Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 7490689, 12 pages
https://doi.org/10.1155/2021/7490689

Research Article

Energy-Efficient Resource Allocation for NOMA-Enabled Internet of Vehicles

Xin Chen,[1] **Zhuo Ma**,[1] **Teng Ma**,[2] **Xu Liu**,[2] **and Ying Chen**[1]

[1]*School of Computer Science, Beijing Information Science & Technology University, Beijing, China*
[2]*School of Automation, Beijing Information Science & Technology University, Beijing, China*

Correspondence should be addressed to Xin Chen; chenxin@bistu.edu.cn

Received 7 May 2021; Revised 10 August 2021; Accepted 25 August 2021; Published 16 September 2021

Academic Editor: Yong Zhang

With the rapid development of Internet of vehicles (IoV) technology, the distribution of vehicles on the highway becomes more dense and the highly reliable communication between vehicles becomes more important. Nonorthogonal multiple access (NOMA) is a promising technology to meet the multiple access volume and the high reliability communication demands of IoV. To meet the Vehicle-to-Vehicle (V2V) communication requirements, a NOMA-based IoV system is proposed. Firstly, a NOMA-based resource allocation model in IoV is developed to maximize the energy efficiency (EE) of the system. Secondly, the established model is transformed into a Markov decision process (MDP) model and a deep reinforcement learning-based subchannel and power allocation (DSPA) algorithm is designed. An event trigger block is used to reduce computation time. Finally, the simulation results show that NOMA can significantly improve the system performance compared to orthogonal multiaccess, and the proposed DSPA algorithm can significantly improve the system EE and reduce the computation time.

1. Introduction

With the rapid development of vehicle wireless communication technology, Internet of vehicles (IoV) has a broad development prospect [1]. Among the various applications generated by IoV, security applications are undoubtedly of the highest priority because they impact on the safety of the vehicles directly [2]. Vehicle-to-Vehicle (V2V) communication as a key technology in intelligent transportation system (ITS) that could meet the strict latency and reliability requirements of safety applications has attracted continuous academic attention [3].

V2V communication is aimed at communicating directly between vehicles with extremely low latency and ultrahigh reliability, which could guarantee the quality of service (QoS) requirements of security applications [4]. In general, device-to-device (D2D) communication provides the principle of direct propagate information between adjacent devices, which could greatly reduce latency and transmission energy consumption. Therefore, D2D technology is commonly used as the basis for V2V communication. That is why the 3rd Generation Partnership Project (3GPP) developed V2V communication principles based on D2D technology [5] in the long-term evolutionary (LTE) system. However, it has been shown that the QoS requirements of V2V communication cannot always be guaranteed under this principle. The reason is D2D communication following this principle is based on orthogonal multiple access (OMA) [6], a technology that does not make full use of spectrum resource and has difficulty in solving interference problems due to the increase in vehicles. When vehicles are deployed densely, IoV system would suffer from severe congestion, which affects the performance of the system.

Such problems have been solved with the rise of 5th generation (5G) mobile networks. 5G introduces nonorthogonal multiple access (NOMA) technology that allows a resource block to be assigned to multiple users, thus greatly expanding the amount of access to the network [7]. In some cases, such as uplink communication intensive scenarios, NOMA-enabled system has a significant performance improvement compared to OMA system. The cost of extended access is that NOMA actively introduces interference information and

requires reducing the impact of interference by successive interference cancelation (SIC) techniques [8]. Compared to OMA system, NOMA is more complex to decode at the receiver side, but after adopting SIC and other technologies, it is beneficial to the whole system performance. SIC technology decodes the received signal level by level and removes it after successful decoding to reduce the interference to the undecoded signal. In NOMA-enabled IoV system, the performance of V2V communication can be significantly improved.

Due to its advantages over OMA, NOMA is widely used in ultradense network (UDN), mobile edge computing (MEC), IoV, and other environments [9, 10]. Currently, NOMA has great potential to expand network access and improve network performance, but there are still some issues that need to be addressed. There have been many works introducing NOMA technology for resource allocation and interference management. In these works, the optimization of system throughput and the QoS requirements for V2V communication have been mainly considered. However, NOMA extends the number of user accesses through channel multiplexing, which increases the difficulty of channel allocation. In addition, the power allocation scheme becomes more complex due to the interference introduced by NOMA, and the overall system power consumption should be considered in the resource allocation scheme. Besides, literature [11] has analyzed the SIC technique and pointed out that, due to the complexity of implementation, normally two users could share the same subchannel at most.

To solve the above problem, we study the resource allocation problem for high energy efficiency (EE) in IoV systems. We describe the scenario of the NOMA-enabled IoV system and present the resource allocation problem for maximizing the system EE. Due to the complexity of the system and the high computational dimensionality of the direct solution, we transform the optimization problem into a Markov decision process (MDP) and use deep reinforcement learning (DRL) method to solve it. The main contributions of this paper are as follows:

(i) We investigate the problem of resource allocation in IoV system. The NOMA technology is introduced to meet the demand for multivehicle access, and the implementation of uplink SIC technology is presented. By allocating the channel and power resources of vehicles, we propose an optimization goal of maximizing the system EE

(ii) We transform the optimization goal into a problem of resource allocation strategy based on MDP and propose a DRL-based subchannel and power allocation (DSPA) algorithm to solve it. Specifically, the deep Q network (DQN) method is used to solve the subchannel selection and the deep deterministic policy gradient (DDPG) method is used to solve the power allocation problem. The event trigger block is used to reduce the computation time

(iii) We simulate and analyze the designed algorithm. The simulation results show that the performance of the NOMA-enabled IoV system is more suitable for multiple vehicle access situations than OMA, and the DSPA algorithm can effectively enhance the system EE and reduce the computation time

The rest of this paper is organized as follows. In Section 2, we analyze the work related to this paper. The system model and problem formulation are given in Section 3. In Section 4, we transform the optimization problem into an MDP model and design the DSPA algorithm for solving it. In Section 5, the proposed resource allocation method is simulated and analyzed. Section 6 is the conclusion.

2. Related Work

Due to the variability of QoS requirements for vehicle users, the resource allocation problem in vehicle networks has attractive research value and has received extensive attention from researchers for years [12, 13]. Since the high speed movement of vehicles in IoV makes it difficult to obtain accurate and fast channel change information, Guo et al. [14] obtained the time delay of V2V link in steady state based on Markov process and determine the optimal transmit power for each possible spectrum and finally allocated the spectrum resource by dichotomous matching method to maximize the system data transmission rate. Chen et al. [15] developed an online network slice resource allocation strategy that can meet the demand for QoS requirements for IoV applications and maximize system capacity. Liang et al. [16] designed a multi-intelligent DQN algorithm to allocate spectrum and power for each V2V link and maximize the total system throughput. Yang et al. [17] studied the design frame structure for V2V communication in IoV and proposed a semipersistent frame scheduling algorithm, which greatly meets the needs of V2V communication.

Resource allocation for IoV system can also be combined with MEC. Chen et al. [18] considered the dynamics of computational task arrival and wireless channel state in the MEC scenario and jointly optimized task and computational resource allocation to minimize system energy consumption while guaranteeing the upper limit of queue length. Zhao et al. [19] studied the collaborative offloading strategy of edge clouds in IoV and designed a distributed computational offloading and resource allocation algorithm to optimize the joint benefits of offloading and resource allocation. The problem of joint allocation of spectrum, computation, and storage resources in MEC-based IoV was studied by Peng et al. [20]. Since the problem has a high computational complexity, the authors transformed the problem using reinforcement learning (RL) method and solved it with a hierarchical learning architecture to obtain the optimal resource allocation decision.

