ICC 2022 - IEEE International Conference on Communications | 978-1-5386-8347-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICC45855.2022.9839283

ADSA: A Multi-path Transmission Scheduling Algorithm based on Deep Reinforcement Learning in Vehicle Networks

Chenyang Yin[1], Ping Dong[1]*, Xiaojiang Du[2], Yuyang Zhang[1], Hongke Zhang[1]
[1]School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, 100044, P. R. China
[2]Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, 07030, USA
Email: {17111009, pdong, zhyy, hkzhang} @bjtu.edu.cn, dxj@ieee.org

***Abstract*—Cognitive Radio (CR) enabled vehicles in Vehicle Networks can use multiple Radio Access Networks (RANs) for data transmission. The simultaneous use of multiple RANs for transmission requires the design of a specific multi-path transmission protocol. Many scholars have studied the scheduling algorithm to improve the quality of multi-path transmission. However, most of the existing scheduling algorithms are difficult to deal with the challenges brought by the diversity and heterogeneity of the vehicle network. To deal with these challenges, this paper proposes an IP layer Deep Reinforcement Learning (DRL) multi-path transmission scheduling algorithm named Adaptive Dynamic Scheduling Algorithm (ADSA), which can dynamically generate the optimal scheduling policy through the interaction between agent and network environment. This paper first models the data packet scheduling strategy of multi-path transmission into an optimization problem of multi-path transmission efficiency. Then this paper transforms the optimization problem into a DRL problem and finds the optimal scheduling strategy through DRL model training. This paper evaluates the network performance of ADSA in different network scenarios compared with traditional scheduling algorithms. Simulation results show that ADSA increases the throughput by 8.9 Mbps compared with the three traditional scheduling algorithms and reduces the transmission delay by 4.3 ms.**

***Index Terms*—Vehicle networks, Cognitive radio, Multi-path transmission, Deep reinforcement learning**

I. Introduction

With the large-scale commercial use of 5G communication technology, vehicular applications have developed rapidly in Vehicle Networks. The vehicle-to-cloud transmission system uses the RAN provided by operators to perform data interaction between the vehicular application and the cloud server. However, the limited spectrum resources are hard to meet the spectrum demands of vehicular applications. To solve this problem, CR has been introduced into vehicular network communication. Operators use heterogeneous CR to optimize spectrum resources. Operators operate multiple RAN through the same or different Radio Access Technology (RAT) protocols and fix the frequency band of each RAN. With the development of the CR system in Vehicle Networks, CR-enabled vehicles in Vehicle Networks will be able to switch the data transmission from one RAN to another RAN to ensure transmission quality according to the service requirement and RAN link condition [1], [2]. However, a large number of studies have shown that the handoff of cellular networks in mobile scenarios will seriously degrade the performance of single RAN transmission [3]–[5]. The simultaneous use of multiple RANs for transmission has gradually become an important research direction for vehicle-to-cloud transmission to meet the increasing throughput and delay requirements of vehicular applications such as autonomous driving, online video, and AR/VR. [6]–[8].

For end-users, each RAN is abstracted as a transmission path. The simultaneous use of multiple paths for data transmission requires the design of a specific multi-path transmission protocol. Multi-path transmission protocols can be roughly classified into transmission layer multi-path transmission protocols and IP layer multi-path transmission protocols [9]–[11]. The deployment of transmission layer multi-path transmission protocol such as Multi-Path TCP (MPTCP) requires modifying the kernel protocol stack of traditional network devices, which will cause a lot of additional costs. This article uses the IP layer multi-path transmission protocol. In the IP layer multi-path transmission protocol, the user data packet is custom-encapsulated by the sending proxy and sent to the receiving proxy through multiple links. The receiving proxy decapsulates the data packet and forwards it to the application server. We refer to the sending proxy as a Smart Mobile Router (SMR) and the receiving proxy as a Smart Aggregation Router (SAR). Multi-path scheduling algorithm plays a vital role in multi-path transmission to improve transmission efficiency and quality. Improper scheduling algorithm will lead to Head-of-Line (HoL) blocking problem in multi-path transmission, which will lead to the long waiting delay of the data packet in SAR and cause excessive delay and worse network throughput.

