Adaptive Digital Twin and Multiagent Deep Reinforcement Learning for Vehicular Edge Computing and Networks

Ke Zhang, *Member, IEEE*, Jiayu Cao, and Yan Zhang, *Fellow, IEEE*

***Abstract*—Technological advancements of urban informatics and vehicular intelligence have enabled connected smart vehicles as pervasive edge computing platforms for a plethora of powerful applications. However, varies types of smart vehicles with distinct capacities, diverse applications with different resource demands as well as unpredictive vehicular topology, pose significant challenges on realizing efficient edge computing services. To cope with these challenges, we incorporate digital twin technology and artificial intelligence into the design of a vehicular edge computing network. It centrally exploits potential edge service matching through evaluating cooperation gains in a mirrored edge computing system, while distributively scheduling computation task offloading and edge resource allocation in an multiagent deep reinforcement learning approach. We further propose a coordination graph driven vehicular task offloading scheme, which minimizes offloading costs through efficiently integrating service matching exploitation and intelligent offloading scheduling in both digital twin and physical networks. Numerical results based on real urban traffic datasets demonstrate the efficiency of our proposed schemes.**

***Index Terms*—Digital twin, multiagent deep deterministic policy gradient (MADDPG), vehicular edge computing.**

I. Introduction

INTELLIGENT transportation system has been a fundamental part both for industry and society. Its activities, such as autonomous driving and smart logistics, depend on environmental information processing and traffic behavior decision, which require intensive computation under strict delay constraints [1]. Although the processing power of smart vehicles is much stronger than portable devices, the complex processing requirements of transportation applications may still pose critical challenges to vehicle processors. Furthermore, different smart vehicles may be equipped with various types of processors that bring different computing performance. An individual smart vehicle with limited onboard computing resource may not meet these processing requirements.

Vehicular edge computing, which can realize the sharing of the computing resource at the edge of vehicular networks, is a dominant paradigm to meet the intensive computation demands. When a vehicle offloads its task to another smart vehicle or a road side units (RSU) with adequate computing power, the communication scheduling for task data delivery is closely related to the computing resource management for task processing, which makes task offloading complicated. Moreover, resource competition between different offloading node pairs as well as time-varying topology of vehicular networks further brings unprecedented challenges in managing vehicular edge computing.

Artificial intelligence (AI) has emerged as a promising approach to cope with abovementioned challenges [2]. Recent advancements in AI and machine learning bring us significant capabilities to aware system environment, determine action strategies, and tackle complex problems that once seemed impossible [3]. However, the effective implementation of AI approach always relies on accurate and real-time system information gathered by learning agents. In vehicular networks characterized by massive connected smart vehicles, highly dynamic topology and limited wireless spectrum, it is impractical to form a centralized AI manager that schedules edge services of the entire network. To address this problem, we turn to multiagent distributed learning empowered vehicular edge management, yet there are still critical challenges in efficient collaboration and joint decision optimization among these multiple agents.

Digital twin is an appealing technology, which brings data-driven representations of physical world to mirrored virtual space [4]. This state mapping between real and virtual dimensions provides users with comprehensive insights of the investigated system, and dramatically reshapes design and engineering process [5]. For instance, in smart grid management, digital twin draws power status data from various types of sensors and offer engineers a virtual grid network diagram for adjusting [6]. In vehicle assisted driving, digital twin of road environment enables drivers to aware complete traffic states and improves driving safety [7].

Manuscript received August 30, 2020; revised May 29, 2021; accepted June 4, 2021. Date of publication June 10, 2021; date of current version October 27, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62071092 and Grant 61941102, in part by the Key R&D Project of Sichuan Province under Grant 2019YFG0520, and in part by the National Key R&D Program of China under Grant 2018YFE0117500. Paper no. TII-20-4150. Paper no. TII-20-4150. *(Corresponding author: Yan Zhang.)*

Ke Zhang and Jiayu Cao are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: zhangke@uestc.edu.cn; liuda945@163.com).

Yan Zhang is with the Department of Informatics, University of Oslo, 0316 Oslo, Norway (e-mail: yanzhang@ieee.org).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2021.3088407.

Digital Object Identifier 10.1109/TII.2021.3088407

It is envisioned that digital twin and AI are two key technologies in industrial Internet of things. Moreover, merging digital twin with AI and machine learning will generate great benefits. On the one hand, digital twin provides AI with comprehensive and accurate system state information, which is exactly what learning process requires [8]. On the other hand, AI brings much intelligence to digital twin, making its information collection and system description smart and efficient. As digital twin technology is at its infant stage, limited research efforts have been made on the collaboration of digital twin and AI. Furthermore, the way of jointly applying these two technologies to vehicular edge computing is still an unexplored question.

To fill this gap, in this article, we propose a new vehicular edge computing network base on digital twin and multiagent learning, which improves agent collaboration and optimizes task offloading efficiency. To the best of our knowledge, this is the first work that focuses on digital twin and multiagent learning in vehicular edge computing networks. The main contributions of this article are as follows.

1) We present a new edge management framework that improves the learning efficiency of multiple agents through digital twin technology while enhances the replicability performance between virtual and physical networks by a learning approach.
2) We leverage digital twin to reveal the potential cooperation between different vehicles and adaptively form multiagent learning groups, which improves edge resource utilization and reduces learning complexity.
3) We propose a distributed multiagent learning scheme, which minimizes vehicular task offloading cost under strict delay constrains in complex vehicular networks and dynamically adjusts digital twin network's state mapping mode.

The rest of this article is organized as follows. In Section II, we review related work. We introduce system model in Section III, and propose digital twin and AI joint scheduling mechanism in Section IV. We present numerical results in Section V. Finally, Section VI concludes this article.

II. Related Work

Being a promising paradigm that enables powerful applications on computation capacity limited vehicles, vehicular edge computing has attracted considerable research interest. Lee *et al.* [9] took parked vehicles as edge servers, and proposed a reinforcement learning combined heuristic algorithm to allocate computing resources to vehicular applications. Zhang *et al.* [10] developed a social-aware mobile edge computing scheme, which maximizes content processing utility through exploiting the social relation between vehicles and roadside units. Wang *et al.* [11] designed an imitation learning enabled online task scheduling algorithm to minimize edge system energy consumption and satisfy task latency constraints. In order to formulate efficient and secure vehicular edge computing networks, Dai *et al.* [12] integrated deep reinforcement learning and permissioned blockchain into vehicular content caching management. Chen *et al.* [13] presented a two-stage meta-learning based mechanism to reserve vehicular computing resources.

As vehicular edge computing networks always consist of massive and large-scale distributed vehicles, centralized learning approach that heavily relies on whole network state collection may not be suitable for efficient edge service management [14]. Multiagent learning, which awares system states and makes designs in a parallel and distributed manner, has been raised to address this problem. Ahrarinouri *et al.* [15] explored multiagent reinforcement learning technology for multicarrier energy management. Cui *et al.* [16] developed a multiagent reinforcement learning framework in unmanned aerial vehicle resource allocation. Kaur and Kumar [17] took multiagent learning to mitigates wireless interference in energy-efficient cognitive radio networks. Applying multiagent learning in vehicular networks, Yuan *et al.* [18] proposed an optimization algorithm to minimize computing service migration cost under delay constraints. Chu *et al.* [19] presented a distributed and adaptive learning-based traffic signal control strategy in complex vehicular traffic networks.

