2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) | 978-1-6654-6483-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/PIMRC56721.2023.10293768

Deep Reinforcement Learning Enabled Power Allocation for Multi-Connectivity C-V2X Downlink

Jianzhe Xue*, Kai Yu*, Tianqi Zhang*, Haibo Zhou*, and Xuemin (Sherman) Shen†
*School of Electronic Science and Engineering, Nanjing University, Nanjing, China, 210023.
†Department of Electrical and Computer Engineering, University of Waterloo, Waterloo Ontario N2L 3G1, Canada.
Email: {jianzhexue, kaiyu, tianqizhang}@smail.nju.edu.cn, haibozhou@nju.edu.cn, and sshen@uwaterloo.ca.

***Abstract*—Cellular vehicle-to-everything (C-V2X) network is a promising solution to support on road diverse quality of services (QoS) such as ultra reliable low latency communication (URLLC) and enhanced mobile broadband (eMBB). However, satisfying the stringent QoS requirements in high-dynamic C-V2X environment is very challenge. In this paper, we leverage the multi-connectivity technology to enhance the reliability of downlink URLLC in C-V2X. Specifically, with the aid of the cloud radio access network (C-RAN), the network controller duplicates each URLLC packet and transmits its replicas over multiple independent wireless links. To ensure the reliability of URLLC links while maximizing the average rate of eMBB links, we design a coordinated multi-agent deep reinforcement learning algorithm for real-time power allocation of multi-connectivity URLLC links. Each URLLC link is treated as an agent here, and its transmit power is its action. The multiple links serving the same URLLC user are coordinated with a three-layer neural network for information sharing, allowing them to cooperatively choose transmit powers in terms of ensuring reliability while minimizing inter-cell interference and energy consumption. Extensive simulation results validate the effectiveness of the proposed power allocation algorithm for multi-connectivity downlink URLLC.**

***Index Terms*—multi-connectivity, packet duplication, URLLC, deep reinforcement learning, C-V2X.**

I. Introduction

The cellular vehicle-to-everything (C-V2X) network is a key infrastructure for intelligent transport systems (ITS), offering various quality-of-service (QoS) communication services such as ultra-reliable low latency communication (URLLC) and enhanced mobile broadband (eMBB) [1], [2]. The 3GPP has defined URLLC as a case for 5G systems, requiring packet reliability of over 99.999% and user plane transmission latency within 1 ms [3]. The downlink URLLC in C-V2X is vital for vehicles to receive safe driving information from ITS controller and has stringent QoS requirements [4]. To reduce air interface transmission delays and achieve more rapid and flexible scheduling, the transmit time interval (TTI) of URLLC is divided into mini-slots, and the URLLC packet is correspondingly downsized [5]. However, a single packet transmission in high dynamic C-V2X network may fail, especially for short blocklength packets, while transmitting the same packet on the same link multiple times will increase the latency [6]. The emergence of the cloud radio access network (C-RAN) allows remote radio heads (RRHs) to support collaborative radio technologys, such as multi-connectivity [7]. To enhance URLLC real-time reliability, we propose a redundant transmission strategy with multi-connectivity architecture, where the same URLLC packet is duplicated and transmitted simultaneously on different independent wireless links to the user equipment (UE), enhancing its reliability in one mini-slot. In this way, the URLLC packet will be successfully transmitted to the UE if one of these replicas from multiple links is decoded successfully.

For ultra reliability, the high signal-to-noise ratio is required by the URLLC link. However, simply increasing transmit powers of URLLC links can lead to excessive power consumption and high inter-cell interference, especially in multi-connectivity scenarios. Moreover, the edge cloud controller have to decide the transmit powers of URLLC links immediately upon packet arrival, necessitating a real-time power allocation algorithm for the complex multi-cellular multi-user C-V2X system. However, traditional optimization-based algorithms are challenging to execute in real-time since they require updating the power allocation variables by solving the non-convex problem of each time slot [8]. Deep reinforcement learning (DRL) can provide efficient solutions for real-time applications by leveraging the fast forward propagation of neural network (NN) [9]. A multi-agent DRL with the sequential actor-critic model has been investigated to optimize the URLLC satisfaction of delay and reliability in a multi-connectivity cellular network [10].