By introducing NOMA technology in IoV scenery, the system performance will be further improved. Di et al. [21] proposed a resource allocation scheme in IoV broadcast scenarios, using NOMA to reduce latency and improve data acceptance probability. The main idea of this scheme is a centralized channel selection strategy and a distributed power allocation strategy. The packet reception probability is significantly improved by this scheme. Liu et al. [22]

studied the optimal power allocation problem in broadcast and multicast transmission schemes in half-duplex NOMA-based IoV scenarios and proposed a bifurcation-based power allocation algorithm that significantly improves the system throughput compared with the OMA scheme.

3. System Model and Problem Formulation

3.1. System Model. We consider a multivehicle highway scenario where one base station is located at the center and the radius of the base station coverage is D, as shown in Figure 1. The time domain is uniformly divided into multiple time slots, and the length of each slot is τ. We denote m as the index for the m-th moving vehicle on the highway where $m \in \{1, 2, \cdots, M\}$, and the maximum travel speed of the vehicle is $v_{\max}$. At each time slot t, there are N $(N < M)$ vehicles that send the required security information to the surrounding vehicles within its communication range through up to one subchannel. Such communications are based on V2V communication, and this transmission vehicles are denoted as VT user; the set of all VT users is $\mathcal{N}$. During each time slot t, the number of VT users $|\mathcal{N}^{(t)}|$ obeys a Poisson distribution

$$\Pr\left\{\left|\mathcal{N}^{(t)}\right| = n\right\} = \frac{(\alpha_{\mathrm{VT}}\tau)^n}{n!}\exp\left(-\alpha_{\mathrm{VT}}\tau\right), \quad n = 0, 1, 2, \cdots, \tag{1}$$

where α_{VT} denotes the arrival intensity of VT users in terms of VTs per second.

A right-angle coordinate system is established with the base station as the origin, and the position of each vehicle is denoted by (a_m, b_m). All vehicles are traveling in one direction with speed v_m, and the coverage radius of V2V communication is $d_{\max}$. The total available bandwidth for the D2D communication is W_{all} and is divided equally into K nonorthogonal subchannels, each bandwidth $W = W_{\mathrm{all}}/K$.

Due to the dense vehicle coverage, when multiple VT users send messages through the same subchannel simultaneously, the receiving vehicles (denoted as VR) located in the common coverage area of these VT users may receive messages with large interference. NOMA allows multiple vehicles to transmit information through the same channel simultaneously, and the VR users use SIC technology to decode the received information and reduce the cochannel interference.

We denote $\mathbb{N}_l$ as the set of all VT users that can be received by the receiving vehicle VR_l, i.e., $\mathbb{N}_l = \{1 \le n \le N \mid d_{n,l} \le d_{\max}\}$, where $d_{n,l} = \sqrt{(b_n - b_l)^2 + (a_n - a_l)^2}$ is the distance between VT_n and VR_l. In time slot t, the signal received by the receiving vehicle VR_l on subchannel k (SC_k) is

$$y_{l,k}^{(t)} = \sum_{n \in \mathbb{N}_l} \alpha_{n,k}^{(t)} \sqrt{p_{n,k}^{(t)}} h_{n,l,k}^{(t)} s_n^{(t)} + z_l^{(t)}, \tag{2}$$

where $\alpha_{n,k}^{(t)}$ is a binary variable that indicates the subchannel selected by VT_n. Specifically, $\alpha_{n,k}^{(t)} = 1$ if VT_n transmits through SC_k, and $\alpha_{n,k}^{(t)} = 0$ otherwise. $p_{n,k}^{(t)}$ is the transmitted power of VT_n in time slot t, $s_n^{(t)}$ denotes the modulation symbol, and z_l^t represents the additive white Gaussian noise (AWGN) for VR_l which obeys the complex Gaussian distribution with variance σ_l^2, that is, $z_l^{(t)} \sim \mathcal{CN}(0, \sigma_l^2)$. $h_{n,l,k}^{(t)}$ denotes the coefficient of SC_k from VT_n to VR_l. Specifically, $h_{n,l,k}^{(t)} = g_{n,l}^{(t)} \cdot \mathrm{PL}^{-1}(d_{n,l}^{(t)})$, where $g_{n,l}^{(t)}$ denotes Rayleigh fading channel gain, and $\mathrm{PL}^{-1}(d_{n,l}^{(t)}) = \beta(d_{n,l}^{(t)})^{-\varphi}$ represents the path loss function with the shadowing component β and the power decay exponent φ.

We map the mobility of the vehicle to the change in the position of the vehicle. Since there is short length of time slots, it can be assumed that the position of the vehicles in time slot t does not change, so the distance of any two vehicles $d_{m,m'}$ remains constant in time slot t. The position of the vehicle needs to be recalculated at the beginning of the time slot $t + 1$. According to Equation (2), the distance between vehicles is further mapped to the change in channel gain, so we assume that the channel gain also remains within one time slot, while it changes in the adjacent time slots. Thus, the SINR between VT_n and VR_l over SC_k in time slot t without SIC technology can be expressed as

$$\Gamma_{n,l,k}^{(t)} = \frac{p_{n,l}^{(t)}\left|h_{n,l,k}^{(t)}\right|^2}{\sigma_l^2 + \sum_{n' \in \mathbb{N}_l, n' \ne n} p_{n',l}^{(t)}\left|h_{n',l,k}^{(t)}\right|^2}, \tag{3}$$

where $\sigma_l^2 = E[|z_l^{(t)}|^2]$ is the noise power on SC_k and $|h_{n,l,k}^{(t)}|^2$ is the channel gain. The data rate of SC_k between VT_n and VR_l without SIC technique can be expressed as

$$R_{n,l,k}^{(t)} = W \cdot \log_2\left(1 + \Gamma_{n,l,k}^{(t)}\right). \tag{4}$$

In the uplink NOMA system, the superimposed signal $y_{n,l}^{(t)}$ received by VR_l needs to have a certain clarity between the different signals in order to eliminate interference. Since the channels between each VT_n and VR_l are different, the signals sent by each VT user in the uplink experience a different channel gain. Therefore, among the superimposed signals $y_{n,l}^{(t)}$, the VT user with the best channel quality may have the strongest received power, and VR_l decodes this VT signal first, i.e., the decoding order of VR_l is from VT users with good channel quality to those with poor channel quality. Otherwise, it has to allocate higher power for VT users with poor channel quality to improve their received power, which will reduce EE. Assuming that there are N VT users sending messages to VR_l over SC_k and the order of the channel gains between each VT user and VR_l is

$$\left|h_{1,l,k}^{(t)}\right|^2 \ge \left|h_{2,l,k}^{(t)}\right|^2 \ge \cdots \ge \left|h_{n,l,k}^{(t)}\right|^2 \ge \left|h_{n+1,l,k}^{(t)}\right|^2 \ge \cdots \ge \left|h_{N,l,k}^{(t)}\right|^2. \tag{5}$$

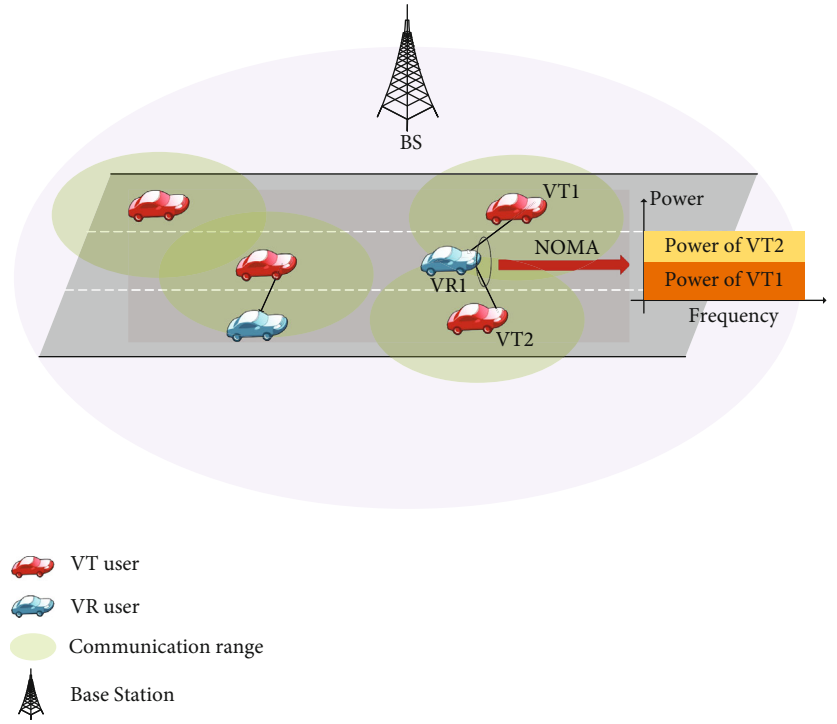

Figure 1: NOMA-based IoV system scenario.