The digital twin concept, which is a key enabler to seamlessly integrate cyber and physical spaces, attracts increasing attention by both academia and industry. Kaigom and Romann [20] utilized this concept to construct robotic digital twin framework that helps physical robot to obtain instruction flexibly and optimize its performance. Altun *et al.* [21] introduced a digital twin based reference model, which ensures home application ownership and promotes human-centric services. Gehrmann and Gunnarsson [22] investigated digital twin replication model and security architectures in securing data dispatch. Focused on communication efficiency of digital twin networks, Lu *et al.* [23] leveraged federated learning to minimize twin model construction costs in industrial IoT environments. Till now, very few works have been done on digital twin enabled mobile edge computing networks. Dong *et al.* [24] integrated digital twin with deep learning to obtain user association strategies and task offloading probabilities. However, the centralized learning approach in this article does not fit for the distributed management requirements of the internet of vehicles.

Although the abovementioned studies have provided some useful insights and promising paradigms for edge service scheduling, digital twin inspired vehicular edge computing has not been investigated. Furthermore, the effective integration between digital twin and multiagent learning techniques remains a critical challenge. To address this gap, in this article, we propose a two-way cooperative mechanism of digital twin and multiagent learning in vehicular edge networks, which optimizes task offloading strategies and minimizes edge service cost.

III. System Model

Fig. 1 shows the framework of a digital twin empowered vehicular edge computing network. There are N smart vehicles running on the roads. These vehicles are equipped with computing power to process tasks and perform learning functions [25]. For vehicle i, where $i \in \mathcal{N}$, its computing capability is denoted as f_i CPU cycles per second. To enable powerful vehicular applications, such as autonomous driving and onboard entertainment, the vehicles generate various types of tasks to be processed. Without loss of generality, we consider vehicle i has J_i types of tasks, and task $w_{i,j}$ is described in the form of three

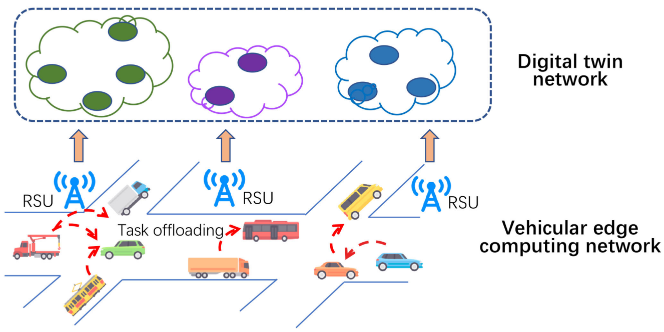


Fig. 1. Digital twin empowered vehicular edge computing networks.

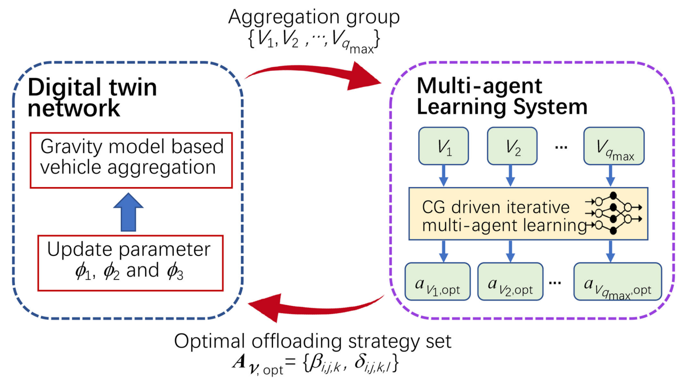


Fig. 2. Incorporation of DTN and multiagent learning for vehicular edge computing.

elements as $w_{i,j} = \{C_{i,j}, D_{i,j}, T_{i,j}^{\max}\}$. Here, $C_{i,j}$ is the amount of computing resources required to execute this task, and $D_{i,j}$ presents the size of the task input data. $T_{i,j}^{\max}$ is the maximum delay that task $w_{i,j}$ can tolerant.

As different vehicles have divers computing capabilities and task processing requirements, parts of vehicles may have sufficient computing resources, while others are tight. Through vehicle-to-vehicle (V2V) communication, a vehicle can offload its tasks to the other ones. We call the target vehicles as vehicular edge servers. Let $\beta_{i,j,k} = 1$ denote vehicle i offloads its task j to vehicular server k, and $\beta_{i,j,k} = 0$ present the vehicle does not offload task j to server k. For each task, it can only be offloaded to one server at most. Thus, there is $\sum_{k=1}^{|V_i|} \beta_{i,j,k} \leq 1$, where V_i is the set of vehicular servers that can be reached by vehicle i.

The time consumed to complete task $w_{i,j}$ is divided into two parts, namely the offloading task transmission time and the task execution time. There are L orthogonal channels for V2V communication, and the bandwidth of each channel is B. We take P_i and $\mathscr{A}_{i,k}$ to denote the transmission power of vehicle i and the transmission gain between vehicles i and k, respectively. Moreover, $\delta_{i,k,l} = 1$ indicates channel l is allocated to task data transmission from vehicle i to target server k. Then, the transmission rate between these two vehicles is calculated as

$$R_{i,k,l} = B\log_2\left(1 + \frac{P_i \mathscr{A}_{i,k,l}}{\sigma^2 + \sum_{i'=1, i'\neq i}^{N} \delta_{i',k,l} P_{i'} \mathscr{A}_{i',k,l}}\right) \quad (1)$$

where σ^2 is white noise of the channel [26]. According to (1), the transmission time of task $w_{i,j}$ from vehicles i to k through channel l is shown as $T_{i,j,k,l}^{\mathrm{tran}} = D_{i,j}/R_{i,k,l}$.

A target vehicular server may receive multiple tasks from the other vehicles, and puts these tasks in a queue. Taking into account task delay constraints, the target server executes the tasks in order according to the length of remaining time, from short to long. Consequently, a task's execution time consists of waiting time in the queue and the time processed in CPU. The execution time of $w_{i,j}$ can be presented as

$$T_{i,j,k}^{\mathrm{exe}} = \sum_{i'=1}^{N} \sum_{j'=1}^{J_{i'}} \mathbf{1}\{T_{i',j'}^{\mathrm{rem}} \leq T_{i,j}^{\mathrm{rem}}\} \beta_{i',j',k} C_{i',j'} / f_{i'} \quad (2)$$

where $\mathbf{1}\{\hat{x}\}$ is an indicator function that equals to 1 if $\hat{x}$ is true and 0 otherwise, and $T_{i,j}^{\mathrm{rem}}$ is the remaining time of task $w_{i,j}$ before deadline.

