process, where BS is located in the center. The crossroad size is 1299 m × 750 m, and each road consists of two lanes in different directions. Four V2I and K V2V transmitters are randomly selected from the vehicle, and each V2V transmitter establishes a V2V link with its adjacent vehicle. The LOS status,

Table 1.
Training Process of V2V Link mm

1. Initialize replay memory $D_m D_m$
2. Initialize action value function $Q\left(s_m^t, a_m^t, \theta_m\right) Q\left(s_m^t, a_m^t, \theta_m\right)$ with random weight $\theta_m \theta_m$
3. **for** each episode **do**
4. Update vehicle locations
5. Reset $z_m = \psi z_m = \psi$, for all $\forall m \in \left\{1,2,...,M\right\} \forall m \in \left\{1,2,...,M\right\}$
6. **for** each step tt **do**
7. **for** V2V link mm **do**
8. Observe state $s_m^t s_m^t$
9. Choose action $a_m^t a_m^t$ from $s_m^t s_m^t$ according to ε-greedy policy
10. **end for**
11. V2V links take action and receive reward $r_t r_t$
12. **for** V2V link mm **do**
13. Observe new state $\tilde{s}_m^t \tilde{s}_m^t$
14. Store $\left(s_m^t, a_m^t, r_t, \tilde{s}_m^t\right)\left(s_m^t, a_m^t, r_t, \tilde{s}_m^t\right)$ in the replay memory $D_m D_m$
15. **end for**
16. **end for**
17. **for** V2V link mm **do**
18. Uniformly sample mini-batches from replay memory $D_m D_m$
19. Optimize loss function $L\left(\theta_m\right) L\left(\theta_m\right)$ between Q network and learning targets
20. **end for**
21. **end for**

path loss, shadowing, and fast fading parameters, which configures in 3GPP TR 37.885 (Meredith et al., 2016). More details simulation parameters are listed in Table 2.

The MADQN algorithm used in the simulation consists of an input layer, three hidden layers, and an output layer. The hyperparameters of MADQN are listed in Table 3. The learning rate is 10−5. The size of the mini-batch is 2,000. The numbers of neurons in the hidden layer are 120, 250, and 500, while ReLu and RMSProp are used as activation functions and optimizers, respectively. Note that the parameters listed are selected from multiple simulation tests to balance the complexity and performance of the MADQN algorithm.

Figure 3 shows the efficiency of the V2I link with the MADQN method under different numbers of V2V links. With the number of V2V links increasing, the efficiency of V2I links first increases and then decreases. When the number of V2V links $M = 2M = 2$, V2V links cooperate with each other to decrease the transmission loss, reducing the impact on the transmission rate $r_n r_n$ of the V2I link. However, the inference between V2V links increases when the number of V2V links $M = 3M = 3$, which affects multiple V2V links and increases the transmission loss. As the transmission power consumption of the V2V link increases, the transmission rate $r_n r_n$ of the V2I link decreases, resulting in a decrease in the efficiency of the V2I link. In addition, with the increase of transmission load $\psi\psi$, the transmission time and transmission power consumption of the V2V link increase, affecting the transmission rate $r_n r_n$ of the V2I link and decreasing the efficiency of the V2I link.

Figure 4 shows the efficiency of the V2V link with different V2V link numbers. Compared with the number of V2V links $M = 1M = 1$ and $M = 2M = 2$, the performance of the MADQN method increased by 7.5% and 5.3% under the number of V2V links $M = 2M = 2$, respectively. Since that

Table 2.
Basic Vehicle Environment Parameters

Parameter	Value
Number of V2V links MM	2
Number of V2I links NN	4
Carrier frequency	2 GHz
Total bandwidth WW	10 MHz
Transmission power of V2I links	23 dBm
Transmission power of V2V links	[23,15,10,0] dBm
Circuit power $p_c p_c$	16 dBm
BS antenna gain	8 dBi
Vehicle antenna gain	3 dBi
Noise power $\sigma^2 \sigma^2$	-114 dBm
Transmission load $\psi\psi$	2 × 1060 bytes
Vehicle speed VV	30 km/h
Lane width	3.5 m

Table 3.
Hyperparameter of DQN

Parameter	Value
Number of hidden layers	3
Learning rate	0.00001
Optimizer	RMSProp
Update steps	100
Batch size	2000
Activation function	ReLU

too many V2V links, resulting in increased interference. This interference reduces the transmission rate $r_n r_n$ of the V2V link so that the comprehensive efficiency of the V2V link continues to decrease. Moreover, with the transmission load $\psi\psi$ increasing, the efficiency of the V2V link decreases continuously mainly because the larger the transmission load is, the more transmission time the V2V link needs. Then, the long-term transmission brings the higher transmission power consumption, thereby decreasing the efficiency of the V2V link.

