Deep Reinforcement Learning-Based RAN Slicing for UL/DL Decoupled Cellular V2X

Kai Yu, *Graduate Student Member, IEEE*, Haibo Zhou, *Senior Member, IEEE*, Zhixuan Tang, *Member, IEEE*, Xuemin Shen, *Fellow, IEEE*, and Fen Hou, *Member, IEEE*

***Abstract*— The emerging uplink (UL) and downlink (DL) decoupled radio access networks (RAN) has attracted a lot of attention due to the significant gains in network throughput, load balancing and energy consumption, etc. However, due to the diverse vehicular service requirements in different vehicle-to-everything (V2X) applications, how to provide customized cellular V2X services with diversified requirements in the UL/DL decoupled 5G and beyond cellular V2X networks is challenging. To this end, we investigate the feasibility of UL/DL decoupled RAN framework for cellular V2X communications, including the vehicle-to-infrastructure (V2I) communications and relay-assisted cellular vehicle-to-vehicle (RAC-V2V) communications. We propose a two-tier UL/DL decoupled RAN slicing approach. On the first tier, the deep reinforcement learning (DRL) soft actor-critic (SAC) algorithm is leveraged to allocate bandwidth to different base stations. On the second tier, we model the QoS metric of RAC-V2V communications as an absolute-value optimization problem and solve it by the alternative slicing ratio search (ASRS) algorithm with global convergence. The extensive numerical simulations demonstrate that the UL/DL decoupled access can significantly promote load balancing and reduce C-V2X transmit power. Meanwhile, the simulation results show that the proposed solution can significantly improve the network throughput while ensuring the different QoS requirements of cellular V2X.**

***Index Terms*— RAN slicing, decoupled access, soft actor-critic, reinforcement learning.**

Manuscript received April 23, 2021; revised August 22, 2021; accepted October 7, 2021. Date of publication November 2, 2021; date of current version May 10, 2022. This work was supported in part by the National Natural Science Foundation Original Exploration Project of China under Grant 62050071; in part by the National Natural Science Foundation of China under Grant 61871211; in part by the Innovation and Entrepreneurship of Jiangsu Province High-Level Talent Program; in part by the Summit of the Six Top Talents Program of Jiangsu Province; in part by the Natural Sciences and Engineering Research Council of Canada (NSERC); in part by the Guangdong-Hong Kong-Macao Joint Laboratory Program under Project 2020B1212030009; and in part by the Science and Technology Development Fund, Macau, under Project SKL-IOTSC(UM)-2021-2023. The associate editor coordinating the review of this article and approving it for publication was C. Huang. *(Corresponding author: Haibo Zhou.)*

Kai Yu, Haibo Zhou, and Zhixuan Tang are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: kaiyu@smail.nju.edu.cn; haibozhou@nju.edu.cn; zhixuantang@smail.nju.edu.cn).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Fen Hou is with the State Key Laboratory of IoT for Smart City and the Guangdong-Hong Kong-Macao Joint Laboratory for Smart Cities, Department of Electrical and Computer Engineering, University of Macau, Macau, China (e-mail: fenhou@um.edu.mo).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2021.3122941.

Digital Object Identifier 10.1109/TWC.2021.3122941

I. Introduction

As 5G cellular networks commercially roll out, the cellular vehicle-to-everything (C-V2X) communications will play a critical role in facilitating the application of automotive and intelligent transportation. The recently expanded 3GPP Release 16 extension is standardized to support more vehicle services, such as platooning, extended sensors, automated driving, and modern intelligent driving experience, etc. To keep up with the ever-increasing mobile traffic demand, 5G cellular networks have evolved from a single-tier homogeneous network to multi-tier heterogeneous networks (HetNets), which consists of several macro base stations (MBSs) with a wide coverage area and a large amounts of small base stations (SBSs). With the trends of network densification and exploding of mobile data usage, one of the most challenging issues of the HetNet topology is the user association, especially for the C-V2X user access with high mobility and stringent communication demand. The uplink and downlink decoupling paradigm has attracted a lot of attentions and a pioneering uplink (UL) and downlink (DL) decoupling paradigm for radio access networks (RAN) is emerging by allowing the access association in uplink and downlink to be different and flexible [1], [2]. It has been proved that the UL/DL decoupled access can obviously reduce the transmit power of UL users [3]. Besides, with flexible cell association, the disruptive uplink and downlink decoupling paradigm has shown great advantages in terms of network throughput enhancing, load balancing, energy efficiency improving, and interference eliminating, etc. [4]. Therefore, as an important part of the 5G and beyond networks, it is of great significance to investigate the impact of UL/DL decoupled access in C-V2X communications. However, flexible access in UL/DL decoupled paradigm will put forward more strict QoS requirements for the dynamic C-V2X communications.