To improve vehicular computing resource utilization, a price-based incentive mechanism is incorporated into resource scheduling. For a vehicular server, the weaker its computing power, the greater the resource demands of its queuing tasks and the tighter the tasks' delay constraints, the higher the price of resources providing for the guest tasks. Thus, the price of an unit computing resource of vehicle i is defined as

$$z_i = \ln\left(1 + \frac{\sum_{j=1}^{J_i} (1 - \beta_{i,j,k}) \eta / T_{i,j}^{\max} (f_i - C_{i,j})}{\sum_{j=1}^{J_i} (1 - \beta_{i,j,k})}\right) \quad (3)$$

where η is a constant coefficient.

In the vehicular edge system, there locates a digital twin network (DTN), which keeps mapping the vehicle states in the physical space, such as communication topology and computing resource demands, to virtual digital space. With the help of DTN, edge service optimization and resource allocation strategies can be efficiently obtained.

IV. Incorporation Schemes of Digital Twin and Multiagent Learning in Vehicular Edge Computing Management

Being merged with digital twin, AI learning gains comprehensive state information and effective guidance for agent learning, while it helps digital twin to accurately model the physical system. In this section, we investigate the incorporation of digital twin and multiagent learning in vehicular edge computing networks, and propose optimal edge service scheduling schemes. The main framework of these schemes is shown in Fig. 2.

A. *Digital Twin Network Aided Edge Vehicle Aggregation*

Due to the large-scale distribution of massive vehicles, it is costly and impractical to globally schedule task offloading of the whole edge network. To address this issue, we leverage DTN and gravity model to design an edge service aggregation scheme, which efficiently aggregates vehicles based on the potential matching relations between supply and demand of computing resources, and greatly reduces the complexity of task offloading scheduling.

To guide the edge service aggregation, digital twin of the vehicular edge network are constructed in the RSUs. Each RSU collects the computing capabilities and communication topology of its surrounding vehicles. Then, they share the gathered information through wired transmission, and form the vehicular edge DTN. A DTN can be regarded as a combination of logical models and parameters recorded in the digital space to characterize the states of the objects in the physical space. Following this consideration, in our work, we define the element of the DTN as $D_s = \{\mathcal{M}, \Phi, \varpi\}$. Here, $\mathcal{M}$ denotes the digital model of the vehicles in the physical system, which is described by vehicle task set $\{w_{i,j}\}$, computing capability set $\{f_i\}$, resource price set $\{z_i\}$, and available transmission rate set $\{R_{i,j}\}$. $\Phi = \{\phi_1, \phi_2, \phi_3\}$ is the modeling parameters, which reflects the importance of the three factors of resource, price, and communication in the DTN modeling, respectively. The value of the parameters updates periodically, and ϖ is the sequence number of the mapping periods.

With the aid of the DTN, we develop a gravity model based vehicle aggregation scheme. Gravity model was derived from Isaac Newton's law of gravity, and has been widely used to estimate the bilateral interactions between objects that are characterized by their mass and distance [27]. Here, we reform the gravity model, and make it suitable to characterize the supply and demand relationship of vehicular edge service. The mass of vehicle i that acts as computing service requester is defined as

$$m_i = \sum_{j=1}^{J_i} C_{i,j}/f_i \tag{4}$$

which measures the demands of vehicle i to alleviate its resource limitation by task offloading. When vehicle i works as an edge server, its mass is expressed as $m_i' = m_i^{-1}$, which shows the availability of surplus resources.

Applying the mass definition into the gravity model, we get the gravitation in the service association between vehicle i and i' as

$$F_{i,i'} = \frac{\phi_1 \max(m_i/m_{i'}, m_{i'}/m_i)}{(\phi_2(z_i/m_i + z_{i'}/m_{i'}) + \phi_3/R_{i,i'})^2}. \tag{5}$$

Since either of vehicles i and i' may become the offloading server of the service requester, in the numerator of (5), we take maximum function to extract the strongest supply-demand relation between the two vehicles. Moreover, in the denominator, the distance element in the traditional gravity model is extended to a joint factor of computing resource price and V2V communication rate, which affect the edge serving performance.

Based on the gravitation obtained in (5), we divide the vehicles into multiple aggregation groups, which are denoted as $\{\mathcal{V}\}$. The main steps of the aggregation scheme are shown in Algorithm 1.

Algorithm 1: Gravity Model Based Edge Vehicle Aggregation.

Require:
Initialize modeling parameters ϕ_1, ϕ_2 and ϕ_3;
Initialize aggregation group set $\mathcal{V} = \emptyset$;

1: Calculate gravitation set $\{F_{i,i'}\}, \forall i, i' \in \mathcal{N}$, according to (5);
2: **while** $\mathcal{N} \neq \emptyset$ **do**
3: Select vehicle pair $\{x, x'\}$, where $F_{x,x'} = \arg\max_{\{i,i'\}}\{F_{i,i'}\}, \forall i, i' \in \mathcal{N}$;
4: **if** $\mathcal{V} == \emptyset$ **then**
5: $\mathcal{V} = \{V_1\}$, where $V_1 = \{x, x'\}$;
6: **else if** $x \parallel x' \in V_q$, where $V_q \in \mathcal{V}$ and $q = \{1, 2, \ldots, q_{\max}\}$ **then**
7: $V_q = V_q \cup \{x' \parallel x\}$;
8: **else if** x & $x' \notin V_q$, where $V_q \in \mathcal{V}$ and $q = \{1, 2, \ldots, q_{\max}\}$ **then**
9: $\mathcal{V} = \mathcal{V} \cup V_{q_{\max}+1}$, where $V_{q_{\max}+1} = \{x, x'\}$;
10: **end if**
11: Take newly updated V_q as a new virtual vehicle, and get its mass as $m_{V_q} = \sum_{i\in V_q}\sum_{j=1}^{J_i} C_{i,j}/\sum_{i\in V_q} f_i$;
12: $\mathcal{N} = \mathcal{N}\backslash\{x, x'\}$;
13: Recalculate gravitation set $\{F_{i,i'}\}$, $\forall i, i' \in \mathcal{N} \cup \{V_1, V_2, \ldots, V_{q_{\max}}\}$;
14: **end while**
15: **return** Aggregation groups $\{V_1, V_2, \ldots, V_{q_{\max}}\}$;

B. Multiagent Learning Empowered Edge Resource Allocation

The DTN aided aggregation algorithm has divided the complex vehicular network into multiple parts according to potential service associations, and effectively reduced edge service scheduling complexity. Based on this aggregation, we leverage multiagent learning approach to optimize edge resource allocation.

