

Article

# Multi-Agent-Deep-Reinforcement-Learning-Enabled Offloading Scheme for Energy Minimization in Vehicle-to-Everything Communication Systems

Wenwen Duan <sup>1</sup>, Xinmin Li <sup>2,3,\*</sup> , Yi Huang <sup>4</sup> , Hui Cao <sup>1</sup> and Xiaoqiang Zhang <sup>1,\*</sup>

<sup>1</sup> School of Information Engineering, Southwest University of Science and Technology, Mianyang 621000, China; caoh@swust.edu.cn (H.C.)

<sup>2</sup> College of Computer Science, Chengdu University, Chengdu 610100, China

<sup>3</sup> Guangdong Provincial Key Laboratory of Future Networks of Intelligence, The Chinese University of Hong Kong, Shenzhen 518172, China

<sup>4</sup> Department of Information and Communication Engineering, Tongji University, Shanghai 201804, China; huangyi718b@tongji.edu.cn

\* Correspondence: lxm\_edu@126.com (X.L.); xqzhang@swust.edu.cn (X.Z.)

**Abstract:** Offloading computation-intensive tasks to mobile edge computing (MEC) servers, such as road-side units (RSUs) and a base station (BS), can enhance the computation capacities of the vehicle-to-everything (V2X) communication system. In this work, we study an MEC-assisted multi-vehicle V2X communication system in which multi-antenna RSUs with liner receivers and a multi-antenna BS with a zero-forcing (ZF) receiver work as MEC servers jointly to offload the tasks of the vehicles. To control the energy consumption and ensure the delay requirement of the V2X communication system, an energy consumption minimization problem under a delay constraint is formulated. The multi-agent deep reinforcement learning (MADRL) algorithm is proposed to solve the non-convex energy optimization problem, which can train vehicles to select the beneficial server association, transmit power and offloading ratio intelligently according to the reward function related to the delay and energy consumption. The improved K-nearest neighbors (KNN) algorithm is proposed to assign vehicles to the specific RSU, which can reduce the action space dimensions and the complexity of the MADRL algorithm. Numerical simulation results show that the proposed scheme can decrease energy consumption while satisfying the delay constraint. When the RSUs adopt the indirect transmission mode and are equipped with matched-filter (MF) receivers, the proposed joint optimization scheme can decrease the energy consumption by 56.90% and 65.52% compared to the maximum transmit power and full offloading schemes, respectively. When the RSUs are equipped with ZF receivers, the proposed scheme can decrease the energy consumption by 36.8% compared to the MF receivers.

**Keywords:** vehicle-to-everything; mobile edge computing; offloading; transmit power; deep reinforcement learning



**Citation:** Duan, W.; Li, X.; Huang, Y.; Cao, H.; Zhang, X. Multi-Agent-Deep-Reinforcement-Learning-Enabled Offloading Scheme for Energy Minimization in Vehicle-to-Everything Communication Systems. *Electronics* **2024**, *13*, 663. <https://doi.org/10.3390/electronics13030663>

Academic Editors: Sherali Zeadally, Wenyu Zhang and Tianwei Hou

Received: 9 January 2024

Revised: 28 January 2024

Accepted: 31 January 2024

Published: 5 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The sixth generation (6G) is expected to enhance energy efficiency and the intelligence level to meet mass connectivity and low latency requirements of wireless communication network [1]. Supported by 6G, vehicle-to-everything (V2X) with the requirements of low communication delay is an instrumental element of future connected autonomous vehicles [2]. This promising technology can obtain the location and speed of vehicles and road hazards, reduce traffic accidents and improve road safety [3,4], and has gained significant interest from vehicle manufacturers, researchers and scientific communities [5]. However, as intelligent applications, i.e., in-vehicle Internet access and cooperative collision warning [6], high mobility of the device [7] and low-latency requirements [8], using V2X technology to handle computation-intensive tasks that require a large amount of data

computation and communication resources is a challenge [9]. By offloading tasks to mobile edge computing (MEC) servers, i.e., base stations (BSs) [7], road-side units (RSUs) [10] and unmanned aerial vehicles (UAVs) [11–13], MEC technology can meet the demand for computation resources in V2X networks [14]. In addition, offloading tasks to the MEC servers that are located close to the vehicles can effectively reduce the computation delays, meeting real-time communication requirements [15,16].

In [17], the authors developed a three-tier offloading architecture where vehicles can offload tasks to the RSUs and BS, and the average latency can be effectively reduced. In [18], vehicles offload tasks to the infrastructures and other vehicles to minimize the communication delay. In [19], the weighted sum of delay and energy consumption is minimized while considering the impact of interference on MEC-assisted V2X systems. In order to reduce the power consumption, the joint connection modes, uplink paths and task assignment optimization are studied in [20]. The communication and computation resource allocation, optimized jointly, is studied in [21] to minimize the cost of the V2X system under delay and energy constraints. In [22], the authors optimized user association and resource allocation jointly to maximize the overall data rate of both cellular users and device-to-device pairs. In [23], the end-to-end latency is minimized by optimizing the user association and resource allocation jointly. But, due to the coupling of many parameters, these optimization problems are highly non-convex and difficult to solve.

Deep reinforcement learning (DRL) is one of the leading research fields of artificial intelligence [24]. It can find the globally optimal or near-optimal solutions for more complex optimization problems even under an unknown environment [25–27], and has been adopted in numerous applications [28]. In [29], the authors present a DRL algorithm to maximize the long-term cache hit rate of the system, which consists of a single BS and several users. In a given time slot, the number of contents that users can request from the BS is fixed, and the BS acts as an agent and makes a content store decision. The simulation results show that the DRL algorithm can improve the cache hit rates and provide significant savings in runtime. In [30], the authors studied the UAV-assisted V2X system, which consists of several BSs and vehicles. In order to minimize the age of information, the DRL algorithm is proposed to jointly optimize the transmit power and computation offloading. The simulation results show that the DRL algorithm can significantly reduce the age of information.

This paper studies an MEC-assisted multi-vehicle V2X system, where the BS and RSUs work as MEC servers jointly to offload the tasks of vehicles. In order to reduce energy consumption while ensuring the requirement of V2X communication delay, the energy consumption minimization problem is formulated under delay and power constraints, and the DRL algorithm is proposed to solve the non-convex energy consumption optimization problem. The abbreviations and acronyms used in this paper are summarized in Table 1 and the main contributions of this work can be summarized as follows:

- (1) To compute the computation-intensive tasks for the computation-limited vehicles, we study an MEC-assisted multi-vehicle V2X communication system, where multi-antenna RSUs with linear receivers and a BS with a zero-forcing (ZF) receiver offload the tasks of vehicles jointly. In order to control the energy consumption, we formulate the energy consumption minimization problem and transform the non-convex optimization problem into a multi-agent decision process. Thus, each vehicle is capable of making intelligent decisions in its own communication environment.
- (2) In order to solve the non-convex optimization problem while satisfying the delay and power constraints, the multi-agent deep reinforcement learning (MADRL)-enabled server association, transmit power and offloading ratio joint optimization scheme is proposed. According to the reward function, which is related to the delay and energy consumption, vehicles are trained to select the beneficial transmit power, server association and offloading ratio. In order to reduce the action space dimensions and complexity of the MADRL algorithm, the improved K-nearest neighbors (KNN) algorithm is used to assign the vehicles that are located within the coverage area of

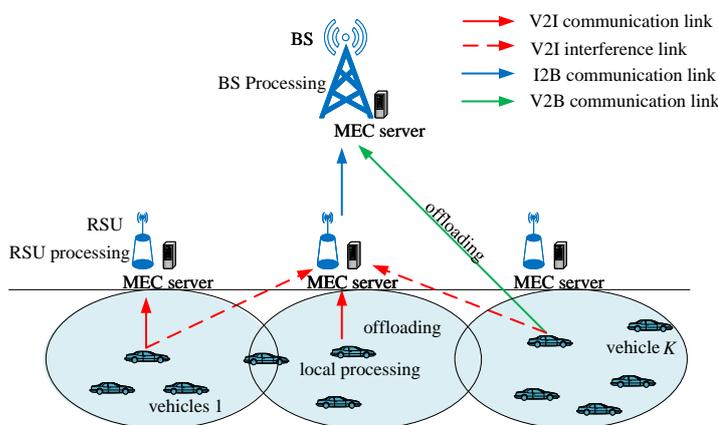
- several RSUs to the specific RSUs. Vehicles that are allocated to the same RSU as a group, and vehicles in the  $l$ -th group, can only offload tasks to the RSU  $l$  or BS.
- (3) Numerical results show that the proposed MADRL scheme can reduce more energy consumption compared with the full offloading and maximum transmit power schemes, and RSUs equipped with ZF receivers can decrease more energy consumption compared to the matched-filtering (MF) receivers. In addition, the proposed DRL scheme has a stable convergence under the different number of vehicles and data packet sizes.