According to the SIC decoding rules, VR_l firstly decodes VT users with $n' < n$ and eliminates $VT_{n'}$ interference symbols when decoding VT_n, but not eliminate $VT_{n''}(n'' > n)$ interference symbols. Therefore, the SINR between VT_n and VR_l over SC_k in time slot t with SIC technology can be expressed as

$$\widetilde{\Gamma}_{n,l,k}^{(t)} = \frac{p_{n,l}^{(t)} \left| h_{n,l,k}^{(t)} \right|^2}{\sigma_l^2 + \sum_{n' \in \mathbb{N}_l'} p_{n',l}^{(t)} \left| h_{n',l,k}^{(t)} \right|^2}, \tag{6}$$

where $\mathbb{N}_l' = \{n' \in \mathbb{N} \mid |h_{n',l,k}^{(t)}| < |h_{n,l,k}^{(t)}|\}$ represents a set of interfering VT users.

Considering the QoS requirements of VT users, VR_l can successfully decode the information delivered by VT_n through subchannel SC_k which also needs to satisfy the transmission rate $R_{n,l,k}^{(t)}$ not below the rate threshold, i.e., $R_{n,l,k}^{(t)} \geq R_{\min}$. Otherwise, VR_l will not be able to decode the information. We assume that the transmission rate $R_{n,l,k}^{(t)} = 0$ in this case. Then, the data rate of SC_k between VT_n and VR_l can be expressed as

$$R_{n,l,k}^{(t)} = \begin{cases} W \cdot \log_2\left(1 + \widetilde{\Gamma}_{n,l,k}^{(t)}\right), & R_{n,l,k}^{(t)} \geq R_{\min}, \\ 0, & \text{otherwise}. \end{cases} \tag{7}$$

Therefore, the total rate of the NOMA-enabled IoV system in time slot t can be expressed as

$$R^{(t)} = \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{n \in \mathbb{N}_l} R_{n,l,k}^{(t)}, \tag{8}$$

where L is the sum of VR users in time slot t.

SIC technique in NOMA-enabled IoV system has been investigated in [11]. At the VR side, as the maximum number of VT users who are multiplexing the same subchannel increases, the difficulty of SIC technology increases dramatically. To avoid excessive SIC complexity for VR users, in this paper, we assume that each VT user delivers information to at most one VR user during each slot. In addition, it also reduces transmission errors.

3.2. Problem Formulation. In NOMA-enabled IoV system, data transmission rate and system power consumption are both important parameters to measure system performance. Our goal is to minimize the overall power consumption of all VT users while maintaining the system transmission rate, i.e., transmitting more bits per unit Joule. Therefore, we set the optimization objective as the ratio of the overall transmission rate to the total transmit power of VT users, i.e., EE, which can be expressed as

$$\mathrm{EE}^{(t)} = \frac{R^{(t)}}{P_{\mathrm{sum}}^{(t)} + P_c}, \tag{9}$$

where $P_{\text{sum}}^{(t)} = \sum_{k=1}^{K} \sum_{n=1}^{N} p_{n,k}^{(t)}$ denotes the sum transmitted power for all VT users in time slot t and P_c is additional circuit power consumption.

Thus, the optimization problem can be expressed mathematically as

$$\begin{aligned}
&\max_{\left\{\alpha_{n,k}^{(t)}, p_{n,l,k}^{(t)}\right\}} \text{EE}^{(t)} \\
\text{s.t.} \quad &\text{C1}: \quad \sum_{k=1}^{K} \left(\alpha_{n,k}^{(t)} + \alpha_{n',k}^{(t)}\right) \le 1 \\
&\left\{n, n'\right\} \in \left\{1 \le n, n' \le N \mid d_{n,n'}^{(t)} < d_{\max}\right\} \\
&\text{C2}: \quad \sum_{n=1}^{N} \alpha_{n,k}^{(t)} \le U_{\max}, \quad \forall k \in \mathscr{SC} \\
&\text{C3}: \quad \alpha_{n,k}^{(t)} \in \{0,1\}, \quad \forall n \in \mathscr{N}, \forall \text{SC}_k \in \mathscr{SC} \\
&\text{C4}: \quad \left|\mathbb{L}_n^{(t)}\right| \equiv 1, \quad \forall n \in \mathscr{N} \\
&\text{C5}: \quad 0 \le \sum_{k=1}^{K} p_{n,k}^{(t)} \le P_{\max}^{VT}, \quad \forall n \in \mathscr{N}.
\end{aligned} \tag{10}$$

Constraint C1 indicates that two vehicles within the communication range cannot pass messages to each other, i.e., VT_n cannot pass messages to $\text{VT}_{n'}$ within its communication range. This is because of the half-duplex nature that no two vehicles can receive a message at the same time as it is passed, according to [21]. To reduce the SIC complexity at the receiver side, we assume that each subchannel SC_k is multiplexed by at most $U_{\max}$ VT users and that each VT user delivers information to at most one VR user within its communication range during slot t, which are reflected in constraints C2, C3, and C4. Constraint C5 limits the threshold of transmit power for VT users.

4. DRL-Based Subchannel and Power Allocation Algorithms

The optimization problem in (10) is nonconvex and NP hard, which has a complex system with high computational dimensionality. The problem requires exponential levels of time complexity for direct computation of all possible subchannel selections and power allocations, which is difficult to implement in practice. Therefore, we use reinforcement learning methods to select the subchannel selection and power allocation strategies of maximizing EE. We first transform the resource allocation problem in NOMA-enabled IoV system into an MDP-based resource allocation problem and then solve the model using DRL methods.

4.1. Optimize Problem Conversion. In the proposed NOMA-enabled IoV system, the system state in each time slot $t+1$ depends only on the actions, including subchannel selection and power allocation, made by the VT users in time slot t. Therefore, we transform the developed model for maximizing EE into a resource allocation model based on MDP and then solve it through the DRL method. The state space **S**, action space **A**, and reward **R** of the MDP model are defined below, respectively.

4.1.1. State Space. The system state information can be described jointly by the system data transmission rate and the energy consumption. Thus, the system state space **S** includes the transmission rates between all VT users and the corresponding VR users, as well as the transmission power of all VT users, and this information is the basis for this resource allocation. Since we assume that each VT user transmits information to only one VR user, during time slot t, the state $s_t \in \mathbf{S}$ can be expressed as follows:

$$s_t = \left\{R_1^{(t)}, R_2^{(t)}, R_3^{(t)}, \cdots, R_N^{(t)}, p_1^{(t)}, p_2^{(t)}, p_3^{(t)}, \cdots, p_N^{(t)}\right\}. \tag{11}$$

4.1.2. Action Space. The action space **A** includes all possible subchannel choices for each VT user, $\alpha_{n,k}^{(t)}$, as well as the choice of transmit power, $p_{n,k}^{(t)}$. In time slot t, action a_t can be expressed as

$$\begin{aligned}
a_t &= \left\{a_t^1, a_t^2\right\}, \\
a_t^1 &= \left\{\alpha_{1,1}^{(t)}, \cdots, \alpha_{n,k}^{(t)}, \cdots, \alpha_{N,K}^{(t)}\right\}, \\
a_t^2 &= \left\{p_{1,1}^{(t)}, \cdots, p_{n,k}^{(t)}, \cdots, p_{N,K}^{(t)}\right\}.
\end{aligned} \tag{12}$$

4.1.3. Reward. We denote the reward for selecting the action a_t under state s_t as EE of the current system, which can be calculated by Equation (9). Specifically, for $r_t \in \mathbf{R}$, it can be expressed as

$$r_t = \text{EE}^{(t)}. \tag{13}$$

The goal of reinforcement learning is to find the optimal policy π^* through multiple iterations to achieve the maximum long-term discounted reward

$$R_t = \mathbb{E}\left[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots\right] = \mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i r_{t+i}\right], \tag{14}$$

where $\gamma \in [0, 1)$ is the discount factor. When γ is equal to 0, only the current reward has been considered, while the subsequent has been ignored. As γ increases, the system will focus more with long-term discount rewards.

