VoI-Driven Joint Optimization of Control and Communication in Vehicular Digital Twin Network

Lei Lei, Kan Zheng, Jie Mei, and Xuemin Shen

Abstract

The vision of sixth-generation (6G) wireless networks paves the way for the seamless integration of digital twins into vehicular networks, giving rise to a Vehicular Digital Twin Network (VDTN). The large amount of computing resources as well as the massive amount of spatial-temporal data in Digital Twin (DT) domain can be utilized to enhance the communication and control performance of Internet of Vehicle (IoV) systems. In this article, we first propose the architecture of VDTN, emphasizing key modules that center on functions related to the joint optimization of control and communication. We then delve into the intricacies of the multitimescale decision process inherent in joint optimization in VDTN, specifically investigating the dynamic interplay between control and communication. To facilitate the joint optimization, we define two Value of Information (VoI) concepts rooted in control performance. Subsequently, utilizing VoI as a bridge between control and communication, we introduce a novel joint optimization framework, which involves iterative processing of two Deep Reinforcement Learning (DRL) modules corresponding to control and communication to derive the optimal policy. Finally, we conduct simulations of the proposed framework applied to a platoon scenario to demonstrate its effectiveness in ensuring the stability and transmission efficiency of VDTN.

Introduction

The evolution of Internet of Vehicle (IoV) is closely connected to the ongoing advancements in the fifth generation (5G) Cellular Vehicle-to-Everything (C-V2X) communication technology. Although 5G C-V2X has ushered in impressive improvements in the communication capabilities for IoV, there are still critical challenges in supporting a wide range of safety and non-safety IoV applications [1]. The inherent promise of sixth generation (6G) lies in its ability to deliver significantly increased throughput and reduced communication latency. These advancements are expected to provide the essential infrastructure needed to overcome the current constraints in IoV. The enhanced capabilities of 6G also open doors for effective integration with edge and cloud computing resources, alleviating the burden on individual vehicles [2].

The rapid advancement of Digital Twin (DT) technologies, especially with robust support by 6G, is poised to revolutionize the IoV [3]. This revolution is exemplified by the emergence of the Vehicular Digital Twin Network (VDTN), characterized by the seamless integration of DT into vehicular networks. In essence, a DT is a real-time evolving digital replica used to emulate a physical object or process, encapsulating its entire history in the virtual space [4]. Within the VDTN, comprehensive virtual representations of objects or entities, such as vehicles in physical space, can be created and maintained to establish the DT. As a result, the VDTN has the potential to leverage DT to enhance communication, control, and overall system performance. It stands as a promising candidate for the next generation of IoV, offering new dimensions of efficiency, reliability, and intelligence through the synergistic integration of DT and 6G technologies.

Among the vast applications of IoV, the efficient and reliable control of autonomous vehicles (AVs) is a critical task, which encompasses both the control and communication subsystems. Traditionally, these subsystems are often designed independently, utilizing classical theories in their respective domains, such as the classical control theory, information theory, and optimization theory. As the AVs can make more informed control decisions when equipped with the V2X communication capability for exchanging sensory information with surrounding vehicles and infrastructure, it is increasingly recognized that the performance of autonomous driving can be significantly enhanced through the joint design of control and communication. Existing efforts in pursuing this holistic approach often adopt either a *control perspective* or *communication perspective*, focusing on designing either "*an efficient AV control (AVC) approach with awareness of the communication nonideality such as delay and packet loss*" [5] or "*an efficient communication approach taking cognizance of the requirements to support the AVC applications*" [6]. In our point of view, the co-design of control and communication

Lei Lei is with the College of Engineering and Physical Sciences, University of Guelph, Guelph, ON N1G 2W1, Canada; Kan Zheng (corresponding author) and Jie Mei are with the College of Electrical Engineering and Computer Sciences, Ningbo University, Ningbo 315211, China; Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Digital Object Identifier: 10.1109/MNET.2024.3522588
Date of Current Version: 16 September 2025
Date of Publication: 25 December 2024

0890-8044 © 2024 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies.
Authorized licensed use limited to: University of Waterloo. Downloaded on October 02,2025 at 13:54:54 UTC from IEEE Xplore. Restrictions apply.

Among the vast applications of IoV, the efficient and reliable control of autonomous vehicles (AVs) is a critical task, which encompasses both the control and communication subsystems.

Category	Symbol	Definition
Control	k	The index of a control interval
	S_k^{CL}	The control state at control interval k
	O_k^{CL}	The impaired observation of control state at control interval k
	$\tilde{S}_k^{CL}$	The augmented control state at control interval k
	u_k^{CL}	The control action at control interval k
	R_{k+1}^{CL}	The control reward at control interval k
	γ	The discount factor
	$\tilde{\pi}_{CL}$	The acting control policy based on impaired observation
	π_{CL}^*	The optimal control policy based on perfect observation
	$\tilde{J}_{CL}$	The control performance of the acting control policy $\tilde{\pi}_{CL}$
	J_{CL}^*	The control performance of the optimal control policy π_{CL}^*
	$A_{\pi_{CL}^*}(S_k^{CL}, u_k^{CL})$	The advantage function of the optimal control policy π_{CL}^*
Communication	t	The index of a communication interval within a control interval
	T	The total number of communication intervals within a control interval
	$S_{(k,t)}^{CM}$	The communication state at communication interval (k, t)
	$u_{(k,t)}^{CM}$	The communication action at communication interval (k, t)
	$R_{(k,t+1)}^{CM}$	The communication reward at communication interval (k, t)
	π_{CM}	The communication policy
	J_{CM}	The communication performance of the communication policy π_{CM}
	P_{CM}	The expected long-term communication performance of traditional services
	$g_{CM}(S_{k,t}^{CM}, u_{k,t}^{CM})$	The instantaneous communication performance of traditional services at communication interval (k, t)
VoI	Ξ	EVoI
	ξ_k	IVoI at control interval k

TABLE 1. Summary of important symbols used.

is loosely-coupled in the above research, since the emphasis is still on the design of individual subsystems.

Although the loosely-coupled co-design approach optimizes one subsystem with awareness of the other subsystem, the level of cognizance is limited since a simplified model or conventional design of the other subsystem is normally considered. The inaccurate or conservative modeling could result in a significant deviation from the optimal solution. For example, the overwhelming majority of existing Radio Resource Allocation (RRA) algorithms were developed under the assumption of a conventional AV controller with little or no consideration on the communication nonideality. Consequently, they tend to allocate more resources than necessary for transmitting AVC messages if an advanced AV controller that is robust to communication nonideality is used in practice. Therefore, the interplay between the control and communication subsystems underscores the need for a *tightly-coupled* co-design approach.