**Table 1.** List of acronyms in alphabetical order.

Acronym	Description	Acronym	Description
BS	Base station	RSU	Road-side unit
DQN	Deep Q-network	SINR	Signal-to-interference-plus-noise-ratio
DRL	Deep reinforcement learning	6G	Sixth generation
I2B	Infrastructure-to-base-station	UAV	Unmanned aerial vehicle
KNN	K-nearest neighbors	V2B	Vehicle-to-base-station
MF	Matched filter	V2X	Vehicle-to-everything
MEC	Mobile edge computing	V2I	Vehicle-to-infrastructure
MADRL	Multi-agent deep reinforcement learning	ZF	Zero forcing
MIMO	Multi-input multi-output		

## 2. System Model

As shown in Figure 1, we consider an uplink MEC-assisted multi-vehicle V2X communication system that consists of a BS equipped with  $N^B$  antennas,  $L$  RSUs equipped with  $N^R$  antennas and  $K$  single-antenna vehicles. In addition, the number of RSUs  $L = 4$  and number of vehicles  $K = 6$ , and the radii of the BS and RSUs are 100 m and 200 m, respectively. In order to ensure real-time communication when the number of vehicles increases or vehicles move at a high speed, the RSUs and BS all work as the MEC servers jointly to offload the computation-intensive task of the vehicles. Thus, the vehicles can process the information locally or offload it to the RSUs via a vehicle-to-infrastructure (V2I) communication link or the BS via a vehicle-to-BS (V2B) communication link for edge computing. Meanwhile, each RSU has two transmission modes: **direct mode**, which transmits the processed task to corresponding vehicles directly, and **indirect mode**, which transmits the task to the wide-coverage BS, which transmits the processed information to the vehicles. For simplicity, the RSUs adopt the indirect transmission mode.



**Figure 1.** The system model of the MEC-assisted multi-vehicle V2X system.

Significant interference caused by the multi-antenna BS and RSUs will increase the computation complexity of the multi-input multi-output (MIMO) system [31]. Thus, we

adopt the linear receivers for the BS and RSUs to reduce the computation complexity [32]. Specifically, the BS is equipped with a ZF receiver and RSUs with linear receivers, respectively.

### 2.1. Channel Model of the Multi-Antenna RSUs

In our vehicular system, the RSUs are equipped with  $N^R$  antennas to serve many vehicles at the same time. Assuming that the number of vehicles in the system  $K = K_1 + K_2 + \dots + K_L$ , with  $K_l$  denoting the number of vehicles in RSU  $l$ , the received signal at the  $l$ -th RSU can be defined as

$$\mathbf{Y}_l^R = (\mathbf{P}_l^R)^{1/2} \mathbf{G}_l^R \mathbf{X}_l^R + \mathbf{Z}_l^R, \tag{1}$$

where the diagonal matrix  $\mathbf{P}_l^R$  is the transmission power matrix of the vehicles in RSU  $l$ , and  $\mathbf{P}_l^R = \text{diag}[p_1^R, p_2^R, \dots, p_{K_l}^R]$ .  $p_k^R$  are the transmit power of the  $k$ -th vehicle in the  $l$ -th RSU, with indexes  $k$  denoting the  $k$ -th vehicle in the  $l$ -th RSU.  $\mathbf{G}_l^R$  and  $\mathbf{X}_l^R$  denote the channel matrix and the transmitted signal matrix between the vehicles and the  $l$ -th RSU, respectively.  $\mathbf{Z}_l^R$  denotes the additive complex Gaussian white noise at the  $l$ -th RSU. The channel matrix between the vehicles and the  $l$ -th RSU can be defined as

$$\mathbf{G}_l^R = [\mathbf{g}_{1,l}^R, \mathbf{g}_{2,l}^R, \dots, \mathbf{g}_{K_l,l}^R] = \mathbf{H}_l^R (\mathbf{D}_l^R)^{1/2}, \tag{2}$$

where  $\mathbf{H}_l^R = [h_{1,l}^R, h_{2,l}^R, \dots, h_{K_l,l}^R]$  and diagonal matrix  $\mathbf{D}_l^R = \text{diag}[\beta_{1,l}^R, \beta_{2,l}^R, \dots, \beta_{K_l,l}^R]$  are the small-scale and large-scale fading matrix between the vehicles and the  $l$ -th RSU, respectively.  $h_{k,l}^R$  and  $\beta_{k,l}^R$  denote the small-scale and large-scale fading between the  $k$ -th vehicle and the  $l$ -th RSU, respectively. Thus, the channel vector between the vehicle  $k$  and the  $l$ -th RSU is given by  $\mathbf{g}_{k,l}^R = \sqrt{\beta_{k,l}^R} h_{k,l}^R$ .

Interference in the MIMO system will increase the analysis complexity of the multi-vehicle V2X system [33]. Thus, the linear equalizers such as the MF and the ZF receivers are applied to reduce the analysis complexity of the system [32]. The MF receiving matrix is represented by  $\mathbf{A}_l^{\text{MF}} = (\mathbf{G}_l^R)^H$ . Thus, the received signal at the  $l$ -th RSU which is equipped with an MF receiver, is given by

$$\mathbf{Y}_l^{\text{MF}} = \mathbf{A}_l^{\text{MF}} \mathbf{X}_l^R = (\mathbf{G}_l^R)^H \mathbf{X}_l^R. \tag{3}$$

The signal-to-interference-plus-noise ratio (SINR) for vehicle  $k$  at the  $l$ -th RSU can be defined as

$$\gamma_{k,l}^{\text{MF}} = \frac{p_k^R |(\mathbf{g}_{k,l}^R)^H \mathbf{g}_{k,l}^R|^2}{I_{k,l}^{\text{MF}}}, \tag{4}$$

where  $I_{k,l}^{\text{MF}} = \sum_{i=1, i \neq k} p_i^R |(\mathbf{g}_{i,l}^R)^H \mathbf{g}_{i,l}^R|^2 + \sigma^2 \|\mathbf{g}_{k,l}^R\|^2$  is the interference between the  $k$ -th vehicle and the  $l$ -th RSU equipped with the MF receiver.

The ZF receiving matrix is represented by  $\mathbf{A}_l^{\text{ZF}} = ((\mathbf{G}_l^R)^H \mathbf{G}_l^R)^{-1} (\mathbf{G}_l^R)^H$ . Thus, the received signal at the  $l$ -th RSU, which is equipped with a ZF receiver, can be defined as

$$\mathbf{Y}_l^{\text{ZF}} = \mathbf{A}_l^{\text{ZF}} \mathbf{X}_l^R = ((\mathbf{G}_l^R)^H \mathbf{G}_l^R)^{-1} (\mathbf{G}_l^R)^H \mathbf{X}_l^R. \tag{5}$$

The SINR for vehicle  $k$  at the  $l$ -th RSU is given by

$$\gamma_{k,l}^{\text{ZF}} = \frac{p_k^R}{\sigma^2 [((\mathbf{G}_l^R)^H \mathbf{G}_l^R)^{-1}]_{k,k}}. \tag{6}$$

### 2.2. Communication Model of the Multi-Antenna BS

In our vehicular system, the BS is equipped with an  $N^B$ -antenna to serve the vehicles and RSUs. Therefore, all the RSUs and vehicles are users of the BS. Let  $M$  denote the BS's

users number, which can be regarded as a combination of  $L$  RSUs and  $K$  vehicles, i.e.,  $M = L + K$ . Thus, the received signal at the BS can be defined as

$$\mathbf{Y}^B = (\mathbf{P}^B)^{1/2} \mathbf{G}^B \mathbf{X}^B + \mathbf{Z}^B, \tag{7}$$

where the diagonal matrix  $\mathbf{P}^B$  is the transmit power matrix of the users, which can be defined as  $\mathbf{P}^B = \text{diag}[p_1^B, p_2^B, \dots, p_M^B]$ , with  $p_m^B$  denoting the transmit power of the  $m$ -th user.  $\mathbf{G}^B$  and  $\mathbf{X}^B$  denote the channel matrix and the transmitted signal matrix between the users and the BS, respectively.  $\mathbf{Z}^B$  denotes the additive complex Gaussian white noise at the BS. The channel matrix between the users and BS is given by

$$\mathbf{G}^B = [\mathbf{g}_1^B, \mathbf{g}_2^B, \dots, \mathbf{g}_M^B] = \mathbf{H}^B (\mathbf{D}^B)^{1/2}, \tag{8}$$

where  $\mathbf{H}^B = [\mathbf{h}_1^B, \mathbf{h}_2^B, \dots, \mathbf{h}_M^B]$  and diagonal matrix  $\mathbf{D}^B = \text{diag}[\beta_1^B, \beta_2^B, \dots, \beta_M^B]$  are the small-scale and large-scale fading matrix between the users and BS, respectively.  $\mathbf{h}_m^B$  and  $\beta_m^B$  denote the small-scale and large-scale fading between the  $m$ -th user and the BS. Thus, the channel vector between the  $m$ -th user and the BS is given by  $\mathbf{g}_m^B = \sqrt{\beta_m^B} \mathbf{h}_m^B$ . In addition, in the vehicular system, the BS is equipped with a ZF receiver to eliminate the interference and provide higher communication rates of the V2B and infrastructure-to-BS (I2B) communication links [33]. The ZF receiving matrix can be defined as  $\mathbf{A}^{\text{ZF}} = ((\mathbf{G}^B)^H \mathbf{G}^B)^{-1} (\mathbf{G}^B)^H$ . Thus, the received signal at the BS can be defined as

$$\mathbf{Y}^{\text{ZF}} = \mathbf{A}^{\text{ZF}} \mathbf{Y}^B = ((\mathbf{G}^B)^H \mathbf{G}^B)^{-1} (\mathbf{G}^B)^H \mathbf{X}^B. \tag{9}$$