Multiagent learning is a promising paradigm that enables agents distributively making decisions according to their local environmental awareness, and achieves satisfactory learning performance in a large-scale complex system. By trial and error without manual engineering of the state space, the learning approach is able to deal with optimization problems of large scale vehicular networks in a general way. Moreover, we can leverage knowledge transfer and strategy reuse to improve the learning efficiency. As the vehicles in the edge network have computing and communication capabilities, they may act as agents to learn the optimal edge scheduling strategies. To minimize the task offloading costs under delay constraints, the optimization problem is given in the following form:

$$\begin{aligned}
&\min_{\{\beta_{i,j,k},\delta_{i,j,k,l}\}} \sum_{V_q\in\mathcal{V}}\sum_{i=1}^{|V_q|}\sum_{j=1}^{J_i}\sum_{k=1}^{|V_q|}\beta_{i,j,k}\sum_{l=1}^{L}\delta_{i,j,k,l}C_{i,j}z_i\\
&\text{C1}: \sum_{k=1}^{V_q}\beta_{i,j,k} = 1, \quad \forall V_q \in \mathcal{V}, i, k \in V_q, j \in J_i\\
&\text{C2}: \beta_{i,j,k} = 0, \quad \forall V_q \in \mathcal{V}, i \in V_q, j \in J, k \notin V_q\\
&\text{C3}: \sum_{k=1}^{V_q}\beta_{i,j,k}(T_{i,j,k}^{\text{tran}} + T_{i,j,k}^{\text{exe}}) \le T_{i,j}^{\max}\\
&\qquad \forall V_q \in \mathcal{V}, i, k \in V_q, j \in J_i
\end{aligned} \tag{6}$$

where constraint C1 ensures that a task can only be offloaded to one vehicle for processing at most. C2 indicates that task offloading only occurs between vehicles belonging to the same aggregation group, and C3 shows the time consumption including transmission and execution time should below the delay constraints of the tasks. Problem (6) is an integer programming problem, and has been proved NP complete.

Let $U_{V_q} = \sum_{i=1}^{|V_q|}\sum_{j=1}^{J_i}\sum_{k=1}^{|V_q|}\beta_{i,j,k}\sum_{l=1}^{L}\delta_{i,j,k,l}C_{i,j}z_i$. The target function of (6) can be written as $\min\sum_{V_q\in\mathcal{V}}U_{V_q}$. According to C2, there is no offloading correlation between different aggregation groups. Thus, to address problem (6), we turn to minimize U_{V_q} by taking a learning approach, where $V_q\in\mathcal{V}$. The number of leaning iterations is represented by time slot t.

For vehicle i that belongs to aggregation group V_q, its action taken at time slot t is shown as $a_i^t = \{\beta_{i,j,k}^t, \delta_{i,j,k,l}^t\}$, where $i,k\in V_q$, $j\in J_i$, and $l\in L$. Then, the action set of the multiple agents is given as $A^t=\{a_i^t\}$. The state at time slot t can be presented as $S^t = \{T_{i,j}^{\mathrm{rem},t}, \Gamma_k^t\}$, where $T_{i,j}^{\mathrm{rem},t}$ and Γ_k^t are the remaining completion time of task $w_{i,j}$ and the set of tasks that have been queued for processing in vehicle k at time slot t, respectively. Taking action A^t in state S^t, the learning system of V_q gains reward

$$Q_q^t(S^t,A^t) = \sum_{i=1}^{|V_q|}\sum_{j=1}^{J_i}\sum_{k=1}^{|V_q|}\beta_{i,j,k}^t\sum_{l=1}^{L}\delta_{i,j,k,l}^t C_{i,j}z_i. \tag{7}$$

The main goal of the multiagent learning in group V_q is to find optimal action strategy for the agents to minimize the group's task offloading costs, which are presented as

$$Q_q(S^0,A) = \mathbb{E}\left[\sum_{t=0}^{\infty}\xi Q_q^t(S^t,A^t)|S^0\right] \tag{8}$$

where ξ is a discount coefficient that indicates the effect of future reward on the current actions, and $0<\xi<1$.

Although taking vehicles of aggregation V_q as learning agents is an appealing approach to obtain the optimal task offloading and resource allocation strategies, there emerges some critical challenges in the learning implementation. The main challenging problem is the nonstationary learning environment caused by concurrent learning processes of multiple agents, which brings much difficulty to the coordination of environmental cognition and strategy acquisition between agents. Moreover, unstable communication topology and limited wireless transmission rate of the Internet of vehicles further restrict the comprehensive information interaction between vehicle agents. Thus, it is imperative to exploit a certain level of correlations among these agents, and design an efficient interagent interaction mechanism.

To address the abovementioned challenges, we resort to coordination graph (CG) technique, which helps to decompose complex coupling relationships of multiple agents into a linear combinations mode [28]. This decomposition can be shown in an undirected graph $G=\{\Psi,E\}$, where each node $\psi\in\Psi$ represents a learning agent and an edge $(\psi,\psi')\in E$ indicates the connected agents should coordinate their actions. There is a theorem of the decomposition.

Theorem 1: In the process of decomposing the complex coupling relationship between multiple agents into linear relationship, it has been proved that the achieved optimal joint actions and the coordination reward do not depend on the agent elimination order [29].

Algorithm 2: Multi-Agent Iterative Learning at Time Slot t in an Aggregation Group.

Require:
Aggregation group V_q, where $q\in\{1,2,\ldots,q_{\max}\}$;
Initialize $Y=|V_q|$;
for $y=Y$ to 1 **do**
  if $y==Y$ **then**
    Obtain agent x_Y's action $a_{x_Y}^{t,\mathrm{opt}}$ according to (11), and then calculate its action reward $Q_{x_Y,V_{q,Y-1}}^{t,\mathrm{opt}}$;
  else if $1<y<Y$ **then**
    Set reward $Q_q^t(S^t,A^t)=\sum_{\varsigma=y+1}^{Y}Q_{x_\varsigma,V_{q,\varsigma-1}}^{t,\mathrm{opt}} + Q_{x_y,V_{q,y-1}}^t(S_{x_y,V_{q,y-1}}^t, a_{x_y}^t, a_{V_{q,y-1}}^t) + \sum_{(x,x')\in E/\{(x_Y,V_{q,y-1})\cup\ldots\cup\{x_{y+1},V_{q,y}\}\}}Q_{x,x'}^t (S_{x,x'}^t, a_x^t, a_{x'}^t)$;
    Get vehicular agent x_y's action $a_{x_y}^{t,\mathrm{opt}} = \arg\min_{a_{x_y}^t} Q_{x_y,V_{q,y-1}}^t(a_{x_y}^t|S_{x_y,V_{q,y-1}}^t, A'^t_{V_{q,y-1}}, a_{x_{y+1}}^{t,\mathrm{opt}}, a_{x_{y+2}}^{t,\mathrm{opt}},\ldots,a_{x_Y}^{t,\mathrm{opt}})$;
    Calculate agent x_y's action reward $Q_{x_Y,V_{q,Y-1}}^{t,\mathrm{opt}}$;
  else if $y==1$ **then**
    Set reward $Q_q^t(S^t,A^t)=\sum_{\varsigma=2}^{Y}Q_{x_\varsigma,V_{q,\varsigma-1}}^{t,\mathrm{opt}} + Q_{x_1}^t,(S_{x_1}^t, a_{x_1}^t, a_{x_2}^{t,\mathrm{opt}},\ldots,a_{x_Y}^{t,\mathrm{opt}})$;
    Derive vehicular agent x_1's action $a_{x_1}^{t,\mathrm{opt}} = \arg\min_{a_{x_1}^t} Q_{x_1}^t(a_{x_1}^t|S_{x_1}^t, a_{x_2}^{t,\mathrm{opt}}, a_{x_3}^{t,\mathrm{opt}},\ldots,a_{x_Y}^{t,\mathrm{opt}})$;
  end if
end for
returnAction set $\{a_{x_1}^{t,\mathrm{opt}}, a_{x_2}^{t,\mathrm{opt}},\ldots,a_{x_Y}^{t,\mathrm{opt}}\}$;