Figure 5 presents the comprehensive efficiency of a V2X communication network with different V2V link numbers. Under three different numbers of V2V links, the comprehensive efficiency of

Figure 3.

The Efficiency of V2I Link $\eta_1\eta_1$ on Different V2V Link Numbers

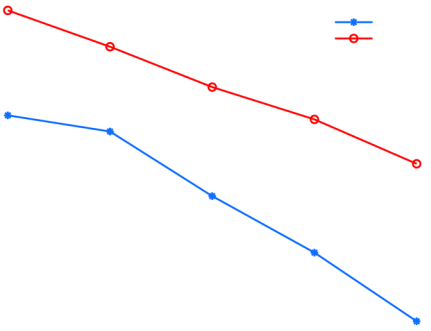


Figure 4.
The efficiency of the V2V link with different V2V link numbers

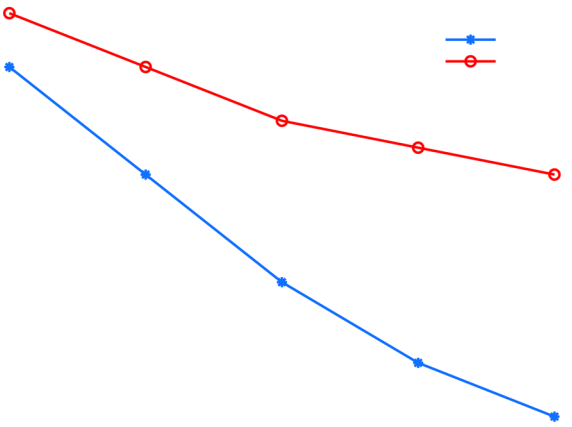
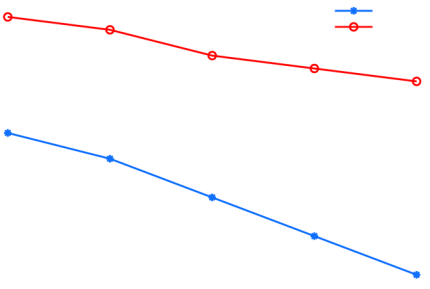


Figure 5.

The Comprehensive Efficiency of V2X Communication Network $\eta_X\eta_X$ on Different V2V Link Numbers



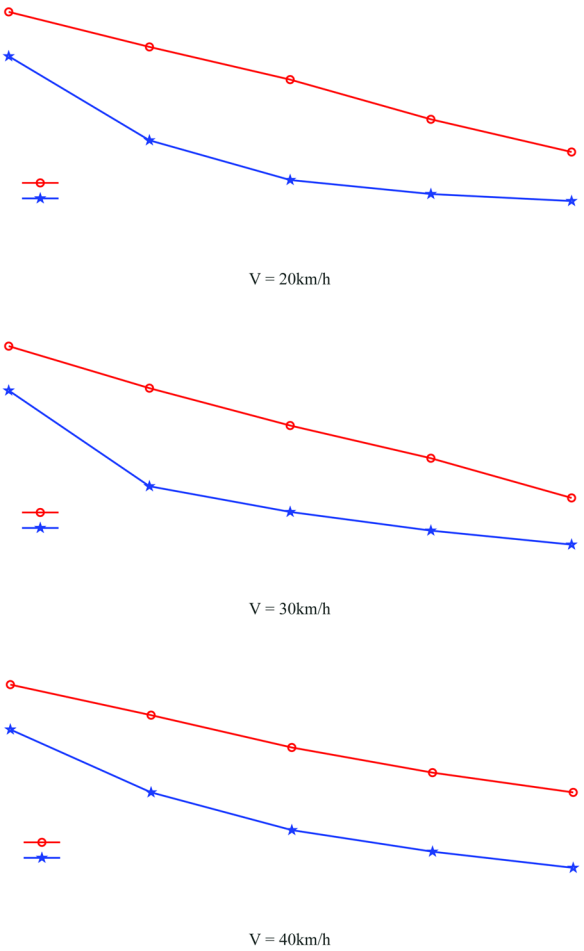
the V2X communication network decreased by 17.2%, 6.8%, and 13%, respectively. Compared with the number of V2V links $M = 1M = 1$, the number of V2V links $M = 2M = 2$ obtains more environmental states by interacting with the environment to help them select the optimal spectrum-allocation $\Gamma_m^n\Gamma_m^n$ and transmission power $p_m^n p_m^n$, thus reducing the transmission power consumption