The SINR for user  $m$  at the BS can be defined as

$$\gamma_m^{\text{ZF}} = \frac{p_m^B}{\sigma^2 [((\mathbf{G}^B)^H \mathbf{G}^B)^{-1}]_{m,m}}. \tag{10}$$

### 3. Computing Model of MEC-Assisted V2X Communication System

The processing capabilities of the vehicles, RSUs and BS executing one bit of data are  $f^V$ ,  $f^R$  and  $f^B$  (in cycles/bit) CPU cycles, respectively. The data packet size generated by the vehicle is  $D_k$ . In our vehicular system, the data model adopts a partial offloading model. Thus, the data packet can be regarded as a combination of local computation data  $D_k^L$  and offloading computation data  $D_k^O$ , i.e.,  $D_k = D_k^L + D_k^O$ . Let  $D_{k,l}^{\text{O,R}}$  and  $D_k^{\text{O,B}}$  denote the data offloading to the  $l$ -th RSU and the BS from the  $k$ -th vehicle, respectively. Thus,

$$D_k^O = \begin{cases} D_{k,l}^{\text{O,R}}, & x_k = 0, \\ D_k^{\text{O,B}}, & x_k = 1, \end{cases} \tag{11}$$

where  $x_k$  denotes the server association of vehicle  $k$ .  $x_k = 0$  and  $x_k = 1$  represent the vehicle offloading tasks to the RSUs or the BS, respectively. It should be noticed that, when the  $k$ -th vehicle offloads a task to the RSUs,  $D_k^{\text{O,B}} = 0$ . Similarly, when the  $k$ -th vehicle offloads a task to the BS,  $D_{k,l}^{\text{O,R}} = 0$ .

#### 3.1. Local Computing Model

When the  $k$ -th vehicle offloads partial data to MEC servers, the vehicle needs to compute partial tasks locally. The computation delay and energy (in Joules, J) at vehicle  $k$  can be defined as [19]

$$T_k^L = \frac{D_k^L}{f_k^V}, \quad E_k^L = c_0 D_k^L (f_k^V)^2, \tag{12}$$

where  $f_k^V$  and  $c_0$  denote the processing capability of the  $k$ -th vehicle and the energy coefficient depending on the chip architecture [34], respectively.

### 3.2. RSU Computing Model

The  $k$ -th vehicle offloading computation-intensive tasks to the  $l$ -th RSU for edge computing consists of three steps: (1) the vehicle offloads tasks to the associated RSU via a V2I communication link, and the RSU computes the offloaded data, (2) the RSU transmits the computed data to the BS via an I2B communication link and (3) the BS transmits downlink data. The feedback data packet size is small; thus, the delay and energy consumption of downlink transmission are ignored.

#### (1) V2I communication link

The MF and ZF receivers are adopted to provide higher communication rates of the V2I communication links. When the RSUs are equipped with the MF receivers, according to Equation (4), the corresponding communication rate between the  $k$ -th vehicle and the specific  $l$ -th RSU can be written as

$$R_{k,l}^{V2I} = B^{V2I} \log_2 \left( 1 + \frac{p_k^V |(\mathbf{g}_{k,l}^R)^H \mathbf{g}_{k,l}^R|^2}{\sum_{i=1, i \neq k} p_i^V |(\mathbf{g}_{k,l}^R)^H \mathbf{g}_{i,l}^R|^2 + \sigma^2 \|\mathbf{g}_{k,l}^R\|^2} \right), \quad (13)$$

where  $B^{V2I}$  is the bandwidth of the V2I communication link and  $p_k^V$  is the transmit power of the  $k$ -th vehicle. When the RSUs are equipped with the ZF receivers, according to Equation (6), the corresponding communication rate between the vehicle  $k$  and the specific  $l$ -th RSU can be defined as

$$R_{k,l}^{V2I} = B^{V2I} \log_2 \left( 1 + \frac{p_k^V}{\sigma^2 [((\mathbf{G}_l^R)^H \mathbf{G}_l^R)^{-1}]_{k,k}} \right). \quad (14)$$

Thus, the transmission delay and energy between vehicle  $k$  and the  $l$ -th RSU can be defined as [21,35]

$$T_{k,l}^{V2I} = \frac{D_{k,l}^{O,R}}{R_{k,l}^{V2I}}, \quad E_{k,l}^{V2I} = p_k^V T_{k,l}^{V2I}. \quad (15)$$

When the  $k$ -th vehicle offloads partial data  $D_{k,l}^{O,R}$  to the  $l$ -th RSU for edge computing, the computation delay and energy at the  $l$ -th RSU is given by [21]

$$T_{k,l}^{O,R} = \frac{D_{k,l}^{O,R}}{f_l^R}, \quad E_{k,l}^{O,R} = c_1 D_{k,l}^{O,R} (f_l^R)^2, \quad (16)$$

where  $f_l^R$  and  $c_1$  denote the processing capability of the  $l$ -th RSU and the energy coefficient.

#### (2) I2B communication link

The ZF receiver is adopted to eliminate the interference so that the communication rate of the I2B communication link can be improved. According to Equation (10), the communication rate between the  $l$ -th RSU and the BS is given by

$$R_l^{I2B} = B^{I2B} \log_2 \left( 1 + \frac{p_l^R}{\sigma^2 [((\mathbf{G}^B)^H \mathbf{G}^B)^{-1}]_{K+l, K+l}} \right), \quad (17)$$

where  $B^{I2B}$  denotes the bandwidth of the I2B communication link, and  $p_l^R$  denotes the transmit power of the  $l$ -th RSU. Let  $D^{\text{Prc}}$  denote the tasks after computing. The transmission delay and energy between RSU  $l$  and the BS are given by

$$T_l^{I2B} = \frac{D^{\text{Prc}}}{R_l^{I2B}}, \quad E_l^{I2B} = p_l^R T_l^{I2B}. \quad (18)$$

The delay for the  $k$ -th vehicle offloading tasks to the  $l$ -th RSU for edge computing includes  $T_{k,l}^{V2I}$ ,  $T_{k,l}^{O,R}$  and  $T_l^{I2B}$ , and energy consumption includes  $E_{k,l}^{V2I}$ ,  $E_{k,l}^{O,R}$  and  $E_l^{I2B}$ . There-

fore, the delay and energy consumption for the  $k$ -th vehicle offloads tasks to the  $l$ -th RSU for edge computing can be written as

$$T_{k,l}^R = T_{k,l}^{V2I} + T_{k,l}^{O,R} + T_l^{I2B}, E_{k,l}^R = E_{k,l}^{V2I} + E_{k,l}^{O,R} + E_l^{I2B}. \quad (19)$$

### 3.3. BS Processing Model

Vehicle  $k$  offloading tasks to the BS for edge computing consists of two steps: (1) the vehicle offloads task to the BS via the V2B communication link and the BS computes the offloaded data, and (2) the BS transmits downlink data. The delay and energy consumption of downlink transmission are ignored. The ZF receiver is adopted to eliminate interference so that the communication rate of the V2B communication link can be improved. According to Equation (10), the corresponding communication rate between vehicle  $k$  and the BS is given by

$$R_k^{V2B} = B^{V2B} \log_2 \left( 1 + \frac{p_k^V}{\sigma^2 [((\mathbf{G}^B)^H \mathbf{G}^B)^{-1}]_{k,k}} \right). \quad (20)$$

Thus, the transmission delay and energy between the  $k$ -th vehicle and the BS can be defined as

$$T_k^{V2B} = \frac{D_k^{O,B}}{R_k^{V2B}}, \quad E_k^{V2B} = p_k^V T_k^{V2B}. \quad (21)$$

When the  $k$ -th vehicle offloads partial data  $D_k^{O,B}$  to the BS for edge computing, the computation delay and energy at the BS is given by

$$T_k^{O,B} = \frac{D_k^{O,B}}{f^B}, \quad E_k^{O,B} = c_1 D_k^{O,B} (f^B)^2, \quad (22)$$

where  $c_1$  is the energy coefficient.

