

# A Flexible Numerology Resource Allocation to Guarantee Delay and Reliability Requirements in 5G NR-V2X Networks

Chen Chen<sup>1,2</sup>, Wenhua Wang<sup>1</sup>, Li Cong<sup>5</sup>, Wei Han<sup>3</sup>, Qizhong Zhang<sup>2</sup>, Ziyi Liu<sup>1</sup>, Haitao Lu<sup>4</sup>, and Qingqi Pei<sup>1</sup>

Corresponding Author: whwang11@stu.xidian.edu.cn

<sup>1</sup> School of Telecommunication Engineering, Xidian University, Xi'an 710071, China.

<sup>2</sup> Xidian Hangzhou Institute of Technology, Hangzhou 311231, China.

<sup>3</sup> Xi'an Beilin University-based Innovation Group Co., Ltd.

<sup>4</sup> Ningbo Joynext Technology Corporation, Ningbo 361000, China.

<sup>5</sup> State Grid Jilin Province Electric Power Company Limited Information Communication Company, Changchun 130000, China.

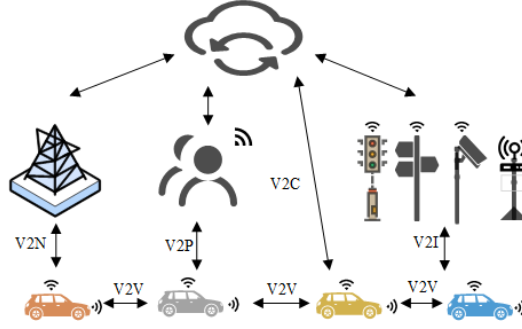
**Abstract.** Aiming at the delay and reliability requirements of different transmission tasks in the New Radio Vehicle-to-Everything (NR-V2X) networks, this paper proposes a flexible numerology resource allocation algorithm for V2V based on deep learning, to solve the problem that traditional resource allocation methods are not flexible enough. Firstly, the flexible frame structure of the physical layer of NR-V2X is used to dynamically configure and allocate spectrum resources, and reasonably plan and allocate spectrum resources for transmitting secure messages to improve spectrum utilization. The deep reinforcement learning method is adopted to carry out the distributed transmission resource selection process at the vehicle end, combining the different requirements of delay and reliability in the process of secure message transmission, and dynamically balancing the influence of delay and packet retransmission when designing the reward function. The experimental simulation proves that the proposed algorithm has good convergence, can effectively reduce the probability of message loss, and can meet the transmission tasks with different delay and reliability requirements.

**Keywords:** NR-V2V · Sidelink · Deep learning · Resource allocation.

## 1 Introduction

As wireless communication technology and Internet of Things (IoT) [1] technology rapidly evolve, Intelligent Transportation System (ITS) [2] has emerged, and rich transportation participants use Vehicle to Everything (V2X) [3] interconnection.

In ITS, C-V2X [4] is the wireless communication protocol for the Internet of vehicles in China. C-V2X is the vehicle-based communication standard using cellular technology proposed by 3GPP. It mainly consists of two communication



**Fig. 1.** V2X Communication types.

modes, one is direct communication mode, and the other is resource scheduling through 4G or 5G base stations. With the continuous development of 5G communication technology, the second generation of NR-V2X based on the 5G new air interface was also proposed, and detailed specifications were given in the subsequent Release 16 [5] and Release 17 [6]. NR-V2X gives the powerful transmission capability of 5G to vehicle-connected communications, meeting the needs of multi-node connectivity in vehicle-connected environments, and greatly reducing the transmission delay for safety-efficient-related applications [7]. Yet, the swift changes in NR-V2X network topology, coupled with high user density and a voluminous amount of vehicle communications, strain the already limited communication resources, like the constrained frequency bands. [8]. With the increasing demand for vehicle-connected communication, it is not only necessary to face the challenges of large-scale terminal device access and massive data transmission [9], but also to meet the requirements of millisecond end-to-end delay and communication reliability [10]. Therefore, the design of a reasonable and efficient resource allocation strategy directly affects the performance of the whole system, which is also an urgent problem to be studied.