The core of the *tightly-coupled* co-design approach lies in the requirement that the communication subsystem optimizes "*how the information is disseminated*" with sufficient and accurate knowledge of "*how the information is leveraged*" in the control subsystem, and vice versa. Moreover, the two subsystems should be *optimized simultaneously* so that each subsystem is aware of the optimal design of the other subsystem to calibrate its own design. As we navigate through the intricacies of this ambitious approach, three fundamental research questions emerge, i.e.,

- What theoretical tool(s) should be harnessed for the co-design?
- How to capture the interplay between control and communication in the conceived optimization problems?
- Where should the co-design and joint optimization be performed, given the computation complexity and the knowledge sharing requirements between the two subsystems?

This article presents a solution for the tightly-coupled co-design of control and communication, addressing the aforementioned questions.

In response to the first question, we propose to design an integrated Deep Reinforcement Learning (DRL)-based decision-making system, where DRL is employed to optimize AVC and RRA in each subsystem [7]. Recent years have witnessed a growing interest in applying data-driven Machine Learning (ML) to autonomous driving as well as vehicular communications. As a combination of deep learning (DL) and reinforcement learning (RL), the DRL approaches have been demonstrated to outperform the traditional model-driven control and optimization approaches, since less restrictive assumptions on the stochastic properties of the system dynamics are required [8].

To implement the tightly-coupled co-design philosophy, we first formulate a multitimescale decision process, wherein a single policy makes control and communication decisions across different time scales to optimize both performance aspects. Given the complexity of solving such a large-scale optimization problem, we decouple the multitimescale decision process into a control MDP and a communication MDP, utilizing DRL algorithms to solve each model. However, since the two MDP models are inherently interrelated, the optimal policy for the control MDP cannot be determined without knowledge of the optimal policy for the communication MDP, and vice versa. To resolve this dilemma, we use an iterative training approach, where both MDP models are sequentially solved in each iteration.

To address the second question, we leverage the Value of Information (VoI) for integrating the control and communication subsystems. The concept of VoI was first introduced by Howard in 1966 [9], which generally accounts for how information impacts the outcome of subsequent decisions. However, Howard did not provide a rigorous definition or mathematical formula for calculating VoI. Building on his work, various definitions of VoI have been proposed to suit different applications and objectives, such

Authorized licensed use limited to: University of Waterloo. Downloaded on October 02,2025 at 13:54:54 UTC from IEEE Xplore. Restrictions apply.

as KL divergence and information quality [10]. Unlike previous studies, this article introduces a systematic approach to defining and evaluating VoI, rooted in the MDP, optimal control, and RL theories. The proposed approach can be applied to any decision making systems that are designed based on the aforementioned theories. In addition, we use VoI as a crucial link between control and communication design. The VoI is estimated by the control module at the information recipient and provided to the communication module at the information source. The VoI is then used to derive the reward function of communication MDP and guide the optimization of communication decisions. In contrast to other information metrics, such as Age of Information (AoI), our proposed VoI has the distinct advantage of accurately reflecting the impacts of information-induced control performance loss [11].

For the third question, we advocate joint optimization performed in the virtual space of the VDTN, fully leveraging its potential of powerful computational capability and massive amount of available data reflecting the physical space. Given the communication nonideality between the physical and virtual spaces, we adopt the Centralized Training Decentralized Execution (CTDE) approach. In this approach, deep neural networks (DNNs) are deployed on physical entities, such as vehicles, to make real-time local decisions in a decentralized manner. Meanwhile, these DNNs are centrally trained in the virtual space, where sensory information from various physical entities for both control and communication subsystems is shared to optimize the models. The trained models are continuously fine-tuned in the virtual space to adapt to non-stationary environment, with model parameters periodically transmitted to the physical space to update the corresponding DNNs.

In the following sections, we will delve deeper into each component of the joint optimization framework, further elucidating their intricacies, interrelationships, and respective contributions to the overarching goal of joint optimization of control and communications. Specifically, this paper first presents the system architecture of the VDTN in the section "System Architecture of the VDTN," laying the groundwork for the subsequent study. Building upon this architecture, we then formally define the VoI in the integrated DRL-based decision making system in the section "DRL-Based VoI for Decision Making," with the target of shedding light on the co-design of control and communication. Next, we propose the framework for VoI-driven joint optimization of control and communication in the VDTN in the section "Joint Optimization of Control and Communication." This proposed framework is characterized by a detailed description, outlining the steps and methodologies involved in achieving the joint optimization objectives. A case study on vehicular platoon applications in the VDTN is performed in the section "A Case Study on Vehicular Platoon Applications in the VDTN" to validate the effectiveness of the proposed framework. Finally, our conclusions and outlooks are given in the section "Conclusion and Outlooks."

System Architecture of the VDTN

As shown in Fig. 1, in the VDTN, various entities including vehicles, road infrastructure such as traffic lights, and communication infrastructure such as base stations collect real-time sensory data, which are then transmitted to the edge or cloud computing servers through 6G network to construct the virtual replicas. In the virtual space, a massive amount of data undergoes pre-processing, completion, prediction, and other operations to generate the virtual representation, which is essential for the subsequent decision-making modules.

In an VDTN ideally powered by 6G, a huge amount of sensors deployed in the physical space reliably and promptly transmit data to the virtual space. Meanwhile, the decisions made in

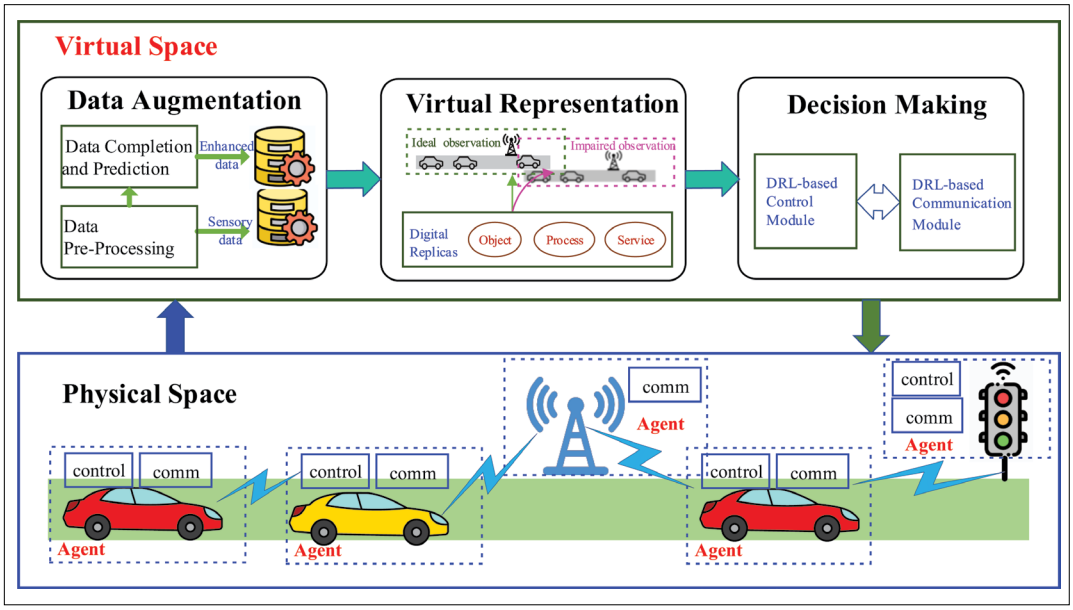

FIGURE 1. Illustration of the system architecture of the Vehicular Digital Twin Network (VDTN).