The reward function can be set to satisfy the requirement of receiving a higher reward when the agent chooses to perform an action that makes the system EE larger and otherwise receives a lower reward or even receives zero reward. After several rounds of iterations, the agent will gradually choose the policy that can obtain higher rewards, i.e., a better resource allocation policy.

4.2. Event Trigger. The framework of the proposed DSPA algorithm is shown in Figure 2. During the process of interacting with the environment, the agent selects and executes an action a_t based on the environment's current state s_t,

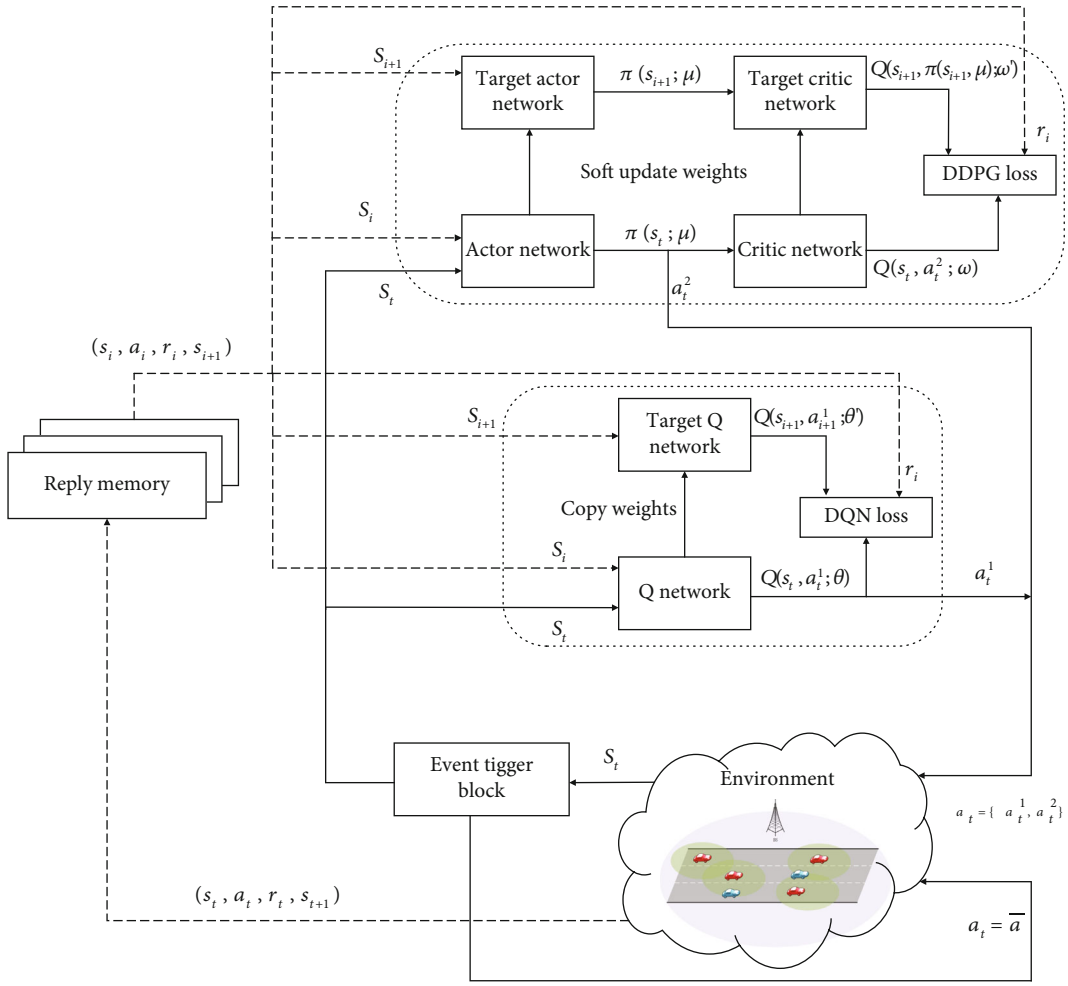

Figure 2: DSPA algorithm framework.

after which the state s_t becomes state s_{t+1}, and the agent gets a reward r_t given by the environment. Then, the agent executes a new action a_{t+1} according to a certain policy π based on the new state and the reward. After a long iterative process, the agent will get an optimal policy π^* that earns the most reward.

Policy π is a mapping of the state space $\mathbf{S}$ to the action space $\mathbf{A}$. Specifically, $\pi = \mathbf{S} \longrightarrow \mathbf{A}$. Considering the state-action value function of the action $Q : \mathbf{S} \times \mathbf{Q} \longrightarrow R$ that represents the expected reward for performing action a with policy π in state s, i.e.,

$$\begin{aligned} Q^{\pi}(s,a) &= \mathbb{E}_{\pi}\left\{\sum_{i=0}\gamma^{i} r_{t+i} \mid s_t = s, a_t = a\right\} \\ &= \mathbb{E}_{\pi}\left\{r_t + \gamma_{t+1} + \gamma^2 r_{t+2} + \cdots \mid s_t = s, a_t = a\right\} \\ &= \mathbb{E}_{\pi}\left\{r_t + \gamma Q^{\pi}(s_{t+1} + a_{t+1}) \mid s_t = s, a_t = a\right\} \\ &= \sum_{s'} P\left(s,a,s'\right)\left(R\left(s,a,s'\right) + \gamma Q^{\pi}\left(s',a'\right)\right). \end{aligned} \tag{15}$$

For the established MDP model, the ultimate goal is to find an optimal policy π^* that can be satisfied as $Q^{\pi^*} \geq Q^{\pi}$ for all policy π. The optimal action-value function can be expressed as

$$Q^*(s,a) = \sum_{s'} P\left(s,a,s'\right)\left(R\left(s,a,s'\right) + \gamma \max_{a'} Q^*\left(s',a'\right)\right). \tag{16}$$

Equation (16) is the Bellman equation, which indicates that when the agent makes an optimal decision, the obtained Q value must be the expected reward for the optimal action in that state.

For the MDP model, the schemes to obtain the optimal policy π^* mainly include model-based approaches and model-free approaches. Since a part of the prior knowledge, such as transfer probability, is unknown in the NOMA-enabled IoV system, it is necessary to use the model-free approach RL to obtain statistical information of the unknown model. DRL combines RL with deep neural networks (DNN) and solves high-dimensional state and action space problems by DNN, which is widely used in IoV systems.

However, solving the MDP model using the DRL method is still time costly, as it takes more time to update the neural network weight parameters, generate the actions,

and calculate the rewards. Several methods have been proposed for reducing the computation time. In [23], the authors propose an event trigger module, which is a controller that updates the neural network parameters only when the system state deviates from a certain level. Such method can effectively reduce the computation time, so we introduce it into our DSPA algorithm.

In NOMA-enabled IoV systems, there may be two adjacent time slots in which the system states are similar or even identical, and then, the action selection corresponding to these two states should also be the same. So when the DNN outputs the action in the first time slot, the same action in the next time slot can be executed directly without the DNN. Referring to Lemma 1 in [24], we give a proof for this consideration.