The delay for vehicle  $k$  offloading tasks to the BS includes  $T_k^{V2B}$  and  $T_k^{O,B}$ , and energy includes  $E_k^{V2B}$  and  $E_k^{O,B}$ . Thus, the delay and energy consumption for vehicle  $k$  offloading tasks to the BS for edge computing can be defined as

$$T_k^B = T_k^{V2B} + T_k^{O,B}, \quad E_k^B = E_k^{V2B} + E_k^{O,B}. \quad (23)$$

Thus, the delay and energy consumption of the  $k$ -th vehicle is related to server association  $\{x_k = 0, 1\}$ , offloading ratio and transmit power, and can be defined as

$$T_k^{\text{total}} = \begin{cases} T_k^L + T_{k,l}^R, & x_k = 0, \\ T_k^L + T_k^B, & x_k = 1, \end{cases} \quad E_k^{\text{total}} = \begin{cases} E_k^L + E_{k,l}^R, & x_k = 0, \\ E_k^L + E_k^B, & x_k = 1. \end{cases} \quad (24)$$

### 3.4. Optimization Problem

In order to reduce the energy consumption of the V2X system while satisfying the requirements of communication delay, the energy consumption minimization problem is formulated under the delay constraint, which can be written as follows:

$$(P1) : \min_{x_k, D_k^O, p_k^V} \sum_{k=0}^K E_k^{\text{total}} \quad (25)$$

$$s.t. T_k^{\text{total}} \leq T_{\max}, k \in \mathcal{K}, \quad (25a)$$

$$x_k \in \{0, 1\}, k \in \mathcal{K}, \quad (25b)$$

$$0 \leq p_k^V \leq P_{\max}^V, k \in \mathcal{K}, \quad (25c)$$

$$0 \leq D_k^O \leq D_k, k \in \mathcal{K}. \quad (25d)$$

To ensure basic communication, the first constraint limits that the processing delay of the tasks generated by vehicle  $k$  must be less than the maximum delay. The second constraint ensures that the tasks can be offloaded to the RSUs or BS for edge computing. Constraint (25c) is to limit  $p_k^V$  to not be greater than the maximum transmit power  $P_{\max}^V$ . Constraint (25d) restricts that, for the vehicles that adopt the partial offloading model, the offloading quantity must be not greater than the maximum value  $D_k$ . It is noted that, because of the coupling of server association, transmit power and offloading ratio, (P1) is a non-convex problem and difficult to solve by using the traditional convex optimization schemes while satisfying the delay and power constraints. Thus, the DRL algorithm is adopted to solve (P1), which can find the globally optimal or near-optimal solution of the complex optimization problem [36].

#### 4. Proposed Scheme for Solving the Problem (P1)

In order to solve the highly non-convex problem, the DRL algorithm is proposed by optimizing the server association, transmit power and offloading ratio jointly. In addition, the improved KNN algorithm is used to assign vehicles and reduce the action space dimensions and complexity of the DRL algorithm.

##### 4.1. Vehicle Grouping Based on the Improved KNN Algorithm

The optimization problem (P1) is aimed at minimizing the energy consumption of the MEC-assisted multi-vehicle V2X system while satisfying the delay constraint, which is related to the transmission rate and SINR of the wireless communication links. A higher SINR can effectively improve the communicate rate in the V2X system [37]. Thus, in order to minimize energy consumption, the SINR must be maximized. According to Equations (4), (6) and (10), the SINR is related to the vehicle's transmit power in the same group. Therefore, assigning vehicles that are located within the coverage area of several RSUs to an RSU with a smaller user number can provide a higher SINR and reduce energy consumption.

The model-free KNN [38] method is one of the commonly used classification algorithms. It calculates the distance between initial center points and all training samples, and then obtains the test sample's nearest neighbors sample and groups them together. But, it is hard for KNNs to correctly classify overlapping samples [39]. Removing an overlapped sample from the training set can increase the accuracy of the KNN classifier [40]. In this paper, the locations of RSUs and vehicles are regarded as initial center points and samples, respectively, and the improved KNN algorithm [41] is adopted to assign the vehicles within the coverage area of multiple RSUs to a specific RSU. The distance between the vehicle and RSUs is denoted by  $d_{k,l}$ , which can be defined as

$$d_{k,l} = \sqrt{(x_k^V - x_l^R)^2 + (y_k^V - y_l^R)^2}, \quad (26)$$

where  $(x_k^V, y_k^V)$  and  $(x_l^R, y_l^R)$  are the position of vehicle  $k$  and the  $l$ -th RSU, respectively. The  $k$ -th vehicle can communicate with the  $l$ -th RSU when  $d_{k,l} \leq R_{\text{RSU}}$ , with  $R_{\text{RSU}}$  denoting the radius of the  $l$ -th RSU. After calculating the distance between vehicles and RSUs, the improved KNN algorithm assigns a vehicle to the RSU with a smaller number of vehicles to decrease system energy consumption and computational complexity. The improved KNN algorithm is shown in Algorithm 1.

**Algorithm 1** The Improved K-nearest Neighbors Algorithm

- 
- 1: **Initial:** The location of vehicles and RSUs. The radius of RSUs and the number of vehicles in each RSU.
  - 2: **for** vehicle  $k = 1 : K$  **do** Initial the set  $\mathcal{I}_k$  and number  $i_k$  of RSUs that can be selected by the  $k$ -th vehicle.
  - 3:     **for** RSU  $l = 1 : L$  **do** Calculate the distance  $d_{k,l}$  in (26). If  $d_{k,l} \leq R_{\text{RSU}}$ ,  $\mathcal{I}_k$  appends RSU  $l$  and  $i_k + 1$ .
  - 4:     **end for**
  - 5:     **if**  $i_k > 1$  **then** Assigning vehicle  $k$  to the  $l$ -th RSU which is in  $\mathcal{I}_k$  and has a lower number of vehicles.
  - 6:     **else** Assigning vehicle  $k$  to corresponding RSU  $l$  which makes  $d_{k,l} \leq R_{\text{RSU}}$ .
  - 7:     **end if**
  - 8: **end for**
- 

#### 4.2. Joint-Optimized Server Association, Offloading Ratio and Transmission Power Scheme Based on DRL

DRL learns from the environment-to-action mappings and develops an optimal strategy via trial and error [24]; according to the reward function, it trains agents to make quick decisions and optimal policies. Due to the non-convexity of objective function (P1), we propose an MADRL scheme. Agents interact with the environment and update the learning strategies. Therefore, the agent can choose the optimal action to minimize the total energy consumption of the V2X system while satisfying the delay constraint.

- (1) **Agent:** In the system, vehicles work as agents. They know their own information—since the BS and RSUs have no direct link to exchange the information, it is difficult to make the optimal decisions for offloading and power. Thus, the training is performed in the optimization problem to make the adjustment to the dynamic environment. We choose the vehicle as the agent to make the decision in the uplink link, and the BS and RSUs will send the corresponding information to the associated vehicles—i.e., the packet size of the task  $D_k$  generated by vehicle  $k$ , set  $\mathcal{I}_k$  and the channel state information.
- (2) **Action space:** Due to the limited communication resources in the system, the server association, offloading ratio and transmit power affect the communication rate and determine the communication energy consumption. Thus, the action space of vehicle  $k$  comprises the server association, offloading ratio and transmit power. The server association action space of the  $k$ -th vehicle is given as  $A_1 = \{x_k = 0, 1\}$ . The offloading and power control action space are given as  $A_2 = \{\frac{D_k}{d_p}, \frac{2D_k}{d_p}, \frac{3D_k}{d_p}, \dots, D_k\}$  and  $A_3 = \{\frac{P_{\max}^V}{d_l}, \frac{2P_{\max}^V}{d_l}, \frac{3P_{\max}^V}{d_l}, \dots, P_{\max}^V\}$ .  $d_p$  and  $d_l$  represent the length of offloading and the power control action space. Therefore, the  $k$ -th agent can choose action from the action space  $\mathbf{a}_k \in A = \{A_1, A_2, A_3\}, k \in \mathcal{K}$ .
- (3) **State space:** According to the communication model established in the system, the system energy consumption is related to transmission power, rate and delay; therefore, the state space comprises the transmit power  $p_k^V$ , transmission rate  $R_k^{\text{des}}$  and communication delay  $T_k^{\text{total}}$  in Equation (24). When  $x_k = 0$ , the transmission rate  $R_k^{\text{des}}$  is calculated by (14), and when  $x_k = 1$ , the transmission rate  $R_k^{\text{des}}$  is calculated by (20). Thus, the state of vehicle  $k$  can be expressed as  $\mathbf{S}_k = [P_k^V, R_k^{\text{des}}, T_k^{\text{total}}], k \in \mathcal{K}$ .
- (4) **Reward:** In the DRL algorithm, the agents are trained to adjust the action choice strategy and obtain the expected value by accumulating the maximum reward return. Aiming to minimize the total energy consumption of the V2X system and meet the delay constraint, the reward function of vehicle  $k$  is modeled as

$$r_k = \begin{cases} \frac{b}{\sum_{k=0}^K E_k^{\text{total}}}, & \text{if satisfied (25a),} \\ -T_k^{\text{total}}, & \text{otherwise,} \end{cases} \quad (27)$$

where  $b$  is a positive number. When the corresponding delay constraint can be satisfied, the agent will obtain a greater reward related with the energy consumption of the V2X system. When the corresponding delay constraint cannot be satisfied, the agent will obtain a negative reward related with the delay. In order to obtain the long-term rewards, agents are trained to choose the action that can minimize the energy consumption and meet the delay constraint.