Generally, there are two important scenarios of C-V2X communications, i.e., cellular vehicle-to-infrastructure (C-V2I) communications and relay-assisted cellular vehicle-to-vehicle (RAC-V2V) communications [5]–[8]. Due to the different requirements of C-V2X communications, the demand-critical vehicular applications have been recognized as an important 5G scenarios of enhanced mobile broadband (eMBB) communications and ultra-reliable low latency communications (URLLC). For example, the cellular vehicle-to-infrastructure communications can support the real-time sharing of 3D HD maps and other media-rich information among vehicles,

TABLE I
SUMMARY OF IMPORTANT SYMBOLS

M	The macro BS	b	The b th small BS
W^{UL}	The total UL bandwidth	W^{DL}	The total DL bandwidth
β_{UL}	The UL slicing ratio	β_{DL}	The DL slicing ratio
f	The slicing ratio index	$\mathcal{H}$	The set of V2I users
$\mathcal{D}$	The set of RAC-V2V users	$\hbar$	The UL transmission power
∂	The DL transmit power	G	The antenna gain
α	The path loss constant	x	The distance from user i to BS l
h	The channel gain	I	The UL/DL interference
$\mathcal{A}$	The UL transmission rate	$\mathcal{B}$	The DL transmission rate
$\Re$	The UL association matrix	$\Im$	The DL association matrix
$\mathcal{P}$	The UL slicing ratio matrix	$\mathcal{Q}$	The DL slicing ratio matrix
$\Lambda(\varpi)$	The effective bandwidth	$\mathfrak{D}_{\max}$	The maximum delay bound
ε	The delay violation probability	$c^{\min}$	The minimum transmission rate for RAC-V2V users
$\mathbb{S}$	The set of states	$\mathbb{A}$	The set of actions
r	The agent reward	a	The action denoted as slicing ratios
π	The policy mapping from states to actions	V_π	The state value
Q_π	The state-action value	$L_V(\psi)$	The squared residual error of value network
$L_Q(\theta)$	The squared residual error of Q network	$L_\pi(\phi)$	The KL-divergence

while the RAC-V2V communications enable longer-distance safety-related vehicular applications with ultra-low-latency requirements [9]. However, how to provide customized C-V2X services with diversified requirements in densified heterogeneous 5G and beyond networks is still challenging. As a widely recognized network architecture of 5G revolution techniques, RAN slicing [10] is emerging as a promising way to provide customized services with differentiated QoS requirements in 5G and beyond C-V2X [11], [12]. With the fast development of advanced AI technologies, the AI-based RAN slicing approaches have become promising solutions to effectively address the optimization problems in dynamic resource allocation with low complexity. Hence, it is meaningful to design a AI-based network slicing framework for C-V2X communications.

For the aforementioned considerations, in this paper, we investigate the feasibility of UL/DL decoupled RAN framework and study the dynamic RAN slicing problem for both V2I communications and RAC-V2V communications. Firstly, the UL/DL decoupled access technology of vehicles is employed in a two-tier heterogeneous cellular network. Secondly, the effective bandwidth and effective capacity theory are leveraged to map the link-layer delay of RAC-V2V communications to different bandwidth requirements [13]. Furthermore, after analyzing the characteristics of the bidirectional traffic flow between UL and DL of RAC-V2V communications, we put forward the QoS metric of C-V2V innovatively. Finally, we design a two-tier RAN slicing solution, including the soft actor-critic (SAC) algorithm based on off-policy deep reinforcement learning and the alternative slicing ratio search (ASRS) algorithm based on block coordinated descent (BCD), respectively. This two-tier RAN slicing algorithm maximizes the network capacity while guaranteeing the different QoS requirements of V2I and RAC-V2V slice. We highlight our contributions in three-fold:

- We investigate the UL/DL decoupled access problem for cellular V2X, which can significantly reduce the transmit power of vehicle users. Through the flexible association pattern, more vehicles will choose to access SBSs, which can significantly release the heavy burden of MBS.
- We innovatively analyze the bidirectional properties between UL and DL traffic and further propose an innovative QoS metric for the RAC-V2V communications, which can ensure stability and stringent delay bound requirement of RAC-V2V communications.
- We propose a two-tier RAN slicing solution. On the first tier, we utilize the off-policy reinforcement learning algorithm SAC for allocating spectrum bandwidth to each slice. On the second tier, we equivalently convert the RAC-V2V QoS metric minimization problem into a non-convex problem and design the ASRS algorithm based on BCD to solve it. The convergence of ASRS algorithm is theoretically proved, and the RAC-V2V QoS metric can be quickly reduced to a minimum value during the actual iteration process.

The remainder of this paper is organized as follows. Section II briefly reviews the related works. Section III introduces the system model. Section IV introduces the problem formulation. Section V presents the two-tier RAN slicing approach. In section VI, the simulation results and detailed analysis of the proposed two-tier algorithm are presented. Finally, section VII concludes this literature. For convenience, Table I summarizes the major notations of this paper.

II. RELATED WORKS

A. UL/DL Decoupled Access

In contrast to traditional user association scenario where UL/DL can only connect to one particular BS, the UL/DL decoupled access proposed by Boccardi *et al.* [4] is emerging as a new flexible cell association paradigm in recent years. The decoupled access technologies which enable mobile users to access different base stations in UL/DL can significantly enhance UL transmission throughput, improve energy efficiency with a relatively low cost, and bring remarkable gains to the communication capacity of cell edge users [14]. To achieve the benefits brought from the UL/DL decoupled access, Lema *et al.* [15] proposed a dual-connectivity

spectrum aggregation for innovative user connection in 5G heterogeneous systems and Bacha *et al.* [16] utilized stochastic geometry to analyze the two-tier heterogeneous networks with multi-antenna base stations. In the context of UL/DL decoupled resource allocation, Chen *et al.* [17] considered an echo state network framework in LTE-U to choose optimal bands with only limited information on network and user states.

B. Radio Access Network Slicing

Network slicing is an evolving resource allocation concept that can be leveraged to meet the diverse requirements of user demands for 5G wireless communications. By slicing the infrastructure network into multiple dedicated logical networks, wireless networks can support a variety of personal user services [18]. Due to the spectrum resource scarcity of radio access networks, RAN slicing can play a crucial role in guaranteeing the QoS requirements of different users. For the application of RAN slicing, Ye *et al.* [19] proposed an alternative concave search algorithm to maximize the aggregated network utility. Focusing on the design of three key network slicing building blocks, Sciancalepore *et al.* [20] proposed to adaptively modify the predicted load based on measurement deviations, which can enable the traffic analysis, network slicing prediction, and admission control decisions for network slicing requests. However, because of the complex and dynamic network environment brought about by the mobility of vehicles, it is difficult for traditional mechanisms to handle heterogeneous vehicular services requirements [21]. Xiong *et al.* [22] modeled this problem as a Markov decision process and proposed a Monte Carlo tree-based smart slice scheduling algorithm in vehicular fog radio access networks. Through the virtualization of multi-dimensional network resources and the design of a machine learning algorithm, the QoS of V2X services can be dramatically enhanced. In addition, Zhang *et al.* [23] investigated the air-ground integrated vehicular network slicing problem and proposed a networking slicing approach to match the multi-resources across different slices. To enhance the quality of experience, Khan *et al.* [24] proposed a joint quality selection and resource allocation technique. By leveraging the clustering algorithm and Lyapunov drift plus penalty method, it can effectively achieve low latency and high-reliability vehicular communications.