of each V2V link. However, when the number of V2V links $M = 3M = 3$, it suffers more interference than the other two cases, resulting in its performance degradation. Furthermore, owing to the increase in transmission load $\psi\psi$, both the efficiency of the V2I link and the efficiency of the V2V link will be affected to a certain extent, resulting in a decrease in the comprehensive efficiency of the V2X communication network. Combining the above two situations, we choose the number of V2V links $M = 2M = 2$ to continue the following experiments.

We compare the performance of the MADQN method with the single-agent reinforcement learning (SARL) method and random baseline method (Rand). The efficiency of the V2I link at different speeds is shown in Figure 6. With the speed of the vehicle increasing, the communication time between the vehicle and the base station is shortened, which is not conducive to data transmission, so the efficiency of the V2I link decreases with the increase of speed. Under the three vehicle speeds, the drop rates of the Rand method are 20.7%, 22.2%, and 27.3%, respectively. The drop rates of the MADQN method are 10.9%, 14.6%, and 17.4%, respectively. Compared with the Rand method, the MADQN method interacts with the environment through multiple V2V links to obtain the environment state and constantly adjusts its own strategy design. The performance of the MADQN method is better than that of the Rand method in a dynamic environment. In addition, when the transmission

Figure 6.

The Efficiency of V2I Link $\eta_I\eta_I$ with Different Vehicle Speed VV