Traditional scheduling algorithms such as lowest Round-Trip Time (RTT) First (lowRTT), Round-Robin (RR), and Weighted Round-Robin (WRR) are modeled for specific network scenarios. For example, lowRTT is suitable for network scenarios with significant link delay difference [12], In the real network environment, lowRTT allocates packets to the path with the shortest RTT for transmission. lowRTT is treated as the approximation of the Earliest Delivery Path First (EDPF) algorithm. WRR algorithm is suitable for network scenarios with significant link bandwidth difference [13]. RR

[1] Ping Dong is the corresponding author.

algorithm transmits data packets through available links in turns, so it is suitable for network scenarios with similar link states [3]. However, traditional scheduling algorithms can not adapt to the diversified vehicle-to-cloud communication network scenarios in Vehicle Networks. When the network scenarios change, traditional scheduling algorithms are difficult to optimize scheduling strategy. To solve the above problems, we propose a multi-path transmission scheduling algorithm based on DRL, which is named Adaptive Dynamic Scheduling Algorithm (ADSA). We first model the multi-path transmission scheduling problem in the vehicle network as a multi-objective optimization problem and then transform the optimization problem into a DRL model. Long Short Term Memory (LSTM) artificial neural network is used to extract features from the network information in multiple time slots. Deep Deterministic Policy Gradient (DDPG) method is used to perform interaction between agent and network environment to generate the corresponding scheduling strategy by maximizing the cumulative reward. Massive simulation experiments indicate that ADSA has increased the average throughput by 8.9 Mbps compared with the three traditional scheduling algorithms and reduced the average delay by 4.3 ms.

Our contributions in this paper can be summarized as follows.

1) We model the scheduling problem of multi-path transmission as a multi-objective optimization problem through queuing theory.
2) We transform the multi-objective optimization problem into a DRL problem and establish the corresponding state, action, and reward spaces. We construct the mapping from state space to action space by maximizing the cumulative rewards.
3) We propose a multi-path transmission scheduling algorithm based on DRL, which overcomes the problem that the traditional algorithm can not dynamically adjust the scheduling strategy in different vehicular network scenarios.

The structure of the paper is as follows. In Section II, we introduce the mathematical model for the packet scheduling of multi-path transmission. In Section III, we transform the multi-objective optimization problem into a DRL problem and introduce the ADSA algorithm in detail. In Section IV, we do a lot of simulation experiments to prove the effectiveness of ADSA in improving network throughput and reducing average delay. Finally, in Section V, we summarize this paper and propose the future research direction.

II. Multi-path Transmission System Model

In this section, we first derive the throughput and average delay of IP layer multi-path transmission and then model the transmission performance of IP layer multi-path transmission as a multi-objective optimization.

As shown in Fig. 1, the IP layer multi-path transmission protocol deploys a pair of SMR and SAR between the user terminal and the application server. The SMR deployed in the vehicle encapsulates the data packets transmitted from the user terminal and sends encapsulated data packets to the SAR deployed in the cloud through heterogeneous wireless links provided by three operators in China. The SAR decapsulates the encapsulated data packets and sends them to the application server. The following network simulation experiments of this paper also adopt the network topology shown in Fig. 1.

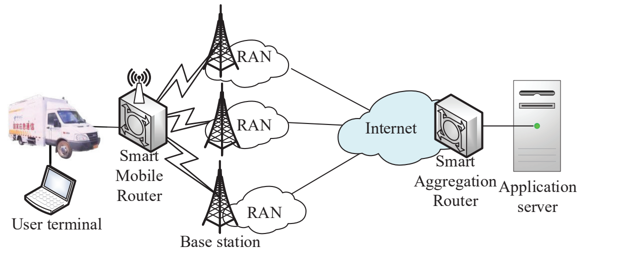


Fig. 1. The network topology of multi-path transmission in Vehicle Network.