In this paper, we propose a multi-connectivity URLLC downlink framework for C-V2X and a corresponding multi-agent DRL algorithm for power allocation. Specifically, URLLC shares the wireless communication resource with eMBB in a dynamic multi-cellular multi-user C-V2X downlink scenario. The non-coherent transmission scheme is used for multi-connectivity URLLC, where the URLLC packet is duplicated and transmitted on multiple independent links simultaneously to the UE. In such case, the URLLC packet transmission will be failed only when all these links are failed. The reliability of each URLLC link is estimated with a quantitative definition of network availability in the short blocklength regime [11]. Besides, we model the power allocation problem as a multi-agent DRL framework, where each URLLC link is regarded as an agent so that it can adopt the dynamic link amount. A three-layer NN is applied for information sharing between the multiple links that are serving the same URLLC user. Our proposed framework is extensively evaluated through simulations, demonstrating its effectiveness in improving URLLC performance. The three main contributions

978-1-6654-6483-3/23/$31.00 ©2023 IEEE

of this paper are summarized as the following:

- We investigate the multi-connectivity packet duplication framework for URLLC downlink in C-V2X network, in which the URLLC packet replicas are transmitted simultaneously through multiple independent wireless links.
- We model the URLLC packet reliability in the short blocklength regime for the multi-connectivity packet duplication case using a quantitative approach that is more suitable for mini-slot TTI.
- We propose a multi-agent DRL algorithm, named coordinated proximal policy optimization (CPPO), for power allocation of multi-connectivity URLLC links to achieve real-time control and adapt dynamic network users.

The rest of this paper is organized as follows. Section II describes the system model of multi-connectivity URLLC with eMBB in C-V2X networks and Section III elaborate the proposed DRL algorithm. The simulation results are demonstrated in section IV. Finally, this work is concluded in Section V.

II. System Model

We consider a multi-connectivity downlink C-V2X network, as shown in Fig. 1, in which each UE can connect to one or more RRHs with multiple independent wireless links for simultaneous data packet transmission. The non-coherent transmission technology is applied for multi-connectivity, in which a wireless link among multi-connectivity is defined as a connection between the UE and a RRH on a sub-carrier and the links serving the same UE are using different frequency sub-carriers. In our model, we consider two specialized QoS services required by C-V2X network UEs, including URLLC and eMBB.

We assume the downlink channel is a multiple input single output (MISO) channel, in which the vehicle has a single receive antenna and the RRH is equipped with N_{T} transmit antennas. The bandwidth of each subcarrier and the duration of one mini-slot transmission time interval (TTI) are denoted by W and τ, respectively. σ^2 is the noise power on each subcarrier. All RRHs are using the same frequency range and each of them has N_{C} sub-carriers in total. We assume the channel state information (CSI) is known at the RRH and the maximum ratio transmission (MRT) is applied for precoding. The useful signal power $y_{k,b,n}$ at the k-th UE from the b-th RRH on the n-th subcarrier can be represented as,

$$y_{k,b,n} = \alpha_{k,b,n}|\mathbf{h}_{k,b,n}\boldsymbol{w}_{k,b,n}|^2 P_{k,b,n}, \tag{1}$$

where $\alpha_{k,b,n}$ is the large scale channel gain, $\mathbf{h}_{k,b,n} \in \mathbb{C}^{N_{\mathrm{T}}}$ is the normalized channel coefficient, $\boldsymbol{w}_{k,b,n}$ is the precoding weights for transmission and $P_{k,b,n}$ is the transmit power. Each element of $\mathbf{h}_{k,b,n}$ follows the complex Gaussian distribution with zero mean and unit variance. The channels are block fading in both frequency and time domains, and the channel coefficient on different sub-carriers allocated to the same UE are independent and identically distributed. The MRT precoding weights at the b-th RRH for the k-th UE on the n-th subcarrier is obtained as,

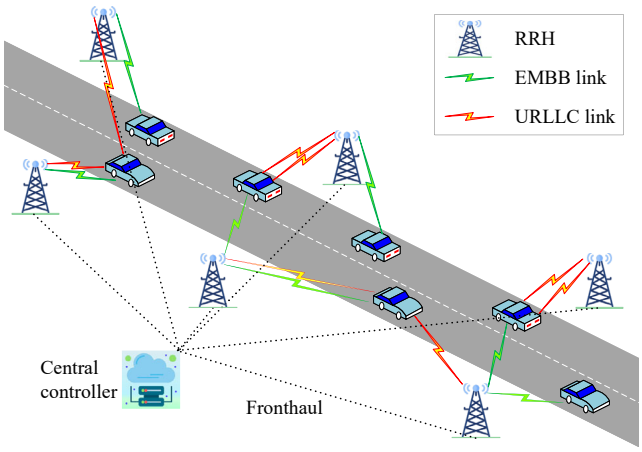


Fig. 1: System model.

