Chinese Journal of Electronics
Vol.32, No.6, Nov. 2023

PLC for In-Vehicle Network: A DRL-Based Algorithm of Diversity Combination of OFDM Subcarriers

CHEN Zhixiong[1,3], ZHANG Zhikun[1], CAO Tianshu[1], and ZHOU Zhenyu[2]

(1. *School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China*)

(2. *School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China*)

(3. *Hebei Key Laboratory of Power Internet of Things Technology, Baoding 071003, China*)

Abstract — For low latency communication service of vehicles, it is critical to improve the delay performance of power line communication (PLC) for in-vehicle network, which can decrease the weight and cost of the vehicle. In order to minimize the total time slots used in a transmission task, an orthogonal frequency-division multiplexing (OFDM) subcarrier diversity combination algorithm of PLC based on the deep reinforcement learning (DRL) is proposed herein. The short packet communication theory is used to develop an optimal combination model with constraints on short packet reliability, transmitting power and the amount of data. The state, action, and reward function of double deep Q-learning network (DDQN) are defined, and diversity combination for OFDM subcarriers is performed using DDQN. An adaptive power allocation algorithm based on the thresholds of error rate and the data amount is used. Simulation results show that the proposed algorithm can effectively improve the delay performance of PLC under the constraints of power and data amount.

Key words — Power line communication, OFDM, Low latency, Diversity, Deep reinforcement learning.

I. Introduction

The 6G integrated sensing and communication technology for the internet of vehicles (IoV) not only has the function of high-precision positioning and sensing, but also can monitor the internal operating conditions of the vehicle in real time to ensure good operation of the vehicle [1]. Therefore, it has attracted extensive attention from academia and industry [2], [3]. Existing research mainly focuses on information interaction such as vehicle to infrastructure (V2I) [4], [5], vehicle to vehicle (V2V) [6], and vehicle to pedestrian (V2P) [7]. The information collected by a large number of sensors in-vehicle needs to be transmitted or aggregated to the gateway via the in-vehicle network, and then be processed or forwarded by the gateway. With the increasing number of sensors and applications, to enhance the performance of in-vehicle network is also crucial for the development of IoV.

Wireless communication has the characteristics of flexible access and strong scalability and can be used for vehicle multimedia service transmission, but it needs to solve the adverse effects such as impulse noise and interference. For orthogonal frequency-division multiplexing (OFDM) systems, the joint sparse theory and machine learning algorithm framework can be used to estimate and eliminate impulse noise to improve communication reliability [8]. In addition, due to the limited space in the vehicle, the wireless sensor network would generate strong mutual interference. A large number of dedicated communication lines, such as twisted pair and coaxial cable are used for transmission in-vehicle network. With the increasing number of 6G sensors and applications, dedicated communication cables will greatly increase the weight and cost of the vehicle. As shown in Fig.1, the use of power lines in-vehicle for data transmission [9] does not require additional communication wiring, which can effectively reduce weight and cost, and has important research value.

Considering the complex communication environment and infrastructure, the delay, throughput and reliability requirements of different services in the IoV are also diverse. Compared with the traditional channel-ori-

Manuscript Received Sept. 29, 2022; Accepted Apr. 25, 2023. This work was supported by the National Natural Science Foundation of China (61601182) and the Fundamental Research Funds for the Central Universities (2021MS070).

 DOI:10.23919/cje.2022.00.331

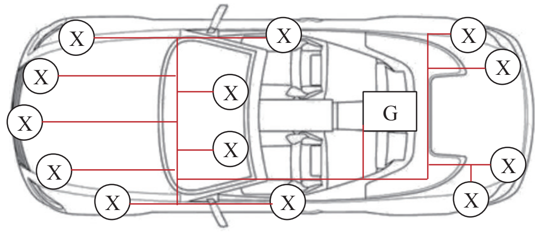


Fig. 1. Schematic diagram of in-vehicle data transmission based on PLC.

ented design ideas, the data-oriented design idea will be more in line with the application requirements of the IoV. Focusing on the data-oriented resource optimization of low latency, Yang *et al.*[10] proposed the delay outage rate (DOR) as an evaluation metric and discussed how optimal rate adaptation (ORA), optimal power and rate adaptation (OPRA) and diversity technologies improve performance. The authors of [11] deduced the DOR for technologies such as adaptive modulation coding and selective diversity, among others, for low latency transmission of small data over fading channels. It investigated the ORA's performance in conjunction with power allocation (ORA-PA). The above literature only analyzes and optimizes the delay performance in wireless communication and proposes theoretical innovative ideas and achievements, rather than provide a specific implementation scheme. Considering that the power line channel in-vehicle is easily affected by factors such as fading and impulse noise, it is challenging to realize the power line transmission with high reliability and low latency. Therefore, the study of how to improve the delay performance of power line communication (PLC) to meet the transmission requirements of low latency services of in-vehicle network has great practical significance.

Diversity technology and massive multiple input multiple output (MIMO) technology are key technologies for implementing ultra-reliable low-latency communication services. In wireless communication systems, massive MIMO technology can be used to achieve data diversity transmission, improving reliability without reducing spectrum utilization, and diversity technology includes spatial diversity, time diversity, frequency diversity, etc. However, PLC is a kind of wired communication, the power line channel can only provide generally 2×2 communication links [12], and the resources are limited, so it cannot use massive MIMO technology like wireless communication. In addition, if the time diversity technology is adopted, the transmission interval of the same symbol should be larger than the coherence time in order to ensure the diversity effect, which will increase the transmission delay. In order to improve reliability on the basis of guaranteeing transmission delay, frequency diversity becomes the primary choice of PLC. In [13], the means of time diversity and frequency diversity are used to improve the ability of PLC to resist periodic impulse noise and narrowband interference to reduce a certain interleaving delay. Iraqi *et al.* [14] proposed a subcarrier diversity-based OFDM smart algorithm. It may be possible to reduce bit error rates while maintaining spectral efficiency and rate by utilizing transmission diversity. Diversity technology is one of the key technologies to low latency. However, the previous studies has not discussed the optimal diversity combination method in detail. For low latency services, the combination should consider the constraints or trade-offs associated with performance indicators such as reliability and delay. As a result, relevant studies pose difficulties and provide theoretical value.

Deep reinforcement learning (DRL) can be used to achieve fast adaptive strategy selection and optimization, and is widely used in communication resource allocation under multi-objective and multi-constraint conditions. Reference [15] proposed an end-edge-cloud collaborative computing framework for 5G-enabled IoV, and designed a distributed service offloading method based on deep learning and DRL. Reference [16] proposed a joint optimization framework for spectrum allocation and power control based on federated DRL. This framework can improve the total transmission rates of all the V2I users under the high-speed movement of the vehicle, and can also guarantee the robustness of the network when new vehicles are connected to the network. Aiming at the bottleneck of spectrum resource shortage and low utilization rate of the air interface in V2X, Reference [17] proposed a method based on distributed multi-agent reinforcement learning, which can allocate resource block sharing in OFDM and vehicle transmission power by optimizing long-term cumulative discount returns, so as to improve the utilization rate of spectrum resources. In the coexistence environment of train to wayside (T2W) communication and train to train (T2T) communication, in order to make full use of limited spectrum resources and solve the problem of co-channel interference caused by frequency multiplexing, Reference [18] proposed an independent channel selection and transmission power selection algorithm for T2T communication based on multi-agent DRL, so as to reduce co-channel interference, improve the throughput of T2T link and system, and ensure successful transmission probability of T2T link within a specified time. Aiming at the problem of unmanned aerial vehicles pair-supported relaying navigation in the Internet of things, Huang *et al.* [19] proposed an algorithm based on a dueling double deep Q-network (dueling DDQN), the algorithm can have better performance without prior knowledge and meet the constraint that the amount of information reaches a certain threshold. In view of the non-reliability of candidate relays in nonorthogonal multi-access networks under power line channels, Reference [20] proposed a robust security transmission scheme that aims to maximize the safety and speed of the system through joint optimal relay selection and power optimization allocation. The solution method of DRL based on quantized channel state information is also presented. The proposed algorithm can effectively alleviate the dimensional disaster problem with low time complexity, and has good scalability and generalization performance. Obviously, under multiple constraints, DRL has certain advantages and exploration value in realizing the dynamic optimal allocation of communication resources.