The detailed design framework based on the deep Q-network (DQN) joint server association, offloading ratio and transmit power optimization scheme in the MEC-assisted V2X system is shown in Figure 2. The framework of the DQN contains two neural networks and an experience replay buffer, and vehicles act as agents interacting with the environment. On the current environment state  $s_k$ , the  $k$ -th agent chooses and executes action  $a_k$  and observes the next state  $s'_k$  and the greatest discount reward  $r_k$ . The experience replay buffer is used to store  $\langle s_k, a_k, r_k, s'_k \rangle$ . DQN trains the agent by randomly selecting a batch of samples from the experience replay buffer during the exploration cycle. Specifically, the main network takes current state  $s_k$  and action  $a_k$  from each data sample to obtain the predicted Q-value  $Q(s_k, a_k)$  of the particular action  $a_k$ . The target network takes next state  $s'_k$  and predicts the action  $a'_k$  with the largest Q-value from all actions that can be taken in that state, obtaining the target Q-value  $Q'(s'_k, a'_k)$ . The predicted and target Q-value is used to calculate the loss function and adjust the parameters of the neural network.

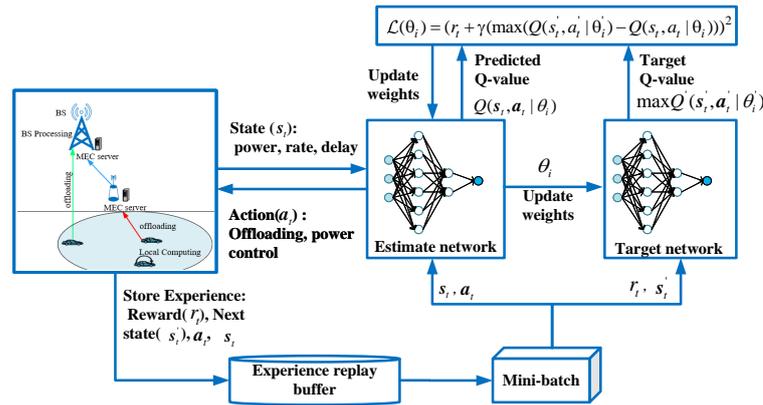


Figure 2. The framework of DQN-based multi-agent transmit power and task offloading scheme.

The loss function of the  $k$ -th agent is used to obtain the best Q-function by adjusting the neural network parameters, which is given by [42]

$$\mathbb{L}(\theta_i) = (r_k + \gamma(\max Q'(s'_k, a'_k | \theta'_i) - Q(s_k, a_k | \theta_i)))^2, \quad (28)$$

where  $\theta_i$  and  $\theta'_i$  denote the weights parameters of the main network and target network during the  $i$ -th training, respectively.  $\gamma \in [0, 1)$  is a discount factor that can weight the future rewards. In order to obtain the maximum long-term reward, the Q-function is updated based on the Bellman equation, which is written as

$$Q(s_k, a_k) = (1 - \alpha)Q(s_k, a_k) + \alpha(r_k + \gamma \max(Q'(s'_k, a'_k))), \quad (29)$$

where  $\alpha$  denotes the learning rate, which directly determines how quickly the agent can adapt to the environment. DRL combines reinforcement learning with the DQN neural network [43] and makes agents observe from the environment and intelligently select the action that interacts with the environment. With repeated iterations, the agent will choose a better server association, offloading ratio and transmit power to minimize the energy consumption of the MEC-assisted multi-vehicle V2X communication system. The DRL-based scheme is show in Algorithm 2.

**Algorithm 2** DRL-based Multi-agent Transmit Power and Offloading Optimization Scheme

- 1: **Inputs:** The state  $s$  of the vehicles.
- 2: **Initial:** DQN-networks for all agents, including of policy strategy  $\pi(s, a)$ ,  $Q(s, a)$ , and  $\{\alpha, \gamma, \epsilon, \theta\}$ .
- 3: **for** each iteration episode **do**
- 4: Agents observe state  $s_t$ , randomly selecting an action with probability  $\epsilon$ , or  $\arg \max_{a_t \in \mathcal{A}} \{Q(s_t, a_t)\}$  from the action space with probability  $1 - \epsilon$ . Interacting with the environment, obtaining the next state  $s'_t$  and reward (27). Storing  $\langle s_t, a_t, r_t, s'_t \rangle$  in the experience replay buffer.
- 5: **for** each training step **do**
- 6: Agents select a random mini-batch data from the experience replay buffer, and transmit it to the neural network. The estimate network calculate the predicted Q-value according to  $s_t, a_t$ , the target network calculate the target Q-value according to  $r_t, s'_t$ . According to the predicted and target Q-value, computing the loss function in (28), update the weights  $\theta_i$  to the target network and Q-value  $Q(s_t, a_t)$  in (29).
- 7: **end for**
- 8: **end for**
- 9: **Outputs:** The action  $a$  of the vehicles used to optimize the offloading decision, offloading ratio and power control.

**4.3. DRL Algorithm Complexity Analysis**

Computational complexity is critical for evaluating the algorithm performance. For the Q-learning algorithm, the Q-table of agents consists of  $(|S|)^K$  rows with  $|S| = 3$  denoting the length of the environment states and  $(|a|_p)^K$  columns with  $|a|_p = 50$  denoting distinct possible agent actions in our simulation. Thus, the whole Q-learning algorithm is  $O((|S|)^K \cdot (|a|_p)^K)$ , and the complexity of the Q-learning algorithm is primarily determined by the dimensions of state space, the distinct possible agent actions and the number of agents. From Algorithm 2, it can be known that the whole MADRL algorithm procedure mainly contains two parts:

- (1) **Calculate the reward function:** Agents, according to the state, select the beneficial action, interact with the environment and obtain the reward. Thus, the computational complexity of an agent calculating the reward is  $O(|s_t|)$ , with  $|s_t|$  denoting the length of the state space for the  $t$ -th training step.
- (2) **Select the beneficial action:** For each agent, the numbers of layers in the DQN network and neurons in each layer are considered. For the DQN network, the number of layers in the DQN network is  $M$ , and neurons in the  $m$ -th layer is  $U_m$ . Thus, the computational complexity of the  $m$ -th layer is  $O(U_{m-1}U_m + U_mU_{m+1})$ , and, for the  $t$ -th step, the computational complexity of an agent selecting the beneficial action is  $O_c = O(|s_t| \cdot U_2 + \sum_{m=3}^M (U_{m-1}U_m + U_mU_{m+1}) + U_{M-1} \cdot |a_t|)$ , with  $|a_t|$  denoting the length of the action space for the  $t$ -th training step. Therefore, the computational complexity of an agent is  $O(O_r + O_c)$ . The computational complexity for a complete episode is  $O(N_{\text{step}}(O_r + O_c))$ , where  $N_{\text{step}}$  is the number of steps in one episode. The computational complexity for the whole algorithm is  $O(N_{\text{episode}} \cdot N_{\text{step}}(O_r + O_c))$ , where  $N_{\text{episode}}$  is the number of iteration episodes. In this paper, all agents have the same DQN network, and the computational complexity of all agents is  $O(K(N_{\text{episode}} \cdot N_{\text{step}}(O_r + O_c)))$ . The complexity of the proposed algorithm is primarily determined by the network architectures of the DQN network structure and the dimensions of the state and action space. In addition, the complexity of the Q-learning algorithm is primarily determined by the number of agents, and the implementation of Q-learning is more complex compared to the MADRL algorithm in the multi-vehicle V2X communication system. Thus, the proposed MADRL algorithm is more suitable for solving the optimization problem (P1) in this work.

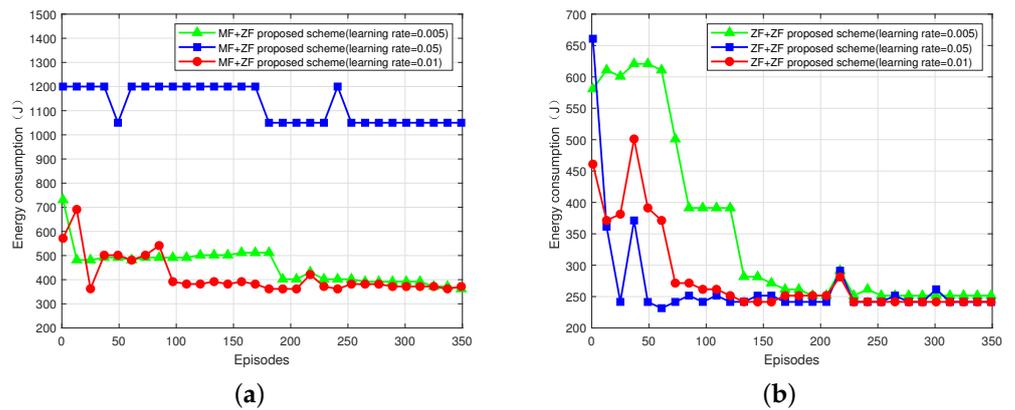
## 5. Simulation Results

Numerical results are provided to prove the effectiveness of the proposed DRL scheme in this section. Unless otherwise stated, the BS is located at the origin and its radius is 400 m. The number of RSUs  $L = 4$  and the radius of the RSUs is 200 m. The number of vehicles  $K = 6$ . Note that the vehicles are in the coverage of the RSUs, which are in the coverage of the BS. Other parameters are shown in Table 2. In addition, in order to reflect the excellent performance of the joint optimization scheme, the maximum transmit power and full offloading schemes are adopted as benchmark schemes. Specifically, the maximum transmit power scheme represents that vehicles use their maximum transmit power offloading task, and the full offloading scheme represents that vehicles offload all tasks to the RSUs or BS for edge computing. The DQN has one input layer, one output layer and two hidden layers, and the number of neurons in each layer is 10. The simulation environment is Python 3.7.4 and TensorFlow 2.1.0.