Authorized licensed use limited to: University of Waterloo. Downloaded on October 02,2025 at 13:54:54 UTC from IEEE Xplore. Restrictions apply.

In an VDTN ideally powered by 6G, a huge amount of sensors deployed in the physical space reliably and promptly transmit data to the virtual space.

the virtual space are delivered in real-time and executed by actuators in the physical space. Consequently, tasks traditionally challenging in conventional IoV due to complexity limitations, such as joint optimization of control and communication, can be effectively performed in the VDTN.

However, in the early stages of 6G development, limited radio resources and non-ideal transmission performance continue to be constraints for the VDTN. There are two common types of wireless connections in the VDTN, i.e., *those between various entities in the physical space* and *those between the physical and virtual spaces*. In this paper, we assume that both types of connections are non-ideal, meaning that data transmission may experience impairments such as delay and packet loss. This assumption reflects real-world scenarios in the near future.

For the sake of illustration, this section provides only a brief description of several DT modules related to AVC and RRA, which are the focus of this paper. However, it does not mean that the VDTN is only limited to these modules and functions.

Data Augmentation

There are mainly two functionalities in this module, i.e.,

- **Data Pre-Processing:** This involves cleaning and organizing the extensive sensory data, thoroughly removing inconsistencies, errors, and redundancies to ensure data integrity. The refined data is systematically stored across various databases, encompassing both structured and unstructured formats. This systematic storage approach enhances the efficiency of data retrieval and utilization in the VDTN.
- **Data Completion and Prediction:** Owing to factors such as measurement errors or communication nonideality, not only the sensory data exchanged between the entities in the physical space but also those received at the virtual space could be impaired. In the virtual space, sophisticated deep learning models such as Generative Nets and Transformers can be harnessed [12], which exploit historical and current data to complete/predict the missing/delayed data, resulting in the *enhanced data*. Subsequently, this new data can be used to achieve a more realistic representation of the physical space.

Virtual Representation

All objects, processes, and services in the physical space of IoV can be represented in the virtual space using the enhanced data. The data used to create these digital replicas includes a diverse set of information, encompassing both road traffic and vehicular communication characteristics. The former comprises vehicle behaviors, road conditions, and so on. The latter includes the radio channel environment, communication protocols, service characteristics, network performance measures, and so forth. From the perspective of control and communication co-design, the virtual representation in VDTN includes the digital replicas for the control subsystem and the communication subsystem, respectively.

From the perspective of RL, the virtual representation provides digital replicas of the environment and agents in the physical space. Despite the non-ideal communication between the physical and virtual spaces, an exact virtual replica of the physical environment could be constructed from the enhanced data as described in the section "Data Augmentation," which means that the states in the physical and virtual spaces are exactly the same for each time step. Moreover, the enhanced data could also be leveraged to learn a model of the physical environment including transition probability and reward function. Based on the learned model, the virtual space could also generate an environment that evolves in the same stochastic manner as that of the physical space. Training agents in the virtual space enjoys much more advantages than in the physical space. For example, the environment can be replicated into multiple instances based on the learned environment model just as the multiple sample realizations of the same stochastic process, while an agent is split into multiple copies that are executed in parallel to obtain different experiences.

In this paper, we consider that two types of environment are conceived for the control subsystem based on these digital replicas, namely, the control environment with *perfect observation* and the control environment with *impaired observation*. The perfect observation refers to the perceived information about the environment that the agent can obtain under the assumption of ideal transmission in V2X links. This type of control environment can be generated based on the enhanced data as described above. However, in the physical space, due to the non-ideal nature of V2X communication between the physical entities, delay or packet loss may occur during transmission, leading to inaccuracies in the information received by the physical entities. Based on such information, the agent can experience impaired observations of its environment. In the virtual space, we can simulate a control environment with impaired observation by deriving it from the control environment with perfect observation, while incorporating statistical insights from the communication environment regarding nonideal communication between physical entities. Both types of environment will be used in the joint optimization of control and communication as elaborated in the section "Joint Optimization of Control and Communication."

Decision Making

Most of IoV services necessitate the simultaneous utilization of communication and control-related information for timely and intelligent decision-making. These decisions span a range of tasks, including intelligent transportation aspects such as AVC, traffic flow control, and traffic route planning, as well as communication-related aspects such as deciding when or whether to share information and RRA. While our focus in this paper revolves around

Authorized licensed use limited to: University of Waterloo. Downloaded on October 02,2025 at 13:54:54 UTC from IEEE Xplore. Restrictions apply.

AVC and RRA decisions, the proposed methodology can be extended to address various other control and communication decisions. Following RL conventions, the decision-makers for AVC and RRA are referred to as the control agent and communication agent, respectively.

Ideally, decisions are made in the virtual space and transmitted to the physical space for execution in real-time. However, considering the strict real-time requirements of safety-related IoV services, the V2X network has difficulty in promptly and accurately acquiring sensory information from and disseminating decisions to the diverse entities in the physical space.

Therefore, adopting the approach of *CTDE* is more practically feasible in the near future. In the physical space, agents are deployed on vehicles and base stations, enabling real-time local control and communication decisions in a decentralized manner. Specifically, each DRL-based agent is usually equipped with a DNN for generating the local actions, where the DNN is either a local actor representing the per-agent policy function or a local Q-network representing the per-agent action-value functions. In the virtual space, these DNNs can be centrally trained by leveraging the large amount of computing power as well as the massive amount of spatial-temporal data. Once essentially converged, the DNN models are swiftly deployed to agents at base stations and vehicles through V2X links. Given the dynamic nature of IoV environments, including road conditions and communication network dynamics, the trained models in the virtual space require periodic and continuous tuning to adapt to these changes. Updated models should be disseminated to the physical space to ensure the effectiveness of decision-making in response to the dynamic environment of IoV. Since models are updated to adapt to changes in the stochastic properties of the environment, the frequency of model updates is much lower than the frequency of decision-making. Consequently, the 6G network is well-suited for disseminating the updated models to the physical space in a timely manner.