Let p_i be the scheduling allocation ratio of the i th link in multi-path transmission and num_i represent the number of data packets that are transmitted through the i th link. Assuming that the number of data packets transmitted from the user terminal to the SMR within time t is $N(t)$, where $N(t)$ obeys the Poisson process, that is, $P\{N(t) = n\} = \frac{(\lambda t)^n}{n!} e^{-(\lambda t)}$. λ is the data packet arrival rate of the SMR. Therefore, num_i can be regarded as a compound Poisson distribution $num_i = \sum_{j=1}^{N(t)} X_j^i$, where X_j^i represents the probability distribution of the j th data packet being allocated to the i th link. X_j^i is assumed to be independent and identically distributed and obeys a Bernoulli distribution $B(x^i, p_i)$. The expectation $E[num_i]$ and variance $D[num_i]$ of num_i within time t can thereby be calculated as $p_i \lambda$. Therefore, num_i can be regarded as a Poisson process with a parameter value of $p_i \lambda t$. Suppose the buffer size of the i th link in the SMR is C_i, and the transmission time interval of each data packet in the i th link obeys the negative exponential distribution. Therefore, the transmission of the i th single link can be regarded as an $M/M/1/C_i/\infty$ queuing system, where ∞ indicates that the number of data packets sent to the i th link is infinite. For convenience of representation, let $p_i \lambda$ be denoted as λ_i. Let $\mu_i = \frac{BW_i}{len_i}$ represent the service rate of the i th link, where len_i represents the average size (bytes) of the data packet. Service intensity can be represented as $\rho_i = \frac{\lambda_i}{\mu_i}$. According to the $M/M/1/C_i/\infty$ queuing model, the number of data packets transmitted through the i th link in multi-path transmission can be calculated as follows:

$$num_i = \frac{\rho_i}{1-\rho_i} - \frac{(C_i+1)\rho_i^{C_i+1}}{1-\rho_i^{C_i+1}}. \tag{1}$$

The transmission delay of the data packet sent to the i th link after being encapsulated by SMR can be calculated as follows:

$$T_i^s = \frac{num_i}{\mu_i(1-p_i^{(0)})}. \tag{2}$$

Where $p_i^{(0)} = \frac{1-\rho_i}{1-\rho_i^{C_i+1}}$ is the probability that the number of data packets allocated to the i th link in multi-path transmission is 0. The total delay of the data packet in the i th link is $T_i = T_i^s + T_i^{RTT}$, where T_i^{RTT} can be expressed as the sum of propagation delay, processing delay and queuing delay. The average delay of multi-path transmission is $T_{ave} = E_{p_i}[T_i]$. The throughput of multi-path transmission can be calculated as follows:

$$throughput = \sum_{i=1}^{|M|} \frac{num_i * (1 - p_i^{PLR}) * len_i * 8}{T_i}. \tag{3}$$

Where M is the set of available links for multi-path transmission and p_i^{PLR} represents the Packet Lost Rate (PLR) of the i th link. We denote the weight parameters of throughput and average delay as α and β, and $\alpha + \beta = 1$. Therefore, the multi-objective optimization can be formulated as:

$$\max_{p_i} \quad \alpha * throughput - \beta * T_{ave} \tag{4}$$

$$\text{s.t.} \quad 0 \le p_i \le 1, \qquad i = 1, 2, \cdots, |M| \tag{5}$$

$$\sum_{i=1}^{|M|} p_i = 1. \tag{6}$$

III. ADSA Scheduling Algorithm

In this section, we first introduce the DDPG model and transform the multi-objective optimization problem into a DRL problem in Section III-A. In Section III-B, we introduce the ADSA algorithm in detail.

A. DDPG Model Transformation

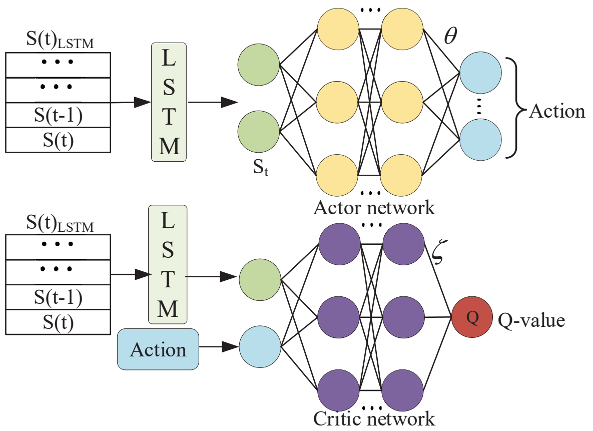


Fig. 2. The structures of actor network and critic network.