Theorem 1. *For two consecutive states s_t and s_{t+1}, their corresponding optimal actions a_t and a_{t+1} should be the same when $s_t = s_{t+1}$.*

Proof. According to Equation (16), after obtaining the optimal state-action value function $Q^*(s, a)$ for all states, by using the greedy strategy, the optimal actions a_t and a_{t+1} corresponding to states s_t and s_{t+1} can be expressed as

$$\begin{aligned} a_t^* &= \arg \max_{a \in \mathbf{A}} Q^*(s_t, a), \\ a_{t+1}^* &= \arg \max_{a \in \mathbf{A}} Q^*(s_{t+1}, a), \end{aligned} \tag{17}$$

where $\mathbf{A}$ represents the action space of two actions. Assuming that $s_t = s_{t+1}$, we can obtain

$$a_t^* = \arg \max_{a \in \mathbf{A}} Q^*(s_t, a) = \arg \max_{a \in \mathbf{A}} Q^*(s_{t+1}, a) = a_{t+1}^*, \tag{18}$$

which proves our assumption. □

Based on the above assumption, we add the event trigger module into the DRL framework as a way to decide whether to output new actions by using the neural network. Specifically, the previous state $\bar{s}$ and the corresponding action $\bar{a}$ are stored in the event trigger. The new state s_t is firstly compared with $\bar{s}$; if the difference between the two is less than a certain threshold, $\bar{a}$ is directly output as the action of state s_t. Otherwise, the DNN outputs the action a according to state s_t, and $\bar{s}$ and $\bar{a}$ are replaced with s_t and a_t. Using the binary variable ζ as the event trigger decision, specifically,

$$\zeta = \begin{cases} 1, & \|s_t - \bar{s}\|^2 \geq \rho, \\ 0, & \text{otherwise}, \end{cases} \tag{19}$$

where ρ is the threshold, $\zeta = 1$ means outputting action a_t through the neural network, and $\zeta = 0$ means obtaining action $\bar{a}$ stored in the event trigger.

4.3. DRL-Based Resource Allocation Framework. In the proposed DSPA algorithm, the subchannel selection action, i.e., a_t^1 in Equation (12), is obtained by the DQN method. Since the transmission power is a continuous interval, we use the DDPG method for power allocation, i.e., a_t^2 in Equation (12).

4.3.1. DQN-Based Subchannel Selection Method. In the DQN algorithm, the Q function is approximated by DNN and the Q value is approximated by the DNN weight parameter θ. The Q value is updated by minimizing the loss function to update the parameter θ; the loss function can be defined as

$$L(\theta) = \mathbb{E}\left[(\text{Target}Q - Q(s, a, \theta))^2\right], \tag{20}$$

where

$$\text{Target}Q = r(s, a) + \gamma \max_{a'} Q\left(s', a', \theta'\right). \tag{21}$$

According to Equations (20) and (21), the gradient descent method can be used to solve for the weight parameter θ. DQN uses the current network to evaluate the current value function and uses the target network to generate the target value in Equation (21). The combination of these two networks can decouple the current Q value and the target Q value to some extent, which in turn improves the stability of the algorithm.

The DQN algorithm further introduces an experience replay mechanism to solve the problem of high sample coupling. At each step, the data of the intelligent body interacting with the environment, i.e., the current state s, action a, reward r, and next state s', are stored in the experience pool. The data can later be drawn from the experience pool for training.

The introduction of the experience replay mechanism makes it easier to store the feedback data and allows training samples to be drawn by random sampling, reducing the high coupling between samples. Furthermore, this mechanism can also solve the problems of nonindependent correlation and nonstationary distribution among data in reinforcement learning, which reduces the convergence difficulty of the network model.

4.3.2. DDPG-Based Power Allocation Method. The DQN method is able to solve large-scale state space problems, but its limitation is that it can only solve discrete action space problems, so it is not feasible to use the DQN method to make choices in continuous power intervals. For this case, we use the DDPG method for power allocation. DDPG is a DRL method based on value function and policy gradient, which can effectively solve the problem of high-dimensional and continuous action space. The method generates a deterministic action directly through a DNN network named actor, i.e.,

$$\mu(s_t \mid \omega^\mu) \approx \mu^*(s_t), \tag{22}$$

where $\mu^*(s_t)$ is the optimal behavior policy and ω^μ is the parameter of actor network. The resulting actions are then evaluated by a DNN network called critic, with the aim of minimizing the loss function. The loss function is

$$L(\omega^Q) = \mathbb{E}\left[\left(Q(s_t, a_t \mid \omega^Q) - y_t\right)^2\right], \tag{23}$$

where

$$y_t = r_t + \gamma Q'\left(s_{t+1}, \mu'\left(s_{t+1} \mid \omega^{\mu'}\right) \mid \omega^{Q'}\right). \tag{24}$$

Similar to DQN, two independent target networks, namely, the target actor network and the target critic network, are introduced to further improve the stability of learning. The parameters of the target network are related to the current network and updated in real time, with the update criterion

$$\begin{aligned} \omega^{Q'} &= \delta\omega^Q + (1-\delta)\omega^{Q'}, \\ \omega^{\mu'} &= \delta\omega^\mu + (1-\delta)\omega^{\mu'}, \end{aligned} \tag{25}$$

where $\delta \ll 1$ is used to limit the change rate of the target value and improve the stability of DNN training.

Based on the above theory, the DSPA algorithm in the NOMA-enabled IoV system is shown in the algorithm.

5. Simulation Experiments and Analysis

5.1. Simulation Environment. In this section, we conduct simulation experiments on the proposed resource allocation scheme and analyze the results. The simulation experiments are conducted on Windows 10 operating platform with Intel i5-8300H CPU, NVIDIA 1050Ti GPU, and 16 G memory size and based on Python 3.7 and use the TensorFlow 1.13 framework. All networks contain two hidden layers with 128 and 64 neurons, respectively. Following the 3GPP standard and existing studies, we set the parameters to meet the simulation requirements of the NOMA-enabled IoV system, as shown in Table 1.

5.2. Parameter Analysis

5.2.1. Learning Rate. In the DSPA algorithm, the learning rate is an extremely important hyperparameter. Generally speaking, the larger learning rate, the faster convergence speed, but will ignore the optimal solution due to premature convergence, and the convergence value is normally lower than the global optimal value. As the learning rate approaches zero, the speed of obtaining the optimal policy π^* decreases gradually and could not obtain the optimal solution quickly. This is because the learning rate controls the size of the optimization gradient step, too large learning rate will lead to too large gradient step, ignoring the optimal solution, while too small learning rate will lead to too small step, requiring more time to converge. Therefore, it is first necessary to choose a suitable learning rate.

We set the values of learning rate as 0.1, 0.01, and 0.001, respectively. The simulation results are shown in Figure 3. When the learning rate is 0.1, the algorithm obtains the maximum EE of 2.8 Mbit/Joule after about 400 iterations. The EE after convergence is not much different between learning rates 0.01 and 0.001, both of which are about 3.2 Mbit/Joule. However, the optimal value is obtained after 500 iterations with the learning rate of 0.01, while the learning rate of 0.001 requires 700 rounds of iterations. In order to take into account the convergence speed and quality, we set the learning rate to 0.01 in the following simulation.

5.2.2. Discount Factor. Figure 4 shows the impact of different discount factors on the convergence of the system EE. We set the values of the discount factor γ as 0.1, 0.5, and 0.9, respectively. As the number of iterations increase, the system EE gradually leveled off. The system EE for each of the three discount factors is maximized after about 500 iterations, when the EE is 3.0 Mbit/Joule, 3.1 Mbit/Joule, and 3.2 Mbit/Joule, respectively. The comparison leads to the conclusion that the smaller γ, the more system focuses on the current reward, and the larger the γ, the more system focuses on the long-term reward. Our goal is to maximize the long-term discounted rewards of the system, so we choose $\gamma = 0.9$ for the following simulation.