Study the combined power management and sub-channel allocation of the underlying Device-to-Device (D2D) and small cell (SC) communications, and solve the throughput maximization problem through the learning method of game theory [11]. However, the problem of energy efficiency in C-V2X networks with reinforcement learning mechanisms is not solved. In [12], in order to fulfill the demands for high reliability and low latency for personalized services in C-V2X networks, a B5G multi-connection personalized resource allocation method based on Lyapunov optimization is proposed. However, with the increasing diversity and complexity of on-board networks [13], the optimization process for this approach becomes highly complex, making real-time execution challenging, while the deep reinforcement learning (DRL) method overcomes the performance limitation of the training data generated by the transmission optimization algorithm, and can markedly enhance communication efficiency. In [14], the transmission efficacy of Vehicle-to-Infrastructure (V2I) and V2V is comprehensively considered, and the sub-channel allocation and power allocation problems are

solved by combining deep Q network and deep deterministic strategy gradient. In literature [15], the Decentralized DRL (DDRL) algorithm is used to optimize the channel capacity of V2I users, while meeting the delay and reliability constraints of V2V communication link sets. However, most of the existing work focuses on the problem of V2V and V2I shared links, and there is not enough research on the resource selection of V2V Sidelink communication. Additionally, there are few investigations into how the flexible numerology of NR influences resource allocation.

To summarize, the key contributions of this paper are as follows: First, this paper analyzes the influence of the flexible numerology of C-V2X on the resource division in the time domain and frequency domain. Then, the DRL method is used to dynamically configure the subcarrier interval for distributed spectrum resource selection in combination with the delay and reliability requirements of secure message transmission. Finally, simulation outcomes demonstrate that the proposed approach exhibits strong and is superior to the Semi-Persistent Scheduling (SPS) resource allocation method and DIstributed Resource Allocation based on multi-agent [16] reinforcement Learning (DIRAL) for secure message transmission with different delay tolerance and reliability requirements.

The rest of this paper is structured as follows: Sect. 2 presents the system models and discusses the optimization of V2V resource allocation. In Sect. 3, a novel scheme for resource allocation is proposed. Sec. 4 presents the experimental results. Finally, the study is concluded in Sect. 5.

## 2 System Model

In NR-V2X, a variety of SCS can support more diverse and flexible multi-service data transmission. Select appropriate numerology ( $\mu$ ) for different V2X applications, and select appropriate SCS according to different services and the current channel and resource pool status to improve communication efficiency, to provide a more extreme user experience on the premise of ensuring driving safety. In this section, we consider Sidelink resource allocation in vehicle-mounted networks and establish a scenario of V2V Sidelink transmission. Next, we model the transmission delay and reliability of the secure message broadcast process and obtain the optimization objectives of the Sidelink resource allocation process.

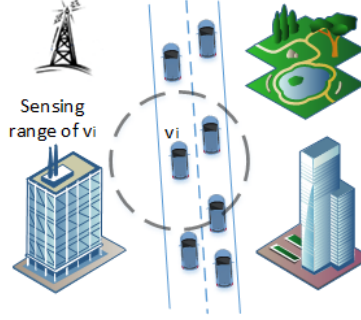
### 2.1 Configure Parameter Sets and Allocate Spectrum Resources

In order to meet diverse transmission requirements, 5G NR-V2X supports configurable and flexible numerology and uses configurable SCS and cyclic prefix (CP) to flexibly divide physical layer resources. Table 1 shows the parameters supported by two frequency ranges (RF) in the NR. NR-V2X supports four different SCS, corresponding to different cycle prefixes and transmission time intervals (TTI).

The size of  $rb$  is  $0.25ms \times 180kHz$ , and all  $rb$  constitute resource pool  $Z$ . This section defines three different types of Resource Block (RB) for SCS and TTI

**Table 1.** Flexible configuration numerology for NR-V2X

$\mu$	SCS $\Delta f = 2\mu \times 15$ [kHz]	Number of slots in the subframe	Slot length [ms]	CP	FR
0	15	1	1	Normal	FR1
1	30	2	0.5	Normal	FR1
2	60	4	0.25	Normal Extended	FR1 FR2
3	120	8	0.125	Normal	FR2

**Fig. 2.** V2V Sidelink transmission scenario.

with different sizes:  $RB_{T1}$ ,  $RB_{T2}$ ,  $RB_{T3}$ . SCS and TTI configurations of  $15\text{kHz}-1\text{ms}$ ,  $30\text{kHz}-0.5\text{ms}$ , and  $60\text{kHz}-0.25\text{ms}$  in the table are respectively adopted. Since each RB encompasses 12 OFDM symbols, in the frequency domain, the size of RB is  $180\text{kHz}$ ,  $360\text{kHz}$ , and  $720\text{kHz}$  respectively. Dynamically configured RBs transmission data is assigned to services arriving at different times of NR-V2X.