In the vehicle communication environment, because the vehicle is moving, the communication environment changes in real time. With the change in vehicle operating environment, the sensor data of each vehicle is different, and the communication environment in the vehicle will still change. If the method of online learning is adopted, a large amount of data needs to be collected for training, and the algorithm processing complexity is relatively large. In contrast, DRL can adopt a combination of offline training and online training to pre-train the model with a large amount of data in advance, and then conduct real-time data collection and online training every time the vehicle runs, which can reduce the amount of data, training time and improve the universality of the algorithm.

Currently, PLC research is also channel-centric, with little attention paid to critical data transmission technologies. Because PLC is susceptible to channel fading and impulse noise, low latency data transmission presents additional challenges. In light of the preceding studies, this paper proposes a novel resource optimization allocation algorithm based on OFDM subcarriers diversity combination and adaptive power allocation to enhance the ability of PLC to transmit low latency services. With a compromise between diversity and multiplexing, subcarriers with varying fading conditions are combined in order to minimize the overall delay of service data transmission, making full usage of subcarriers with poor performance. The main contributions of this paper are summarized as follows:

1) In order to minimize transmission delay, an optimization model for PLC systems with impulse noise by combining short packet communication (SPC) and OFDM technology is established.

2) A DRL-based resource optimization algorithm is proposed. The DDQN algorithm is used to determine the optimal channel diversity combination. An adaptive power allocation algorithm is improved and used, considering the SPC error rate and data amount threshold.

3) Under the multiple constraints of short-packet reliability, subcarrier, and power, the model and its optimization performance were compared, and the effects of diversity order, power distribution factor, and DDQN parameters on deterministic delay performance were analyzed, which provides theoretical support for the application of PLC in the in-vehicle network under the background of 6G.

The remaining of this paper is as follows. The system model is introduced in Section II. Section III proposes a DRL-based algorithm of diversity combination of OFDM subcarriers approach. Section IV provides simulation results and the conclusion is finally remarked in Section V.

II. System Model

In the vehicular network, the data volume of different services is different. For example, the control information is usually only tens or hundreds of bits, and the image information is several Mbits. The required transmission time varies from one transmission slot to multiple transmission slots. The data-centric transmission scheme is to analyze and design from the transmission perspective of a single service according to different service requirements and resource constraints, to carry out dynamic resource allocation according to the real-time status of the power line channel and provide a stable and effective transmission scheme for data services that can meet boundary requirements.

1. Channel model

This subsection mainly introduces the OFDM communication system model and the optimization model. Fig.2 is the schematic diagram of the OFDM system transmission framework based on diversity combination. Based on the quality of service (QoS) requirements of low-delay services and constraints of physical layer resources (number of subcarriers, maximum transmission power, etc.), the transmitter performs resource optimization allocation within a transmission time slot in combination with the subcarrier state information obtained by perfect channel estimation: Firstly, subcarriers with different fading degrees are combined by diversity, that is, the same information is transmitted on multiple subcarrier channels with different fading estimates, so as to fully integrate transmission resources of poor subcarriers and improve the boundary short-board performance of multi-carrier transmission. Then, under multiple constraints such as resource and performance, algorithms of estimated error modeling, adaptive power allocation,

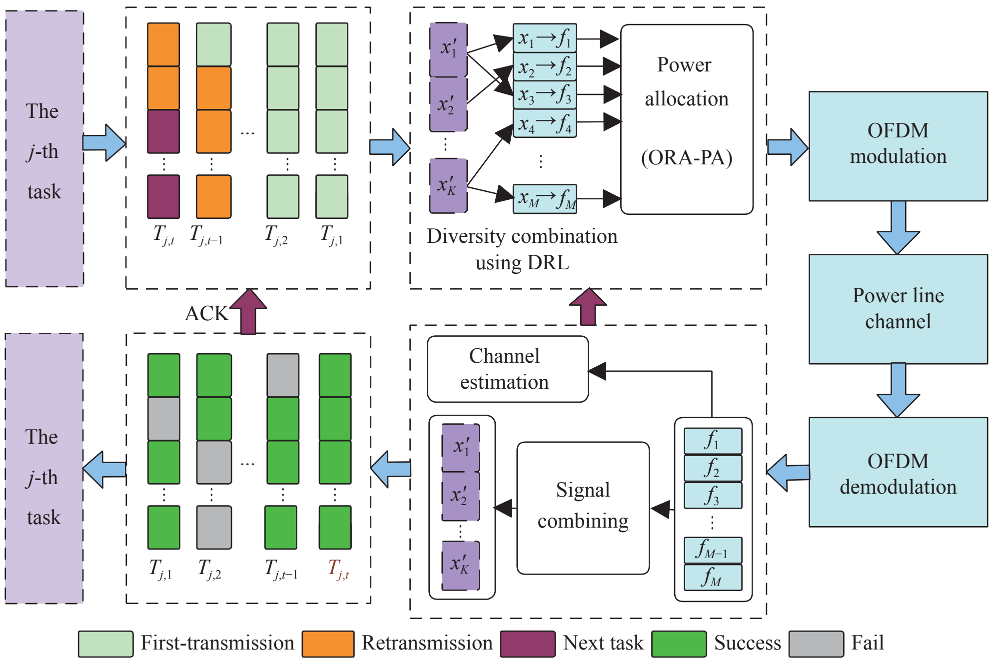


Fig. 2. OFDM system model using DRL-based diversity combination.

and machine learning are used to minimize service transmission delay, so as to realize the effective tradeoff between diversity and multiplexing.

According to the service requirements, In the t-th transmission slot, the transmitter obtains K complex signals $\boldsymbol{X}_t' = [x_{t,1}', x_{t,2}', \dots, x_{t,K}']^{\mathrm{T}}$ by serial-to-parallel conversion and mapping. The system performs diversity combination and power allocation according to the channel estimation values to obtain transmission signal $\boldsymbol{X}_t = [x_{t,1}, x_{t,2}, \dots, x_{t,M}]^{\mathrm{T}}$, where M denotes the number of subcarriers and $(\cdot)^{\mathrm{T}}$ denotes the transpose. Then, the modulated symbol by OFDM is transmitted in the power line channel. After discrete Fourier transform (DFT) demodulation at the receiver, the equivalent receiving sequence $\boldsymbol{Y}_t$ is produced:

$$\boldsymbol{Y}_t = \boldsymbol{H}_t \sqrt{\boldsymbol{P}_t} \boldsymbol{X}_t + \boldsymbol{F} \boldsymbol{n}_t \tag{1}$$

where $\boldsymbol{H}_t$ denotes the power line channel frequency domain fading matrix, $\boldsymbol{P}_t$ represents the transmission power matrix, $\boldsymbol{F}$ denotes the operator of discrete Fourier transform, and $\boldsymbol{n}_t$ represents the time-domain noise vector.

For the PLC system of OFDM modulation, it is assumed that the channel delay spread is less than the cyclic prefix length, and the influence of inter symbol interference can be ignored, so $\boldsymbol{H}_t$ is a diagonal matrix, then $\boldsymbol{H}_t = \mathrm{diag}\{h_{t,1}, h_{t,2}, \dots, h_{t,M}\}$, where $h_{t,i}, i = 1, 2, \dots, M$ denotes the channel fading coefficient of the i-th subcarrier in the t-th transmission slot, and it can be obtained using a bottom-up deterministic modeling approach based on transmission line theory [21] or a top-down statistical modeling approach that relies on channel measurement data [22], [23]. This paper focuses on the channel combination and power allocation algorithm under the condition of instantaneous fading, so without loss of generality, it is assumed that the channel fading $h_{t,i}$ satisfies the log-normal distribution [24], [25], which can be expressed as $h_{t,i} = \log\mathrm{N}(\mu_{\mathrm{p}}, \sigma_{\mathrm{p}}^2)$, and the subcarrier channel is a flat block fading channel, that is, the channel fading coefficient $h_{t,i}$ remains unchanged in a transmission time slot, and its probability density function (PDF) is

$$f_{h_{t,i}}(h_{\mathrm{p}}) = \frac{1}{h_{\mathrm{p}}\sqrt{2\pi\sigma_{\mathrm{p}}^2}} \exp\left(-\frac{(\ln h_{\mathrm{p}} - \mu_{\mathrm{p}})^2}{2\sigma_{\mathrm{p}}^2}\right) \tag{2}$$

In order to ensure that the channel fading does not change the average power of the received signal, this paper normalizes the channel fading energy [24], so that $\mathrm{E}(h_{\mathrm{p}}^2) = \exp(2\mu_{\mathrm{p}} + 2\sigma_{\mathrm{p}}^2) = 1$, then $\mu_{\mathrm{p}} = -\sigma_{\mathrm{p}}^2$.