5.2.3. Transmission Rate Thresholds. We compare the effect of different transmission rate thresholds $R_{\min}$ on the system EE, as shown in Figure 5. According to Equation (7), when the transmission rate $R_{n,l,k}^{(t)} < R_{\min}$, VR_l cannot successfully decode the information from VT_n, and we set $R_{n,l,k}^{(t)} = 0$ in this case. That is, $P_{n,l,k}^{(t)} > 0$ but $R_{n,l,k}^{(t)} = 0$, which will seriously affect the system EE. We set $R_{\min}$ as 0 Mbps, 0.1 Mbps, 0.5 Mbps, and 1 Mbps, respectively. Simulation results show that the system EE is maximum when $R_{\min} = 0$. In this case, all messages are decoded successfully as valid messages. However, this setting is not reasonable considering the QoS demand of VT users. The increase of $R_{\min}$ indicates that the QoS demand of VT users becomes more strict, and more messages are discarded as invalid messages because they cannot meet the QoS requirement; the system EE gradually decreases as a result. In the following simulations, we choose $R_{\min} = 0.1$ Mbps because the QoS demand of most VT users can be satisfied.

5.3. Comparison Experiments

5.3.1. Comparison on SIC Technology. We compare the EE of the NOMA-enabled IoV system with SIC technology, the NOMA-enabled IoV system without SIC technology, and the OMA IoV system with different vehicles, as shown in Figure 6. It can be seen that when the system contains only 10 vehicles, whether to use SIC technology has less impact on the system EE, while OMA system has the lowest EE. This is because when there are fewer vehicles, the probability of two VT users occupying the same subchannel is lower and only a small amount of interference is generated at the receiving end. The increase of the total number of vehicles means that there are more VT users that need to transmit

1: Initialize the Q network weight parameters θ
2: Initialize the actor and critic network weight parameters ω^Q and ω^μ
3: Initialize the weight parameters of the target network $\theta' \longleftarrow \theta$, target actor network $\omega^{Q'} \longleftarrow \omega^Q$, and target critic network $\omega^\mu \longleftarrow \omega^{\mu'}$
4: Initialize replay memory $\mathscr{D}$ and event trigger block $\bar{s}, \bar{a}$
5: **for** episode = 1, **M do**
6: Initialize random noise ϱ_t
7: Initialize the state of the NOMA-enabled IoV system s_1
8: **for** $t = 1$, T **do**
9: Calculate the difference between $\bar{s}$ and s_t according to Equation (19)
10: **if** $\zeta = 1$ then
11: Select action a_t^1 according to the DQN method
12: Select action a_t^2 according to the DDPG method
13: Replace $\bar{s}$ and $\bar{a}$ in the event trigger with s_t and $a_t = \{a_t^1, a_t^2\}$
14: **else**
15: Output the action $a_t = \bar{a}$
16: **end if**
17: Perform a_t, get reward r_t and new state s_{t+1}
18: Store sample (s_t, a_t, r_t, s_{t+1}) into replay memory $\mathscr{D}$
19: Sampling samples (s_i, a_i, r_i, s_{i+1}) from replay memory $\mathscr{D}$
20: Update the Q network, actor network, and critic network weight parameters θ, ω^Q, and ω^μ
21: Update the target network, target actor network, and target critic network weight parameters $\theta' \longleftarrow \theta$, $\omega^{Q'} \longleftarrow \omega^Q$, and $\omega^\mu \longleftarrow \omega^{\mu'}$
22: **end for**
23: **end for**

Algorithm 1: DRL-based resource allocation algorithm.

Table 1: Parameter setting in simulation.

Parameter	Value
K	5
M	10 MHz
τ	1 ms
D	500 m
$d_{\max}$	50 m
$P_{\max}^{VT}$	23 dBm
$v_{\max}$	36 km/h
σ^2	-114 dBm
Selectable power levels	10
Pathloss model	LOS in WINNER +B1
Fast fading	Rayleigh fading

information; under the condition of a certain number of subchannels, the EE of all three approaches gradually decreases, the EE of the system with SIC technology is always the highest, and the EE of the system without SIC technology is gradually lower than the EE of the OMA approach. The reason is that NOMA actively introduces interference, the large number of VT users multiplexing the same subchannel, the stronger interference received of VR user, and not using SIC technology can lead to disastrous results.

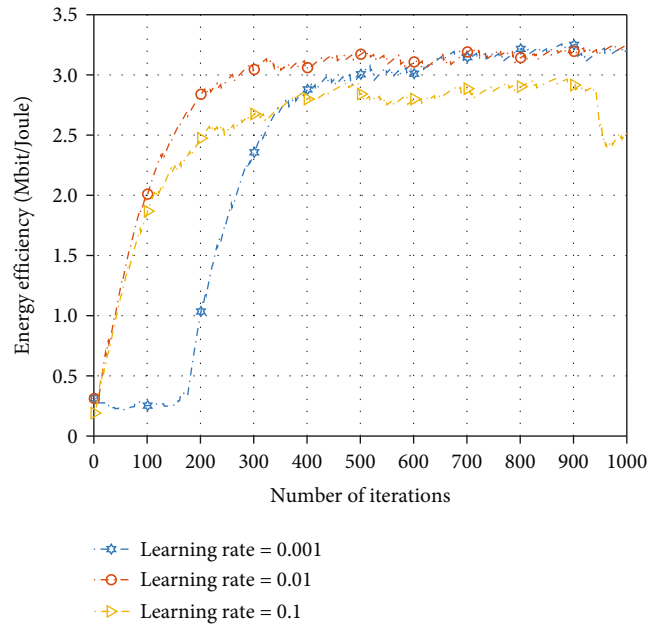

Figure 3: Impact of learning rate.

5.3.2. Comparison on Event Trigger Block. Next, we analyze the event trigger block by comparing the impact of the event trigger block on the system EE. The threshold ρ of the event trigger module is set to 0.1, and the results are shown in Figure 7. In a variety of different situations, the event trigger block has little impact on the system EE.

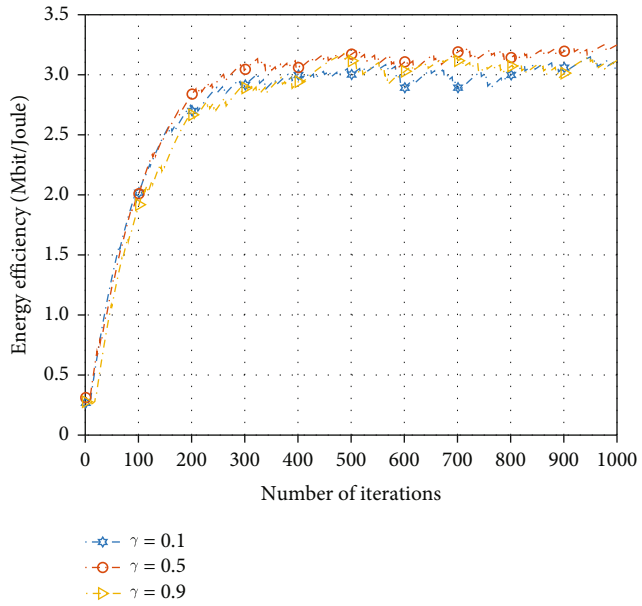

Figure 4: Impact of discount factor γ.