## 2.2 System Model

As shown in Fig. 2, V2V transmission of NR-V2X is the process of exchanging information between vehicles. Research focuses on ensuring low latency and high reliability when broadcasting secure messages. Delay refers to the time from sending a safety message to receiving and decoding it, which is essential for real-time safety decisions such as sudden braking. Reliability requirements refer to the stability of secure message broadcasting. Data is transmitted through a reasonable selection of time-frequency resources to avoid resource contention during transmission and ensure that messages can be correctly received by the receiver.

For V2V broadcast, the transmission delay mainly includes the following four parts: queue delay  $T_{que}$ , the processing delay  $T_{proc}^{vt}$  of the security message sending vehicle node, propagation delay  $T_{tran}$  and the processing delay  $T_{proc}^{vr}$  of the security message receiving vehicle node. This section considers the propagation delay  $T_{tran} = TTI$  of a hop in the vehicle broadcast range and assumes that all of the above delays are multiples of  $TTI$ . Therefore, the total time delay of

NR-V2V is:

$$\begin{aligned} T_{de}^{v_t} &= T_{que}^{v_t} + T_{proc}^{v_t} + T_{proc}^{v_r} + T_{trans}^{v_t} \\ &= \alpha \cdot T_{que} + (\theta + 1) \cdot TTI, \end{aligned} \quad (1)$$

where  $\alpha, \theta \in \mathbb{Z}^{++}$  is a parameter of the total processing delay, generally speaking,  $\theta = 4$  [17][18].

This chapter regards the security message processing process as the  $M/G/1$  model [19], and the queue delay is:

$$T_{que} = \frac{TTI(3 - 2\lambda TTI)}{2(1 - \lambda TTI)} \quad (2)$$

$PL(d_{t,r})$  represents the path loss of the channel over the transmission distance  $d_{t,r}$  [20]:

$$PL(d_{t,r}) = 32.4 + 20 \log 10(d_{t,r}) + 20 \log 10(f_c) \quad (3)$$

In V2V transmission,  $d_{t,r}$  represents the distance between the message-sending vehicle and the message-receiving vehicle,  $g_{t,r}[i]$  represents small-scale fading during transmission and follows a lognormal distribution. And we use  $h_{t,r}[i]$  to represent the channel gain:

$$|h_{t,r}[i]|^2 = |g_{t,r}[i]|^2 PL(d_{t,r}) \quad (4)$$

Based on this, the average Signal to Interference plus Noise Ratio (SINR) is:

$$SINR_{t,r}[i] = \frac{P_{t,r}[i] |h_{t,r}[i]|^2}{I_{t,r}[i] + \sigma^2} \quad (5)$$

$$SINR_{t,r} = \frac{1}{M_n} \sum_{i \in M_n} SINR_{t,r}[i] \quad (6)$$

Equation 6 shows that the more the sending node  $v_t$  shares the  $rb$  with other interfering nodes, the lower the SINR of the receiving node. NR uses the Hybrid Automatic Repeat Request (HARQ) mechanism to determine the validity of the message according to the SINR threshold. If the SINR is higher than the threshold, an Acknowledgment (ACK) is sent. Otherwise, a Negative ACK (NACK) is sent. Therefore, the packet delivery rate during a broadcast in  $v_t$  is:

$$\mathbb{P}_t = \frac{N_{t,ack}}{N_{t,ack} + N_{t,nack}} \quad (7)$$

To meet the reliability and real-time requirements of different V2V service broadcast processes, the vehicle set  $V = \{v_1, v_2, \dots, v_N\}$ . Based on reliability threshold  $\sigma$  and delay value  $\tau$ , the optimization problem of V2V resource allocation is modeled as follows:

$$\begin{aligned} \max \quad & \sum_{t=1}^N \psi_p \mathbb{P}_t - \psi_t T_{de}^t \\ \text{s.t.} \quad & \mathbb{P}_t \geq \sigma \\ & T_{de}^t \leq \delta \end{aligned} \quad (8)$$