$$\boldsymbol{w}_{k,b,n} = \frac{\sqrt{\alpha_{k,b,n}}\mathbf{h}^*_{k,b,n}}{||\sqrt{\alpha_{k,b,n}}\mathbf{h}_{k,b,n}||_2}, \tag{2}$$

where $\mathbf{h}^*_{k,b,n}$ is the conjugate transpose of $\mathbf{h}_{k,b,n}$. Similarly, the interference signal power at the k-th UE on the n-th subcarrier is,

$$\gamma_{k,n} = \sum_{l \in \mathcal{B}^{\mathrm{I}}} \alpha_{k,l,n}|\mathbf{h}_{k,l,n}\boldsymbol{w}_{l,n}|^2 P_{l,n}\rho_{l,n}, \tag{3}$$

where $\alpha_{k,l,n}$ is the large scale channel gain, $\mathbf{h}_{k,l,n} \in \mathbb{C}^{N_{\mathrm{T}}}$ is the normalized channel coefficient from the l-th interference RRH to the k-th UE on the n-th subcarrier, $\boldsymbol{w}_{l,n}$ is the precoding weights, $P_{l,n}$ is the transmit power at the l-th interference RRH on the n-th subcarrier, and $\rho_{l,n}$ is a binary spectrum allocation indicator with $\rho_{l,n} = 0$ implying the l-th interference RRH is vacant on the n-th subcarrier and $\rho_{l,n} = 1$ implying it is used.

For the eMBB service, we assume the data flow is continuous and the transmit power of the eMBB link is constant. The eMBB service always occupies one sub-carrier to download information. The demand for eMBB is to maximize the long-term average transmission rate of UEs. Based on the Shannon theorem, the downlink rate of the eMBB link for the k-th UE connecting to the b-th RRH at time t is calculated as,

$$c_k = W\log_2(1 + \frac{y_{k,b,n}}{\gamma_{k,n} + \sigma^2}). \tag{4}$$

For URLLC, it has a short packet size and a short blocklength of channel coding. We assume the URLLC packet arrival process follows a Bernoulli process and the packet is transmitted immediately after its arrival to guarantee its latency demand. In addition, the reliability evaluation of URLLC packets cannot be satisfied by the Shannon capacity alone, as decoding errors in small packets cannot be ignored. The achievable rate of the URLLC link can be approximated

according to the normal approximation of the achievable rate in the short blocklength regime as [11],

$$c^{\mathrm{U}}_{k,b,n} \approx \frac{W}{\ln 2}\left\{\ln(1+\frac{y_{k,b,n}}{\gamma_{k,n}+\sigma^2}) - \sqrt{\frac{\Omega}{\tau W}}\mathcal{Q}_G^{-1}(\delta_{k,b,n})\right\}, \tag{5}$$

where τ is the data transmission duration, $\delta_{k,b,n}$ is the packet decoding error probability for the k-th UE connecting to the b-th RRH at time t, $\mathcal{Q}_G(\cdot)$ is the Gaussian Q-function and Ω is the channel dispersion given by [12],

$$\Omega = 1 - \frac{1}{(1+\frac{y_{k,b,n}}{\gamma_{k,n}+\sigma^2})^2}. \tag{6}$$

Note that Ω is accurately approximate to 1 when the SNR is higher than 5 dB, which can easily achieved in cellular network especially for URLLC service [13]. Then, the decoding error probability of an URLLC transmission packet between the k-th UE from the b-th RRH on the n-th subcarrier can be derived as,

$$\delta_{k,b,n} = \mathcal{Q}_G\left\{\sqrt{\tau W}\left[ln(1+\frac{y_{k,b,n}}{\gamma_{k,n}+\sigma^2}) - \frac{a^{\mathrm{U}}_{k,b} ln2}{\tau W}\right]\right\}, \tag{7}$$

where $a^{\mathrm{U}}_{k,b}$ is the packet length of the URLLC packet.