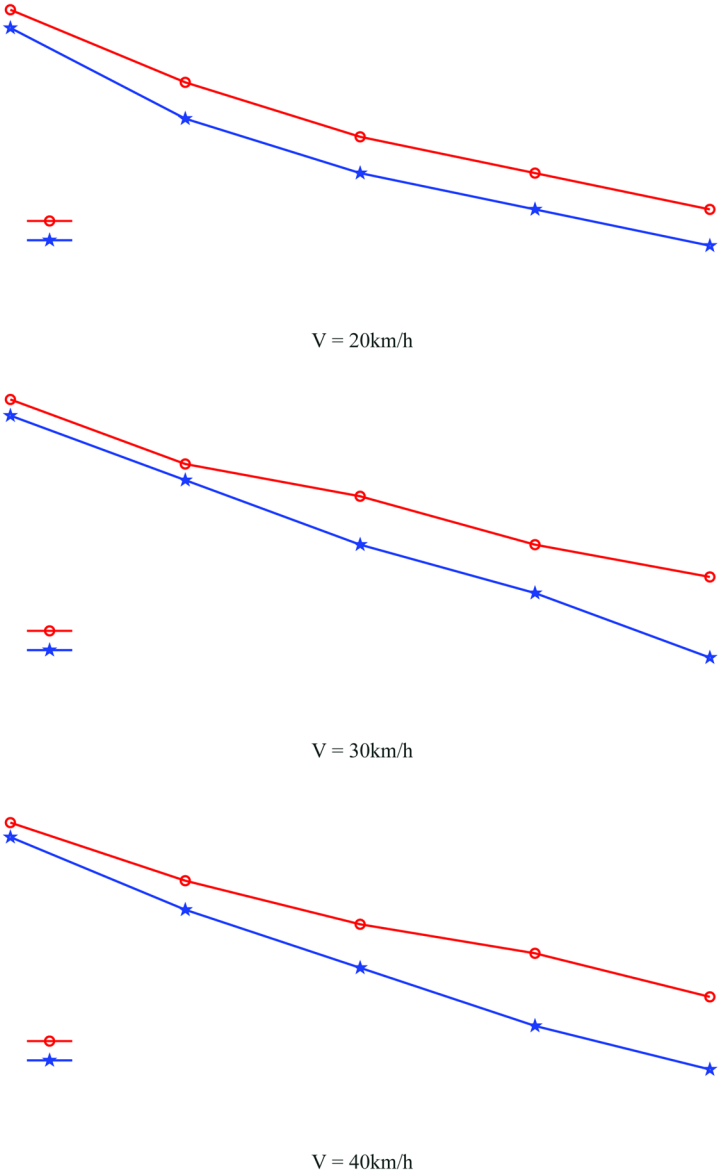


load $\psi\psi$ of the V2V link increases, the transmission time of the V2V link increases. Long-term transmission increases the transmission power consumption of the V2V link and the interference to the V2I link, thus reducing the transmission rate of the V2I link. Thus, in these three cases, the efficiency of the V2I link decreases as the transmission load $\psi\psi$ increases.

Figure 7 describes the functional relationship between the efficiency of V2V link $\cdot_{V}$ and the transmission load $\psi\psi$ under different speeds. The efficiency of the V2V link decreases with the increase of the transmission load $\psi\psi$. Because the transmission load $\psi\psi$ is smaller, the lower the transmission loss is generated. Figure 7 shows that when the transmission load $\psi\psi$ increases, the performance

Figure 7.

The Efficiency of V2V Link $\eta_V\eta_V$ with Different Vehicle Speed VV

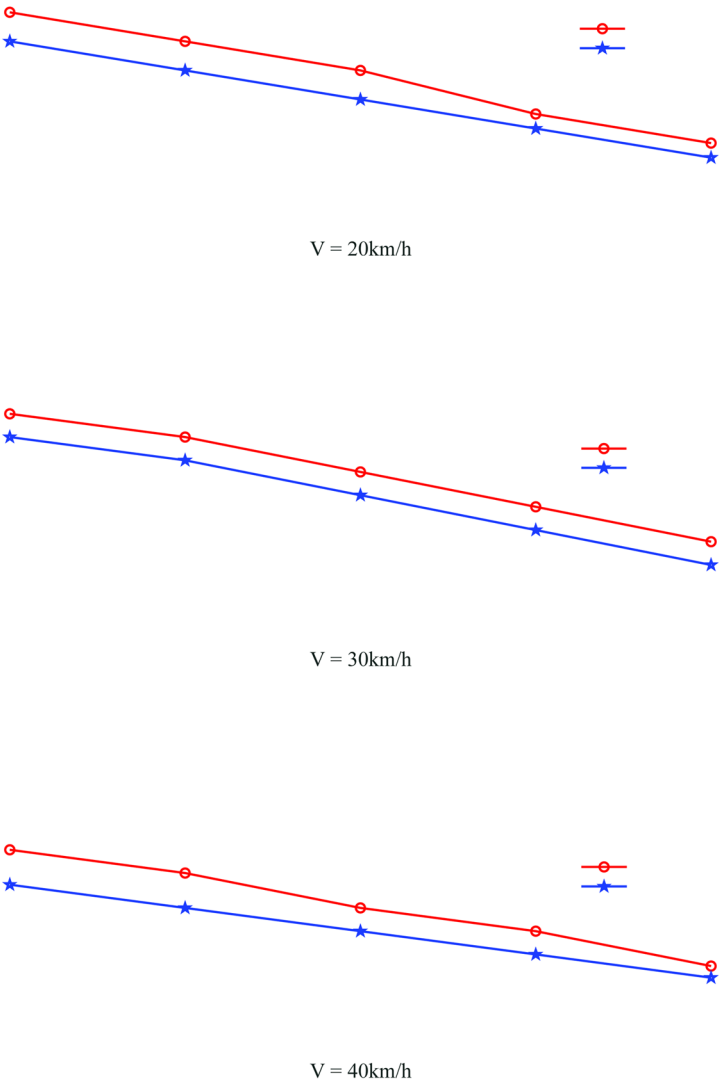


degradation of the SARL and the Rand methods is much higher than that of the MADQN method. In addition, the transmission time of the V2V link decreases as the speed increases, resulting in a decrease in the link connection time, affecting the transmission performance of the V2V link. At the vehicle speeds $V = 20km / hV = 20km / h$, the performance gap is the largest between the MADQN method and the other two methods when the transmission load $\psi\psi$ is 3. However, with the increase of vehicle speed, the gap gradually decreases, but the MADQN method still has the highest performance.