As a paradigm of artificial intelligence, reinforcement learning is widely used in solving optimization problems [14]. Reinforcement learning can generate corresponding action according to the state that the agent obtains from the environment, and continuously optimize the action strategy by maximizing the cumulative reward.

In this paper, we use the probability distribution $P = (p_1, p_2, \cdots, p_{|M|})$ to represent the scheduling action of the IP layer multi-path transmission, where p_i represents the probability that the data packet obtained by the SMR from the user terminal is transmitted to the SAR through the i th link. Due to the value range of p_i is 0 to 1, so the reinforcement learning model needs to be able to process continuous actions. Compared with Deep Q-Leaning Network (DQN), which can only process discrete actions, the DDPG model is a combination of Policy-Based and Value-Based, which can process the continuous action space more effectively. At the same time, DDPG improves the convergence ability of model training through the method of experience playback and dual network architecture.

Before introducing the DDPG network architecture, we first give the definition of state S, action A, and reward R in DRL according to the optimization goals and optimization variables in the previous section. The action $A = \{A_1, A_2, \cdots, A_{|M|}\}$ corresponds to the optimization variable p_i in the multi-objective optimization, and the size of the action vector is $|M| \times 1$.

The input state S contains 5 column vectors, which are the link bandwidth column vector S_1, the link RTT column vector S_2, the link data packet average size column vector S_3, the link packet loss rate column vector S_4, and the transmission performance column vector S_5. The size of S_1, S_2, S_3, and S_4 are all $|M| \times 1$, and the size of S_5 is 2×1. We use $S_1(i)$, $S_2(i)$, $S_3(i)$, and $S_4(i)$ to represent the i th element in the column vector, which corresponds to the bandwidth, RTT, average packet size, and packet loss rate of the i th link in the previous Section II, respectively. The first element $S_5(1)$ in S_5 corresponds to the throughput of multi-path transmission $throughput$, the second element $S_5(2)$ corresponds to the delay expectation of multi-path transmission T_{ave} in the previous Section II. S can be expressed as the following equation:

$$S = \{S_1^\top, S_2^\top, S_3^\top, S_4^\top, S_5^\top\}_{1\times(4*|M|+2)}. \tag{7}$$

We use $S(t)$ to represent the state collected by the agent in time slot t, and $S(t)_{LSTM} = \{S(t), S(t-1), \cdots S(t-Step+1)\}$ to represent the input of the LSTM network in Fig. 2, $Step$ represents the step size of LSTM input. The output of LSTM corresponds to the input S_t of the actor network in the DDPG model. The output of LSTM and the output of the actor network in the DDPG model together form the input of the critic network in the DDPG model.

R corresponds to the action A generated by agent under the state S at the moment of t, and the model receives R at the moment of $t+1$. We set $R = \alpha*(S_{5,1} - thr_{max}) + \beta*thr_{max}*(delay_{min} - S_{5,2})/(delay_{max} - delay_{min})$, where thr_{max} indicates the maximum throughput of IP layer multi-path transmission. $delay_{max}$ and $delay_{min}$ indicate the maximum total delay and the minimum total delay in IP layer multi-path transmission. Strategy refers to the probability that the agent will take action A under state S, that is $\pi(A|S) = P(A_t|S_t)$, where A_t and S_t represent the scheduling action and network state in time t. Therefore, $Q_\pi(S, A)$ is formulated as follows:

$$Q_\pi(S, A) = \mathbb{E}_\pi(\sum_{i=1} \gamma^{i-1} R_{t+i} | S_t, A_t). \tag{8}$$

γ is the reward attenuation factor, the range of γ is $[0, 1]$. $\gamma^{i-1} R_{t+i}$ indicates the influence of subsequent rewards on the value function. The farther the subsequent rewards are from the current time slot, the smaller the influence of subsequent rewards on the value function. $Q_\pi(S, A)$ is the output of the critic network in the DDPG model. We usually approximate strategy $\pi(A|S)$ to $\pi_\theta(S, A)$ and approximate current Q value $Q_\pi(S, A)$ to $Q(\zeta, S, A)$ in DDPG model. θ and ζ represent the parameters of actor network and critic network.