In the case of packet duplication multi-connectivity URLLC, each URLLC packet is duplicated and transmitted on multiple independent wireless links simultaneously and independently to improve the packet reliability in the time of one mini-slot. The URLLC packet will be lost only when all transmissions on different links fail at the same time. Since the decoding failure of each link is an independent and uncorrelated event, the transmission failure probability of a multi-connectivity URLLC packet can be obtained as [14],

$$e_k = \prod_{\mathrm{m}=1}^{\mathrm{M_k}} \delta^m_{k,b,n}, \tag{8}$$

where M_k is the number of links that are used to transmit an URLLC packet simultaneously for k-th UE, $\delta^m_{k,b,n}$ is transmission failure probability of the m-th link of the k-th UE. The URLLC reliability is expected to be above 99.999%, which requires e_k to be less than 10^{-5}.

III. DRL for Power Allocation

The structure of our DRL framework for power allocation of multi-connectivity URLLC is shown in Fig. 2. The environment is a simulator of the communication system that includes eMBB and URLLC. These multi-connectivity URLLC links transmit replicas of the packet to the target UE with transmit powers given by the DRL algorithm. Then, the environment gives the next state and the reward to the DRL algorithm.

The number of active URLLC links in each time slot is dynamic since the demand of URLLC transmission is dynamic. However, the input and output sizes of the NN are fixed, which cannot adapt to the state and action space dynamics. Therefore, each URLLC link is regarded as an agent to obtain the fixed state and action space. So that the challenge

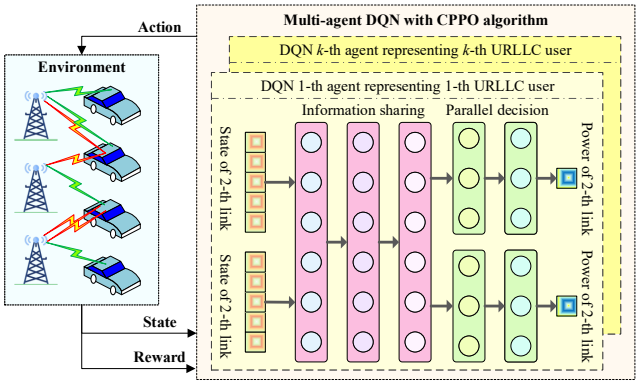


Fig. 2: Deep reinforcement learning framework.

Algorithm 1 CPPO for Power Allocation

Input:
The trained CPPO model
Total states, $\mathcal{S}_t = \{S_1, ..., S_j, ..., S_J\}$
Output:
Total actions, $\mathcal{A}_t = \{a_1, ..., a_j, ..., a_J\}$

```
Concatenate the state vectors of links serving the same UE into a vector
for all URLLC UEs do in parallel
    Pass through the three-layer NN for information sharing
    Cut the output vector into multiple vector pieces, each representing a URLLC link
    for all URLLC links do in parallel
        Pass through the three-layer NN to get action
    end for
end for
Match actions with URLLC links
return Total actions A_t = {a_1, ..., a_j, ..., a_J}
```

of dynamic URLLC link amount is solved by converting it to a multi-agent DRL, in which the changes in the number of active URLLC links only requires to change the number of agents. In our proposed CPPO algorithm, the agents of the URLLC links that serve the same UE are integrated as a small group, in which agents exchange information with each other while making decisions. The process of CPPO for power allocation is given in Algorithm 1. At first, for all the active URLLC links, the links that are serving the same UE are combined into a group and their state vectors are concatenated together. Then, a three-layer NN is used to share the information between these links. After that, the output vector of information sharing block is cut into multiple pieces and pass through a same two-layer NN in parallel to get the action of each agent.

A. Agent Observation State

The state of each URLLC link includes both the transmission demand and the channel gain of the same frequency sub-carriers of all RRHs. Assuming a URLLC link is using the n-th sub-carrier frequency to serve the k-th UE, the state of the b-th RRH can be denoted as $s_{b,n} = \{d_{b,n}, g_{k,b,n}\}$, where