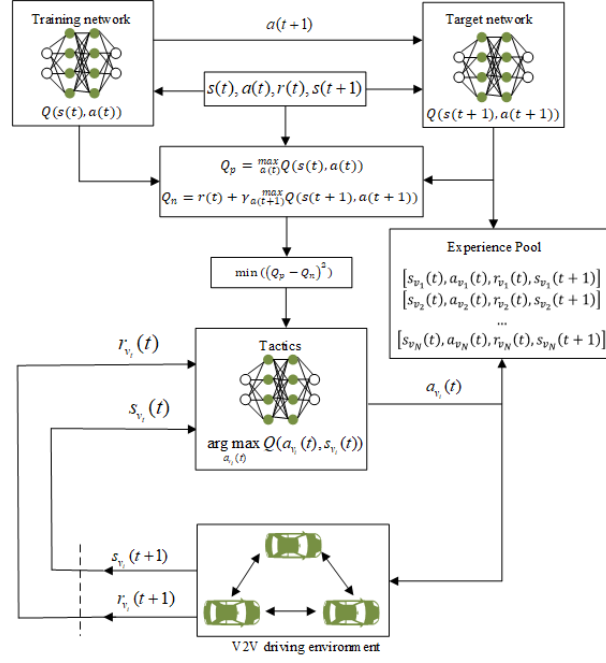


Fig. 3. DL-VRA framework.

### 3 Proposed Scheme

The basic reinforcement learning framework consists of interacting agents, actions, and environments. As shown in Fig. 3, the DL-VRA framework is comprised of agents and environments representing the vehicle, each of which is trained centrally and performs action selection in a distributed manner. The training phase begins with a random environmental state, in which the agent chooses an action based on the state space, and the environment responds with a reward to evaluate the action value.

The Q-learning algorithm is used to derive the optimal strategy, where the Q value is the expected cumulative discount reward obtained when the strategy  $\pi(t)$  is used to take action  $a(t)$  under the state and action  $(s(t), a(t))$ , and the real-time update of the reward and the selection of new actions are realized through the update iteration of Q value [21]:

$$Q(s(t), a(t)) \leftarrow Q(s(t), a(t)) + \theta \left[ r(t) + \gamma \max_{a(t+1)} Q(s(t+1), a(t+1)) - Q(s(t), a(t)) \right] \quad (9)$$

where  $\theta$  is the learning rate,  $r(t)$  is the real-time reward in the current action and state, and  $\max_{a(t+1)} Q(s(t+1), a(t+1))$  represents the maximum reward in the new state and state. If each action in the action space is performed countless

times in each state in the state space, with the adjustment of  $\theta$ , the Q value will converge to the optimal value with probability 1 [22]. In this paper, a greedy algorithm is used to explore the action space:

$$a(t+1) = \begin{cases} \operatorname{argmax}_{a(t+1)} Q(s(t+1), a(t+1)), & P = 1 - \epsilon \\ a_{\text{random}} \in \mathcal{A}, & P = \epsilon \end{cases} \quad (10)$$

The formula 10 indicates that to prevent converging to a local optimal solution, every time the action is updated, the action is chosen randomly from the action space with a probability of  $\epsilon$ , and the action is updated based on the Q value with another probability of  $1 - \epsilon$ . Due to the rich spectrum resources of NR V2V broadcast and large action space and state space, to accelerate the convergence speed, this section introduces a Deep Q network instead of a Q table, and approximated Q value by Deep Neural Networks (DNN), such as the loss function of formula 11. This section uses a DNN network with a weight of  $\delta$ . Given a reward  $r(t)$ , DNN updates the network parameters by minimizing errors:

$$L(\delta) = \left[ r(t) + \gamma \max_{a(t+1)} Q(s(t+1), a(t+1); \delta_{t+1}) - Q(s(t), a(t); \delta) \right]^2 \quad (11)$$