Applying CG technique in this multiagent learning system and incorporating gravitation to model the coordination between agents, we take edges to connect vehicle pairs $\{x,x'\}$ given in Algorithm 1 and form a coordination graph. Based on this graph, $Q_q^t(S^t,A^t)$ in (7) can be transformed into the linear sum of the rewards of the coordinated agent pairs, and be shown as

$$Q_q^t(S^t,A^t) = \sum_{(x,x')\in E}Q_{x,x'}^t(S_{x,x'}^t, a_x^t, a_{x'}^t) \tag{9}$$

where x and x' can either be an learning agent or a virtual entity composed of multiple agents, and $S_{x,x'}^t$ is the relevant state of x and x' at time slot t.

Next we investigate the coordination action strategy of agent pairs (x,x'). In order to lower the learning complexity and reduce the amount of information exchanged between agents, so as to cater for the large number of vehicles and limited bandwidth of vehicular communication, we propose an iterative agent learning algorithm to solve the coordination problem and obtain optimal action sets $\{a_x^t, a_{x'}^t\}$ in (9).

Recall that in the vehicle aggregation process as shown in Algorithm 1, the vehicles join an aggregation group sequentially. Let x_y denote the yth vehicle to join gourp V_q, where $y=\{1,2,\ldots,Y\}$ and $Y=|V_q|$. When vehicle x_y joins the group, it is specifically associated to a previous group $V_{q,y-1}$ that has been already formed with vehicles $\{x_1,x_2,\ldots,x_{y-1}\}$. Following the CG approach, there exists an edge $(x_y, V_{q,y-1})$ to present the learning relation between vehicular agent x_y and agents $\{x_1, x_2, \ldots, x_{y-1}\}$. Moreover, the sequential aggregation characteristic indicates that the decision strategy of vehicular agent x_Y, which is the last one to join group V_q, only depends on the states and actions of the previously aggregated agents. Thus, we can rewrite (9) as

$$Q_q^t(S^t, A^t) = Q^t_{x_Y, V_{q,Y-1}}(S^t_{x_Y, V_{q,Y-1}}, a^t_{x_Y}, A^t_{V_{q,Y-1}}) + \sum_{(x,x')\in E/(x_Y, V_{q,Y-1})} Q^t_{x,x'}(S^t_{x,x'}, a^t_x, a^t_{x'}). \tag{10}$$

In the multiagent learning process, vehicular agent x_Y collects the possible states and actions related to edge $(x_Y, V_{q,Y-1})$, and obtains its best-response action $a^{t,\mathrm{opt}}_{x_Y}$ following a conditional reward function as:

$$a^{t,\mathrm{opt}}_{x_Y} = \arg\min_{a^t_{x_Y}} Q^t_{x_Y, V_{q,Y-1}}(a^t_{x_Y} | S^t_{x_Y, V_{q,Y-1}}, A'^t_{V_{q,Y-1}}) \tag{11}$$

where $A'^t_{V_{q,Y-1}}$ is the historical action strategy of $V_{q,Y-1}$ stored in agent x_Y.

Based on the obtained $a^{t,\mathrm{opt}}_{x_Y}$, we calculate $Q^{t,\mathrm{opt}}_{x_Y, V_{q,Y-1}}$ and take it into (10). Then, we have

$$Q_q^t(S^t, A^t) = \sum_{(x,x')\in E/(x_Y, V_{q,Y-1})} Q^t_{x,x'}(S^t_{x,x'}, a^t_x, a^t_{x'}) + Q^{t,\mathrm{opt}}_{x_Y, V_{q,Y-1}}. \tag{12}$$

In (12), among all the agents whose action has not been determined, the last one to join group V_q is x_{Y-1}. Similar to the way to derive best-response action $a^{t,\mathrm{opt}}_{x_Y}$, we present (12) as

$$Q_q^t(S^t, A^t) = Q^t_{x_{Y-1}, V_{q,Y-2}}(S^t_{x_{Y-1}, V_{q,Y-2}}, a^t_{x_{Y-1}}, A^t_{V_{q,Y-2}}) + \sum_{(x,x')\in E/\{(x_Y, V_{q,Y-1}),(x_{Y-1}, V_{q,Y-2})\}} Q^t_{x,x'}(S^t_{x,x'}, a^t_x, a^t_{x'}) + Q^{t,\mathrm{opt}}_{x_Y, V_{q,Y-1}} \tag{13}$$

and then get $Q^{t,\mathrm{opt}}_{x_{Y-1}, V_{q,Y-2}}$.

In this iterative agent learning approach, we can derive the optimal action set of all the vehicular agents belong to V_q at time slot t. The action set will be spreaded among these agents, and prepared as strategic information for the learning in the next time slot. The main steps of the proposed iterative learning algorithm are shown in Algorithm 2. This algorithm leverages the sequential characteristic of the gravity-based aggregation process in DTN to construct a linear iterative learning mechanism in the physical vehicular network, which greatly reduces the learning complexity of multiple agents with complex relations. In addition, the interaction between the agents is limited to their chosen actions, which significantly reduces the amount of data transmission and adapts to the spectrum resource constrained vehicular networks.

Based on Algorithm 2 that describes the learning approach for an aggregation group at one time slot, we propose a CG-driven multiagent deep deterministic policy gradient (MADDPG) learning scheme for all the agents of the whole vehicular edge system. It is noteworthy that unlike the traditional DDPG approach, whose input of critic network only has state-action information, in the proposed learning scheme, the input further consists of the action data spreaded among the agents belongs to the same group. The main steps of the MADDPG learning scheme are described in Algorithm 3, and the joint process of Algorithm 2 and Algorithm 3 is shown in Fig. 3.

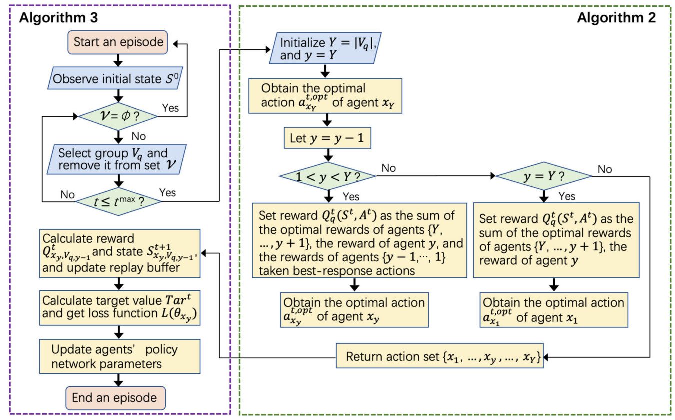


Fig. 3. Flow chart of Algorithm 2 and Algorithm 3.

Algorithm 3: CG-Driven MADDPG Learning.