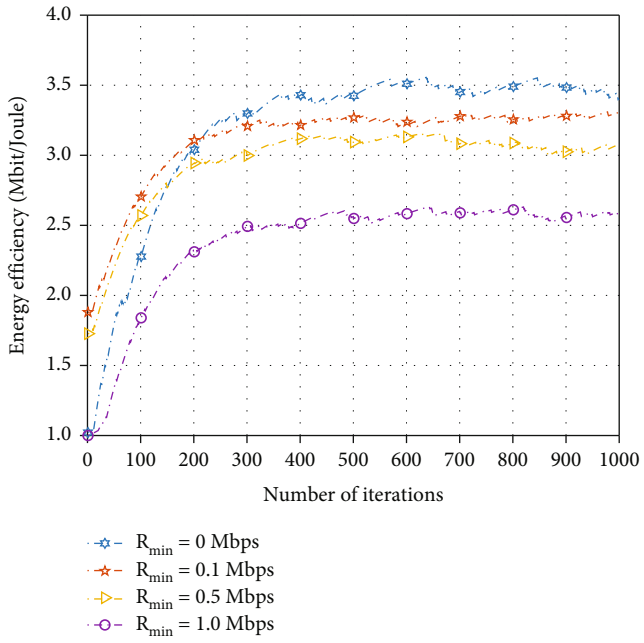

Figure 5: Impact of transmission rate threshold.

Figure 8 reflects the average computation time for the three comparisons. As can be seen from the figure, the average computation time per execution increases as the number of vehicles increases, and the event trigger block effectively reduces the computation time. Such result shows that although the event trigger block costs extra time to compute the environment similarity, it can reduce some unnecessary neural network computations, which take more time.

We further compared the impact of event trigger thresholds ρ on the system EE, and the results are shown in Figure 9. It can be seen that when the threshold ρ is equal

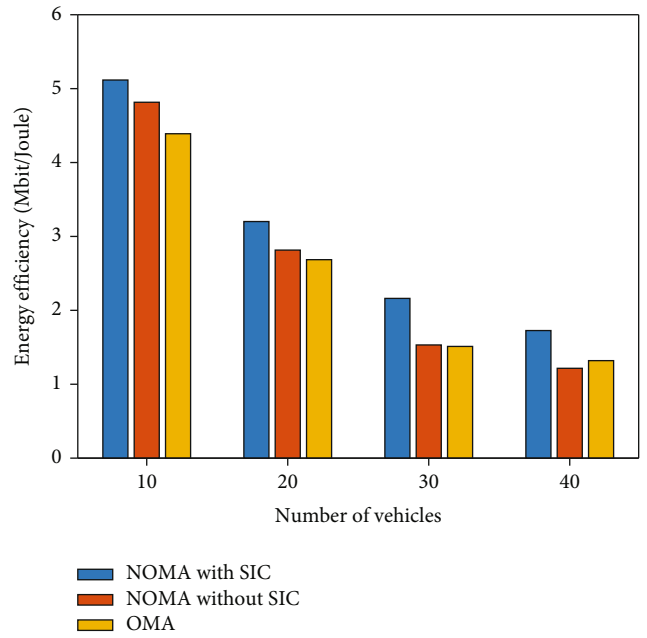

Figure 6: The comparison of different access methods.

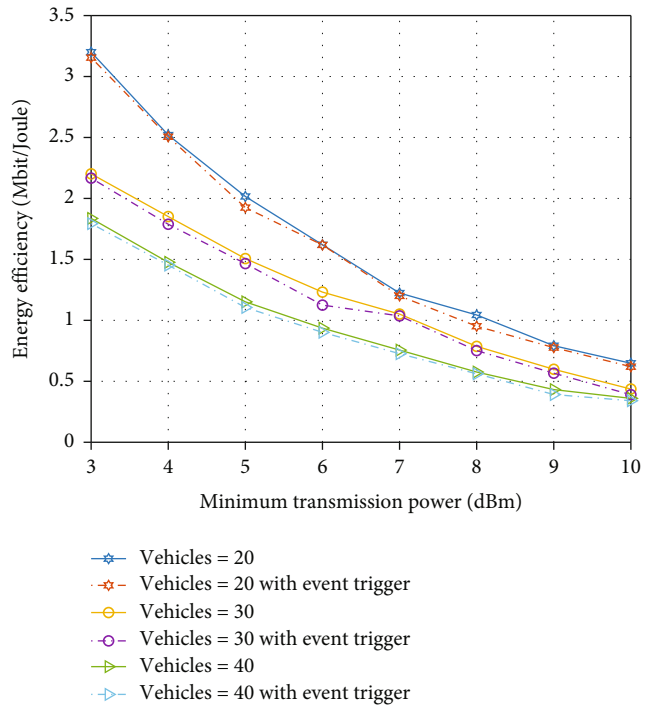

Figure 7: Impact of the event trigger.

to 0.1, it only slightly decreases the system EE. Combining Figures 7–9, choosing an appropriate threshold ρ can reduce the computation time of the DSPA algorithm with a slight reduction in system performance.

5.3.3. Comparison with Other Algorithms. Finally, we compare the system EE of the DSPA, DQN, and random method under different numbers of vehicles, and the results are

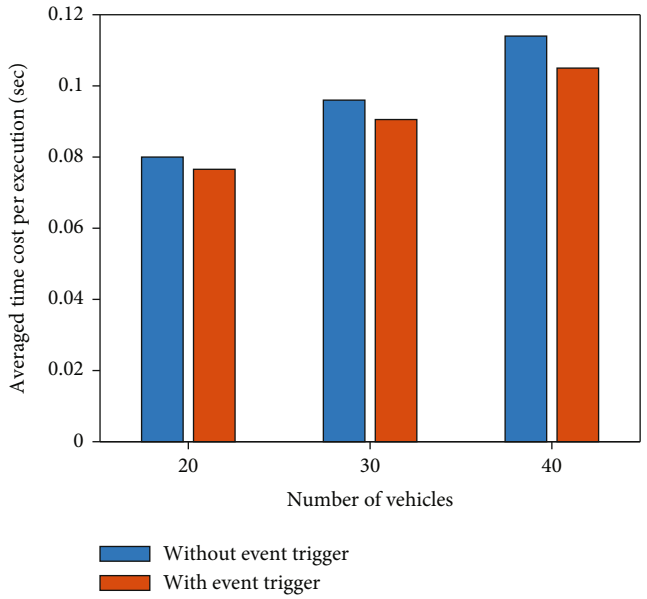

Figure 8: The comparison of averaged time cost.

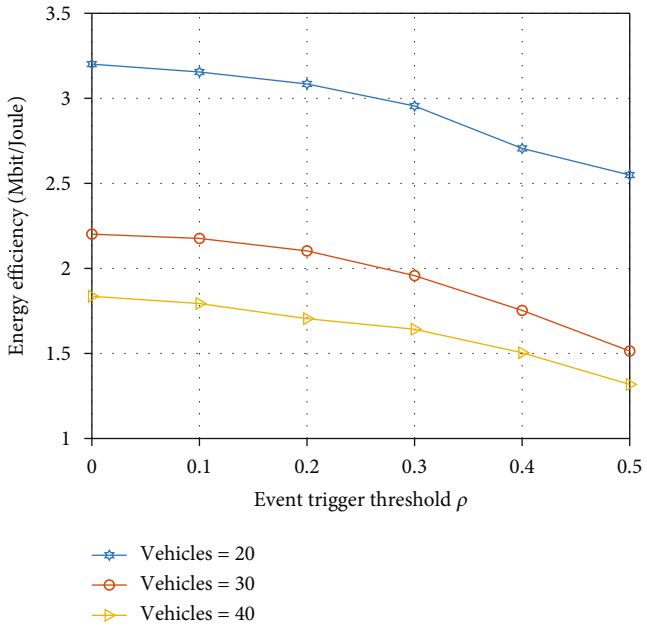

Figure 9: The comparison of different threshold ρ.

shown in Figure 10. In the DQN method, we discrete the transmission power uniformly into 10 levels to meet the demand of DQN for discrete action space. The random algorithm indicates that the VT user randomly selects the channel and transmit power each time. As shown in Figure 10, we can see that the system EE decreases for all three algorithms. For both the DSPA algorithm and the DQN algorithm, the system EE decreases faster when the number of vehicles first increases and then gradually decreases. The reason is that, when the system interference is low, adding vehicles causes a significant change in system