---

**Algorithm 1** DLFN Algorithm

---

**Require:**  $RB_T$  type set, channel status information  $RB_M$

**Ensure:**  $\mu$  Select policy and Resource select policy

- 1: Start the environment simulator and generate the vehicle according to the environment configuration
  - 2: Randomly initialize the Q network of all agents
  - 3: **for** training round  $e = \{0, 1, \dots, L\}$  **do**
  - 4:   Update the channel parameters
  - 5:   **for** slot  $\{t = 0, 1, \dots, T\}$  **do**
  - 6:     **for** agent  $v_t = \{v_1, v_2, \dots, v_N\}$  **do**
  - 7:       Initializes the state space  $s_t^{v_t}$
  - 8:       Observe the state space
  - 9:       Select the action  $a_t^{v_t}$  based on the policy
  - 10:       Run the action  $a_t^{v_t}$  to interact with the environment
  - 11:       Determine whether retransmission is required according to the delay and reliability, and calculate the reward  $r_t^{v_t}$
  - 12:     **end for**
  - 13:     **if** do needs to be retransmission **then**
  - 14:       Execute the retransmission policy and update the state space
  - 15:     **end if**
  - 16:     **for** agent  $v_t = v_1, v_2, \dots, v_N$  **do**
  - 17:       Sample and update the Q network from the sample pool
  - 18:     **end for**
  - 19:   **end for**
  - 20: **end for**
-

The whole training process is shown in Algorithm 1. In the execution stage, each agent makes action decisions according to the local state space and selects the optimal transmission action according to the action network output during the training process to ensure the transmission delay and reliability requirements of each vehicle.

## 4 Simulation Results

In this section, to assess the performance of the proposed algorithm, we compare it with the SPS resource allocation algorithm [23] and DIRAL method [24] through Matlab and SUMO co-simulation.

### 4.1 Simulation Settings

The experimental simulation environment adopts the road scene of a networked highway. The lane width is 6.5 meters, divided into three lanes and the lane length is 500m. Vehicles are randomly distributed on the highway and run in a straight line at a random speed of 80km/h-90km/h. Table 2 lists the simulation parameters used in the experiment and the parameters of the training process.

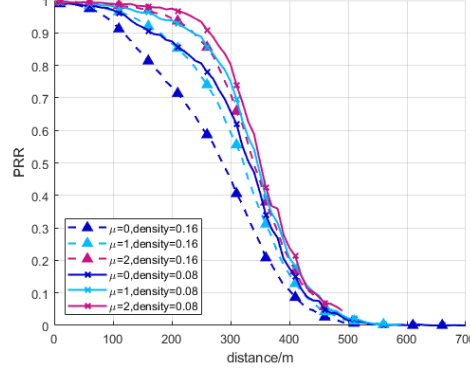
**Table 2.** Simulation parameter

Parameter	Value
Carrier frequency	2GHz
bandwidth	4Mhz
Antenna Gain(dBi)	3
Vehicle Transmission Power(dBm)	23
Noise power	-114dBm
$\mu$	0,1,2
Number of vehicles(units)	6-12
Discount factor $\gamma$	0.9
$\epsilon$	1.0-0.05
episode	2500
Learning rate	0.001
Sample pool size	12
MCS	21

### 4.2 Performance Evaluation

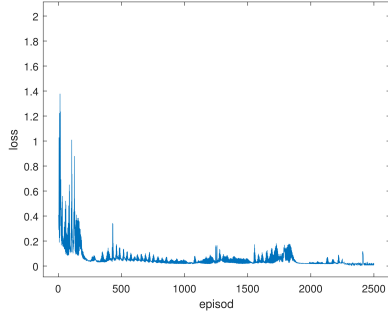
Fig. 4 shows that under different vehicle densities, the larger the  $\mu$ , the higher the Paket Reception Ratio (PRR) of security messages. This is because in NR-V2X transmission, the time slot length is inversely proportional to the subcarrier interval, but the use of a larger  $\mu$  will account for more bandwidth, and the number of UE that can receive orthogonal transmission in the same time slot is reduced, which is more likely to cause packet collision when transmitting multiple UE. Therefore, it is very important to dynamically adjust  $\mu$  to optimize spectrum resource allocation.



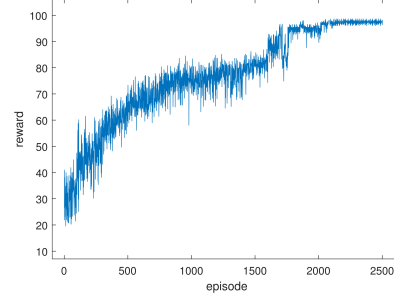


**Fig. 4.** Influence of  $\mu$  on PRR in NR C-V2X resource selection process.

Fig. 5 and Fig. 6 use centralized training and distributed execution to conduct experiments. The experiments show that the loss curve of the DLFN resource allocation algorithm first decreases, then increases, and finally converges during the training process, and the change is slight after 300 iterations. Simultaneously, the normalization reward grows as the increase of the number of iterations and becomes stable after 1600 iterations. This demonstrates the algorithm's effectiveness and its ability to enable vehicles to efficiently select spectrum resources.