The power line channel is susceptible to the mixed influence of background noise and impulse noise, and the two are independent. In order to study the influence of noise in PLC system, the power line noise is modeled as a Bernoulli-Gaussian model [26], which consists of two parts: background noise and impulse noise. Its PDF has the following form:

$$f(n) = (1 - P_{\mathrm{I}})\mathcal{CN}(0, \sigma_{\mathrm{B}}^2) + P_{\mathrm{I}}\mathcal{CN}(0, \sigma_{\mathrm{I}}^2) \tag{3}$$

where $\mathcal{CN}(0, \sigma_{\mathrm{B}}^2)$ and $\mathcal{CN}(0, \sigma_{\mathrm{I}}^2)$ denote the normal distribution, P_{I} is the probability of the occurrence of impulse noise, σ_{B}^2 and σ_{I}^2 denote the power of background noise and impulse noise, respectively. So the average total noise power is $N_0 = (1 - P_{\rm I})\sigma_{\rm B}^2 + P_{\rm I}\sigma_{\rm I}^2$.

2. Optimization model

It is known that there are M subcarriers at the transmitter and receiver, and $\mathcal{M} = \{1, 2, \ldots, M\}$ indicates the sequence number set of the subcarriers. This paper discusses the use of diversity to transmit the same data over multiple subcarriers. When a rate constraint exists, the channel with poor performance has a high probability of being interrupted and retransmitted. As a result, the paper effectively combines the channel with low performance with other channels in order to balance the performance of each subcarrier diversity combination. Finally, the utilization rate and delay performance of the channel are improved. At the receiver, the merging algorithm is adopted to process the received signals, so the equivalent signal-to-noise ratio (SNR) of the k-th group of signals after merging in the t-th transmission slot is

$$\gamma_{t,k}^{\rm C} = \sum_{i\in\mathcal{M}_k} \gamma_{t,i} \tag{4}$$

where $\gamma_{t,i} = \frac{p_{t,i}|h_{t,i}|^2}{N_0 W_{t,i}}$ denotes the SNR of the i-th subcarrier in t-th time slot, $p_{t,i}$ denotes the transmit power and $W_{t,i}$ denotes the channel bandwidth, $\mathcal{M}_k$ denotes the sequence number set of the subcarriers corresponding to the k-th group.

This paper assumes that the same diversity order is used in an OFDM transmission slot, so that $||\mathcal{M}_{k_1}||_0 = ||\mathcal{M}_{k_2}||_0 = L_{\rm D}$, $k_1 \neq k_2$, where $L_{\rm D}$ denotes the diversity order, and $||\cdot||_0$ is the 0-norm, that is the number of nonzero elements (In this paper represents the number of subcarriers that are divided into the same group). So the transmission rate of SPC [27] can be approximated as

$$R_{t,k} = W_{t,k}\left[\log_2\left(1+\gamma_{t,k}^{\rm C}\right) - \sqrt{\frac{V_{t,k}}{L_{{\rm S},t,k}}}\frac{Q^{-1}(\varepsilon_{t,k})}{\ln 2}\right] \tag{5}$$

where $V_{t,k} = 1 - (1+\gamma_{t,k}^{\rm C})^{-2}$ is channel dispersion, $L_{{\rm S},t,k}$ is packet size, $Q^{-1}(\varepsilon_{t,k})$ is the inverse of Q-function, and $\varepsilon_{t,k}$ is error rate and it is taked $\varepsilon_{t,k} = 10^{-6}$ in this paper.

For low latency services, the amount of data $D_{\rm th}^{\rm all}$ that needs to be transmitted is small. Following the data-oriented approach in this paper, the rate formula is converted into the amount of data formula. Then the amount of data that the k-th packet transmitted at the t-th transmission slot in OFDM is

$$D_{t,k} = T_{{\rm C},t} W_{t,k}\left[\log_2\left(1+\gamma_{t,k}^{\rm C}\right) - \sqrt{\frac{V_{t,k}}{L_{{\rm S},t,k}}}\frac{Q^{-1}(\varepsilon_{t,k})}{\ln 2}\right] \tag{6}$$

Then, the total amount of data transmitted by all subcarriers is

$$D_t = \sum_{k=1}^{K} D_{t,k} \tag{7}$$

where K denotes the total number of groups in the t-th OFDM transmission slot. Considering the system reliability, the system interruption will occur when $D_{t,k}$ is lower than a certain threshold. Therefore, the actual amount of data transmitted is

$$D_t = D_{\rm th}\sum_{k=1}^{K} {\rm U}(D_{t,k} - D_{\rm th}) \tag{8}$$

where ${\rm U}(\cdot)$ is the unit step function, and $D_{\rm th}$ is the amount of data threshold, which is equal to the packet length $L_{{\rm S},t,k}$. Given that the t-th transmission slot is $T_{{\rm C},t}$, if the amount of data $D_{\rm th}^{\rm all}$ needs $L_{\rm t}$ transmission slots to complete transmission, the total duration is $\sum_t^{L_{\rm t}} T_{{\rm C},t}$. Without loss of generality, let the transmission slot duration of OFDM be equal, $T_{{\rm C},t} = T_{\rm C}$. That means our goal is to minimize the total number of transmission slots for service transmission. Hence, the optimization problem is formulated as

$$\min_{\{\mathcal{M}_1,\ldots,\mathcal{M}_K\}} \sum_{t=1}^{L_{\rm t}} T_{{\rm C},t} \tag{9}$$

$$\text{s.t.}\begin{cases} \text{C1: } D_{\rm th}\sum_{t=1}^{L_{\rm t}}\sum_{k=1}^{K}{\rm U}({\rm D_{t,k}} - {\rm D_{th}}) \geq {\rm D_{th}^{all}} \\ \text{C2: } \sum_{t=1}^{L_{\rm t}}\sum_{k=1}^{K}\sum_{i\in\mathcal{M}_k} p_{t,k,i} \leq p_{\rm total}, \forall t,k,i \\ \text{C3: } 0 \leq p_{t,k,i} \leq p_{\max}, \forall t,k,i \\ \text{C4: } \mathcal{M}_{k_1}\cap\mathcal{M}_{k_2} = \emptyset, k_1\neq k_2, k_1,k_2\in\{1,2,\ldots,K\} \\ \text{C5: } \mathcal{M}_1\cup\mathcal{M}_2\cup\cdots\cup\mathcal{M}_K = \{1,2,\ldots,M\} \end{cases}$$

where C1 represents the total data constraint of a single service. C2 represents the power constraint of a single service, $p_{\rm total}$ is the total power, $p_{\max}$ is the maximum transmit power of a single subcarrier. C3 is the transmit power constraint on a single subcarrier; and C4 and C5 are constraints on channel diversity combination, indicating that within a $T_{\rm C}$, all subcarriers participate in resource allocation, and each subcarrier can only be used once.

It can be known from formula (9) that the optimization model has high computational complexity. Taking diversity combination as an example, when the number of subcarriers is M, the optimal combination scheme is $\frac{M!}{\frac{M}{2}!\times 2^{\frac{M}{2}-1}}$. For instance, when $M = 20$, there are 1.3×10^9 combinations. It also involves parameter optimization. Obviously, the traditional algorithm is difficult and lacks closed analytic expression. In order to solve the above problems, it proposes a low complexity resource optimization algorithm based on DDQN to obtain an approximate optimal solution. DDQN is a kind of DRL that mainly solves the problems of continuous state space and discrete action space. Herein, the parameters are continuous, and the diversity combination is discrete, so DDQN is the most suitable for resource optimization.