$d_{b,n}$ is the transmitting URLLC packet size and $g_{k,b,n}$ is the channel gain from the b-th RRH to the k-th UE on the n-th sub-carrier frequency. Specially, if the n-th sub-carrier of the b-th RRH is vacant, $d_{b,n}$ and $g_{k,b,n}$ are both set as 0. Or, if it is used for eMBB service, $d_{b,n}$ is set as 0 and $g_{k,b,n}$ is the channel gain. The observation state of the agent for the j-th URLLC link can be obtained by grouping all $s_{b,n}$ as,

$$S_j = \{s_{1,n}, ..., s_{b,n}, ..., s_{B,n}\}. \tag{9}$$

These $s_{b,n}$ in S_j will then be sorted from smallest to largest by their distance between the b-th RRH and the j-th URLLC link service RRH. After that, the completed state at time t of the environment can be obtained as,

$$\mathcal{S}_t = \{S_1, ..., S_j, ..., S_J\}, \tag{10}$$

which is a combination of all the observation states S_j of active URLLC links.

B. Action

The action a_j of the j-th URLLC link is in a continuous number within a range $a_j \in [a^{\min}, a^{\max}]$, which represents its transmit power at the RRH. It is calculated based on the observation state S_j and the policy π. Then, the completed action at time t can be denote as,

$$\mathcal{A}_t = \{a_1, ..., a_j, ..., a_J\}, \tag{11}$$

which is a combination of all the transmit power of active URLLC links.

C. Reward and Problem Formulation

The reward function plays a crucial role in problem formulation and solving in DRL. In our multi-agent DRL framework, all agents share the same reward, which indicates the overall system performance. The benefit is that the agents will have a full consideration about the whole system performance rather than falling into a vicious competition during the training process. Consequently, the agents can learn a cooperative policy that not only meets their individual QoS requirements, but is also considerate of other agents, leading to an optimal solution for the entire system.

For the C-V2X network system, our objectives are twofold: maximizing the reliability guarantee of URLLC and maximizing the average eMBB rate. The reward function is made up of three components and each of them is a sub-reward mathematically described by the linear piecewise function, $\mathcal{F}(\cdot)$, with the maximum value as 1. The reward about URLLC reliability of the k-th UE at time t is denoted as $R^{\mathrm{U}}_{t,k}$, which is an linear increasing and then decreasing function of $-\log_{10}(e_k)$, and it has the maximum value 1 when $-\log_{10}(e_k) = 7$. Denoting the total number of active URLLC UE at time t as K^{U}_t, the average reward of all URLLC UEs is,

$$\overline{R^{\mathrm{U}}_t} = \frac{1}{K^{\mathrm{U}}_t} \sum\nolimits_{k=1}^{K^{\mathrm{U}}_t} \mathcal{F}^{\mathrm{U}}(-\log_{10}(e_{t,k})). \tag{12}$$

The reward about eMBB rate of the k-th UE at time t is denoted as $R^{\mathrm{E}}_{t,k}$, which is in linear positive correlation with $c_{t,k}$ and equals to 1 when $c_{t,k} = 6\mathrm{MB/s}$. Denoting the total number of eMBB UE at time t as K^{E}_t, the average reward of eMBB UEs is,

$$\overline{R^{\mathrm{E}}_t} = \frac{1}{K^{\mathrm{E}}_t} \sum\nolimits_{k=1}^{K^{\mathrm{E}}_t} \mathcal{F}^{\mathrm{E}}(c_{t,k}). \tag{13}$$

The reward about the j-th URLLC link's transmit power of at time t is denoted as $R^{\mathrm{P}}_{t,j}$, which is a punishment for using high transmit power and it decrease from 1 when $a_{t,j}$ is greater than 27 dBm. Denoting the total number of active URLLC links at time t as J_t, the average transmit power reward is,

$$\overline{R^{\mathrm{P}}_t} = \frac{1}{J_t} \sum\nolimits_{j=1}^{J_t} \mathcal{F}^{\mathrm{P}}(a_{t,j}). \tag{14}$$

Then, the total reward at time t can be obtained by summing of the above three partial rewards as,

$$\mathcal{R}_t = \lambda_{\mathrm{U}} \overline{R^{\mathrm{U}}_t} + \lambda_{\mathrm{E}} \overline{R^{\mathrm{E}}_t} + \lambda_{\mathrm{P}} \overline{R^{\mathrm{P}}_t}, \tag{15}$$

where λ_{U}, λ_{E} and λ_{P} are the weights of each partial reward. Finally, the power allocation problem is formulated with the goal of maximizing the total reward.