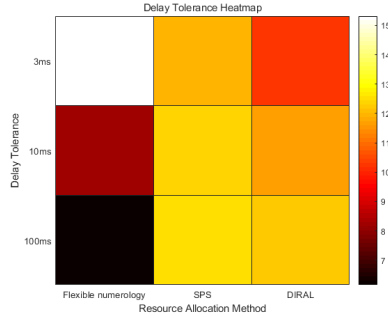
**Fig. 5.** Loss function curve.



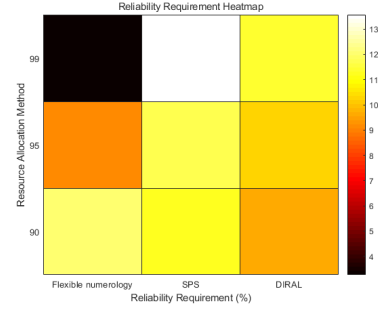
**Fig. 6.** Normalized reward curve.

On the premise of ensuring network convergence, the loss probability of the proposed algorithm is further compared under different delay tolerance and reliability requirements. As shown in Fig. 7 and Fig. 8, the results show that for small delay-tolerant tasks, the packet loss rate of the DLFN algorithm is large. The probability of loss decreases as the delay tolerance rises, correlating with an increase in the number of service retransmissions. Compared with SPS and DIRAL methods, the proposed algorithm reduces the average packet loss rate by about 3% and 2%, respectively. Fig. 8 shows that as reliability requirements increase, the probability of loss decreases. This is due to the fact that tasks

with higher reliability requirements can be retransmitted more times, thereby increasing the probability of packet reception.



**Fig. 7.** Message loss probability under different delay tolerance.



**Fig. 8.** Message loss probabilities under different reliability requirements.

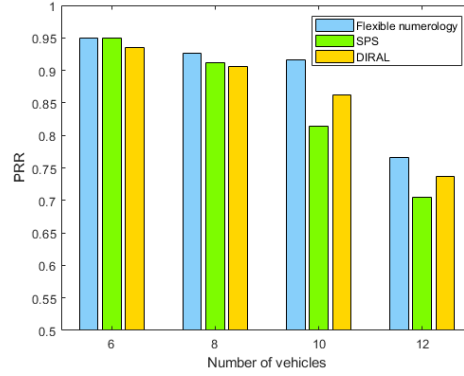
Fig. 9 compares the average PRR of 6, 8, 10, and 12 vehicles on a 500-meter-long road. Experimental results show that when the number of vehicles is small, the PRR of each method is similar, and the resource competition is not significant. As the number of vehicles increases, the performance superiority of the proposed method becomes more apparent, with the average PRR declining gradually, yet consistently outperforming the SPS and DIRAL methods.

## 5 Conclusion

In this paper, a DLFN resource allocation method is proposed to solve the problem that traditional resource allocation methods are not flexible enough, optimize network performance, and ensure the reliability and real-time transmission of secure messages. The algorithm adopts a deep reinforcement learning algorithm to carry out adaptive resource allocation, including algorithm flow, model structure, training process, etc., and uses highly flexible numerology V2V, which can make adaptive adjustments according to different communication environments and Vehicle dynamics.

The training process of reinforcement learning may be affected by environmental changes, initial parameter setting, and reward design, and further research is needed to improve the stability and robustness of the algorithm in the future.

**Acknowledgement.** This work was supported by the 2023 science and technology project ‘Research and application of multi communication system fusion networking technology for typical scenarios in power grids’2023JBGS-11 of State Grid Jilin Electric Power Company.