The DDPG model has four network architectures, namely evaluation-actor network, evaluation-critic network, target-actor network, and target-critic network. The evaluation-actor network generates scheduling action A with the state S. According to the sampled next state $\hat{S}$, target-actor network generates next scheduling action $\hat{A}$. The evaluation-critic network calculates current Q value $Q(\zeta, S, A)$ with the state S and scheduling action A. The target-critic network calculates target Q value $\hat{Q}(\hat{\zeta}, \hat{S}, \hat{A})$ with the sampled next state $\hat{S}$ and next scheduling action $\hat{A}$.

For the evaluation-critic network, its loss function is similar to DQN, which is mean square errors. The specific formula of the loss function of the evaluation-critic network is $J(\zeta) = \frac{1}{n} \sum_{k=1}^{n} (V_k - Q(\zeta, S_k, A_k))^2$. n is the number of training samples for gradient descent. We use $terminated_k$ to indicate whether the k th interaction is over, normally set $terminated_k$ to 0, V_k thereby is calculated as follows:

$$V_k = \begin{cases} R_k & terminated_k \ is \ 1, \\ R_k + \gamma \hat{Q}(\hat{\zeta}, \hat{S}_k, \pi_{\hat{\theta}}(\hat{S}_k)) & terminated_k \ is \ 0. \end{cases} \tag{9}$$

The role of the loss function of the evaluation-actor network can be understood as increasing the probability of action corresponding to a larger Q value and reducing the probability of action with a smaller Q value. The specific formula of the loss function of the evaluation-actor network is $J(\theta) = -\frac{1}{n} \sum_{k=1}^{n} Q_(\zeta, S_k, A_k)$. When the evaluation network has been optimized a certain number of times, the target network will exchange parameters with the evaluation network. Let τ be denoted as soft exchange factor, the exchange formula thereby is as follows:

$$\begin{aligned} \hat{\zeta} &\leftarrow \tau\zeta + (1-\tau)\hat{\zeta}, \\ \hat{\theta} &\leftarrow \tau\theta + (1-\tau)\hat{\theta}. \end{aligned} \tag{10}$$

B. The Specific Design of ADSA Algorithm

The proposed ADSA algorithm includes two parts: model training and multi-path transmission scheduling. The model training part mainly trains the DDPG model by collecting relevant network states. There are E episodes in the whole training process, and each episode contains TS training steps. We set the soft update factor as τ and the number of samples for batch gradient descent as n. Let C be denoted as the parameter exchange frequency of target network, and $\widehat{N}$ represents random noise to increase the exploration ability of the generated scheduling action A. The specific algorithm is listed in Algorithm 1.

Algorithm 1: Model training part of ADSA Algorithm

```
 1: Parameters $\{\zeta, \theta, \hat{\zeta}, \hat{\theta}\}$ initialization
 2: for each episode $e \in$ [1,E] do
 3:    for each trian step $ts \in$ [1,TS] do
 4:       Initialize $S$ and input $S$ into evaluation-actor
          network
 5:       Generate scheduling action $A$ from evaluation-actor
          network and add noise $\widehat{N}$ to $A$
 6:       Agent executes $A$ and gets $\hat{S}$, $R$, $terminated$
 7:       Add $\{S, A, \hat{S}, R, terminated\}$ into experience
          playback buffer $D$ set and $S = \hat{S}$
 8:       Get $n$ samples from $D$ and calculate $J(\zeta)$, $J(\theta)$
 9:       Update $\zeta$ and $\theta$
10:       if $ts$ mod $C$==0 then
11:          parameters soft exchange in Equation (10)
12:       end if
13:       if $terminated$ is 1 then
14:          The current round of iteration is completed
15:       else
16:          Go to Step 5
17:       end if
18:    end for
19: end for
```