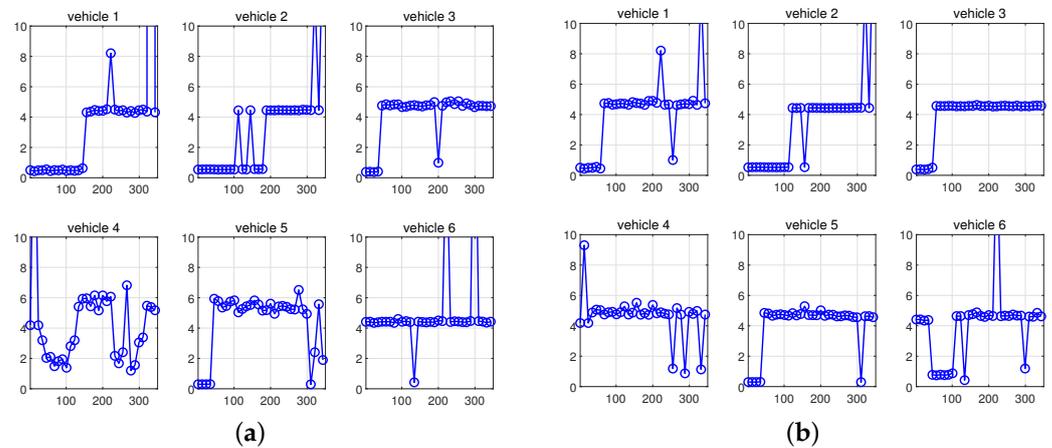
The energy consumption of the MEC-assisted multi-vehicle V2X system versus the number of training episodes under different learning rates, i.e.,  $\alpha = 0.005, 0.01$  and  $0.05$ , is shown in Figure 3. Figure 3a,b show that the energy consumption decreases when the RSUs are equipped with MF and ZF receivers, respectively. It can be seen that, whether the RSUs are equipped with MF or ZF receivers, as the training episode increases, the energy consumption decreases and is finally stable under the proposed scheme with different learning rates, especially when  $\alpha = 0.01$ . Thus,  $\alpha = 0.01$  is adopted to train vehicles to select the beneficial server association, offloading ratio and transmit power to minimize the energy consumption. In addition, energy consumption when the RSUs are equipped with the ZF receivers is lower than when equipped with the MF receivers. This is because the ZF receivers eliminating the inter-user interference of the V2I, I2B and V2B communication links results in a higher transmission rate. The delay of each vehicle versus the number of training episodes under the proposed scheme when  $\alpha = 0.01$  is shown in Figure 4. Figure 4a,b show the delay of each vehicle when the RSUs are equipped with MF and ZF receivers, respectively. It is noted that the delay of each vehicle is less than the maximum vehicle delay; thus, the proposed scheme can decrease the energy consumption while satisfying the delay constraint.

**Table 2.** Simulation parameters.

Symbol	Description	Value
$T_{\max}$	Maximum vehicle delay	8 s
$f^V$	Processing capability of the vehicles	0.1 Mbit/s
$f^R$	Processing capability of the RSUs	5 Mbit/s
$f^{BS}$	Processing capability of the BS	10 Mbit/s
$D_{\max}$	Data packet	2 M
$D^{\text{Prc}}$	Processed data packet size	1 M
$B^{V2I}$	Bandwidth of the V2I communication link	1 MHz
$B^{I2B}$	Bandwidth of the I2B communication link	5 MHz
$B^{V2B}$	Bandwidth of the V2B communication link	5 MHz
$P_{\max}^V$	Maximum transmit power of the vehicles	23 dBm
$P^R$	Transmit power of the RSUs	40 dBm
$d_o$	Length of offloading action space	5
$d_p$	Length of power control action space	5
$\sigma^2$	Noise power	−90 dBm
$\alpha$	Learning rate	0.01
$\gamma$	Discount factor	0.9
$\epsilon$	Exploration possibility	0.96
$N_{\text{episode}}$	Iteration episode	350
$N_{\text{step}}$	Training step in one episode	200

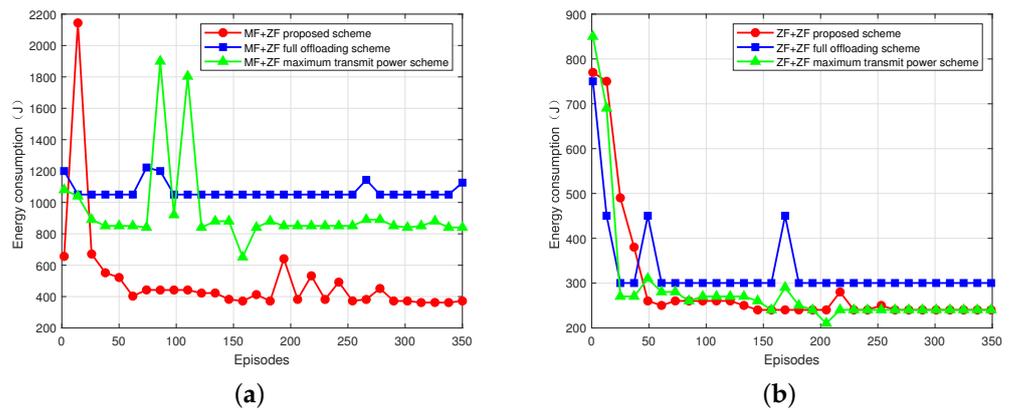


**Figure 3.** Energy consumption under the proposed scheme with different learning rate when  $L = 4$ . (a) RSU equipped with MF receivers. (b) RSU equipped with ZF receivers.



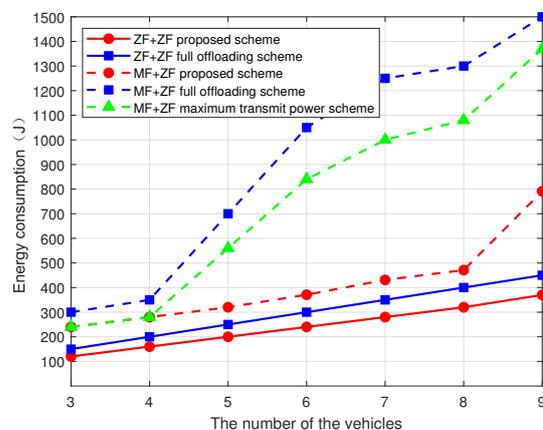
**Figure 4.** Delay of each agents under the proposed scheme when  $\alpha = 0.01$  and  $L = 4$ . (a) RSU equipped with MF receivers. (b) RSU equipped with ZF receivers.

The energy consumption versus the number of training episodes under different schemes is shown in Figure 5. Figure 5a,b show the energy consumption under different schemes when the RSUs are equipped with MF and ZF receivers, respectively. It is noted that the energy consumption decreases and is finally stable under different schemes. When the RSUs are equipped with the MF receivers, the proposed scheme can decrease the energy consumption by 55.89% and 64.67% more than the maximum transmit power and full offloading schemes, respectively. This is because the proposed scheme optimizes the offloading ratio and the transmission power jointly, which both have an impact on the energy consumption. When the RSUs are equipped with the ZF receivers, the proposed scheme can decrease the energy consumption by 20.00% more than the full offloading scheme, but has a similar performance compared to the maximum transmit power scheme. This is because ZF receivers can eliminate inter-user interference, making vehicles choose the maximum transmit power to offload tasks. In addition, RSUs equipped with ZF receivers under the proposed scheme can decrease the energy consumption by 35.04% more than MF receivers.



**Figure 5.** Energy consumption versus training episodes under different schemes when  $L = 4$ . (a) RSU equipped with MF receivers. (b) RSU equipped with ZF receivers.

Figure 6 shows the impact of the number of vehicles on the energy consumption. It can be seen that as the number of vehicles increases, the energy consumption increases under the same scheme, which is independent of the RSUs equipped with the receiver. This is because the number of vehicles increases, the interference caused by other vehicles increases and transmission rate decreases. In order to satisfy the delay constraint, vehicles need to offload more tasks to the MEC server for edge consumption, which makes the energy consumption increase. When the number of vehicles  $K$  is lower than the number of RSUs  $L$ , the energy consumption of the maximum transmit power scheme is the same as the proposed scheme. This is because the improved KNN algorithm assigns vehicles to different RSUs, and the vehicles will use the maximum transmit power in the offloading task but not cause interference to each other. When the number of vehicles is larger than the number of RSUs, the proposed scheme has a lower energy consumption than the benchmark schemes. Specifically, when  $K = 5$  and RSUs are equipped with MF receivers, the proposed scheme can decrease the energy consumption by 42.86% and 54.29% compared with the maximum transmission power and full offloading schemes, respectively. When  $K = 9$ , the energy consumption of the proposed scheme reduces by 42.26% and 47.27% compared with the maximum transmit power and the offloading schemes, respectively. When  $K = 9$  and RSUs are equipped with ZF receivers, the proposed scheme can decrease the energy consumption by 17.78% compared with the full offloading schemes, respectively. Thus, the proposed scheme has a better performance in reducing the energy consumption of the system when the number of vehicles increases.