**Fig. 9.** PRR under different numbers of vehicles.

## References

1. H. Li, C. Chen, H. Shan, P. Li, Y. C. Chang, and H. Song, "Deep deterministic policy gradient-based algorithm for computation offloading in iov," *IEEE Transactions on Intelligent Transportation Systems*, 2023.
2. A. H. Sodhro, M. S. Obaidat, Q. H. Abbasi, P. Pace, S. Pirbhulal, G. Fortino, M. A. Imran, M. Qaraqe *et al.*, "Quality of service optimization in an iot-driven intelligent transportation system," *IEEE Wireless Communications*, vol. 26, no. 6, pp. 10–17, 2019.
3. S. Islam, A. Iqbal, M. Marzband, I. Khan, and A. M. Al-Wahedi, "State-of-the-art vehicle-to-everything mode of operation of electric vehicles and its future perspectives," *Renewable and Sustainable Energy Reviews*, vol. 166, p. 112574, 2022.
4. 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.
5. 3GPP, "Architecture Enhancements for 5G System (5GS) to Support Vehicle-to-Everything (V2X) Services(V16.2.0, Release 16)", Std., 2020.
6. —, "Technical Specification Group Radio Access Network; NR", *Physical layer procedures for data (Release 17)*", Std., 2021.
7. C. Chen, C. Wang, C. Li, M. Xiao, and Q. Pei, "A v2v emergent message dissemination scheme for 6g-oriented vehicular networks," *Chinese Journal of Electronics*, vol. 32, no. 6, pp. 1179–1191, 2023.
8. G. Abbas, Z. H. Abbas, M. Bilal, S. C. Shah, H. Song *et al.*, "Reliability analysis of cognitive radio networks with reserved spectrum for 6g-iot," *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 2726–2737, 2022.
9. L. Liu, C. Chen, Q. Pei, S. Maharjan, and Y. Zhang, "Vehicular edge computing and networking: A survey," *Mobile networks and applications*, vol. 26, pp. 1145–1168, 2021.
10. M. Hao, D. Ye, S. Wang, B. Tan, and R. Yu, "Urrlc resource slicing and scheduling for trustworthy 6g vehicular services: A federated reinforcement learning approach," *Physical Communication*, vol. 49, p. 101470, 2021.
11. H. Dai, Y. Huang, R. Zhao, J. Wang, and L. Yang, "Resource optimization for device-to-device and small cell uplink communications underlaying cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1187–1201, 2017.

12. Z. Lu, T. Zhang, X. Ji, B. Qian, L. Jiao, and H. Zhou, "Personalized wireless resource allocation in multi-connectivity b5g c-v2x networks," in *2022 14th International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE, 2022, pp. 1–6.
13. T. Xiao, C. Chen, Q. Pei, Z. Jiang, and S. Xu, "Sfo: An adaptive task scheduling based on incentive fleet formation and metrizable resource orchestration for autonomous vehicle platooning," *IEEE Transactions on Mobile Computing*, no. 01, pp. 1–18, 2023.
14. Y. Yuan, G. Zheng, K.-K. Wong, and K. B. Letaief, "Meta-reinforcement learning based resource allocation for dynamic v2x communications," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 8964–8977, 2021.
15. I. Rasheed, "Dynamic mode selection and resource allocation approach for 5g-vehicle-to-everything (v2x) communication using asynchronous federated deep reinforcement learning method," *Vehicular Communications*, vol. 38, p. 100532, 2022.
16. T. Xiao, C. Chen, M. Dong, K. Ota, L. Liu, and S. Dustdar, "Multi-agent reinforcement learning-based trading decision-making in platooning-assisted vehicular networks," *IEEE/ACM Transactions on Networking*, 2023.
17. H. Yin, L. Zhang, and S. Roy, "Multiplexing urllc traffic within embb services in 5g nr: Fair scheduling," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 1080–1093, 2020.
18. M. H. C. Garcia, A. Molina-Galan, M. Boban, J. Gozalvez, B. Coll-Perales, T. Şahin, and A. Kousaridas, "A tutorial on 5g nr v2x communications," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1972–2026, 2021.
19. C. Guo, L. Liang, and G. Y. Li, "Resource allocation for vehicular communications with low latency and high reliability," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 3887–3902, 2019.
20. 3GPP, *Study on evaluation methodology of new Vehicle-to-Everything (V2X) use cases for LTE and NR (v15.3.0 Release 15)*, Std., 2019.
21. S. S. Shinde and D. Tarchi, "Collaborative reinforcement learning for multi-service internet of vehicles," *IEEE Internet of Things Journal*, vol. 10, no. 3, pp. 2589–2602, 2022.
22. R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
23. N. Bonjorn, F. Foukalas, F. Cañellas, and P. Pop, "Cooperative resource allocation and scheduling for 5g ev2x services," *Ieee Access*, vol. 7, pp. 58 212–58 220, 2019.
24. A. Gündoğan, H. M. Gürsu, V. Pauli, and W. Kellerer, "Distributed resource allocation with multi-agent deep reinforcement learning for 5g-v2v communication," in *Proceedings of the Twenty-First International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, 2020, pp. 357–362.