Require:
Aggregation group V_q, where $q \in \{1, 2, \ldots, q_{\max}\}$;
Initialize strategy parameter set Θ and experience replay buffer.

for Each episode **do**
  Observe initial state S^0.
  for Each vehicular aggregation group V_q **do**
    for $t = 1$ to $t^{\max}$ **do**
      for Each agent $x_y \in V_q$ **do**
        Get action $a^{t,\mathrm{opt}}_{x_y}$ according to Algorithm 2;
        Calculate reward $Q^t_{x_y, V_{q,y-1}}$ and state set $S^{t+1}_{x_y, V_{q,y-1}}$;
        Store $(S^t_{x_y, V_{q,y-1}}, a^{t,\mathrm{opt}}_{x_y}, Q^t_{x_y, V_{q,y-1}}, S^{t+1}_{x_y, V_{q,y-1}})$ in the experience replay buffer;
        Calculate target value $Tar^t = Q^t_{x_y, V_{q,y-1}} + \xi Q_{x_y, V_{q,y-1}}(a^{t+1}_{x_y}, S^{t+1}|\theta_{x_y})$;
        Get loss function $L(\theta_{x_y}) = \mathbb{E}[(Tar^t - Q_{x_y, V_{q,y-1}}(S^t, a^t_{x_1}, \ldots, a^t_{x_Y}|\theta_{x_y}))^2]$;
        Obtain policy gradient as $\nabla_{\theta_{x_y}} J = E[\nabla_{a^t_{x_y}} Q_{x_y, V_{q,y-1}}(S^t, a^t_{x_1}, \ldots, a^t_{x_Y}|\theta'_{x_y}) \nabla_{\theta_{x_y}} \pi(S^t|\theta_{x_y})]$, where $Q_{x_y, V_{q,y-1}} (S^t, a^t_{x_1}, \ldots, a^t_{x_Y}|\theta'_{x_y}) = \mathbb{E}[Q^t_{x_y, V_{q,y-1}} + \xi Q_{x_y, V_{q,y-1}}(S^{t+1}, \pi(S^{t+1}|\theta_{x_y}))]$;
        Update the policy network parameters of each agent following $\theta'_{x_y} = \lambda\theta_{x_y} + (1-\lambda)\theta'_{x_y}$, where λ is an update factor.
      end for
    end for
  end for
end for

C. Mutual Assisted Interaction Between Digital Twin Evolution and Multiagent Learning

The DTN and the multiagent learning system operate corporately in scheduling the vehicular edge service. On the one hand, the DTN determines the distributed learning environments of the multiple agents through aggregating vehicular groups under the guidance of parameters $\Phi = \{\phi_1, \phi_2, \phi_3\}$. This aggregation improves the supply and demand matching of edge resources and reduces multiagent learning complexity. On the other hand, the multiagent learning results, i.e., the task offloading target selection and edge resource allocation, affect vehicular edge service performance and the performance indicators can be used in turn to evaluate the pros and cons of the aggregation mechanism, so as to adjust aggregation parameter set Φ. These two parts iteratively interact and update to make themselves adapt to the changes in application scenarios.

In each mapping period of the DTN, parameter set Φ updates as follows. Let $\Phi_\varpi = \{\phi_{1,\varpi}, \phi_{2,\varpi}, \phi_{3,\varpi}\}$ denote the value of the parameters in period ϖ. As $\phi_{1,\varpi}$, $\phi_{2,\varpi}$, and $\phi_{3,\varpi}$, respectively, reflect the influence of various factors on the aggregation operation, they each have a different update strategy. For ϕ_1, its update approach in period ϖ can be expressed as

$$\phi_{1,\varpi} = \phi_{1,\varpi-1} \cdot \frac{1}{|\mathcal{V}|} \sum_{V_q \in \mathcal{V}} \left(\sum_{i=1}^{|V_q|} \sum_{j=1}^{J_i} C_{i,j} \Big/ \sum_{i=1}^{|V_q|} \sum_{j=1}^{J_i} \sum_{k=1}^{|V_q|} (\beta_{i,j,k} f_k + (1-\beta_{i,j,k})) f_i T_{i,j}^{\max} \right). \tag{14}$$

The correction factor of $\phi_{1,\varpi-1}$ in (14) is the average ratio of demand to supply of computing resources of the aggregated groups in the last period. If the ratio is greater than 1, then ϕ_1 needs to be increased to strengthen the role of computing resource matching in the aggregation process.

Parameter ϕ_2 indicates the influence of computing resource prices on the multiple vehicle aggregation, and its adjustment approach is given as

$$\phi_{2,\varpi} = \phi_{2,\varpi-1} \cdot \frac{1}{|\mathcal{V}|} \sum_{V_q \in \mathcal{V}} \left(\sum_{i=1}^{|V_q|} \sum_{j=1}^{J_i} \sum_{k=1}^{|V_q|} \beta_{i,j,k} f_k T_{i,j}^{\max} / C_{i,j} \right) \tag{15}$$

where the second term on the right side is the average ratio of the computing resources obtained by task offloading to the required resources of the tasks. The ratio higher than 1 indicates that the allocated computing resources exceed the demand. In this case, to reduce the offloading cost of the vehicular edge system, we need to enhance the price sensitivity of customer vehicles by increasing ϕ_2.

Task data transmission is an indispensable part of edge computation offloading. Parameter ϕ_3 determines the impact of communication capabilities between vehicles on the group aggregation. The update of ϕ_3 depends on the satisfaction degree of the task data transmission requirements in the previous period,

Fig. 4. Area division for test verification.

which is presented as

$$\phi_{3,\varpi} = \phi_{3,\varpi-1} \cdot \frac{1}{|\mathcal{V}|} \sum_{V_q \in \mathcal{V}} \sum_{l=1}^{L} \sum_{i=1}^{|V_q|} \sum_{k=1}^{|V_q|} \left(\sum_{j=1}^{J_i} \frac{\beta_{i,j,k} \delta_{i,j,k,l} D_{i,j}}{R_{i,k,l} \tilde{T}_{i,k,\varpi-1}} \right) \tag{16}$$

where $\tilde{T}_{i,k,\varpi-1}$ is the historical transmission time of channel l taken by vehicles i and k in period $\varpi - 1$.

V. Numerical Results

In this section, we evaluate the performance of our proposed vehicular edge task offloading schemes based on real traffic datasets, which are extracted from the historical mobility traces of taxi cabs in San Francisco Bay area. There are approximately 500 cabs, and the average time interval for their GPS coordinates update is less than 10 s [30]. To investigate the influence of traffic environment characteristics on the offloading scheme performance, we further divide Bay area into 6 square areas, as shown in Fig. 4.

We consider a scenario where the computation capacities of vehicles are randomly taken from (10, 20) units. The computation resource requirements, data size and maximum tolerable latency of the tasks are randomly chosen from (30, 50) units, (5, 10) MB, and (0.5, 2) s, respectively [10]. In addition, there are 5 orthogonal channels for offloading transmission and the bandwidth of each channel is 0.3 MHz.