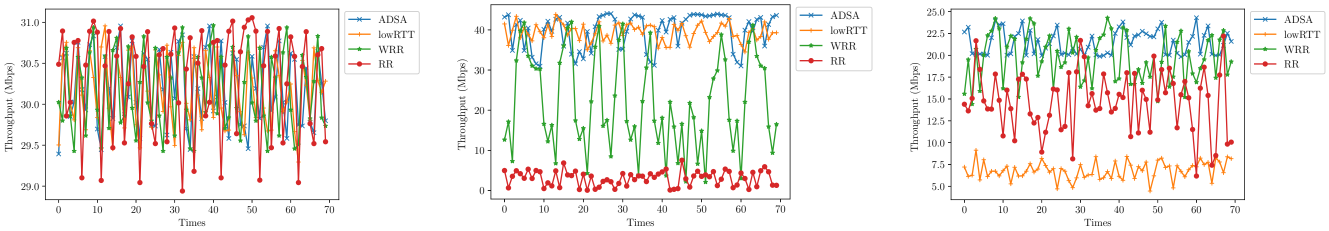
The multi-path transmission scheduling part schedules packets into different available links with the help of the trained model. Algorithm 2 generates a random number l $(0 \leq l \leq 1)$ for each packet transmitted by SMR to determine which link k the packet belongs to by comparing the size of the random number and the sum of different link allocation proportions p_{sum}.

Algorithm 2: Packet scheduling part of ADSA Algorithm

```
for each packet $u$ do
   Trained model obtains action $A = \{p_1, p_2, .., p_{|M|}\}$
   based on state $S$
   Generate random number $l$ in [0,1], $k = 1$, $flag = 1$,
   $p_{sum} = p_1$
   while flag==1 do
      if $l \leq p_{sum}$ then
         $flag = 0$
      else
         $k = k + 1$, $p_{sum} += p_k$
      end if
   end while
   return $k$
end for
```

IV. Transmission Performance Evaluation

We evaluate ADSA algorithm from two aspects: throughput and average delay. Throughput refers to the amount of successfully transmitted data per unit of time for a network device. We count the network throughput every 100 ms. Average



(a) Throughput comparison of scheduling algorithms in the communication environment with no significant difference between different links.

(b) Throughput comparison of different scheduling algorithms in the communication environment with significant link delay difference between different links.

(c) Throughput comparison of different scheduling algorithms in the communication environment with significant link bandwidth difference between different links.

Fig. 3. The throughput performances of different scheduling algorithms in different network scenarios.

delay refers to the average value of the time interval for the transmission of data packets from the sender to the receiver. After each experiment, we calculate the average delay of multi-path transmission with different scheduling algorithms. We use NS3 gym version 3.29 as the simulation platform, which is composed of Network Simulator 3 (NS3) network simulator and openAI gym framework. NS3 is an Internet system discrete-event network simulator. IP layer multi-path transmission protocol is deployed in the NS3 platform. The openAI gym framework is a toolkit for developing and comparing reinforcement learning algorithms. The DDPG model is deployed in the openAI gym framework. NS3 gym is used to seamlessly integrate the two frameworks to obtain observations from NS3 and use it for deep reinforcement learning in the openAI gym framework [15].

TABLE I
DIFFERENT CONFIGURATIONS OF BANDWIDTH AND DELAY.

Config.	Link A	Link B	Link C
1	100 Mbps 15 ms	100 Mbps 15 ms	100 Mbps 15 ms
2	100 Mbps 10 ms	100 Mbps 15 ms	75 Mbps 500 ms
3	100 Mbps 20 ms	10 Mbps 15 ms	10 Mbps 10 ms

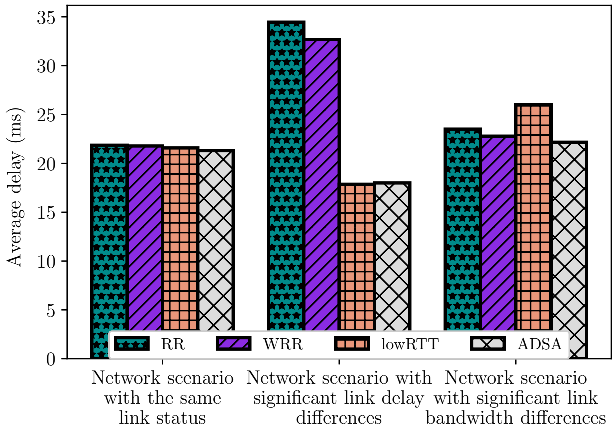


Fig. 4. The delay comparisons of different scheduling algorithms in different network scenarios.