III. DRL-Based Resource Optimization Algorithm

This section proposes a DDQN allocation (DDQN-AL) algorithm for the low latency service's SPC to optimize the allocation of OFDM subcarriers and power, so as to minimize the transmission delay.

1. Double deep Q-learning network

The key feature of standard Q learning and DQN is to use the same Q value to evaluate and select the best action, which makes agent overestimate the Q value of the selected action. To solve this problem, the most effective way is to separate the Q value of action selection from the Q value of action evaluation. Using the network structure shown in Fig.3, DDQN conducts training by maximizing the Q value of the selected action into two network models: action selection and action evaluation, so as to reduce the overestimation of Q value through unbiased estimation of Q value. Its working principle is that if the value network overestimates an action, as long as the target network does not overestimate the action, accurate Q value can still be calculated according to the target network.

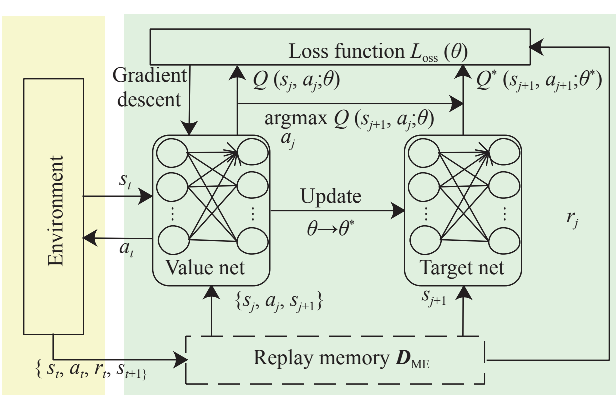


Fig. 3. DDQN principle.

When training the parameters of DDQN model, the weight of its value network is updated and the weight of target network remains unchanged, while the weight of the target network θ^* is the parameter through the value network θ replace the original weight parameters get. Since the weight parameters of the value network and the target network are not updated synchronously, and the update frequency of the weight parameters of the target network is lower than that of the value network, the model parameters are more stable when converging. The experimental research of Van DDQN can effectively reduce overestimation and thus obtain more stable learning in the experimental study by van Hasselt *et al.* [28], and the DDQN algorithm can achieve better performance than DQN in Atari 2600 games.

2. Solution process of DDQN-AL algorithm

The system is assumed to use the ideal channel estimation algorithm. After receiving feedback about the channel state, the transmitter uses DDQN to determine the optimal diversity combination. Then, ORA is used in conjunction with a power allocation algorithm to minimize the total transmission slots for service data blocks. To obtain the optimal diversity combination using the DDQN algorithm, the state space, action, and reward functions of DDQN should be defined in accordance with the optimization formula (9), as illustrated in Fig.3.

1) State: The environmental state s_t describes the key characteristics or parameters of the system for different transmission slots t. In DRL, the definition of s_t should be able to provide necessary reference for the agent's decision-making, reduce redundant information and avoid interference to the agent's decision-making.

In this paper, first calculate the channel-to-noise ratio as $g_{t,i} = \frac{|h_{t,i}|^2}{N_0 W_{t,i}}$ according to the channel fading coefficient $h_{t,i}$ and the noise parameter N_0, then sort $g_{t,i}$, $i = 1, 2, \ldots, M$ in descending order to get $\{g'_{t,1}, g'_{t,2}, \ldots, g'_{t,M}\}$. Finally, the expression obtained by normalizing $g'_{t,i}$ is as follows:

$$\beta_{t,i} = \frac{g'_{t,i} - \min(g'_{t,i})}{\max(g'_{t,i}) - \min(g'_{t,i})},\ i = 1, 2, \ldots, M \tag{10}$$

Therefore, this paper defines s_t as the normalized channel-to-noise ratio and noise power N_0 of all PLC subcarriers in the t-th transmission time slot:

$$s_t = \{\beta_{t,1}, \beta_{t,2}, \ldots, \beta_{t,M}, N_0\} \tag{11}$$

2) Action: Action is the agent's response to the environment in the current transmission slot t according to the current state information. The action space $\boldsymbol{A}$ in this paper consists of two parts: all possible combinations of subcarriers $\boldsymbol{A}_{\mathcal{M}}$ and the power quantization set $\boldsymbol{A}_{\mathcal{P}}$.

The action of subcarrier combining of the t-th transmission slot can be expressed as

$$a_{\mathcal{M},t} = \{\mathcal{M}_{t,1}, \mathcal{M}_{t,2}, \ldots, \mathcal{M}_{t,K}\} \tag{12}$$

The equation (26) in reference [11] is combined with SPC, and an improved adaptive power allocation algorithm based on error rate and the amount of data threshold is adopted. Finally, the total power $p_{t,k}$ of the k-th group in the t-th transmission slot is

$$p_{t,k} = \frac{N_0 W_{t,k}}{\sum_{i \in \mathcal{M}_k} |h_{t,k,i}|^2} \Big[\exp(\frac{D_{\mathrm{th}} \ln 2}{T_{\mathrm{C},t} W_{t,k}} + \eta_{t,k}) - 1 \Big] \quad (13)$$

where $\eta_{t,k} = \sqrt{\frac{\overline{V}_{t,k}}{\overline{L}_{\mathrm{S},t,k}}} Q^{-1}(\varepsilon_{t,k})$, $\overline{V}_{t,k} = 1-(1+\overline{\gamma}^{\mathrm{C}}_{t,k})^{-2}$, and $\overline{\gamma}^{\mathrm{C}}_{t,k}$ denotes average SNR. In the k-th group, in order to satisfy the constraint of subchannel and total power, if equal power allocation for each subcarrier, i.e., $p_{t,k,i} = \frac{p_{t,k}}{||\mathcal{M}_{t,k}||_0}$, is larger than $p_{\max}$ or the total power, the power will be restricted according to the limit. In this paper, it represents the number of subcarriers in the k-th group. Then the set of possible actions for the power allocation of the t-th transmission slot is

$$a_{\mathcal{P},t} = \{\mathcal{P}_{t,1}, \mathcal{P}_{t,2}, \ldots, \mathcal{P}_{t,K}\} \quad (14)$$

where $\mathcal{P}_{t,k}, k = 1, 2, \ldots, K$ denotes the power set of the k-th group. Therefore, the action obtained in the t-th transmission slot is

$$a_t = \big\{\{\mathcal{M}_{t,1}, \mathcal{P}_{t,1}\}, \{\mathcal{M}_{t,2}, \mathcal{P}_{t,2}\}, \ldots, \{\mathcal{M}_{t,K}, \mathcal{P}_{t,K}\}\big\} \quad (15)$$

In order to strike a tradeoff between exploration and exploitation, the agent chooses actions according to ϵ-greedy algorithm.

3) Reward: In this model, after completing a transmission, the agent will get feedback from the environment, which is a set of rewards. The reward represents the optimization goal of the DRL task. The total transmission slots in the optimization model is determined by the total amount of data, the amount of data correctly transmitted at a single time, etc. When the amount of data transferred is determined, the number of transmitted time slots is inversely proportional to the amount of data successfully transmitted by a single time slot. To meet the rate and delay requirements for different services, the reward function r_t consists of three parts, which is defined as

$$r_t = \delta_1 r_{t,1} + \delta_2 r_{t,2} + \delta_3 r_{t,3} \quad (16)$$

where δ_1, δ_2, δ_3 denote weight coefficients of the three sub-reward functions, respectively, and $\delta_1+\delta_2+\delta_3 = 1$. $r_{t,1}$ denotes the rate corresponding to equation (5) after the diversity combination. $r_{t,2}$ in the second part denotes the equivalent transmission rate corresponding to the worst subchannel group in OFDM frame at the t-th transmission slot that is the lower limit of system rate:

$$r_{t,2} = \min\{R_{t,1}, R_{t,2}, \ldots, R_{t,K}\} \quad (17)$$

Obviously, the higher the worst sub-channel grouping rate is, the more data the OFDM system can successfully communicate in a fixed time.