It is noteworthy that CTDE is a widely adopted paradigm for addressing multi-agent DRL problems, where a DNN, namely the centralized critic representing the joint action-value function, is typically trained to assist in training multiple decentralized per-agent DNNs [13]. In our context, CTDE not only tackles the multi-agent problems for AVC and RRA. More importantly, it is essential for the co-design of AVC and RRA, where the VoI estimated by the control subsystem of information recipient can be shared with the information disseminator to guide the optimization of communication subsystem, as elaborated in the subsequent sections of the paper.

DRL-Based VoI for Decision Making

Multitimescale Decision Process

In the VDTN, the control and communication subsystems make decisions at different time scales. For example, the sampling and decision-making intervals for AVC normally range between 0.01 s to 0.1 s. Meanwhile, the RRA decisions for vehicular communication are made per subframe, e.g., with a duration of 1 ms in the LTE-V2X systems [14]. Therefore, the time steps in the MDP model for AVC and RRA have different lengths, and are therefore referred to as the control intervals and communication intervals, respectively, whenever it is necessary to distinguish between them. We consider that a control interval consists of T communication intervals. In the rest of the paper, the control intervals are indexed by $k \in \{0, 1, 2, \ldots\}$, while the communication intervals in control interval k are indexed by (k, t), where $t \in \{0, 1, 2, \ldots, T-1\}$.

As such, the co-design problem of control and communication can be conceived as a multitimescale decision process, where a single policy prescribes the control and communication actions at different time scales, with the goal of optimizing the control performance as well as the communication performance. However, such a full-state approach yields a large-scale optimization problem, which is computationally infeasible even for the VDTN.

To overcome the above challenge, we decouple the single full-state MDP model into two separate MDP models, i.e., a control MDP and a communication MDP, where each MDP model is solved by the DRL algorithms for making the control and communication decisions, respectively. Thus, the integrated decision making system comprises a DRL-based control module and a DRL-based communication module. However, as the environment of the communication MDP is affected by the control policy and vice versa, the two MDP models are interrelated as will be explained in detail below, where the optimal policy of one model depends on that of the other model.

Interplay Between Control and Communication

1) Impact of the RRA Policy on Control: In RL, the environment state includes all the information that the environment uses to determine the next state and reward. However, the environment state is usually not fully and accurately visible to the agent. Therefore, the agent has to choose actions based on its internal representation of the state, which is normally called the observations.

As discussed in the section "System Architecture of the VDTN," the control agents in VDTN are deployed in the vehicles for making real-time decisions in a decentralized fashion, after they are centrally trained in the virtual space. Thus, the observations of each control agent are formed by the aggregation of the local information sampled by the sensors equipped on its own vehicle, as well as the V2X information sampled by and transmitted from the remote sensors on the other vehicles and the infrastructure. Since the wireless transmission in IoV is usually not ideal, the sensory data via V2X links could be delayed or lost. In such a case, the control agent has to make decisions based on the impaired observations O_k^{CL} instead of the perfect observations of the state S_k^{CL}. Therefore, the vehicle control problem becomes a partial observable MDP (POMDP).

The POMDP can normally be converted into an MDP using state augmentation, where the augmented control state $\tilde{S}_k^{\mathrm{CL}}$ includes not only the impaired observation O_k^{CL}, but also the

Authorized licensed use limited to: University of Waterloo. Downloaded on October 02,2025 at 13:54:54 UTC from IEEE Xplore. Restrictions apply.

action-observation history as well as some auxiliary information such as the delay of the current observation. The optimization objective of the DRL-based control module can be written as

$$\max_{\tilde{\pi}_{\mathrm{CL}}} \tilde{\mathbf{J}}_{\mathrm{CL}} = \max_{\tilde{\pi}_{\mathrm{CL}}} \mathbb{E}_{\tilde{\pi}_{\mathrm{CL}},\pi_{\mathrm{CM}}}\left[\sum_{k=0}^{\infty} \gamma^k R_{k+1}^{\mathrm{CL}}\right], \quad (1)$$

where R_{k+1}^{CL} is the immediate reward that the control agent receives at control interval k, and $\gamma \in [0, 1]$ is the discount factor. The subscript of $\mathbb{E}$ denotes the expected value of a random variable when the control agent follows policy $\tilde{\pi}_{\mathrm{CL}}$ and the communication agent follows policy π_{CM}. Equation (1) reveals that the control performance $\tilde{\mathbf{J}}_{\mathrm{CL}}$ is not only affected by the control policy $\tilde{\pi}_{\mathrm{CL}}$ but also the RRA policy π_{CM}. The impact of $\tilde{\pi}_{\mathrm{CL}}$ on $\tilde{\mathbf{J}}_{\mathrm{CL}}$ is obvious, since given the augmented state $\tilde{S}_k^{\mathrm{CL}}$ at control interval k, the control decision on the action u_k^{CL} is made according to the control policy $\tilde{\pi}_{\mathrm{CL}}(u_k^{\mathrm{CL}} \mid \tilde{S}_k^{\mathrm{CL}})$. Meanwhile, given the observation $\tilde{S}_k^{\mathrm{CL}}$ and action u_k^{CL}, the next state $\tilde{S}_{k+1}^{\mathrm{CL}}$ is affected by the RRA policy, i.e., $\Pr(\tilde{S}_{k+1}^{\mathrm{CL}} \mid \tilde{S}_k^{\mathrm{CL}}, u_k^{\mathrm{CL}}, \pi_{\mathrm{CM}})$. This is because the RRA policy affects the stochastic properties of communication nonideality, such as delay distribution, which in turn impacts the impaired observation received by the control agent at the next control interval.

Since the RRA policy is an integral part of the control environment, the DT must be aware of the RRA policy or the resulting data from communication nonidealities to emulate a realistic environment when training the control agent.

2) Impact of the AVC Policy on Communication: The optimization objective of the DRL-based communication module in the VDTN is typically twofold - optimizing network performance metrics for traditional services and minimizing control performance loss due to communication nonidealities for control-related services. For this purpose, the communication agent should obviously allocate the scarce radio resources to carry the more meritorious information with higher priority, so that the information is subject to less impairment during transmission. To quantify which information is more "meritorious" for the control system, we will formally define the VoI in the section "VoI Definition." *The VoI serves as a bridge between the two MDP models for control and communication, which is estimated by the control agent who receives and acts upon the V2X information; and optimized by the communication agent who transmits the V2X information*. Note that the merit of the information depends on how it is leveraged by its recipient. Therefore, the VoI depends on AVC policy, as the more robust the policies are to the communication nonideality, the less important it is to transmit the information without impairment.

VoI Definition

The term VoI has been studied in different research communities for various decision-making tasks. It is generally used to measure the benefit of acquiring additional information at the time of making decisions. For the purpose of joint control and communication optimization in the VDTN, we use VoI to quantify the performance loss that would result from making AVC decisions based on the impaired V2X information instead of the perfect information, where the impairment is caused by the communication nonideality such as packet delay and loss.