We simulate three different types of network scenarios by changing the bandwidth and delay of different links for IP layer multi-path transmission in simulation. As shown in Tab. I, Config. 1 represents a network environment without significant link difference. The bandwidths and delays of the three transmission links for multi-path transmission are set to 100 Mbps and 15 ms. Config. 2 represents the network environment with significant link delay difference. The bandwidths of the three transmission links are set to 100, 100, and 75 Mbps respectively, and the delays of the three transmission links are set to 10, 15, and 500 ms respectively. Config. 3 represents the network environment with significant link bandwidth difference. The bandwidths of the three transmission links are set to 100, 10, and 10 Mbps respectively, and the delays of the three transmission links are set to 20, 15, and 10 ms respectively. Under three different network environments, we evaluate the throughput performance of IP layer multi-path transmission deployed with lowRTT, WRR, RR, and ADSA scheduling algorithms, respectively. The test results are shown in Fig. 3. At the same time, we also evaluate the average delay of IP layer multi-path transmission deployed with lowRTT, WRR, RR, and ADSA scheduling algorithms. The delay performances are shown in Fig. 4.

As shown in Fig. 3(a), when there is no significant difference in link bandwidth and delay, the transmission throughput performances of ADSA, lowRTT, WRR, and RR algorithms are the same, and the average throughput is 30 Mbps. When the link delay difference is significant, the throughput performances of different scheduling algorithms are shown in Fig. 3(b). Since the WRR algorithm uses the link bandwidth as the weight parameter of data packet allocation proportion for different links, WRR is not sensitive to the link delay difference. Some data packets are allocated to link C with the largest delay for transmission, resulting in a certain degree of out-of-order packets problem. Therefore, the throughput performance of WRR decreases, and the average throughput of WRR is about 20 Mbps. Since the RR algorithm does not interact with the network environment, $1/3$ of the data packets are allocated to each link for transmission, so the transmission throughput also decreases. Moreover, the data packets allocated to link C are more than those allocated to link C by WRR, so the throughput performance of RR is worse than that of WRR, and the average throughput of RR is 3 Mbps. The lowRTT algorithm only transmits through the link with the smallest link delay, so lowRTT is not affected by the delay difference between links, and the average throughput of lowRTT is 38 Mbps. ADSA trains the model through the information obtained by interacting with the network environment to continuously optimize the scheduling strategy and avoid allocating data packets to link C for transmission. Therefore, the average throughput of ADSA is 38 Mbps. When the link bandwidth difference is significant, the throughput performances of different algorithms are shown in Fig. 3(c). In contrast to Fig. 3(b), WRR is sensitive to link bandwidth difference, while the lowRTT algorithm does not consider link bandwidth difference. Therefore, the average throughput of the WRR algorithm is 17 Mbps and that of lowRTT is 6 Mbps. The average throughput of the RR algorithm is 15 Mbps. ADSA optimizes the scheduling strategy in this scenario through model training and reduces the number of data packets transmitted through links B and C. Therefore, the average throughput of ADSA is 22 Mbps.

As shown in Fig. 4, the abscissa represents different network scenarios and the ordinate represents the average delays of different algorithms in different network scenarios. The first histogram represents a network scenario with no significant difference between links. The second histogram represents a network scenario with significant delay difference between links. The third histogram represents a network scenario with significant bandwidth difference between links. The average delay of ADSA is always at a low level in different network scenarios. However, the average delays of WRR, RR, and lowRTT algorithms will increase due to changes in network scenarios. For example, in a network scenario where the link delay difference is significant, the average delay of lowRTT is as small as 18 ms. When the network scenario changes and the bandwidth difference between links becomes more significant, the average delay of lowRTT increases to 26 ms. In the network scenario with significant difference in link bandwidth, the average delay of WRR is 22 ms. When the network scenario changes and the delay differences between links becomes more significant, the average delay of WRR increases to 32 ms.

From the above two sets of simulation experiments, it can be found that ADSA can effectively guarantee transmission performance in different vehicle network scenarios. Considering the average transmission performance of scheduling algorithms in three network scenarios, ADSA has increased the average throughput by 8.9 Mbps compared with the three traditional algorithms and reduced the average delay by 4.3 ms.