In equation (16), $r_{t,3}$ is the data successfully transmitted after using the power allocation algorithm and can be expressed as

$$r_{t,3} = \frac{D_{\mathrm{th}}}{K} \sum_{k=1}^{K} \mathrm{U}(D_{t,k} - D_{\mathrm{th}}) \quad (18)$$

When the weights are different, the performance and focus of optimization are also different. When the value of δ_1 is large, the algorithm is suitable for high-speed transmission. For example, when $\delta_1 : \delta_2 : \delta_3 = 1 : 0 : 0$, the selection of DDQN action is only related to the average transmission rate. In order to maximize the transmission rate, DDQN will imitate the actions of the water-filling algorithm, combine subcarriers with better channel conditions, and allocate more power to maximize the transmission rate. However, due to ignoring subcarriers with poor performance, a large amount of erroneous data may need to be retransmitted during data transmission, which increases the transmission delay. When the value of δ_2 is large, the algorithm is suitable for services that require deterministic performance. If $\delta_1 : \delta_2 : \delta_3 = 0 : 1 : 0$, DDQN pays attention to the performance improvement of poorer subcarriers to eliminate the short board effect, but the highest transmission rate obtained is lower than that of the water-filling algorithm. In order to facilitate theoretical analysis and simulation, without loss of generality, their values are set to $\delta_1 = \delta_2 = \delta_3 = \frac{1}{3}$.

Since the DDQN model maximizes the cumulative reward function for training and action optimization, DDQN have a certain correlation in time through discount factors. For data services, the objective of minimizing transmission delay is to ensure the amount of data transmitted in a single pass and the number of retransmissions of erroneous data, and to accumulate rewards for several slots through a discount factor.

4) Loss function: In order to ensure the stability of the algorithm, experience playback and quasi-static target network techniques can be used. For experience replay, instead of using a single experience to train the QNN at the end of each execution step, we can pool many experiences for batch training. Specifically, the loss function is defined as

$$L_{\mathrm{oss}}(\theta) = \mathrm{E}\big[(q_{\mathrm{target}} - Q(s_j, a_j; \theta))^2\big] \quad (19)$$

$$q_{\mathrm{target}} = r_j + \mathit{\Gamma} Q^* \big[s_{j+1}, \arg\max_{a_j} Q(s_{j+1}, a_j; \theta); \theta^*\big] \quad (20)$$

Iteratively, q denotes the target Q^* value and Γ denotes the discount factor. Additionally, Q^* and Q denotes the weights of the target and value network. Based on reference [29], the parameter θ can be optimized by minimizing the loss function $L_{\rm oss}(\theta)$. The steps of the DDQN algorithm in this paper are shown in the pseudocode, and the DDQN-AL can be expressed as Algorithm 1.

Algorithm 1 Agent update DDQN-AL algorithm

Input: Environment simulator, Q network, minibatch size.
Initialize: Q network with random weights θ and θ^*, replay memory $\boldsymbol{D}_{\rm EM}$;
for each training step do
  Repeat
    Observes state $s_{t,0}=\{\beta_{t,1},\beta_{t,2},...,\beta_{t,M},N_{t,0}\}$;
    for $k=1:K$ do
      Select action $a_{t,k}=\{\mathcal{M}_{t,k},\mathcal{P}_{t,k}\}$ according to $s_{t,k}$ and the probability ϵ;
      Observe a new state $s_{t,k+1}$ with $s_{t,k}$,$a_{t,k}$;
    end for
    Obtain action $a_t=\{\{\mathcal{M}_{t,1},\mathcal{P}_{t,1}\},\{\mathcal{M}_{t,2},\mathcal{P}_{t,2}\},\dots,\{\mathcal{M}_{t,K},\mathcal{P}_{t,K}\}\}$ at t-th;
    Execute a_t, calculate dynamic reward r_t by (16);
    Save experience $(s_{t,k},a_{t,k},r_{t,k},s_{t,k+1})$ into the $\boldsymbol{D}_{\rm EM}$ until constraint (C1) in (9) sample a random minibatch data $(s_j,a_j,r_j,s_{j+1}),j=1,2,\dots$ from $\boldsymbol{D}_{\rm EM}$;
    Perform a gradient descent step and Update θ;
    Every ζ steps, update the target network $\theta\rightarrow\theta^*$;
end for
Return: Trained DDQN and state-action values.

Complexity analysis of the algorithm in this paper: The main solution steps in this paper are to use DDQN to complete the optimal combination of subcarriers and use (13) for power allocation, and the complexity of (13) is low, so the complexity of the algorithm in this paper is mainly influenced by DDQN. Because DDQN adopts a fully connected layer network, the time complexity of each resource cluster is $O(\sum_{l=1}^{L_{\rm F}}(X_lY_l))$, where $L_{\rm F}$ denotes the number of connection layers, and X_l and Y_l are respectively the input and output dimensions of the l-th fully connected layer.

IV. Experimental Evaluation

1. Parameter setting of simulation scenario

In order to verify the performance of the proposed algorithm, this paper conducts simulation analysis of the algorithm through Python 3.0, TensorFlow 1.14.0 and Keras 2.3.1. The gain of the power line channel is randomly generated, that is, the channel gain corresponding to each subcarrier in the algorithm processing process is different. Reference [8], unless otherwise specified, the parameter settings during system simulation and calculation are shown in Table 1, and the parameters of DDQN are shown in Table 2.

Table 1. The parameters of system simulation

Parameters	Value
Number of sampling points	256
Total number of subcarriers	128
Number of data subcarriers M	20
Sub-carrier wave band width W (kHz)	100
Power line fading coefficient $\sigma_{\rm P}^2$ (dB)	2.5
Channel coherence time (ms)	2
Impulse noise probability $P_{\rm I}$	0.01
$\sigma_{\rm I}^2/\sigma_{\rm B}^2$ (dB)	15

Table 2. The parameters of DDQN

Parameters	Value
Memory size	1024
Batch size	64
Learning rate	0.002
Discount factor	0.99
Max-ε-epsilon	1
Min-ε-epsilon	0.05
Attenuation factor	0.995
Hidden layers	3
Neurons	256
Activation function	Relu
θ^* Update frequency ζ	5
Optimizer	Adam

In this section, we analyze the performance of the proposed DDQN-AL. The ORA-PA [11] and water-filling are selected to compare with the algorithm in this paper. The ORA-PA algorithm maximizes the amount of effective data in a fixed transmission time by adaptive modulation coding and power allocation according to the fading value of the channel. In addition, to better reflect the performance of the algorithm in this paper, the referenced ORA-PA algorithm also uses diversity technology, but the subcarrier packets are randomly combined.

2. Diversity orders analysis

Figs.4 and 5 are the comparison diagrams of the lower limit of system rate and overall transmission delay corresponding to different SNR and diversity orders ($L_{\rm D}=1,2,3,4$). Fig.4 shows that when SNR = 8.5 dB, the $L_{\rm D}=1$ case is the best, and the performance of the $L_{\rm D}=2$ case is better than that of the $L_{\rm D}=3$ and the $L_{\rm D}=4$ when SNR = 2.5–8 dB. Based on equation (5), the transmission rate of a single packet follows the logarithmic growth trend. When SNR is large, the benefits brought by diversity will gradually decrease. When the SNR is small, the higher the diversity order is, the better the performance is. Hence, $L_{\rm D}=2$ has relatively better performance in the range of SNR = 2.5–8 dB, and the performance of $L_{\rm D}=2$ is worse in other ranges.

Fig.5 shows the average transmission delay of the system corresponding to different SNR and diversity orders when transmitting short packets of 10 Kbits. When SNR = 9.5 dB, it is not necessary to combine the subcarriers (uncombined). In the range of 6–10 dB, the performance of the $L_{\mathrm{D}} = 2$ case is better. It can be seen from Figs.4 and 5 that the SNR of the best performance of $L_{\mathrm{D}} = 2$ is different for different evaluation indexes, because the optimization schemes of machine learning are different for different indexes. Based on the simulation analysis in the following paper, The following analysis will be based on $L_{\mathrm{D}} = 2$.