Fig. 5 shows the output of the DTN aided vehicle aggregation scheme implemented in different areas. Diverse vehicular topologies and edge service supply-demand relationships in these areas lead to different aggregation characteristics, which affect the offloading costs illustrated in the following figures. It is noteworthy that although areas 5 and 6 have similar number of vehicles, their aggregation output is significantly different, which indicates that the DTN aided aggregation scheme has effective edge environment feature identification capabilities.

Fig. 6 presents the convergence of the CG-driven MADDPG learning scheme. We randomly select two agents from area 3 and area 5, respectively. All the agents' learning converges around 3300 iterations. Furthermore, this figure demonstrates that the difference of edge network characteristics and aggregation groupings between the two areas has little effect on the convergence performance.

Fig. 7 shows the offloading costs with different scheduling schemes. Compared to the other two schemes, our proposed CG-driven MADDPG obtains the lowest cost. In the independent learning scheme, each vehicle works as an agent to aware

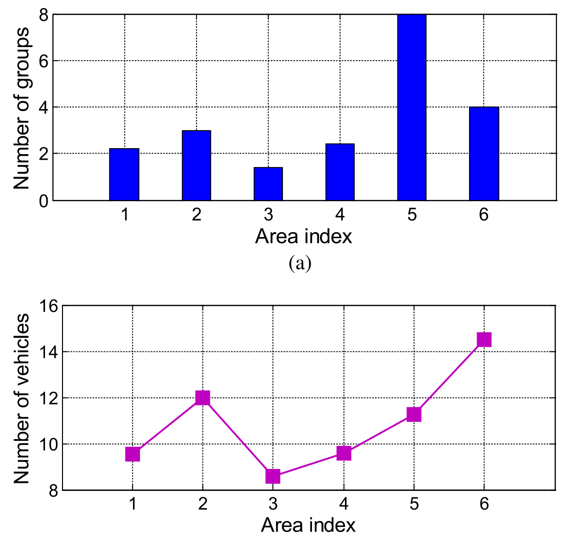


Fig. 5. Digital twin aided vehicle aggregations in different areas. (a) Average number of aggregation groups in an area. (b) Average number of vehicles in an aggregation group.

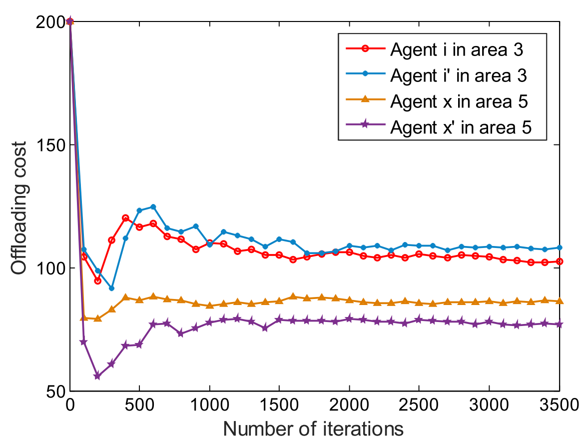


Fig. 6. Convergence of CG-driven MADDPG learning.

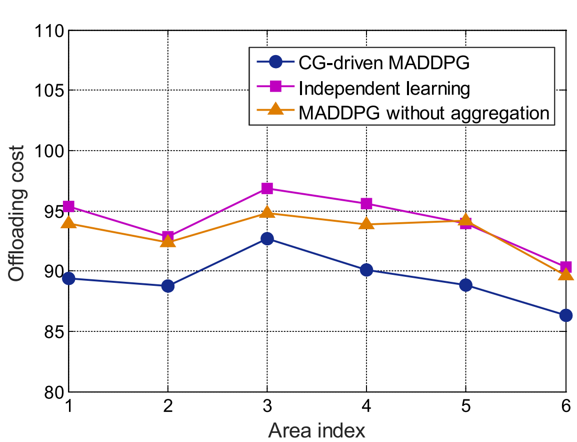


Fig. 7. Comparison of offloading costs with different schemes.

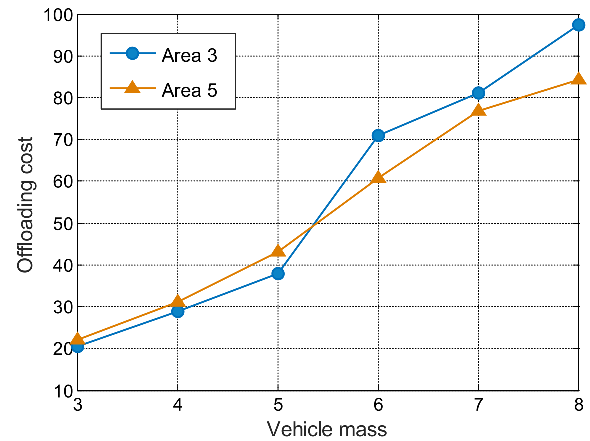


Fig. 8. Offloading costs of vehicles with different mass values in selected areas.

edge service environment and makes self-interested offloading actions without interaction among agents. This independent decision-making approach may cause resource surplus or shortage between some vehicular service pairs, thereby undermining offloading efficiency of the whole system. In the MADDPG without aggregation, all agents in the same area adopt joint decision making. Due to the complexity of vehicle topology and potential service relations, this scheme is difficult to reach the optimal offloading strategy under constrained learning iterations. In contrast to the previews two schemes, the CG-driven MADDPG scheme aggregates vehicular agents based on the DTN-aided edge service matching, which helps the scheme to realize low-complexity multiagent collaborative learning under the premise of efficient resource utilization and get the lowest cost.

Fig. 8 compares the offloading costs of vehicles with different mass values in selected areas. According (4), parameter mass indicates the eagerness of a vehicle's computing resource demand. Higher mass value drives the vehicle willing to pay a greater cost to offload its computation tasks. It is noteworthy that as the mass value increases, the comparison of the cost growth of areas 3 and 5 has changed. When the mass value reaches 6, the cost of area 3 has a significant increase, and the cost keeps higher than that of area 5 as the mass continues to increase. The reason is that when vehicle mass is small, a few vehicular edge servers around it can meet its task offloading demand. However, when the mass reaches to 6, the vehicle needs more computing resources. The sparse distribution of vehicles in area 3 makes the vehicular requester do not have much choice of the servers. To meet the task demands, requesters have to choose servers with higher resource prices, which incurs higher costs.

VI. Conclusion

In this article, we incorporated digital twin technology and artificial intelligence to design an efficient vehicular edge computing network, where the digital twin helps to reveal the potential edge service matching among massive vehicle pairs and reduce the complexity of service management, while the learning approach enables the vehicles to obtain their task offloading strategies. Moreover, we developed a gravity model based vehicle aggregation scheme in the digital twin side, and proposed a corresponding multiagent learning algorithm to optimize edge resource scheduling in the physical vehicular networks. We evaluated the performance of the proposed schemes based on real traffic datasets. The numerical results demonstrated that our schemes have lower costs compared to the benchmark schemes.

References

[1] C. L. Stergiou, K. E. Psannis, and B. B. Gupta, "IoT-based big data secure management in the fog over a 6 G wireless network," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5164–5171, Apr. 2021.