In contrast to traditional communication performance metrics like throughput and delay, along with widely-used control-oriented metrics such as AoI, VoI emerges as a more apt gauge for optimizing the communication system. This suitability stems from its ability to well capture the influence of communication nonidealities on control performance. Next, we define two types of VoI, which are used to formulate the MDP model for the communication systems.

Expected Cumulative VoI (EVoI)

The EVoI $\bar{\Xi}$ is defined as the difference in the performance of the acting control policy $\tilde{\pi}_{\mathrm{CL}}$ based on the impaired observation and the optimal control policy π_{CL}^* based on the perfect observation, respectively, i.e., $\bar{\Xi} = \tilde{\mathbf{J}}_{\mathrm{CL}} - \mathbf{J}_{\mathrm{CL}}^*$.

The EVoI indicates the performance loss as a consequence of the control agent consistently receiving the impaired observations at every control interval. In the terminology of RL, EVoI corresponds to the regret of the acting control policy $\tilde{\pi}_{\mathrm{CL}}$. It is worth noting that the EVoI is a non-positive value, with a larger EVoI indicating less performance loss. When the EVoI reaches its maximum value of zero, no performance loss is experienced.

Immediate VoI (IVoI)

Given the perfect observation S_k^{CL} and impaired observation O_k^{CL} at control interval k, the IVoI ξ_k is defined as the difference in the performance starting from control interval k, where the control decision at control interval k is made by the acting control policy $\tilde{\pi}_{\mathrm{CL}}$ based on the impaired observation and the optimal control policy π_{CL}^* based on the perfect observation, respectively, while the rest of the sequential control decisions are made by π_{CL}^* based on the perfect observations $S_{k+1}^{\mathrm{CL}}, S_{k+2}^{\mathrm{CL}}, \ldots$, i.e., $\xi_k = A_{\pi_{\mathrm{CL}}^*}(S_k^{\mathrm{CL}}, u_k^{\mathrm{CL}})\big|_{u_k^{\mathrm{CL}} \sim \tilde{\pi}_{\mathrm{CL}}(\cdot \mid \tilde{S}_k^{\mathrm{CL}})}$, where $u_k^{\mathrm{CL}} \sim \tilde{\pi}_{\mathrm{CL}}(\cdot \mid \tilde{S}_k^{\mathrm{CL}})$ indicates that the action u_k^{CL} is sampled according to $\tilde{\pi}_{\mathrm{CL}}$.

The IVoI indicates the performance loss as a consequence of the control agent receiving the impaired observations only at a specific control interval k, while consistently obtaining perfect observations throughout the remaining time. Moreover, IVoI is calculated for a specific state and observation pair at control interval k. In other words, IVoI reveals that if the perfect observation at control interval k is $S_k^{\mathrm{CL}} = \mathrm{s}$, what is the performance loss if the agent only has access to the impaired observation O_k^{CL} instead? In the terminology of RL, IVoI corresponds to the advantage function of the optimal policy π_{CL}^*. Similar to EVoI, IVoI is also a non-positive value, where maximizing IVoI leads to minimizing the performance loss.

It is proved in [15] that the EVoI is equivalent to the expected cumulative IVoI, similar to the relationship between the expected cumulative reward and immediate reward in RL. As a result, we name the two types of VoI as the expected

Authorized licensed use limited to: University of Waterloo. Downloaded on October 02,2025 at 13:54:54 UTC from IEEE Xplore. Restrictions apply.

cumulative VoI and immediate VoI, respectively. However, it is essential to clarify that the term immediate VoI doesn't imply its sole focus on quantifying the disparity in the immediate rewards at control interval k when the control decisions are made based on the impaired versus perfect observations, respectively. Since the decision at control interval k not only affects the immediate reward at the current control interval but also the future rewards through its impact on the next state, the IVoI measures the performance loss in terms of the expected cumulative reward just as the EVoI does.

Joint Optimization of Control and Communication

In general, the objective is to maximize communication performance as well as the control performance through the optimization of the AVC and RRA policy in the VDTN. As discussed in the section "Joint Optimization of Control and Communication," we can conceive two MDP models for the control subsystem and communication subsystem, respectively, utilizing DRL algorithms for their solutions.

Due to the inherent interplay between control and communication, the optimal policy of one MDP model cannot be derived without knowing the optimal policy of the other model. To address this dilemma, an iterative training approach is adopted to find the optimal strategy of the control MDP and communication MDP simultaneously [15]. In each iteration, both agents are sequentially trained, and their optimal policies under the current environments are obtained. Notably, the control agent is consistently trained before the communication agent, facilitating the update of the VoI and providing it to the communication agent for further optimization of communication policy. In the initial iteration, a random communication policy can be employed for training the control agent.

The iterative joint optimization framework is illustrated in Fig. 2. In this framework, the control policy derived by the DRL-based control module is provided as input to the DRL-based communication module. Similarly, the communication policy derived by the DRL-based communication module is provided as input to the DRL-based control module. In the following, we delve into further details on how the communication and control agents are trained to learn their respective optimal policies in each iteration.

Learning of Communication-Aware Control Policy

From the above discussion, the DRL-based control module needs to be communication-aware, which means the control policy is learned under the assumption that a communication policy is given. Note that the communication policy is an integrated part of the control environment, since it affects communication nonideality statistics, which in turn influences the impaired observations at the control agent. As discussed in the section "Learning of Control-Aware RRA Policy," an MDP model for the AVC problem with the optimization objective as shown in (1) can be formulated. Subsequently, state-of-the-art DRL algorithms can be used to acquire the desired control policy. Additionally, to further engage in joint optimization with the communication subsystem, it is necessary to obtain the VoI associated with the control performance.

To obtain either EVoI or IVoI, the DRL-based control module has to train the optimal control policy π^*_{CL} in addition to the acting control policy $\tilde{\pi}_{\mathrm{CL}}$. As discussed in the section "Virtual Representation," two types of environment are created in the virtual representation. As shown in Fig. 2, the