**Figure 6.** Energy consumption versus number of vehicles when  $L = 4$ .

The energy consumption versus the number of RSUs  $L$  under different schemes when the number of vehicles  $K = 6$  is shown in Figure 7. It can be seen that as the number of

RSUs increases, the energy consumption decreases under the same scheme. This is because the number of RSUs increases and more vehicles can offload tasks to the RSUs and have a lower energy consumption. When the number of RSUs  $L = 8$ , the proposed scheme can decrease energy consumption by 19.93% more than the full offloading scheme and has a similar performance compared to the maximum transmit power scheme. This is because the improved KNN algorithm assigns vehicles to different RSUs when  $L > K$ , making vehicles choose the maximum transmit power to offload tasks.

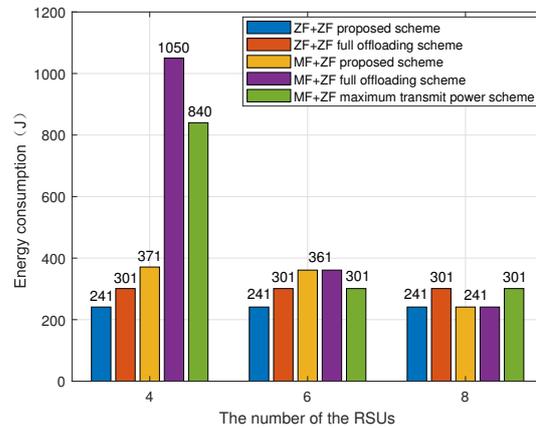


Figure 7. Energy consumption versus number of RSUs.

The energy consumption versus packet size increase under different schemes is shown in Figure 8. It is noted that as the packet size increases, the energy consumption increases under the same scheme, no matter what receivers the RSUs are equipped with. This is because the packet size increases and the process delay increases, and so the vehicles need to offload more tasks to the MEC server for edge consumption to satisfy the delay constraint. This makes the energy consumption increase. In addition, the energy consumption of the proposed scheme is lower than the benchmark schemes, and when the RSU is equipped with a ZF receiver, it can reduce more energy consumption. Specifically, when the packet size is 1.2 M and RSUs are equipped with MF receivers, the energy consumption of the proposed scheme reduces by 52.38% and 80.95% compared with the maximum transmit power and offloading scheme, respectively. When the packet size is 3M, the energy consumption of the proposed scheme reduces by 52.30% and 61.84% compared with the benchmark schemes. Thus, the proposed scheme has a better performance in reducing the energy consumption of the system when the packet size increases.

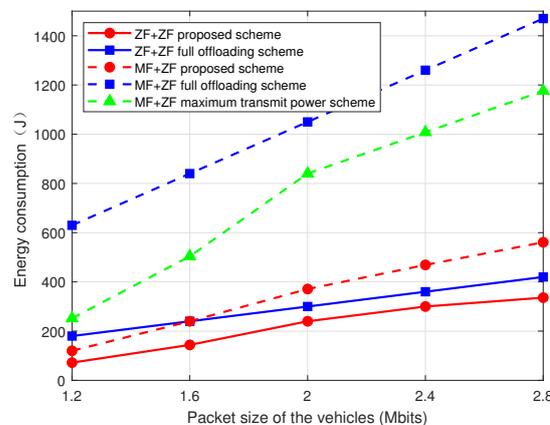


Figure 8. Energy consumption versus the packet size.

The DRL scheme is proposed to solve the non-convex energy consumption minimization problem. From the above simulation results, it can be concluded that the proposed

DRL scheme can obtain superior performance in reducing the energy consumption of the system under a different number of vehicles, number of RSUs and packet size.

## 6. Conclusions

This paper studied the uplink MEC-assisted multi-vehicle V2X communication system in which vehicles can offload tasks to the RSUs and BS for edge computing. To minimize the energy consumption and satisfy the delay requirements, an energy consumption minimization problem was formulated under a delay constraint. Due to the coupling of the server association, offloading ratio and transmit power, as well as delay constraints, the minimization problem is highly non-convex and difficult to solve. Thus, the DRL algorithm was proposed to solve the optimization problem. Vehicles as agents were trained according to the reward function, which is related to the energy consumption of the system and processing delay, to obtain a better action containing server association, offloading ratio and transmit power. In addition, an improved KNN algorithm was used to assign vehicles within the coverage area of multiple RSUs to reduce the action space and complexity of the DRL scheme. Numerical results have shown that the proposed scheme outperforms the maximum transmit power and full offloading schemes regarding minimizing the energy consumption with a different number of vehicles and packet size. In addition, the RSUs equipped with ZF receivers can decrease more energy consumption than MF receivers. However, the collaborative RSUs for computation processing are not considered in this work due to the complicated action space and the computation complexity. It is interesting to investigate the collaborative RSUs for computation processing to reduce the energy consumption and reduce the latency, especially for the resource-limited V2X network. Joint offloading and RSU allocation optimization is difficult with communication delay and power constraints due to the action space being related to a discrete RSU.

**Author Contributions:** Conceptualization, W.D., Y.H. and X.Z.; methodology, W.D., X.L. and Y.H.; software, H.C.; writing—original draft, X.L. and X.Z.; writing—review and editing, W.D. and X.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China under Grant 62101386, the Natural Science Foundation of Sichuan Province under Grant 2023NSFSC1388, the Guangdong Provincial Key Laboratory of Future Networks of Intelligence, the Chinese University of Hong Kong, Shenzhen under Grant 2022B1212010001-OF04, the National Natural Science Foundation of China under Grant 62201479 and the Fundamental Research Funds for the Central Universities under Grant 22120230311.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. You, X.; Wang, C.; Huang, J.; Gao, X. Towards 6G Wireless Communication Networks: Vision, Enabling Technologies, and New Paradigm Shifts. *Sci. China Inf. Sci.* **2021**, *64*, 110301. [[CrossRef](#)]
2. Moubayed, A.; Shami, A.; Heidari, P.; Larabi, A.; Brunner, R. Edge-Enabled V2X Service Placement for Intelligent Transportation Systems. *IEEE Trans. Mob. Comput.* **2021**, *20*, 1380–1392. [[CrossRef](#)] [[CrossRef](#)]
3. Noor-A-Rahim, M.; Liu, Z.; Lee, H.; Khyam, M.O.; He, J.; Pesch, D.; Moessner, K.; Saad, W.; Poor, H.V. 6G for Vehicle-to-Everything (V2X) Communications: Enabling Technologies, Challenges, and Opportunities. *Proc. IEEE* **2022**, *110*, 712–734. [[CrossRef](#)] [[CrossRef](#)]
4. Parada, R.; Vázquez-Gallego, F.; Sedar, R.; Vilalta, R. An Inter-Operable and Multi-Protocol V2X Collision Avoidance Service Based on Edge Computing. In Proceedings of the IEEE Vehicular Technology Conference (VTC-Spring), Helsinki, Finland, 19–22 June 2022; pp. 1–5. [[CrossRef](#)]
5. Vladyko, A.; Elagin, V.; Spirikina, A.; Muthanna, A.; Ateya, A.A. Distributed Edge Computing with Blockchain Technology to Enable Ultra-Reliable Low-Latency V2X Communications. *Electronics* **2022**, *11*, 173. [[CrossRef](#)] [[CrossRef](#)]
6. Amrita, G.; Mauro, C. Security Issues and Challenges in V2X: A Survey. *Comput. Netw.* **2020**, *169*, 107093. [[CrossRef](#)]
7. Wang, J.; Lv, T.; Huang, P.; Mathiopoulos, P.T. Mobility-Aware Partial Computation Offloading in Vehicular Networks: A deep Reinforcement Learning Based Scheme. *China Commun.* **2020**, *17*, 31–49. [[CrossRef](#)] [[CrossRef](#)]