The comprehensive efficiencies of the V2X communication network under different vehicle speeds are shown in Figure 8. When the transmission load $\psi\psi$ increases from 4 to 5, the reduction rates of the MADQN method under the vehicle speed $V = 20km / hV = 20km / h$ and $V = 30km / hV = 30km / h$ are 1.4% and 1.45%, respectively. The reduction rate of the MADQN method under the vehicle speed $V = 40km / hV = 40km / h$ is 6.2%. In these three cases, the

Figure 8.

The Comprehensive Efficiency of V2X Communication Network $\eta_X\eta_X$ with Different Vehicle Speed VV



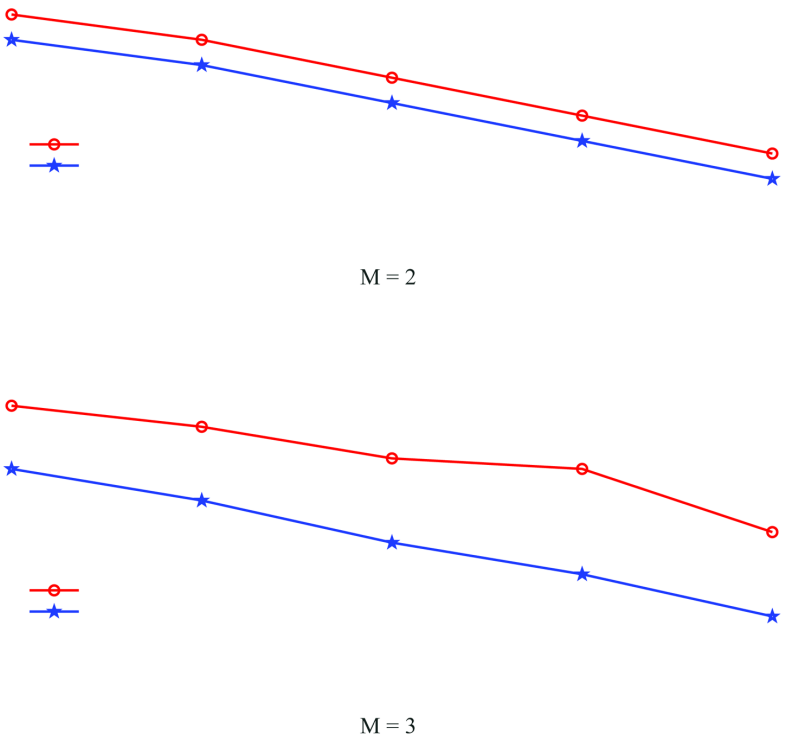
comprehensive efficiencies of the V2X communication network under all methods decrease with the increase of transmission load $\psi\psi$ because when the transmission load $\psi\psi$ is large, the transmission time of the V2V link also increases gradually, thereby increasing the transmission power consumption and interference. Compared with the other two methods, the MADQN method can better adapt to the dynamic environment and obtain the higher performance of the V2X communication network.

Figure 9 shows the comprehensive efficiency comparison of the V2X communication network on different V2V link numbers. Figure 9 indicates that when the number of V2V links $M = 2M = 2$, the comprehensive efficiency of the V2X communication network decreases with the increase of the transmission load $\psi\psi$. Compared with the SARL method, the MADQN method has multiple V2V links to explore the vehicle environment to obtain the environmental state; consequently, the MADQN method can better guide its own policy design. Thus, the performance of the MADQN method is higher than that of the SARL method.

Figure 9 indicates that the performance of the MADQN method is still better than other two methods. In addition, when the number of V2V links $M = 2M = 2$ and $M = 3M = 3$, the comprehensive efficiency of the V2X communication network under the MADQN method drops by 6.8% and 13%, respectively because as the number of V2V links increases, the interference between V2V links also intensifies, resulting in a performance degradation of the MADQN method. Compared with the number of V2V links $M = 3M = 3$, the MADQN method can achieve the higher comprehensive efficiency of the V2X communication network under the number of V2V links $M = 2M = 2$.

Figure 9.