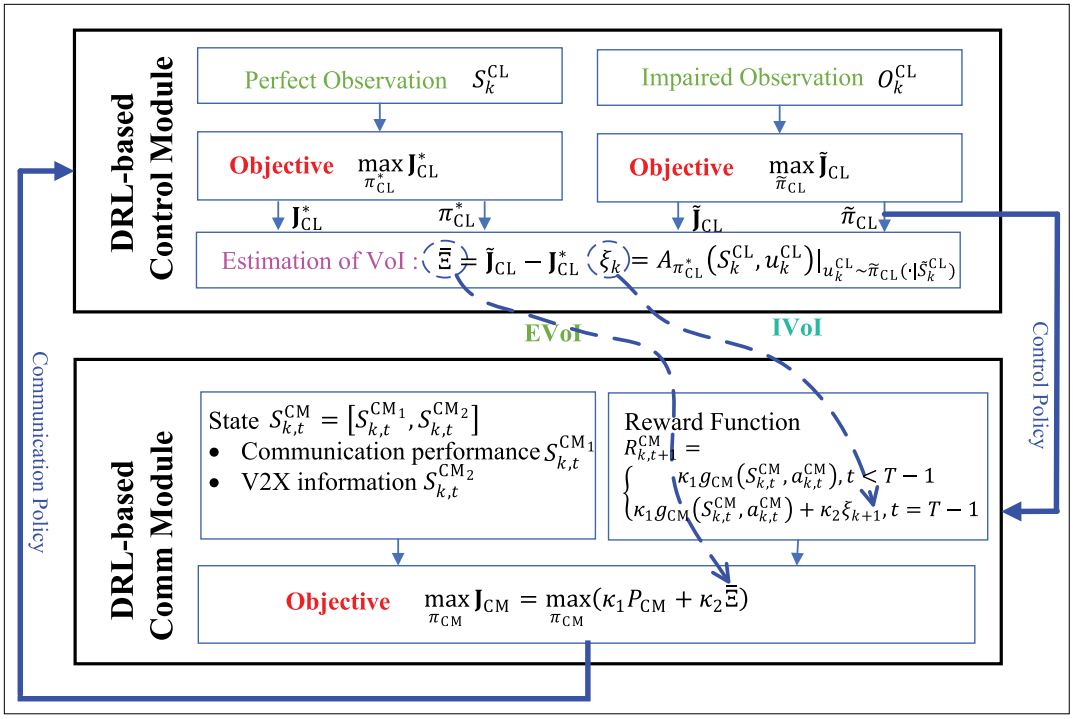

FIGURE 2. Framework of Joint Optimization of Control and Communication in VDTN.

Authorized licensed use limited to: University of Waterloo. Downloaded on October 02,2025 at 13:54:54 UTC from IEEE Xplore. Restrictions apply.

acting control policy $\tilde{\pi}_{\mathrm{CL}}$ is learned in the environment with impaired observations, while the optimal control policy π^{*}_{CL} is learned in the environment with perfect observations. After training the two control policies, the EVoI can be derived by simply calculating the performance difference between two policies.

The IVoI corresponds to the advantage function of the optimal control policy with perfect observations, where different methods can be used to estimate the advantage function of a policy in RL, such as by the Temporal Difference (TD) error $\delta_{\tilde{\pi}_{\mathrm{CL}},k} = R^{\mathrm{CL}}_{k+1} + v_{\pi^{*}_{\mathrm{CL}}}(S^{\mathrm{CL}}_{k+1}) - v_{\pi^{*}_{\mathrm{CL}}}(S^{\mathrm{CL}}_{k})$. In this case, the IVoI can be derived for each control interval based on the reward R^{CL}_{k+1} as a result of implementing action $u_k^{\mathrm{CL}} \sim \tilde{\pi}_{\mathrm{CL}}(\cdot|\tilde{S}^{\mathrm{CL}}_{k})$, as well as the value function $v_{\pi^{*}_{\mathrm{CL}}}$ of the optimal control policy for the current state S^{CL}_{k} and the next state S^{CL}_{k+1}.

Learning of Control-Aware RRA Policy

Similarly, the DRL-based communication module has to be control-aware, implying that the communication policy is learned under the assumption that a control policy is given. Specifically, the VoI is estimated within the control module of the information recipient and is subsequently used to guide the optimization within the communication module of the information disseminator.

As discussed in the section "Learning of Control-Aware RRA Policy," the objective of the DRL-based communication module is to optimize the traditional network performance metrics as well as to minimize the control performance loss due to communication nonideality, which is captured by the EVoI as defined in the section "VoI Definition." As a result, the optimization objective is given as

$$\max_{\pi_{\mathrm{CM}}} \mathbf{J}_{\mathrm{CM}} = \max_{\pi_{\mathrm{CM}}} \left(\kappa_1 \mathrm{P}_{\mathrm{CM}} + \kappa_2 \Xi\right), \tag{2}$$

where P_{CM} is the expected long-term communication performance of traditional services, such as the expected cumulative throughput over the considered time horizon, and Ξ is the EVoI. The weighting coefficients κ_1 and κ_2 reflect the relative importance of traditional network services and control services in the optimization objective. They can be conveniently set according to control performance requirements, availability of radio resources, and other factors, satisfying the constraints $0 \le \kappa_1, \kappa_2 \le 1$ and $\kappa_1 + \kappa_2 = 1$.

Correspondingly, the reward function of the MDP model for the communication subsystem can be designed as

$$R^{\mathrm{CM}}_{k,t+1} = \begin{cases} \kappa_1 g_{\mathrm{CM}}\left(S^{\mathrm{CM}}_{k,t}, u^{\mathrm{CM}}_{k,t}\right) & \text{if } t < T-1 \\ \kappa_1 g_{\mathrm{CM}}\left(S^{\mathrm{CM}}_{k,t}, u^{\mathrm{CM}}_{k,t}\right) + \kappa_2 \xi_{k+1} & \text{if } t = T-1 \end{cases} \tag{3}$$

where $g_{\mathrm{CM}}(S^{\mathrm{CM}}_{k,t}, u^{\mathrm{CM}}_{k,t})$ represents the instantaneous communication performance of traditional services at communication interval (k, t), such as the instantaneous data rate. The $g_{\mathrm{CM}}(S^{\mathrm{CM}}_{k,t}, u^{\mathrm{CM}}_{k,t})$ component is present in the reward of each communication interval, while the IVoI component, i.e., ξ_{k+1}, is only present in the last communication interval of each control interval when $t = T - 1$. Note that $\xi_{k+1} = A_{\pi^{*}_{\mathrm{CL}}}(S^{\mathrm{CL}}_{k+1}, u^{\mathrm{CL}}_{k+1})\big|_{u^{\mathrm{CL}}_{k+1} \sim \tilde{\pi}_{\mathrm{CL}}(\cdot|\tilde{S}^{\mathrm{CL}}_{k+1})}$ represents the amount of performance loss that the information recipient will suffer from, as a consequence of its control agent having to make a decision at the next control interval $k + 1$ based on the impaired observation $\tilde{S}^{\mathrm{CL}}_{k+1}$ according to its acting policy $\tilde{\pi}_{\mathrm{CL}}$ instead of based on the perfect observation S^{CL}_{k+1} according to its optimal policy π^{*}_{CL}. The impaired observation received at control interval $k + 1$ is affected by the degree of communication nonideality it has experienced, which is in part a result of the sequence of communication actions $u^{\mathrm{CM}}_{k,0}, u^{\mathrm{CM}}_{k,1}, \ldots, u^{\mathrm{CM}}_{k,T-1}$ that are taken during control interval k. Since EVoI equals the expected cumulative IVoI, the expected discounted sum of the immediate reward in (3) equals the optimization objective.