8. Amrita, G.; Mauro, C. Efficient Anchor Point Deployment for Low Latency Connectivity in MEC-Assisted C-V2X Scenarios. *IEEE Trans. Veh. Technol.* **2023**, *72*, 16637–166495.
9. Prathiba, S.B.; Raja, G.; Anbalagan, S.; Dev, K.; Gurumoorthy, S.; Sankaran, A.P. Federated Learning Empowered Computation Offloading and Resource Management in 6G-V2X. *IEEE Trans. Netw. Sci. Eng.* **2022**, *9*, 3234–3243. [[Crossref](#)] [[CrossRef](#)]
10. Zhang, K.; Leng, S.; Peng, X.; Pan, L.; Maharjan, S.; Zhang, Y. Artificial Intelligence Inspired Transmission Scheduling in Cognitive Vehicular Communications and Networks. *IEEE Internet Things J.* **2019**, *6*, 1987–1997. [[Crossref](#)] [[CrossRef](#)]
11. Pratik, T.; Anurag, T.; Atul, K. GREENSKY: A Fair Energy-Aware Optimization Model for UAVs in Next-Generation Wireless Networks. *Green Energy Intell. Transp.* **2024**, *6*, 100130. [[Crossref](#)]
12. Li, X.; Li, J.; Yin, B.; Yan, J.; Fang, Y. Age of Information Optimization in UAV-Enabled Intelligent Transportation System via Deep Reinforcement Learning. In Proceedings of the IEEE Vehicular Technology Conference (VTC-Fall), London, UK, 26–29 September 2022; pp. 1–5. [[Crossref](#)]
13. Hwang, R.H.; Islam, M.M.; Tanvir, M.A.; Hossain, M.S.; Lin, Y.D. Communication and Computation Offloading for 5G V2X: Modeling and Optimization. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), Taipei, Taiwan, 7–11 December 2020; pp. 1–6. [[Crossref](#)]
14. Huang, Y.; Fang, Y.; Li, X.; Xu, J. Coordinated Power Control for Network Integrated Sensing and Communication. *IEEE Trans. Veh. Technol.* **2022**, *71*, 13361–13365. [[Crossref](#)] [[CrossRef](#)]
15. Nguyen, P.L.; Hwang, R.H.; Khiem, P.M.; Nguyen, K.; Lin, Y.D. Modeling and Minimizing Latency in Three-Tier V2X Networks. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), Taipei, Taiwan, 7–11 December 2020; pp. 1–6. [[Crossref](#)]
16. Bréhon-Grataloup, L.; Kacimi, R.; Beylot, A.L. Mobile Edge Computing for V2X Architectures and Applications: A Survey. *Comput. Netw.* **2022**, *206*, 108797. [[CrossRef](#)]
17. Dinh, H.; Nguyen, N.H.; Nguyen, T.T.; Nguyen, T.H.; Nguyen, T.T.; Le Nguyen, P. Deep Reinforcement Learning-Based Offloading for Latency Minimization in 3-Tier V2X Networks. In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA, 10–13 April 2022; pp. 1803–1808. [[Crossref](#)]
18. Wang, H.; Lin, Z.; Guo, K.; Lv, T. Computation Offloading Based on Game Theory in MEC-Assisted V2X Networks. In Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops), Montreal, QC, Canada, 14–23 June 2021; pp. 1–6. [[Crossref](#)]
19. Zhang, Y.; Dong, X.; Zhao, Y. Decentralized Computation Offloading over Wireless-Powered Mobile-Edge Computing Networks. In Proceedings of the IEEE International Conference on Artificial Intelligence and Information Systems (ICAIS), Dalian, China, 20–22 March 2020; pp. 137–140. [[Crossref](#)]
20. Xiong, R.; Zhang, C.; Yi, X.; Li, L.; Zeng, H. Joint Connection Modes, Uplink Paths and Computational Tasks Assignment for Unmanned Mining Vehicles? Energy Saving in Mobile Edge Computing Networks. *IEEE Access* **2020**, *8*, 142076–142085. [[Crossref](#)] [[CrossRef](#)]
21. Zhang, J.; Hu, X.; Ning, Z.; Ngai, E.C.H.; Zhou, L.; Wei, J.; Cheng, J.; Hu, B. Energy-Latency Trade-Off for Energy-Aware Offloading in Mobile Edge Computing Networks. *IEEE Internet Things J.* **2018**, *5*, 2633–2645. [[Crossref](#)] [[CrossRef](#)]
22. Kai, C.; Meng, X.; Mei, L.; Huang, W. Deep Reinforcement Learning Based User Association and Resource Allocation for D2D-Enabled Wireless Networks. In Proceedings of the IEEE/CIC International Conference on Communications in China (ICCC), Xiamen, China, 28–30 July 2021; pp. 1172–1177.
23. Sun, Y.; Xu, J.; Cui, S. Joint User Association and Resource Allocation Optimization for MEC-Enabled IoT Networks. In Proceedings of the IEEE International Conference on Communications (ICC), Seoul, Republic of Korea, 16–20 May 2022; pp. 4884–4889. [[Crossref](#)]
24. Lyu, L.; Shen, Y.; Zhang, S. The Advance of Reinforcement Learning and Deep Reinforcement Learning. In Proceedings of the IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), Changchun, China, 25–27 February 2022; pp. 644–648. [[Crossref](#)]
25. Lin, Y.; Zhang, Y.; Li, J.; Shu, F.; Li, C. Popularity-Aware Online Task Offloading for Heterogeneous Vehicular Edge Computing Using Contextual Clustering of Bandits. *IEEE Internet Things J.* **2022**, *9*, 5422–5433. [[Crossref](#)] [[CrossRef](#)]
26. Lin, Y.; Zhang, Z.; Huang, Y.; Li, J.; Shu, F.; Hanzo, L. Heterogeneous User-Centric Cluster Migration Improves the Connectivity-Handover Trade-Off in Vehicular Networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 16027–16043. [[Crossref](#)] [[CrossRef](#)]
27. Liang, T.; Lin, Y.; Shi, L.; Li, J.; Zhang, Y.; Qian, Y. Distributed Vehicle Tracking in Wireless Sensor Network: A Fully Decentralized Multiagent Reinforcement Learning Approach. *IEEE Sens. Lett.* **2021**, *5*, 1–4. [[Crossref](#)] [[CrossRef](#)]
28. Luong, N.C.; Hoang, D.T.; Gong, S.; Niyato, D.; Wang, P.; Liang, Y.C.; Kim, D.I. Applications of Deep Reinforcement Learning in Communications and Networking: A Survey. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3133–3174. [[CrossRef](#)]
29. Zhong, C.; Gursoy, M.C.; Velipasalar, S. A Deep Reinforcement Learning-Based Framework for Content Caching. In Proceedings of the Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 21–23 March 2018; pp. 1–6. [[Crossref](#)]
30. Yin, B.; Li, X.; Yan, J.; Zhang, S.; Zhang, X. DQN-Based Power Control and Offloading Computing for Information Freshness in Multi-UAV-Assisted V2X System. In Proceedings of the IEEE Vehicular Technology Conference (VTC-Fall), London, UK, 26–29 September 2022; pp. 1–6. [[Crossref](#)]
31. Hossain, T.; Ali, M.Y.; Mowla, M.M. Energy Efficient Massive MIMO 5G System with ZF Receiver. In Proceedings of the International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, Bangladesh, 26–28 December 2019; pp. 133–136.

32. Das, D.; Shbat, M.; Tuzlukov, V. Employment of Generalized Receiver with Equalization in MIMO Systems. In Proceedings of the IET International Conference on Information Science and Control Engineering (ICISCE), York, UK, 15–17 May 2012; pp. 1–5. [[Crossref](#)]
33. Louie, R.H.Y.; McKay, M.R.; Collings, I.B. Spatial Multiplexing with MRC and ZF Receivers in Ad Hoc Networks. In Proceedings of the IEEE International Conference on Communications (ICC), Dresden, Germany, 14–18 June 2009; pp. 1–5. [[Crossref](#)]
34. Wang, H.; Li, X.; Ji, H.; Zhang, H. Federated Offloading Scheme to Minimize Latency in MEC-Enabled Vehicular Networks. In Proceedings of the IEEE Globecom Workshops (GC Wkshps), Abu Dhabi, United Arab Emirates, 9–13 December 2018; pp. 1–6. [[Crossref](#)]
35. Yang, T.; Li, X.; Shao, H. Federated Learning-Based Power Control and Computing for Mobile Edge Computing System. In Proceedings of the IEEE Vehicular Technology Conference (VTC-Fall), Norman, OK, USA, 27–30 September 2021; pp. 1–6. [[Crossref](#)]
36. Osman, R.A. Optimizing Autonomous Vehicle Communication through an Adaptive Vehicle-to-Everything (AV2X) Model: A Distributed Deep Learning Approach. *Electronics* **2023**, *12*, 4023. [[Crossref](#)] [[CrossRef](#)]
37. Wu, X.; Ma, Z.; Wang, Y. Joint User Grouping and Resource Allocation for Multi-User Dual Layer Beamforming in LTE-A. *IEEE Commun. Lett.* **2015**, *19*, 1822–1825. [[Crossref](#)] [[CrossRef](#)]
38. Wang, Q.; Wang, C.; Feng, Z.; Ye, J.F. Review of K-Means Clustering Algorithm. *Electron. Des. Eng.* **2012**, *20*, 21–24.
39. Zhang, N.; Karimoune, W.; Thompson, L.; Dang, H. A Between-Class Overlapping Coherence-Based Algorithm in KNN classification. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; pp. 572–577. [[Crossref](#)]
40. Eick, C.; Zeidat, N.; Vilalta, R. Using Representative-Based Clustering for Nearest Neighbor Dataset Editing. In Proceedings of the IEEE International Conference on Data Mining (ICDM), Brighton, UK, 1–4 November 2004; pp. 375–378. [[Crossref](#)]
41. C, C. Prediction of Heart Disease using Different KNN Classifier. In Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 6–8 May 2021; pp. 1186–1194. [[Crossref](#)]
42. Li, X.; Yin, B.; Yan, J.; Zhang, X.; Wei, R. Joint Power Control and UAV Trajectory Design for Information Freshness via Deep Reinforcement Learning. In Proceedings of the IEEE Vehicular Technology Conference (VTC-Spring), Helsinki, Finland, 19–22 June 2022; pp. 1–5. [[Crossref](#)]
43. Chen, W.; Qiu, X.; Cai, T.; Dai, H.N.; Zheng, Z.; Zhang, Y. Deep Reinforcement Learning for Internet of Things: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 1659–1692. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.