To further facilitate the flexibility and capability of wireless resource allocation while guarantee QoS requirements of various services, intelligent allocation mechanisms based on machine learning are widely investigated in network slicing [12], [25]. Aijaz [26] adopted a post-decision state reinforcement learning (RL) method to allocate radio resources for different slices. Albonda and Prez-Romero [27] proposed an efficient Deep Q-learning based RAN slicing algorithm to ensure the availability of spectrum resources and fulfill the V2I and V2V slice QoS requirements. Although these aforementioned model-free RL algorithms have achieved a certain effect, they are hindered from widespread use in network slicing by the following two bottlenecks. On the one hand, the on-policy based RL algorithms, like trust region policy optimization (TRPO) and proximal policy optimization (PPO) need more samplings to calculate the gradient while the huge amount of sampling process in the actual environment turns out to be a great burden [28].On the other hand, the deterministic-policy-based RL algorithms such as deep deterministic policy gradient (DDPG) can often result in strong hyperparameter sensitivity and low sampling efficiency [29].

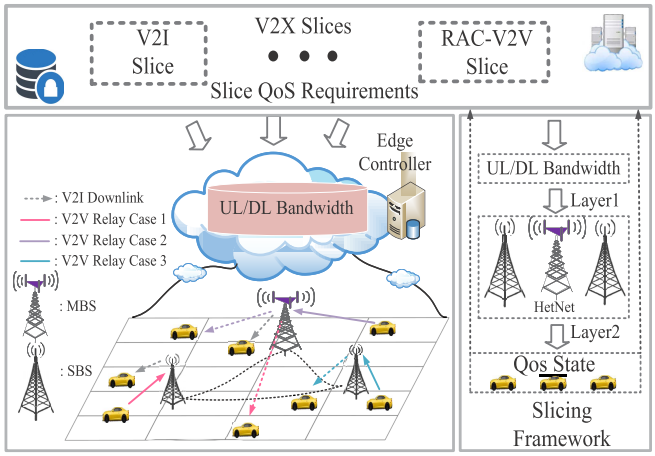


Fig. 1. RAN slicing scenario for decoupled access.

III. System Model

A. Network Model

As illustrated in Fig. 1, we consider a dynamic RAN slicing framework for C-V2X communications underlaid a two-tier HetNet [30]. Under the decoupled rules, the UL and DL RAC-V2V users can be freely connected to MBS or SBS independently. Generally, there are three different UL/DL association cases (i.e., the two-hop RAC-V2V communications can be achieved through one or two base stations) in Fig. 1. In cases 2 and 3, the two-hop relay can be achieved through one base station. However, in case 1, the relay which is realized by establishing UL/DL connections with two base stations should be supported by the interface between BSs. The inter-BS communication (i.e., the communication between MBS and SBS) can be achieved by the standardized X2 interface while the data will be forwarded to another BS via a common public radio interface [4]. Specifically, in the slicing framework, we consider both V2I and RAC-V2V slices. Meanwhile, the cloud-RAN (C-RAN) technique is leveraged to aggregate UL/DL bandwidth at the edge cloud. Since the communication requirements of V2I and RAC-V2V slices are different, in this paper, we design a two-tier bandwidth slicing algorithm to maximize the sum capacity while meeting different QoS requirements of different slices. As shown in Fig. 1, at the first slicing tier (i.e., tier 1), resource orchestration between V2I and RAC-V2V slice is achieved by allocating bandwidth among MBS and SBSs. At the second tier (i.e., tier 2), bandwidth will be precisely allocated to each vehicle user.

Let M and b denote the macro BS and small BS, respectively, where $b \in \{1, \cdots, b, \cdots, \mathrm{B}\}$. In addition, $\mathbb{M}$ and Φ_b

represents UEs associated with MBS and SBSs, respectively. The total UL and DL bandwidths are denoted as W^{UL} and W^{DL}, respectively. In order to improve the efficiency of spectrum bandwidth utilization, bandwidth resources are reused between small BSs. Besides, the spectrum is orthogonally allocated between macro BS and small BSs. Therefore, we can denote $\beta_f W^f$, $\left(1-\beta_f\right) W^f$, $f \in \{\mathrm{UL},\mathrm{DL}\}$ as the bandwidth allocated to the macro cell M and a small cell b, respectively. By deciding the optimal slicing ratios β^*_{UL} and β^*_{DL}, the purpose of this slicing framework is to maximize bandwidth utilization and ensure different QoS of V2I and RAC-V2V slice. There are two kinds of vehicle users, one is the downlink V2I users $\mathcal{H} = \{1, \cdots, h, \cdots \mathrm{H}\}$ and the other is the pairs of RAC-V2V users $\mathcal{D} = \{1, \cdots, d, \cdots \mathrm{D}\}$.