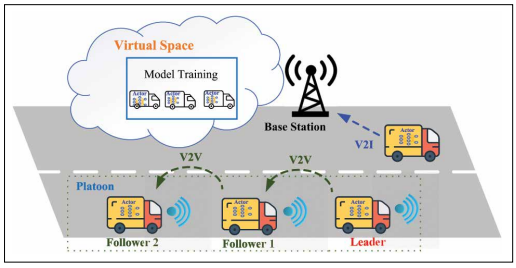

FIGURE 3. Simulation scenario of the vehicular platoon supported by VDTN.

As illustrated in Fig. 2, the EVoI and IVoI are estimated by the DRL-based control module and provided to the DRL-based communication module to formulate the communication MDP.

The communication state $S^{\mathrm{CM}}_{k,t}$ consists of two components, i.e., $S^{\mathrm{CM}}_{k,t} = [S^{\mathrm{CM}_1}_{k,t}, S^{\mathrm{CM}_2}_{k,t}]$. $S^{\mathrm{CM}_1}_{k,t}$ is related to g_{CM}, including Channel State Information (CSI), Queue State Information (QSI), etc., which are typically defined in the MDP models for RRA. Meanwhile, $S^{\mathrm{CM}_2}_{k,t}$ is related to IVoI, which theoretically should include any information that is useful in predicting the IVoI at the information recipient, such as the V2X information that is to be transmitted. Therefore, the formulation for $S^{\mathrm{CM}_2}_{k,t}$ differs depending on the specific AVC tasks.

Finally, the DRL algorithms can be employed to determine the optimal communication policy for the DRL-based communication module in each iteration.

A Case Study on Vehicular Platoon Applications in the VDTN

In this section, we assess the effectiveness of the proposed joint optimization of control and communication (JOCC) by conducting simulations in a vehicular platoon supported by the VDTN. As shown in Fig. 3, the AVs in the platoon communicate using V2V links, while V2I links establish connections between the AVs and the Base Station (BS) for high-throughput services. The Predecessors Following (PF) information typology (IFT) is adopted in the platoon, where the Collaborative

Authorized licensed use limited to: University of Waterloo. Downloaded on October 02,2025 at 13:54:54 UTC from IEEE Xplore. Restrictions apply.

Parameter Description	Value
Number of vehicles in the platoon	4
Number of V2I links	4
Control interval of vehicle in the platoon	50 ms
Size of CAM packet	1800 Bytes
Number of communication intervals within each control interval	50
Carrier frequency	2 GHz
Bandwidth of sub-channel	180 kHz
Maximum transmit power of vehicle	23 dBm

TABLE 2. Main parameters of a platoon system supported by VDTN.

Adaptive Message (CAM) of the preceding vehicle is transmitted to its following vehicle. The AVC decisions are made at a control interval of 50 ms, while the RRA decisions are executed at a communication interval of 1 ms. For comparison purpose, the performance of the Delay-aware RRA (i.e., DRRA) method is also given. DRRA is an improved DRL-based RRA algorithm only aiming to minimize the latency and maximize the success probability of delivering the CAM generated at the beginning of the control interval. Both DRL algorithms are trained/tested where the velocity profile of leading vehicle is obtained from the open-source driving data[1]. The main parameters of the platoon system are given in Table 2.

As illustrated in Fig. 4, the control performance of JOCC is minimally impacted in terms of gap-keeping error and velocity error when compared to those of the DRRA. On the other hand, JOCC exhibits a higher average throughput on each V2I link compared to DRRA, achieving 0.86 Mbps compared to 0.91 Mbps, respectively. This indicates the effective performance of the proposed JOCC in the case of platooning in VDTN.

Conclusion and Outlooks

This paper has proposed a VDTN architecture with designed key functional modules for joint control and communication optimization. Aiming to fully utilize the computing capabilities and massive data of the virtual space in digital twin as well as satisfying the real-time requirements, the CTDE framework is adopted in VDTN. Then, we have established a multitimescale decision process for joint optimization of control and communication, incorporating VoI to reflect control performance degradation due to communication nonideality. Furthermore, a joint optimization framework based on DRL has been proposed to address AVC and RRA simultaneously. Finally, simulation results in a platoon scenario have validated the effectiveness of the proposed framework.

The anticipated application scope of the proposed VDTN architecture, coupled with the VoI-driven joint optimization approach, is extensive, encompassing a spectrum of vehicle control and communication tasks. The former includes tasks like car following, lane change, overtaking, ramp merging, and negotiating intersections. Concurrently, the latter involves considerations of what to communicate, when to communicate, and how to communicate. In the context of 'how-tocommunicate', the RRA function discussed in this article plays a pivotal role. The exploration of VoI estimation and VoI-driven optimization in VDTN presents compelling avenues for future research. Some potential research directions include:

- VoI Definition: The VoI definitions in this article can be expanded. For example, VoI can be defined to quantify control performance loss based on whether information is present or absent, rather than whether it is impaired or not.
- VoI Estimation in the Control Subsystem: Accurate evaluation of VoI is crucial for guiding the optimization of communication decisions. However, existing RL methods for estimating the advantage function encounter challenges when applied to IVoI evaluation. For instance, the TD error, which serves as a sample of the advantage function, can be noisy. Thus, improving the accuracy of VoI estimation remains an important open issue.
- VoI-Driven Optimization in the Communication Subsystem: Diverse communication tasks are performed at various time scales, and could, therefore, leverage either the EVoI and/or IVoI as appropriate. Effectively utilizing VoI for optimizing different communication functions is an interesting area for future exploration.

Additionally, several significant challenges must be addressed in the DRL, such as

- Multi-Agent RL (MARL) Issues: Solving both the control and communication MDPs relies on MARL algorithms. However, state-of-the-art MARL algorithms like QMIX and Multiagent Proximal Policy Optimization (MAPPO) are not fully effective in addressing

Aiming to fully utilize the computing capabilities and massive data of the virtual space in digital twin as well as satisfying the real-time requirements, the CTDE framework is adopted in VDTN.