D. Training algorithm of CPPO

The CPPO is trained with the proximal policy optimization (PPO) algorithm, which is based on the actor-critic mechanism. Specifically, the actor network outputs the probability distribution of actions while the critic network evaluates actions with a value function. Since the packet arrival process of URLLC packets is random and the states of active URLLC links are uncorrelated in time, the critic's input is set to be the same as the actor's input state.

PPO improves the problem of DRL sample efficiency by employing surrogate objectives to regularize policy updates and enable reuse of training data. Surrogate objectives prevent the new policy from deviating far from the old one. The probability ratio term defined as,

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}, \tag{16}$$

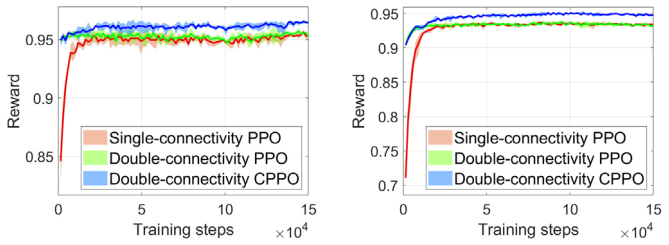
where π_θ is a stochastic policy and θ represents NN parameters. It is an importance sampling estimator used to compensate for the gap between the distribution of training data and the distribution of current policy state. Besides, the clipped probability ratio is used to avoid policy changes moving $r_t(\theta)$ away from 1, and it is defined as,

$$\begin{aligned} r^{\mathrm{clip}}_t(\theta) &= \mathrm{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) \\ &= \begin{cases} 1-\epsilon, & r_t(\theta) < 1-\epsilon \\ 1+\epsilon, & r_t(\theta) > 1+\epsilon \\ r_t(\theta), & \text{otherwise} \end{cases}, \end{aligned} \tag{17}$$

where ϵ is a hyper-parameter. It makes sure that the new and old policies are at the very least near to one another and avoids any abrupt adjustments. Our goal is to minimize the training loss given as,

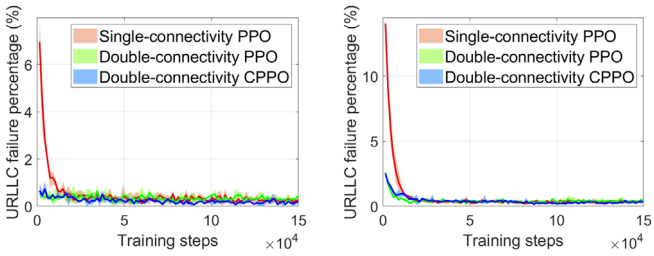
$$L(\theta) = -\mathbb{E}_t[min(r_t(\theta)A_t, r^{\mathrm{clip}}_t(\theta)A_t)], \tag{18}$$

where A_t is an estimated advantage function that equal to reward given by environment minus value given by the critic.



(a) 10 vehicles. (b) 20 vehicles.

Fig. 3: Reward variation of training.



(a) 10 vehicles. (b) 20 vehicles.

Fig. 4: URLLC failure percentage variation of training.

IV. Simulation and Results

A. Experiment Setting

The considered scenario is a 1 km highway section and 3 RRHs are symmetrically located on each side of the road. The base stations are positioned 400m apart from each other, and 100m away from the road. The vehicles are randomly distributed on the road according to a spatial Poisson process. We assume there are 30 sub-carriers for each RRH in total. The path loss model is $128.1 + 37.6\mathbf{log}_{10}(d)$, where d is the distance in km. The shadowing follows a log-normal distribution with a standard deviation of 5 dB. The sub-carriers with the best channel gain are selected for both eMBB and URLLC services. For each step during the training process, we randomly store the information of two agents into memory rather than storing all the information. The NNs of actor and critic are updated 10 times with the data memory of every 1500 steps. The learning rate starts from 0.001 and decays with 0.95 for every 1500 steps. The total training step is set as 1.5×10^5. All the input and output features are linearly normalized to a range of -1 to 1 for consistency in scale. Our simulation is carried out on i7-12700H with PyTorch 1.10. Other simulation parameters are listed in Tab. I [3].

TABLE I: Simulation Parameters.