We consider the queues at the base station and user side, which are responsible for the uplink and downlink transmission. We analyze different QoS requirements of flow of data packets from different vehicle users. The packet transmission rate from the BSs to the V2I user is λ_h packets/s, the length of each packet is a constant L_h bits. For RAC-V2V communications, considering its strict delay requirements and link instability, we model the process of the arrival of both UL and DL packets of RAC-V2V users as Poisson processes with the same average rate λ_d packet/s, and the length of the packet is L_d bits.

B. Capacity Analysis of Vehicle Users

We assume all vehicle users are equipped with a single antenna. Traditional UL/DL access is based on the maximum average received signal power (RSP). However, in the decoupled scenario, vehicular users choose the nearest BS as UL service provider while DL users choose the BS with the largest received power.

For a typical vehicle user $i \in \{\mathcal{H}, \mathcal{D}\}$ connected to a cell l in UL, the distance is $x_{i,l}$, $l \in \{\Phi_b, \mathbb{M}\}$, if and only if:

$$\hbar_{i,l} G_l \|x_{i,l}\|^{-\alpha_l} \geq \hbar_{i,k} G_k \|x_{i,k}\|^{-\alpha_k}, \quad \forall l,k \in \{\Phi_b, \mathbb{M}\}. \tag{1}$$

Similarly, a user connected to a cell l in DL with distance $x_{i,l}$, if and only if:

$$\partial_l G_l \|x_{i,l}\|^{-\alpha_l} \geq \partial_k G_k \|x_{i,k}\|^{-\alpha_k}, \quad \forall l,k \in \{\Phi_b, \mathbb{M}\}, \tag{2}$$

where G_k and α_k represents the antenna gain and path loss constant of user i connected to BS k, respectively. $\hbar_{i,\mathrm{M}}$ and ∂_{M} indicates UL transmit power of vehicle users and DL transmit power of macro cell, respectively [31]. Similarly, $\hbar_{i,\Phi_b}$ and ∂_{Φ_b} denotes UL and DL transmit power of users associated with small cells and transmit power of small BSs, respectively.

We denote the SNR of communication link between vehicle user and connected macro cell as:

$$\begin{aligned} \mathrm{SNR}^i_{\mathrm{UL,M}} &= \frac{\hbar_{i,\mathrm{M}} G_{\mathrm{M}} h \|x_{i,\mathrm{M}}\|^{-\alpha_{\mathrm{M}}}}{\sigma^2}; \\ \mathrm{SNR}^i_{\mathrm{DL,M}} &= \frac{\partial_{\mathrm{M}} G_{\mathrm{M}} h \|x_{i,\mathrm{M}}\|^{-\alpha_{\mathrm{M}}}}{\sigma^2}. \end{aligned} \tag{3}$$

We consider the spectrum bandwidth can be reused between small BSs and allocated orthogonally in one small BS. Due to the co-channel interference, the SINR of communication link between vehicle user and a small BS can be shown as:

$$\begin{aligned} \mathrm{SINR}^i_{\mathrm{UL},\Phi_b} &= \frac{\hbar_{i,\Phi_b} G_{\Phi_b} h \|x_{i,\Phi_b}\|^{-\alpha_{\Phi_b}}}{\sum\limits_{j \in \Phi_b' \neq \Phi_b} \hbar_{j,\Phi_b} G_{\Phi_b} h \|x_{j,\Phi_b}\|^{-\alpha_{\Phi_b}} + \sigma^2}; \\ \mathrm{SINR}^i_{\mathrm{DL},\Phi_b} &= \frac{\partial_{\Phi_b} G_{\Phi_b} h \|x_{i,\Phi_b}\|^{-\alpha_{\Phi_b}}}{\sum\limits_{\Phi_b' \neq \Phi_b} \partial_{\Phi_b'} G_{\Phi_b'} h \|x_{i,\Phi_b'}\|^{-\alpha_{\Phi'}} + \sigma^2}. \end{aligned} \tag{4}$$