[2] H. Fatemidokht, M. K. Rafsanjani, B. B. Gupta, and C. Hsu, "Efficient and secure routing protocol based on artificial intelligence algorithms with UAV-assisted for vehicular ad hoc networks in intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, to be published, doi: 10.1109/TITS.2020.3041746.

[3] Y. Dai, K. Zhang, S. Maharjan, and Y. Zhang, "Deep reinforcement learning for stochastic computation offloading in digital twin networks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 7, pp. 4968–4977, Jul. 2021.

[4] W. Sun, H. Zhang, R. Wang, and Y. Zhang, "Reducing offloading latency for digital twin edge networks in 6 G," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12240–12251, Oct. 2020.

[5] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, "Low-latency federated learning and blockchain for edge association in digital twin empowered 6G networks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 7, pp. 5098–5107, Jul. 2021.

[6] A. Saad, S. Faddel, T. Youssef, and O. A. Mohammed, "On the implementation of IoT-based digital twin for networked microgrids resiliency against cyber attacks," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5138–515150, Nov. 2020.

[7] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2405–2415, Apr. 2019.

[8] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, "Communication-efficient federated learning and permissioned blockchain for digital twin edge networks," *IEEE Intern. Things J.*, vol. 8, no. 4, pp. 2276–2288, Feb. 2021.

[9] S. Lee *et al.*, "Resource allocation for vehicular fog computing using reinforcement learning combined with heuristic information," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10450–10464, Oct. 2020.

[10] K. Zhang, J. Cao, H. Liu, S. Maharjan, and Y. Zhang, "Deep reinforcement learning for social-aware edge computing and caching in urban informatics," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5467–5477, Aug. 2020.

[11] X. Wang, Z. Ning, S. Guo, and L. Wang, "Imitation learning enabled task scheduling for online vehicular edge computing," *IEEE Trans. Mobile Comput.*, to be published, doi: 10.1109/TMC.2020.3012509.

[12] Y. Dai, D. Xu, K. Zhang, S. Maharjan, and Y. Zhang, "Deep reinforcement learning and permissioned blockchain for content caching in vehicular edge computing and networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4312–4324, Apr. 2020.

[13] D. Chen, Y. Liu, B. Kim, J. Xie, C. S. Hong, and Z. Han, "Edge computing resources reservation in vehicular networks: A meta-learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5634–5646, May 2020.

[14] B. B. Gupta and M. Quamara, "An overview of Internet of Things (IoT): Architectural aspects, challenges, and protocols," *Concurrency Comput. Pract. Experience*, vol. 32, no. 21, pp. 1–24, Nov. 2020.

[15] M. Ahrarinouri, M. Rastegar, and A. R. Seifi, "Multiagent reinforcement learning for energy management in residential buildings," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 659–666, Jan. 2021.

[16] J. Cui, Y. Liu, and A. Nallanathan, "Multi-agent reinforcement learning-based resource allocation for UAV networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 729–743, Feb. 2020.

[17] A. Kaur and K. Kumar, "Energy-efficient resource allocation in cognitive radio networks under cooperative multi-agent model-free reinforcement learning schemes," *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 3, pp. 1337–1348, Sep. 2020.

[18] Q. Yuan, J. Li, H. Zhou, T. Lin, G. Luo, and X. Shen, "A joint service migration and mobility optimization approach for vehicular edge computing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9041–9052, Aug. 2020.

[19] T. Chu, J. Wang, L. Codec, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1086–1095, Mar. 2020.

[20] E. G. Kaigom and J. Romann, "Value-driven robotic digital twins in cyber-physical applications," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3609–3619, May 2021.

[21] C. Altun, B. Tavli, and H. Yanikomeroglu, "Liberalization of digital twins of IoT-enabled home appliances via blockchains and absolute ownership rights," *IEEE Commun. Mag.*, vol. 57, no. 12, pp. 65–71, Dec. 2019.

[22] C. Gehrmann and M. Gunnarsson, "A digital twin based industrial automation and control system security architecture," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 669–680, Jan. 2020.

[23] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, "Communication-efficient federated learning for digital twin edge networks in industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5709–5718, Aug. 2021.

[24] R. Dong, C. She, W. Hardjawana, Y. Li, and B. Vucetic, "Deep learning for hybrid 5G services in mobile edge computing systems: Learn from a digital twin," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4692–4707, Oct. 2019.

[25] C. Esposito, M. Ficco, B. B. Gupta, "Blockchain-based authentication and authorization for smart city applications," *Inf. Process. Manage.*, vol. 58, no. 2, pp. 1–16, 2021.

[26] F. Mirsadeghi, M. K. Rafsanjani, and B. B. Gupta, "A trust infrastructure based authentication method for clustered vehicular ad hoc networks," *Peer-to-Peer Netw. Appl.*, Oct. 2020. [Online]. Available: https://doi.org/10.1007/s12083-020-01010-4

[27] J. E. Anderson, "The gravity model," *Annu. Rev. Econ.*, vol. 3, no. 1, pp. 133–160, 2011.

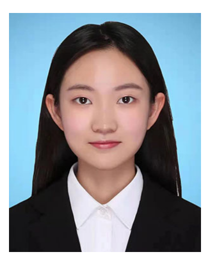
[28] C. Yu *et al.*, "Distributed multiagent coordinated learning for autonomous driving in highways based on dynamic coordination graphs," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 2, pp. 735–748, Feb. 2020.

[29] C. Guestrin, M. Lagoudakis and R. Parr, "Coordinated reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2002, pp. 227–234.

[30] [Online]. Available: http://crawdad.org/epfl/mobility/20090224.

Ke Zhang (Member, IEEE) received the Ph.D. degree in communication and information systems from the University of Electronic Science and Technology of China, Chengdu, China, in 2017.

He is currently an Associate Professor with the School of Information and Communication Engineering, University of Electronic Science and Technology of China. His research interests include scheduling of mobile edge computing, design and optimization of next-generation wireless networks, smart grid, and the Internet of Things.



Jiayu Cao received the B.S. degree in 2018 from the University of Electronic Science and Technology of China, Chengdu, China, where she is currently working toward the Ph.D. degree with the School of Information and Communication Engineering, both in communication and information systems.

Her research interests include the vehicular networks, mobile edge computing, and multiagent system.

Yan Zhang (Fellow, IEEE) received the B.S. degree from the Nanjing University of Post and Telecommunications, Nanjing, China, the M.S. degree from Beihang University, Beihang, China, and the Ph.D. degree from the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore.

He is currently a Full Professor with the Department of Informatics, University of Oslo, Oslo, Norway. His research interests include next-generation wireless networks leading to 6G and green and secure cyber-physical systems (e.g., smart grid and transport).

Dr. Zhang is a Fellow of IET, an Elected Member of Academia Europaea, the Royal Norwegian Society of Sciences and Letters (DKNVS), the Norwegian Academy of Technological Sciences. Since 2018, he has been a recipient of the global "Highly Cited Researcher" Award (Web of Science top one).