The Comprehensive Efficiency of V2X Communication Network $\eta_X\eta_X$ with Different V2V Link Numbers



CONCLUSION

In this paper, the MADQN method is proposed for the better efficiency resource management in V2X communication. Considering the vehicle dynamics and the randomness of V2V link transmission data, we studied the optimization of the comprehensive efficiency of the V2X communication network. In addition, aiming at the non-convexity of the optimization problem, we modeled it as an MDP. Then, we used the MADQN method to the spectrum-allocation and transmission power of V2V links selects. Based on TargetNet and experience replay, the MADQN method can efficiently learn optimization strategies. Numerical results show that MADQN method has better performance than other two methods. In the future, we will comprehensively consider user requirements and bandwidth allocation scenarios, and further analyze the robustness of MADQN-based algorithms to better improve the comprehensive efficiency of V2X networks.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Minghu Wu: Project administration, funding acquisition. Nan Zhao and Jiaye Wang: Writing original draft, methodology, conceptualization, software, validation, formal analysis. Bo Jin: Supervision, software, validation, writing and editing. Ru Wang, Yiyang Pei, and Lufeng Zheng: Validation, writing and editing.

ACKNOWLEDGMENT

This work was supported by the Key Research and Development Plan of Hubei Province, China (2021BGD013), the Science and Technology Research Program of Hubei Provincial Department of Education (T201805), the Natural Science Foundation of Hubei Province (2022CFA007) and the Knowledge Innovation Special Project of Wuhan Science and Technology Bureau, China (No. 2022010801010255).

REFERENCES

Bahonar, M. H., Omidi, M. J., & Yanikomeroglu, H. (2021). Low-complexity resource allocation for dense cellular vehicle-to-everything (C-V2X) communications. *IEEE Open Journal of the Communications Society*, *2*, 2695–2713. doi:10.1109/OJCOMS.2021.3135290

Bischoff, D., Schiegg, F. A., Schuller, D., Lemke, J., Becker, B., & Meuser, T. (2021). Prioritizing relevant information: Decentralized V2X resource allocation for cooperative driving. *IEEE Access: Practical Innovations, Open Solutions*, *9*, 135630–135656. doi:10.1109/ACCESS.2021.3116317

Chen, S., Hu, J., Shi, Y., Peng, Y., Fang, J., Zhao, R., & Zhao, L. (2017). Vehicle-to-everything (V2X) services supported by LTE-based systems and 5G. *IEEE Communications Standards Magazine*, *1*(2), 70–76. doi:10.1109/MCOMSTD.2017.1700015

Choi, J.-Y., Jo, H.-S., Mun, C., & Yook, J.-G. (2021). Deep reinforcement learning-based distributed congestion control in cellular V2X networks. *IEEE Wireless Communications Letters*, *10*(11), 2582–2586. doi:10.1109/LWC.2021.3108821

Haapola, J., & Samarasinghe, T. (2021). The effect of concurrent multi-priority data streams on the MAC layer performance of IEEE 802.11 p and C-V2X mode 4. *IEEE Transactions on Communications*, *70*(1), 592–605. doi:10.1109/TCOMM.2021.3119703

Han, S., Huang, Y., Meng, W., Li, C., Xu, N., & Chen, D. (2018). Optimal power allocation for SCMA downlink systems based on maximum capacity. *IEEE Transactions on Communications*, *67*(2), 1480–1489. doi:10.1109/TCOMM.2018.2877671

Lee, C.-H., Chang, R. Y., Lin, C.-T., & Cheng, S.-M. (2017). Sum-rate maximization for energy harvesting-aided D2D communications underlaid cellular networks. In *Proceedings of the IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)* (vol. 1, pp. 1–6). IEEE. doi:10.1109/PIMRC.2017.8292471

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. doi:10.1109/JSAC.2019.2933962

Liu, Z., Han, Y., Fan, J., Zhang, L., & Lin, Y. (2020). Joint optimization of spectrum and energy efficiency considering the C-V2X security: A deep reinforcement learning approach. In *Proceedings of the 2020 IEEE 18th International Conference on Industrial Informatics (INDIN)* (pp. 315–320). IEEE. doi:10.1109/INDIN45582.2020.9442103

Meredith, J. M. (2016). *Study on LTE-based V2X services* (Release 14). 3GPP Portal. https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2934

Pang, S., Wang, N., Wang, M., Qiao, S., Zhai, X., & Xiong, N. N. (2021). A smart network resource management system for high mobility edge computing in 5G Internet of vehicles. *IEEE Transactions on Network Science and Engineering*, *8*(4), 3179–3191. doi:10.1109/TNSE.2021.3106955