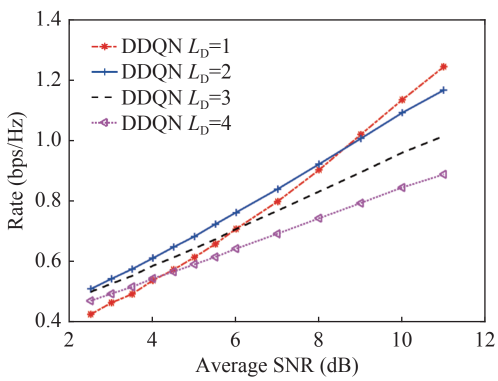


Fig. 4. Lower limit of system rate.

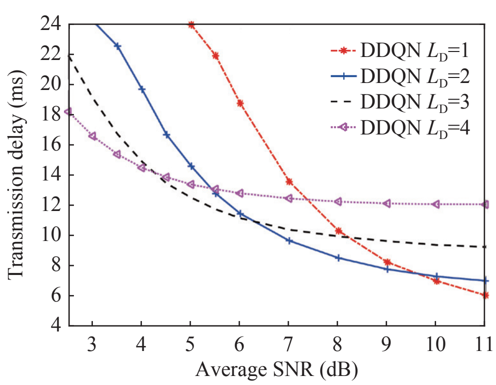


Fig. 5. Performance of delay.

3. Convergence analysis

Fig.6 compares the effects of various optimizers on the DDQN convergence rate. As can be seen, the Adam optimizer has the best performance and a high rate of convergence, which is close to optimal. When the root mean square prop (RMSprop) optimizer is used, the system exhibits significant volatility. The convergence rate of the stochastic gradient descent (SGD) optimizer is slower than that of the Adam optimizer, and it does not converge to the optimal.

Fig.7 illustrates the performance of the trained model when the environment changes. Fig.7(a) shows the training times required for model re-convergence at various SNRs when the channel fading σ_{p}^2 changes from 1dB to 2dB, while Fig.7(b) shows the training times re-

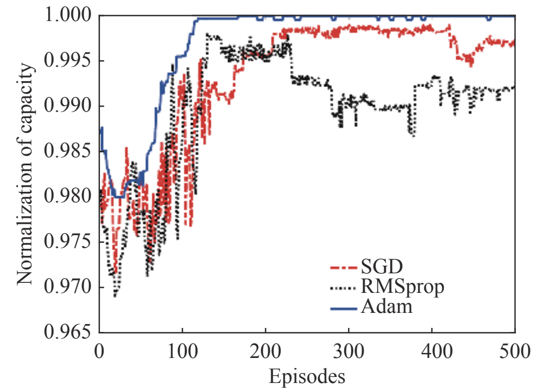


Fig. 6. Convergence comparisons of different optimizers.

quired for model reconvergence when the SNR changes. As can be seen, the training times required for re-convergence of the trained model are less than those required for initialization, and the volatility of the convergence process is significantly less than that of the initialization process. This is because when DDQN is initialized, the internal parameters are randomly distributed, and the parameters have converged after offline training. However, when the environment changes, although the input parameters also change, the relevant definitions of the optimization model and DDQN do not change. Only a little adjustment of the original neural network parameters is needed to ensure rapid convergence again. This demonstrates that the proposed algorithm has a higher convergence rate and a better robust performance, as well as the ability to rapidly adapt to changes in environmental parameters. This performance is critical to multi-dimensional power line channel as well as and optimization model and service adaptation for algorithm parameters.

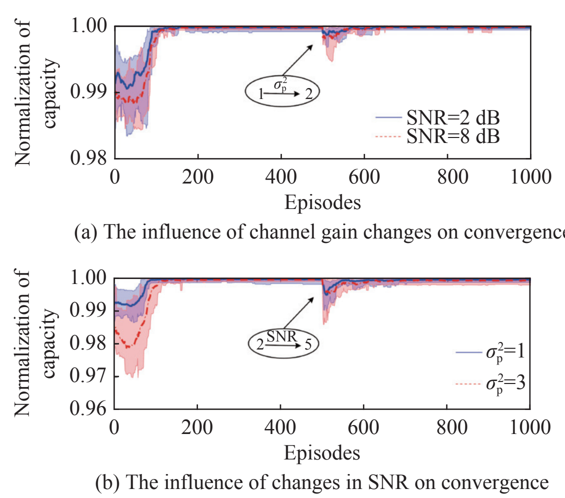
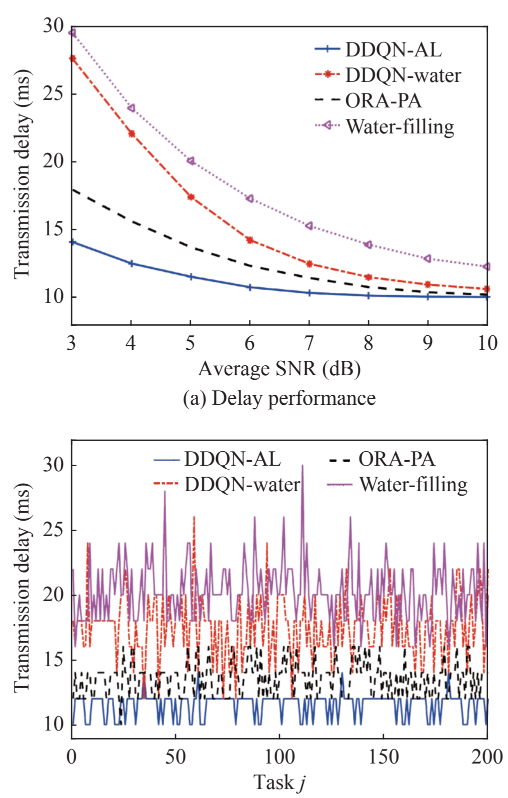


Fig. 7. Convergence of algorithm with parameters change.

4. Analysis of delay outage rate

Fig.8(a) illustrates the system average transmission delay for various algorithms at various SNRs when transmitting a short 10 Kbits packet. The transmission delay of each algorithm decreases as the SNR increases, but the delay required by the algorithm proposed in this paper is always kept to a minimum. Fig.8(b) illustrates the transmission delay sampling values for various algorithms when SNR = 5 dB. Not only is the average transmission delay of the algorithm proposed in this paper the shortest possible, but also the volatility is as small as possible. Meanwhile, the average transmission delay and volatility of the water-filling algorithm are at their maximum. This is because the data with transmission error needs to be retransmitted under the constraint of ensuring high reliability, so the resources of subcarriers with poor performance will be wasted. In this paper, the algorithm makes full use of subcarrier resources with different performance through the diversity transmission of subcarrier groups to achieve the balance and overall improvement of subcarrier groups' performance. However, the water-filling algorithm only allocates power to subcarriers with good performance and does not pay attention to transmission interruption of poor subcarriers in time slots, so the transmission performance may fluctuate greatly.

For the data-centric optimization (9) in this paper and Reference [12], this paper uses the delay outage rate as an evaluation indicator of system reliability, where DOR is defined as the probability that the delay



(a) Delay performance

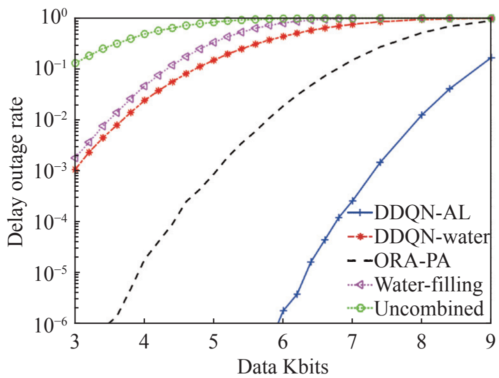
(b) Delay performance (SNR=5 dB)

Fig. 8. Delay performance of different algorithms.