V. Conclusion

To solve the problem that traditional multi-path transmission scheduling algorithms are difficult to adapt to different network scenarios in Vehicle Networks, we proposed a scheduling algorithm based on deep reinforcement learning named ADSA. We converted the scheduling problem of multi-path transmission into a deep reinforcement learning model. With the interaction of network scenarios, the scheduling algorithm is continuously optimized to improve the transmission performance of the multi-path transmission. The simulation results showed that ADSA can adaptively change the scheduling strategy in different network scenarios to improve the transmission throughput and reduce the average transmission delay of the multi-path transmission. In the future, we will study the online training and rapid deployment of DRL models in multi-path transmission.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61872029 and 61972026, and in part by the National Key R&D Program of China under Grant 2018YFB1800305.

References

[1] X. Du, D. Wu, W. Liu, and Y. Fang, “Multiclass routing and medium access control for heterogeneous mobile ad hoc networks,” *IEEE Transactions on Vehicular Technology*, vol. 55, no. 1, pp. 270–277, 2006.

[2] D. Wang, P. Qi, Q. Fu, N. Zhang, and Z. Li, “Multiple high-order cumulants-based spectrum sensing in full-duplex-enabled cognitive iot networks,” *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 9330–9343, 2021.

[3] C. Paasch, S. Ferlin, O. Alay, and O. Bonaventure, “Experimental evaluation of multipath tcp schedulers,” in *Proceedings of the 2014 ACM SIGCOMM workshop*. Chicago,IL,USA: ACM, 2014, pp. 27–32.

[4] L. Li, K. Xu, T. Li, K. Zheng, C. Peng, D. Wang, X. Wang, M. Shen, and R. Mijumbi, “A measurement study on multi-path tcp with multiple cellular carriers on high speed rails,” in *Proceedings of the 2018 ACM SIGCOMM*. Budapest, Hungary: ACM, 2018, pp. 161–175.

[5] Y. Xiao, K. K. Leung, Y. Pan, and X. Du, “Architecture, mobility management, and quality of service for integrated 3g and wlan networks,” *Wireless Communications and Mobile Computing*, vol. 5, no. 7, pp. 805–823, 2005.

[6] P. Dong, B. Song, H. Zhang, and X. Du, “Improving onboard internet services for high-speed vehicles by multipath transmission in heterogeneous wireless networks,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 9493–9507, 2016.

[7] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, “Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey,” *Computer networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[8] G. Sun, Y. Li, H. Yu, A. V. Vasilakos, X. Du, and M. Guizani, “Energy-efficient and traffic-aware service function chaining orchestration in multi-domain networks,” *Future Generation Computer Systems*, vol. 91, pp. 347–360, 2019.

[9] M. Li, A. Lukyanenko, Z. Ou, A. Ylä-Jääski, S. Tarkoma, M. Coudron, and S. Secci, “Multipath transmission for the internet: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2887–2925, 2016.

[10] X. Du, “Qos routing based on multi-class nodes for mobile ad hoc networks,” *Ad Hoc Networks*, vol. 2, no. 3, pp. 241–254, 2004.

[11] X. Du and F. Lin, “Designing efficient routing protocol for heterogeneous sensor networks,” in *Proceedings of the 24th IEEE International PCCC*. Phoenix, AZ, USA: IEEE, 2005, pp. 51–58.

[12] K. Chebrolu and R. Rao, “Communication using multiple wireless interfaces,” in *Proceedings of the 2002 IEEE WCNC*. Orlando, FL, USA: IEEE, 2002, pp. 327–331.

[13] L. B. Le, E. Hossain, and A. S. Alfa, “Service differentiation in multirate wireless networks with weighted round-robin scheduling and arq-based error control,” *IEEE Transactions on Communications*, vol. 54, no. 2, pp. 208–215, 2006.

[14] W. Zhang, D. Yang, H. Peng, W. Wu, W. Quan, H. Zhang, and X. Shen, “Deep reinforcement learning based resource management for dnn inference in industrial iot,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7605–7618, 2021.

[15] P. Gawłowicz and A. Zubow, “Ns-3 meets openai gym: The playground for machine learning in networking research,” in *Proceedings of the 22nd ACM MSWIM*. Miami, FL, USA: ACM, 2019, pp. 113–120.