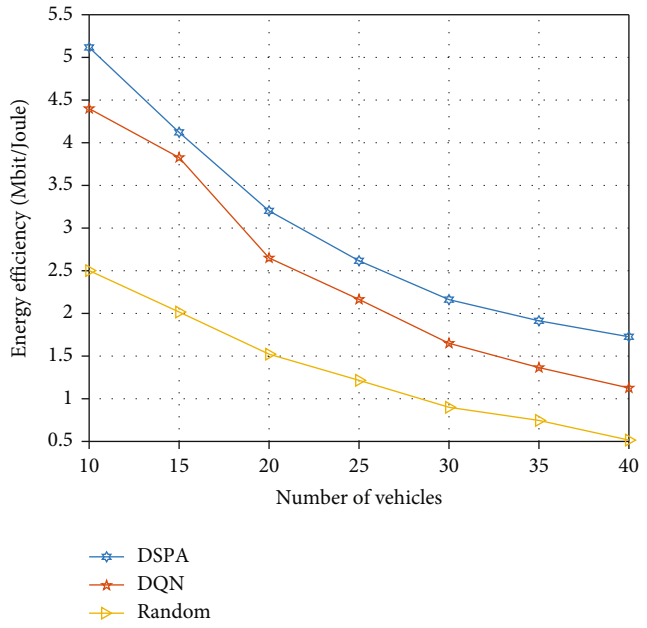

Figure 10: The comparison of different algorithms under different vehicle numbers.

interference; with the gradual increase of vehicles, the change of system interference gradually flattens out. The system EE of the DQN algorithm is lower than that of our proposed DSPA framework because in the DSPA algorithm we use the DDPG method to select among continuous power intervals, while in the DQN algorithm we can only select among discrete 10 power levels. We believe that the performance of the DQN algorithm will be improved if the power selection levels in the DQN algorithm are increased. However, this would increase the action dimension of the DQN algorithm and take a lot of time. The system EE using the random algorithm is always the lowest due to the random selection of subchannels and transmission power at each step, which can produce catastrophic results.

6. Conclusion

In this paper, we study the NOMA-enabled resource allocation problem in IoV system. Firstly, we have maximized the system EE by allocating channel resources and power resources for VT users to reduce transmission power consumption on the basis of guaranteed system transmission rate. Secondly, we have transformed the resource allocation problem of maximizing EE into an MDP model. Finally, we designed a DSPA algorithm to obtain the subchannel selection and power allocation strategies for maximizing system EE and used the event trigger block to reduce the computation time. Simulation results show that the NOMA-enabled IoV system outperforms the OMA system, and the proposed resource allocation scheme can significantly improve the system EE compared to other schemes and reduce the computation time. In future work, we will study other NOMA-enabled resource allocation strategies and consider the introduction of mobile edge computing in IoV.

Data Availability

Data is available on request from the corresponding authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is partly supported by the National Natural Science Foundation of China (No. 61902029 and No. 61872044), the Excellent Talents Projects of Beijing (No. 9111923401), and the Scientific Research Project of Beijing Municipal Education Commission (No. KM202011232015).

References

[1] L. Qiao, "Mobile data traffic offloading through opportunistic vehicular communications," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 3093581, 12 pages, 2020.

[2] J. Huang, C. Zhang, and J. Zhang, "A multi-queue approach of energy efficient task scheduling for sensor hubs," *Chinese Journal of Electronics*, vol. 29, no. 2, pp. 242–247, 2020.

[3] H. Gao, C. Liu, Y. Li, and X. Yang, "V2VR: reliable hybrid-network-oriented V2V data transmission and routing considering RSUs and connectivity probability," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, pp. 3533–3546, 2020.

[4] Y. Chen, N. Zhang, Y. Zhang, X. Chen, W. Wu, and X. S. Shen, "Energy efficient dynamic offloading in mobile edge computing for Internet of Things," *IEEE Transactions on Cloud Computing*, vol. 9, no. 3, pp. 1050–1060, 2019.

[5] 3rd Generation Partnership Project: Technical Specification Group Radio Access Network, *Study LTE-Based V2X Services*, (Release 14), Standard 3GPP TR, 2016.

[6] T. T. T. Dao and P. N. Son, "Cancel-decode-encode processing on two-way cooperative NOMA schemes in realistic conditions," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 8828443, 15 pages, 2021.

[7] M. Bello, W. Yu, A. Chorti, and L. Musavian, "Performance analysis of NOMA uplink networks under statistical QoS delay constraints," in *ICC 2020-2020 IEEE international conference on communications (ICC)*, pp. 1–7, Dublin, Ireland, 2020.

[8] Z. Liu, G. Hou, Y. Yuan, K. Y. Chan, K. Ma, and X. Guan, "Robust resource allocation in two-tier NOMA heterogeneous networks toward 5G," *Computer Networks*, vol. 176, p. 107299, 2020.

[9] A. Kiani and N. Ansari, "Edge computing aware NOMA for 5G networks," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1299–1306, 2018.

[10] J. Huang, S. Li, and Y. Chen, "Revenue-optimal task scheduling and resource management for IoT batch jobs in mobile edge computing," *Peer-to-Peer Networking and Applications*, vol. 13, no. 5, pp. 1776–1787, 2020.

[11] Z. Ding, R. Schober, and H. V. Poor, "Unveiling the importance of SIC in NOMA systems—part 1: state of the art and recent findings," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2373–2377, 2020.

[12] X. Chen, X. Liu, Y. Chen, L. Jiao, and G. Min, "Deep Q-network based resource allocation for UAV-assisted ultra-dense networks," *Computer Networks*, vol. 196, article 108249, 2021.

[13] Z. Ma, X. Chen, T. Ma, and Y. Chen, "Deep deterministic policy gradient based resource allocation in Internet of vehicles," in *International Symposium on Parallel Architectures, Algorithms and Programming*, pp. 295–306, Springer, Singapore, 2020.

[14] C. Guo, L. Liang, and G. Y. Li, "Resource allocation for vehicular communications with low latency and high reliability," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 3887–3902, 2019.

[15] Y. Chen, Y. Wang, M. Liu, J. Zhang, and L. Jiao, "Network slicing enabled resource management for service-oriented ultra-reliable and low-latency vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7847–7862, 2020.

[16] L. Liang, H. Ye, and G. Y. Li, "Spectrum sharing in vehicular networks based on multi-agent reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2282–2292, 2019.

[17] H. Yang, K. Zhang, K. Zheng, and Y. Qian, "Joint frame design and resource allocation for ultra-reliable and low-latency vehicular networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3607–3622, 2020.

[18] Y. Chen, N. Zhang, Y. Zhang, X. Chen, W. Wu, and X. S. Shen, "TOFFEE: task offloading and frequency scaling for energy efficiency of mobile devices in mobile edge computing," *IEEE Transactions on Cloud Computing*, 2019.

[19] J. Zhao, Q. Li, Y. Gong, and K. Zhang, "Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7944–7956, 2019.

[20] H. Peng and X. S. Shen, "Deep reinforcement learning based resource management for multi-access edge computing in vehicular networks," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 2416–2428, 2020.

[21] B. Di, L. Song, Y. Li, and G. Y. Li, "NOMA-based low-latency and high-reliable broadcast communications for 5G V2X services," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–6, Singapore, 2017.

[22] G. Liu, Z. Wang, J. Hu, Z. Ding, and P. Fan, "Cooperative NOMA broadcasting/multicasting for low-latency and high-reliability 5G cellular V2X communications," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7828–7838, 2019.

[23] X. Yang, H. He, and D. Liu, "Event-triggered optimal neuro-controller design with reinforcement learning for unknown nonlinear systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 9, pp. 1866–1878, 2017.

[24] X. Wang, Y. Zhang, R. Shen, Y. Xu, and F. C. Zheng, "DRL-based energy-efficient resource allocation frameworks for uplink NOMA systems," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7279–7294, 2020.