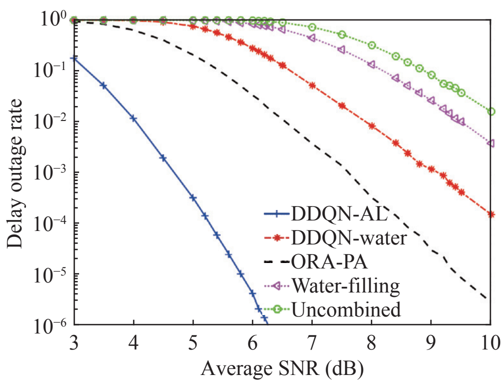
of transmitting a specific amount of data volume is greater than the threshold duration, and the threshold usually results from the delay requirement of the considered data traffic. Therefore, compared to the traditional physical layer outage probability, DOR focuses on the overall performance of service over multiple time slots transmission.

Fig.9 compares and analyzes the DOR of different algorithms. In Fig.9(a), when SNR = 5 dB, the DOR of the four algorithms increases with the increase of the amount of data. When the same volume of data is transmitted, the DOR of the algorithm in this paper is the minimum. Taking DOR = 10^{-6} as an example, the algorithm proposed in this paper can transmit 6 Kbits, which is obviously superior to other algorithms. Fig.9(b) shows the relationship between DOR and SNR when the system transmits 7 Kbits. The algorithm in this paper can reach 10^{-6} when SNR $\approx$ 6.3 dB, which also has the best performance. It can be seen that under the constraint of deterministic delay, the transmitted data usually only experiences a few states of the channel when the amount of data transmitted by a single service is small. Therefore, the randomness of the channel has a great impact on DOR. The algorithm in this paper first adopts DDQN to optimize the combination of subcarriers, maximize the performance of the worst subcarriers, and reduce the randomness between each group. Then, it satisfies the basic requirements of the lowest transmission rate of each group through power distribution and provides a stable channel environment for the transmission of services. While other algorithms mainly maximize the performance of better subcarriers, so there are no corresponding measures are given to reduce the bit error rate when the channel is in a poor state. Therefore, the DOR performance of the proposed algorithm is the best compared with other algorithms.

Taking the maximum transmitted amount of data $5KD_{\rm th}$ of $5T_{\rm C}$ as a reference, we can use the formula $\left[\sum_{t=1}^{5} D_{\rm th} \sum_{k=1}^{K} {\rm U}(D_{t,k} - D_{\rm th})\right] / (5KD_{\rm th})$ to perform normalization for the data actually transmitted, and the normalized amount of data of the algorithm can be obtained, this indicator shows the percentage of successfully transmitted data volume in the total transmitted data volume. Fig.10 compares the mean value, and the root mean square error (RMSE) of the normalized data set for various algorithms. In this paper, the DDQN-AL algorithm has the largest average normalized data. With increasing SNR, the volatility or RMSE of normalized data for the DDQN-AL algorithm decreases, remaining superior to other algorithms. In five consecutive time slots, the data cannot experience all the states of the channel, so the algorithm needs to analyze and



(a) Delay outage rate of different $D_{\rm th}^{\rm all}$



(b) Delay outage rate under different SNRs

Fig. 9. Comparison of DOR of different algorithms.

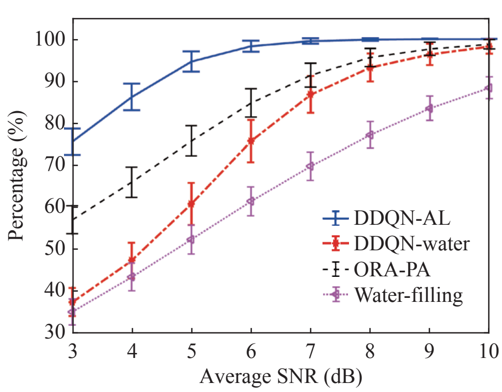


Fig. 10. Comparison of data packets transmitted by different algorithms.

predict the five consecutive states to make more reasonable use of power resources. However, the DRL algorithm is trained with the objective of maximizing the cumulative reward of continuous multiple environmental states, so as to realize the effective utilization of power in different time slots. Therefore, the DDQN-AL algorithm in this paper can achieve the best performance.

In order to analyze the effect of DDQN-AL in suppressing impulse noise, we simulate the delay outage rate of these three algorithms (DDQN-AL, ORA-PA [11] and Blanking [30] combined with ORA-PA (Blank ORA-PA)) under different fading channel conditions. It can be seen from Fig.11 that when $\sigma_{\rm p}^2 = 1$ dB, because the peak-to-average power ratio (PAPR) in OFDM is very small, the performance of the Blank ORA-PA is better than DDQN-AL at a relatively high average SNR. However, when $\sigma_{\rm p}^2 = 2$ dB or $\sigma_{\rm p}^2 = 3$ dB, the PAPR of OFDM increases, which leads to the performance degradation of detection probability and false alarm probability of Blanking algorithm. Although DDQN-AL does not specifically handle impulse noise, its performance is significantly better than the other two methods. All in all, the effect of DDQN-AL in suppressing impulse noise is better than other algorithms in most cases, especially when the channel fading is greater, its effect is more significant.

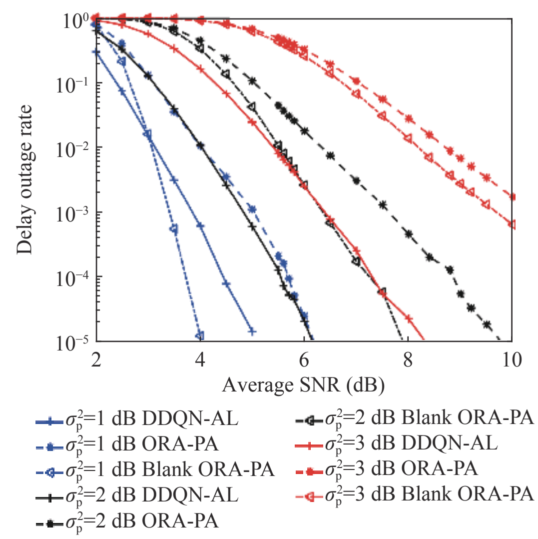


Fig. 11. DOR performance analysis.

V. Conclusions and Future Work

According to the actual needs of in-vehicle communication, enhancing the communication capacity of the power line to support low latency service has a certain role in the development of the IoV. In order to solve this complex optimization problem, this paper proposes a resource optimization algorithm applied for PLC systems with OFDM modulation based on the diversity combination of OFDM subcarriers. This algorithm first combines SPC and OFDM technology to establish an optimization model, and then solves the relevant model by defining the state, role, and reward function of DDQN and improving the adaptive power allocation algorithm, in order to minimize the transmission delay.

Moreover, the effectiveness of this method is proved by simulation data analysis. The DDQN-AL algorithm can adjust the weight of the reward function according to different environmental conditions and optimization objectives, so as to meet the requirements of service performance indicators. For DOR under deterministic delay, the performance of the DDQN-AL algorithm surpasses that of the ORA-PA algorithm and DDQN-water algorithm. Specifically, under the constraint of 10 ms and 20 MHz, the DDQN-AL algorithm can transmit about 2 Kbits more data.

The DDQN-AL algorithm used in this paper achieves good results in deterministic time delay. But mainly for the end-to-end communication system. In future work, we will combine in-vehicle communication and V2X communication to explore the delay performance of mixed multi-hops in IoV. Additionally, considering the resource scheduling problem between different users in the system, the multi-user carrier resources can be jointly optimized in the future.

References

[1] X. Cheng, H. T. Zhang, Z. H. Yang, *et al.*, "Integrated sensing and communications for Internet of vehicles: Current status and development trend," *Journal on Communications*, vol.43, no.8, pp.188–202, 2022. (in Chinese)

[2] X. C. Xu, K. Liu, C. H. Liu, *et al.*, "Potential game based channel allocation for vehicular edge computing," *Acta Electronica Sinica*, vol.49, no.5, pp.851–860, 2021. (in Chinese)

[3] G. Thandavarayan, M. Sepulcre, and J. Gozalvez, "Generation of cooperative perception messages for connected and automated vehicles," *IEEE Transactions on Vehicular Technology*, vol.69, no.12, pp.16336–16341, 2020.