Prathiba, S. B., Raja, G., & Kumar, N. (2021). Intelligent cooperative collision avoidance at overtaking and lane changing maneuver in 6G-V2X communications. *IEEE Transactions on Vehicular Technology*, *71*(1), 112–122. doi:10.1109/TVT.2021.3127219

Simsek, M., Bennis, M., & Czylwik, A. (2012). Dynamic inter-cell interference coordination in HetNets: A reinforcement learning approach. In *Proceedings of the 2012 IEEE Global Communications Conference (GLOBECOM)* (pp. 5446–5450). IEEE. doi:10.1109/GLOCOM.2012.6503987

Thunberg, J., Bischoff, D., Schiegg, F. A., Meuser, T., & Vinel, A. (2021). Unreliable V2X communication in cooperative driving: Safety times for emergency braking. *IEEE Access: Practical Innovations, Open Solutions*, *9*, 148024–148036. doi:10.1109/ACCESS.2021.3124450

Wu, C., Yoshinaga, T., Ji, Y., & Zhang, Y. (2018). Computational intelligence inspired data delivery for vehicle-to-roadside communications. *IEEE Transactions on Vehicular Technology*, *67*(12), 12038–12048. doi:10.1109/TVT.2018.2871606

Xiang, P., Shan, H., Wang, M., Xiang, Z., & Zhu, Z. (2021). Multi-agent RL enables decentralized spectrum access in vehicular networks. *IEEE Transactions on Vehicular Technology*, *70*(10), 10750–10762. doi:10.1109/TVT.2021.3103058

Yan, S., Zhang, X., Xiang, H., & Wu, W. (2019). Joint access mode selection and spectrum allocation for fog computing based vehicular networks. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 17725–17735. doi:10.1109/ACCESS.2019.2895626

Ye, H., Li, G. Y., & Juang, B.-H. F. (2019). Deep reinforcement learning based resource allocation for V2V communications. *IEEE Transactions on Vehicular Technology*, *68*(4), 3163–3173. doi:10.1109/TVT.2019.2897134

Zhang, X., Peng, M., Yan, S., & Sun, Y. (2019). Deep-reinforcement-learning-based mode selection and resource allocation for cellular V2X communications. *IEEE Internet of Things Journal*, *7*(7), 6380–6391. doi:10.1109/JIOT.2019.2962715

Zhao, N., Liang, Y.-C., & Pei, Y. (2018). Dynamic contract incentive mechanism for cooperative wireless networks. *IEEE Transactions on Vehicular Technology*, *67*(11), 10970–10982. doi:10.1109/TVT.2018.2865951

Zhao, N., Liu, Z., Cheng, Y., & Tian, C. (2020). Multi-agent actor critic for channel allocation in heterogeneous networks. *International Journal of Mobile Computing and Multimedia Communications*, *11*(1), 23–41. doi:10.4018/IJMCMC.2020010102

Zhao, N., Ren, F., Du, W., & Ye, Z. (2021). Deep reinforcement learning for task offloading and power allocation in UAV-assisted MEC system. *International Journal of Mobile Computing and Multimedia Communications*, *12*(4), 1–20. doi:10.4018/IJMCMC.289163

Zhou, H., Ma, T., Xu, Y., Cheng, N., & Yan, X. (2021). Software-defined multi-mode access management in cellular V2X. *Journal of Communications and Information Networks*, *6*(3), 224–236. doi:10.23919/JCIN.2021.9549119

Nan Zhao received B.S., M.S., and Ph.D. degrees from Wuhan University, Wuhan, China, in 2005, 2007, and 2013, respectively. She is currently a professor at the Hubei University of Technology, Wuhan. Her current research involves machine learning in wireless communications. She also is currently studying electronic information at the Hubei University of Technology's School of Electrical and Electronic Engineering. She is proficient in data communication, big data analysis, data management, Internet of Things, cloud computing, and other technologies.

Minghu Wu received a B.S. degree in electronic information engineering from the Communication University of China, Beijing, China, in 1998; an M.S. degree in communication information systems from the Huazhong University of Science and Technology, Wuhan, China, in 2002; and a Ph.D. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2013. He is currently a professor at the Hubei University of Technology. His major research interests include communication signal processing and video coding. He also has worked in the field of Internet of Things research and engineering for more than 10 years.