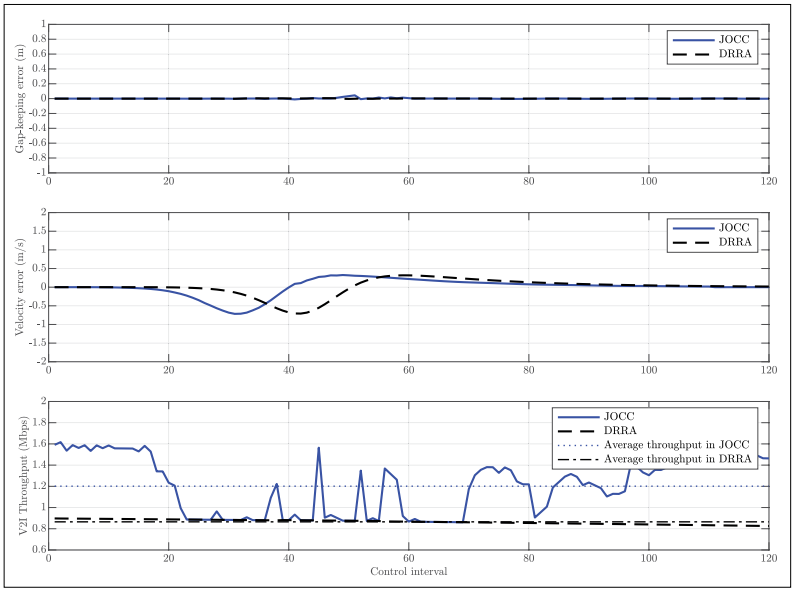

FIGURE 4. Performance comparison of AVs in the platoon system supported by VDTN.

[1] Next Generation Simulation (NGSIM) Vehicle Trajectories and Supporting Data provided by ITS DataHub through data.transportation.gov.

Authorized licensed use limited to: University of Waterloo. Downloaded on October 02,2025 at 13:54:54 UTC from IEEE Xplore. Restrictions apply.

well-known challenges such as non-stationary environment, partial observability, and credit assignment problems. Moreover, the popular CTDE paradigm, used by many MARL algorithms, faces scalability issue attributed to centralized training. Finally, addressing the open ad-hoc team problem is a promising direction for the highly dynamic and non-stationary vehicular environment, where an agent can collaborate with a variable number of other agents without prior coordination.

- Long-Horizon Sparse-Reward Problem : The time horizon of the communication MDP spans multiple control intervals, with each control interval further divided into multiple communication intervals. Additionally, the VoI reward is solely accessible during the last communication interval of each control interval in (3). This long-horizon, sparse-reward environment presents significant challenges for designing RL algorithms, as highly specific action sequences must be executed prior to observing any non-trivial feedback.

Finally, VDTN faces several deployment issues in real-world scenarios. The first challenge to deploy VDTN is determining the optimal placement of virtual entities, data, and models within the DT domain - specifically, whether to position them on cloud servers or edge servers. For instance, ones that require extensive data processing and computational power are generally better hosted in the cloud, while those with strict real-time requirements are ideally deployed on the edge. Furthermore, the transmission of data and models through various V2X links is another challenge, necessitating an effective strategy for allocating wireless resources to these links.

Acknowledgment

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Discovery Grant.

References

[1] H. Bagheri et al., "5G NR-V2X: Toward connected and cooperative autonomous driving," *IEEE Commun. Stand. Mag.*, vol. 5, no. 1, pp. 48–54, Mar. 2021.
[2] H. Li, K. Ota, and M. Dong, "Learning IoV in 6G: Intelligent edge computing for Internet of Vehicles in 6G wireless communications," *IEEE Wireless Commun.*, vol. 30, no. 6, pp. 96–101, Mar. 2023.
[3] X. Shen et al., "Holistic network virtualization and pervasive network intelligence for 6G," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 1–30, 1st Quart., 2022.
[4] X. Lin et al., "6G digital twin networks: From theory to practice," *IEEE Commun. Mag.*, vol. 61, no. 11, pp. 72–78, Nov. 2023.
[5] F. Ma et al., "Distributed control of cooperative vehicular platoon with nonideal communication condition," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8207–8220, Aug. 2020.
[6] M. Parvini, "AoI-aware resource allocation for platoon-based C-V2X networks via multi-agent multi-task reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 9880–9896, Aug. 2023.
[7] L. Lei et al., "Deep reinforcement learning for autonomous Internet of Things: Model, applications and challenges," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1722–1760, 3rd Quart., 2020.
[8] C. Jiang et al., "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.
[9] R. A. Howard, "Information value theory," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-2, no. 1, pp. 22–26, Feb. 1966.
[10] F. Alawad and F. A. Kraemer, "Value of information in wireless sensor network applications and the IoT: A review," *IEEE Sensors J.*, vol. 22, no. 10, pp. 9228–9245, May 2022.
[11] T. Soleymani et al., "Value of information in feedback control: Quantification," *IEEE Trans. Autom. Control*, vol. 67, no. 7, pp. 3730–3737, Jul. 2022.
[12] B. Erman and C. D. Martino, "Generative network performance prediction with network digital twin," *IEEE Netw.*, vol. 37, no. 2, pp. 286–292, Mar. 2023.
[13] R. Lowe, et al., "Multiagent actor-critic for mixed cooperative-competitive environments," in *Proc. NIPS*, 2017, pp. 6382–6393.
[14] K. Sehla et al., "Resource allocation modes in C-V2X: From LTE-V2X to 5G-V2X," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 8291–8314, Jun. 2022.
[15] L. Lei et al., "Multitimescale control and communications with deep reinforcement learning—Part II: control-aware radio resource allocation," *IEEE Internet Things J.*, vol. 11, no. 9, pp. 15475–15489, May 2024.

Biographies

Lei Lei (Senior Member, IEEE) (leil@uoguelph.ca) received the B.S. and Ph.D. degrees in telecommunications engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2001 and 2006, respectively. She is currently an Associate Professor with the College of Engineering and Physical Sciences, University of Guelph, Canada. Her research interests include machine learning/deep reinforcement learning, the Internet of Things/the Internet of Vehicles, mobile edge computing, and smart grid.

Kan Zheng (Fellow, IEEE) (zhengkan@nbu.edu.cn) received the Ph.D. degree from the Beijing University of Posts and Telecommunication (BUPT), China, in 2005. He is currently a Full Professor with Ningbo University, China. His research interests include vehicular networks and the Internet of Things (IoT).

Jie Mei (Member, IEEE) received the Ph.D. degree from the Beijing University of Posts and Telecommunication (BUPT) in 2019. He is currently an Associate Professor with Ningbo University, China. His research interests include vehicular networks and the IoT.

Xuemin (Sherman) Shen (Fellow, IEEE) (sshen@uwaterloo.ca) is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo. His research focuses on network resource management, wireless network security, social networks, 5G and beyond, and vehicular ad hoc networks. He is a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, an Engineering Institute of Canada Fellow, and a Chinese Academy of Engineering Foreign Fellow. He received the President's Excellence in Research from the University of Waterloo in 2022, the Canadian Award for Telecommunications Research from the Canadian Society of Information Theory (CSIT) in 2021, the R. A. Fessenden Award in 2019 from IEEE, Canada, and the Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) in 2019.

Authorized licensed use limited to: University of Waterloo. Downloaded on October 02,2025 at 13:54:54 UTC from IEEE Xplore. Restrictions apply.