Parameter	Value
Carrier frequency	2 GHz
Bandwidth of each subcarrier W	1 MHz
Length of mini-slot TTI τ	0.125 ms
Noise spectral density	-174 dBm/Hz
Transmit power of URLLC link	[0, 35] dBm
Transmit power of eMBB link	20 dBm
URLLC data packet size	[32,64] bytes
URLLC data packet demand	1000 packets/s
RRH antenna height	25 m
Vehicular antenna height	1.5 m
PPO clip parameter ϵ	0.2
Mini-batch size	128

B. Results and Analysis

In this subsection, we test the system performance with the different number of vehicles ranging from 10 to 25 and all these cases use a common reward function. We compare the performance of three different DRL approaches for URLLC link power allocation:

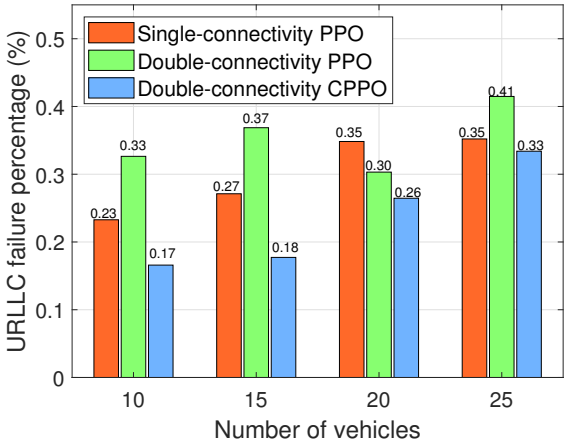


Fig. 5: URLLC failure percentage vs number of vehicles.

1) Single-connectivity PPO: Each URLLC packet is transmitted once on one single link. Each URLLC link is an agent and its actor uses the same 4-layer NN.
2) Double-connectivity PPO: Each URLLC packet is duplicated and transmitted on two independent links simultaneously. Each URLLC link is an agent and its actor uses the same 4-layer NN. There is no information sharing among agents who are serving the same UE.
3) Double-connectivity CPPO: Each URLLC packet is duplicated and transmitted on two independent links simultaneously. The transmit power is computed by the proposed CPPO algorithm.

The variation of reward and the URLLC failure percentage are shown in Fig. 3 and Fig. 4, respectively. For rewards, the initial reward of single-connectivity PPO is lower than that of double-connectivity PPO, but eventually reached a similar value. In double-connectivity cases, the CPPO is able to achieve higher rewards after training, despite starting at a similar level as PPO. For URLLC failure percentage, the single-connectivity PPO had a high failure rate at the beginning, while double-connectivity CPPO exhibited the best performance after training.

The C-V2X network system performances including the URLLC failure percentage and the eMBB average rate are shown in Fig. 5 and Fig. 6, respectively. For URLLC failure percentage, we can see that double-connectivity PPO cannot

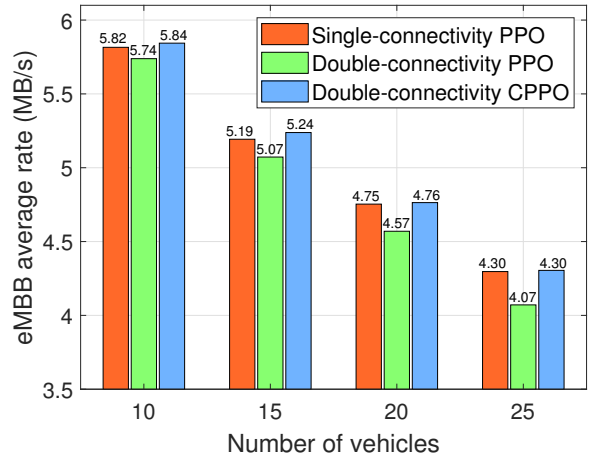


Fig. 6: eMBB average rate vs number of vehicles.

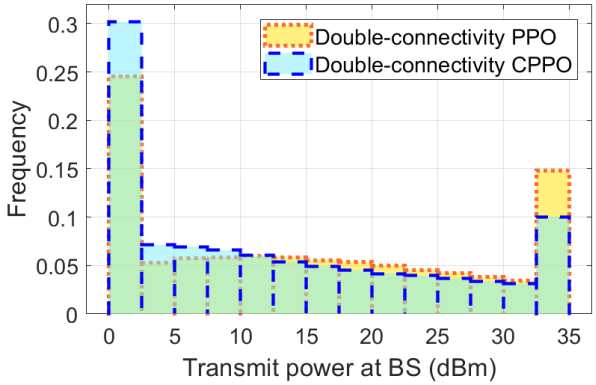


Fig. 7: eMBB average rate vs number of vehicles.

improve the performance comparing with single-connectivity PPO. Besides, double-connectivity CPPO has the lowest failure percentage for all four cases. For eMBB average rate, there is a decrease in the rate of double-connectivity PPO compared to single-connectivity PPO and the decrease is greater as the number of vehicles increases. Meanwhile, double-connectivity CPPO has the highest eMBB average rate in all cases.