[4] Y. Liao, X. Y. Tian, Z. R. Cai, *et al.*, "Intelligent channel estimation based on edge computing for C-V2 I," *Acta Electronica Sinica*, vol.49, no.5, pp.833–842, 2021. (in Chinese)

[5] H. H. Yang, J. C. Shang, J. J. Li, *et al.*, "Multi-traffic targets tracking based on an improved structural sparse representation with spatial-temporal constraint," *Chinese Journal of Electronics*, vol.31, no.2, pp.266–276, 2022.

[6] H. Z. Du, Q. Y. Wen, S. S. Zhang, *et al.*, "A pairing-free certificateless signcryption scheme for vehicular Ad hoc networks," *Chinese Journal of Electronics*, vol.30, no.5, pp.947–955, 2021.

[7] D. B. Rawat, R. Doku, A. Adebayo, *et al.*, "Blockchain enabled named data networking for secure vehicle-to-everything communications," *IEEE Network*, vol.34, no.5, pp.185–189, 2020.

[8] S. C. Liu, L. Xiao, L. F. Huang, *et al.*, "Impulsive noise recovery and elimination: A sparse machine learning based approach," *IEEE Transactions on Vehicular Technology*, vol.68, no.3, pp.2306–2315, 2019.

[9] A. Pittolo, M. De Piante, F. Versolatto, *et al.*, "In-vehicle power line communication: Differences and similarities among the in-car and the in-ship scenarios," *IEEE Vehicular Technology Magazine*, vol.11, no.2, pp.43–51, 2016.

[10] H. C. Yang and M. S. Alouini, "Data-oriented transmission in future wireless systems: Toward trustworthy support of advanced internet of things," *IEEE Vehicular Technology Magazine*, vol.14, no.3, pp.78–83, 2019.

[11] H. C. Yang, S. Choi, and M. S. Alouini, "Ultra-reliable low-latency transmission of small data over fading channels: A data-oriented analysis," *IEEE Communications Letters*, vol.24, no.3, pp.515–519, 2020.

[12] N. Shlezinger, R. Shaked, and R. Dabora, "On the capacity of MIMO broadband power line communications channels," *IEEE Transactions on Communications*, vol.66, no.10, pp.4795–4810, 2018.

[13] S. C. Liu, F. Yang, and J. Song, "An optimal interleaving scheme with maximum time-frequency diversity for PLC systems," *IEEE Transactions on Power Delivery*, vol.31, no.3, pp.1007–1014, 2016.

[14] Y. Iraqi and A. Al-Dweik, "Efficient information transmission using smart OFDM for IoT applications," *IEEE Internet of Things Journal*, vol.7, no.9, pp.8397–8409, 2020.

[15] X. L. Xu, Z. J. Fang, L. Y. Qi, *et al.*, "A deep reinforcement learning-based distributed service offloading method for edge computing empowered internet of vehicles," *Chinese Journal of Computers*, vol.44, no.12, pp.2382–2405, 2021. (in Chinese)

[16] X. C. Wang, F. Wu, Y. Z. Sun, *et al.*, "Internet of vehicles resource management based on federal deep reinforcement learning," *Electronic Measurement Technology*, vol.44, no.10, pp.114–120, 2021. (in Chinese)

[17] J. H. Liu, J. H. Zhao, and X. K. Sun, "Deep reinforcement learning based wireless resource allocation for V2X communications," in *Proceedings of the 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP)*, Beijing, China, pp.1–5, 2021.

[18] J. H. Zhao, J. Liu, Y. W. Nie, *et al.*, "Location-assisted beam alignment for train-to-train communication in urban rail transit system," *IEEE Access*, vol.7, pp.80133–80145, 2019.

[19] F. Huang, G. X. Li, H. C. Wang, *et al.*, "Navigation for UAV pair-supported relaying in unknown IoT systems with deep reinforcement learning," *Chinese Journal of Electronics*, vol.31, no.3, pp.416–429, 2022.

[20] H. H. Pu, X. S. Liu, and D. G. Xu, "Deep reinforcement learning based robust secure transmission for cooperative non-orthogonal multiple access networks," *Proceedings of the CSEE*, vol.42, no.13, pp.4760–4774, 2022. (in Chinese)

[21] J. Anatory, N. Theethayi, and R. Thottappillil, "Power-line communication channel model for interconnected networks—part I: Two-conductor system," *IEEE Transactions on Power Delivery*, vol.24, no.1, pp.118–123, 2009.

[22] M. Tlich, A. Zeddam, F. Moulin, *et al.*, "Indoor power-line communications channel characterization up to 100 MHz—part I: One-parameter deterministic model," *IEEE Transactions on Power Delivery*, vol.23, no.3, pp.1392–1401, 2008.

[23] A. Dubey and R. K. Mallik, "PLC system performance with AF relaying," *IEEE Transactions on Communications*, vol.63, no.6, pp.2337–2345, 2015.

[24] Z. X. Chen, L. J. Wang, D. S. Han, *et al.*, "A unified performance analysis of relaying communication system for IoT application with hybrid fading," *IEEE Internet of Things Journal*, vol.7, no.1, pp.570–583, 2020.

[25] K. M. Rabie, B. Adebisi, A. M. Tonello, *et al.*, "Two-stage non-orthogonal multiple access over power line communication channels," *IEEE Access*, vol.6, pp.17368–17376, 2018.

[26] S. P. Herath, N. H. Tran, and T. Le-Ngoc, "Optimal signaling scheme and capacity limit of PLC under Bernoulli-Gaussian impulsive noise," *IEEE Transactions on Power Delivery*, vol.30, no.1, pp.97–105, 2015.

[27] L. H. Xi, Y. Wang, Y. Wang, *et al.*, "Deep reinforcement learning-based service-oriented resource allocation in smart grids," *IEEE Access*, vol.9, pp.77637–77648, 2021.

[28] H. van Hasselt, A. Guez, and R. D. Silver, "Deep reinforcement learning with double Q-learning, " in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, AZ, USA, pp.2094–2100, 2015.

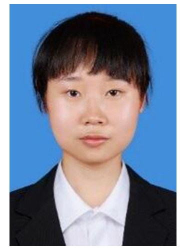
[29] Z. J. Zhang, "Improved Adam optimizer for deep neural networks," in *Proceedings of the 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, Banff, AB, Canada, pp.1–2, 2018.

[30] S. V. Zhidkov, "Analysis and comparison of several simple impulsive noise mitigation schemes for OFDM receivers,"

IEEE Transactions on Communications, vol.56, no.1, pp.5–9, 2008.

CHEN Zhixiong was born in 1983. He received the M.S. degree from Harbin Institute of Technology of China in 2007 and the Ph.D. degree in electrical engineering and its automation from North China Electric Power University of China in 2010. He is an Associate Professor at School of Electrical and Electronic Engineering, North China Electric Power University, China. His research interests include power line communications, smart grid communications and Internet of things. (Email: zxchen@ncepu.edu.cn)

ZHANG Zhikun (corresponding author) was born in 1995. He received the B.E. degree in communication engineering from North China Electric Power University in 2020. He is pursuing the M.S. candidate in information and communication engineering from North China Electric Power University. His current research interests include power line and wireless hybrid Communication.
(Email: zhikun_zhang@126.com)



CAO Tianshu was born in 1997. She received the B.E. degree in communication engineering from North China Electric Power University in 2021. She is currently pursuing the M.S. candidate in information and communication engineering from North China Electric Power University. Her current research interests include power line and wireless hybrid communication. (Email: 201703000401@ncepu.edu.cn)

ZHOU Zhenyu was born in 1983. He received the M.S. degree and Ph.D. degree from Waseda University, Tokyo, Japan in 2008 and 2011 respectively. From September 2012 to April 2019, he was an Associate Professor at School of Electrical and Electronic Engineering, North China Electric Power University, China. Since April 2019, he has been a Full Professor at the same university. His research interests mainly focus on resource allocation in device-to-device (D2D) communications, machine-to-machine (M2M) communications, smart grid communications, and Internet of things (IoT). He is a Senior Member of IEEE, Chinese Institute of Electronics (CIE), and China Institute of Communications (CIC).
(Email: zhenyu_zhou@ncepu.edu.cn)