To better understand the source of performance gains, we plot the frequency distribution comparison of the chosen transmit power for the 20 vehicles case under different learning algorithms in Fig. 7. We can see the proposed CPPO is more likely to choose lower transmit powers. Specifically, compared to PPO, CPPO consumes only 71.50% of the energy transmitted. This is the benefit of the information sharing between multiple links serving the same UE. In multi-connectivity, there is no need for every URLLC link to go beyond the reliability requirement but a need for their synergy reliability. CPPO enables each link to know the CSI of the others, so the transmit power of multiple links can be concertedly allocated, avoiding that all the links choose high transmit power. Therefore, information sharing is a key point in power allocation of multi-connectivity URLLC links.

V. Conclusions

In this paper, we have investigated the multi-connectivity packet duplication for URLLC in the C-V2X network and proposed a multi-agent DRL algorithm for real-time power allocation. Extensive simulation results have demonstrated the effectiveness of proposed algorithm which provides a cooperative solution for multi-connectivity power allocation. For future work, we will study the multi-connectivity URLLC with imperfect CSI.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China under Grant 62271244, the National Natural Science Foundation Original Exploration Project of China under Grant 62250004, the Natural Science Fund for Distinguished Young Scholars of Jiangsu Province under Grant BK20220067.

References

[1] S. Chen, J. Hu, Y. Shi, L. Zhao, and W. Li, "A vision of C-V2X: Technologies, field testing, and challenges with chinese development," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 3872–3881, 2020.

[2] N. Cheng, F. Lyu, J. Chen, W. Xu, H. Zhou, S. Zhang, and X. Shen, "Big data driven vehicular networks," *IEEE Network*, vol. 32, no. 6, pp. 160–167, 2018.

[3] 3GPP TR 38.913, "Study on scenarios and requirements for next generation access technologies," *Release 14*.

[4] X. Ge, "Ultra-reliable low-latency communications in autonomous vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 5005–5016, 2019.

[5] Z. Li, M. A. Uusitalo, H. Shariatmadari, and B. Singh, "5G URLLC: Design challenges and system concepts," in *2018 15th international symposium on wireless communication systems (ISWCS)*. IEEE, 2018, pp. 1–6.

[6] C. She, Z. Chen, C. Yang, T. Q. S. Quek, Y. Li, and B. Vucetic, "Improving network availability of ultra-reliable and low-latency communications with multi-connectivity," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5482–5496, 2018.

[7] M. Labana and W. Hamouda, "Advances in CRAN performance optimization," *IEEE Network*, vol. 35, no. 3, pp. 140–146, 2021.

[8] R. Dong, C. She, W. Hardjawana, Y. Li, and B. Vucetic, "Deep learning for radio resource allocation with diverse quality-of-service requirements in 5G," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2309–2324, 2021.

[9] M. Li, J. Gao, L. Zhao, and X. Shen, "Deep reinforcement learning for collaborative edge computing in vehicular networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 4, pp. 1122–1135, 2020.

[10] Q. Zhao, S. Paris, T. Veijalainen, and S. Ali, "Hierarchical multi-objective deep reinforcement learning for packet duplication in multi-connectivity for URLLC," in *2021 Joint European Conference on Networks and Communications and 6G Summit (EuCNC/6G Summit)*, 2021, pp. 142–147.

[11] C. She, C. Yang, and T. Q. S. Quek, "Cross-layer optimization for ultra-reliable and low-latency radio access networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 127–141, 2018.

[12] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[13] C. Sun, C. She, C. Yang, T. Q. S. Quek, Y. Li, and B. Vucetic, "Optimizing resource allocation in the short blocklength regime for ultra-reliable and low-latency communications," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 402–415, 2019.

[14] J. Rao and S. Vrzic, "Packet duplication for URLLC in 5G: Architectural enhancements and performance analysis," *IEEE Network*, vol. 32, no. 2, pp. 32–40, 2018.