with σ^2 being the power of the additive white Gaussian noise. The interference $I^i_{\mathrm{UL},\Phi_b} = \sum\limits_{j \in \Phi_b' \neq \Phi_b} \hbar_{j,\Phi_b} G_{\Phi_b} h \|x_{j,\Phi_b}\|^{-\alpha_{\Phi_b}}$ and $I^i_{\mathrm{DL},\Phi_b} = \sum\limits_{\Phi_b' \neq \Phi_b} \partial_{\Phi_b'} G_{\Phi_b'} h \|x_{i,\Phi_b'}\|^{-\alpha_{\Phi_b'}}$ represents the uplink and downlink inter-cell interference, respectively.

The load of each cell is determined by the number of devices associated with the same BS and the achievable rates of corresponding users. There are two different users, i.e., V2I and RAC-V2V users. The matrix $\mathcal{A}$ and $\mathcal{B}$ represents the Shannon capacity with full bandwidth associated with macro and small BSs in UL and DL, respectively. The index $i \in \{1, \cdots, \mathrm{H}\}, j \in \{1, \cdots, 1+\mathrm{B}\}$ denotes the rows and columns of matrix $\mathcal{A}$, respectively. Therefore, when the UL bandwidth of each BS is given, the Shannon capacity of each user $\mathcal{A}_{i,j}$ can be expressed as:

$$\mathcal{A}_{i,j} = \begin{cases} \beta_{\mathrm{UL}} \mathrm{W}^{\mathrm{UL}} \log(1 + \mathrm{SNR}^{i,j}) & j \leq 1; \\ (1-\beta_{\mathrm{UL}}) \mathrm{W}^{\mathrm{UL}} \log(1 + \mathrm{SINR}^{i,j}) & \text{otherwise.} \end{cases} \tag{5}$$

For the DL association, the index $i \in \{1, \cdots, \mathrm{H}+\mathrm{D}\}, j \in \{1, \cdots, 1+\mathrm{B}\}$ denotes the rows and columns of matrix $\mathcal{B}$. When the DL bandwidth of each BS is given, the Shannon capacity of each user matrix $\mathcal{B}_{i,j}$ can be calculated by:

$$\mathcal{B}_{i,j} = \begin{cases} \beta_{DL} \mathrm{W}^{\mathrm{DL}} \log(1 + \mathrm{SNR}^{i,j}) & j \leq 1; \\ (1-\beta_{\mathrm{DL}}) \mathrm{W}^{\mathrm{DL}} \log(1 + \mathrm{SINR}^{i,j}) & \text{otherwise.} \end{cases} \tag{6}$$

C. Effective Bandwidth Theory for RAC-V2V Communications

The effective bandwidth theory derived from the large deviation theory represents the lowest service rate under the queuing delay constraint of a given data rate of source flow, which is often used to obtain the optimal resource allocation strategy. We consider the end-to-end delays can be calculated between a packet from its arrival to departure, e.g., a packet is generated at the user side or base station side, and then to its destination. For the two-hop RAC-V2V communications, to simplify the problem formulation, we consider both the uplink and the downlink packets of the relay are Poisson arriving, and the effective bandwidth is the same [32]. Therefore, we can derive a minimum transmission rate with delay violation probabilities of RAC-V2V communications.

We first denote the QoS index ϖ, and the effective bandwidth of RAC-V2V communications can be expressed as:

$$\Lambda(\varpi) = \lim_{t \to \infty} \frac{1}{t} \frac{1}{\varpi} \log\left(E\left[e^{\varpi O(t)}\right]\right), \tag{7}$$

where $O(t)$ represents the number of packets arriving over $[0, t)$ for RAC-V2V communications, $E[\cdot]$ represents the

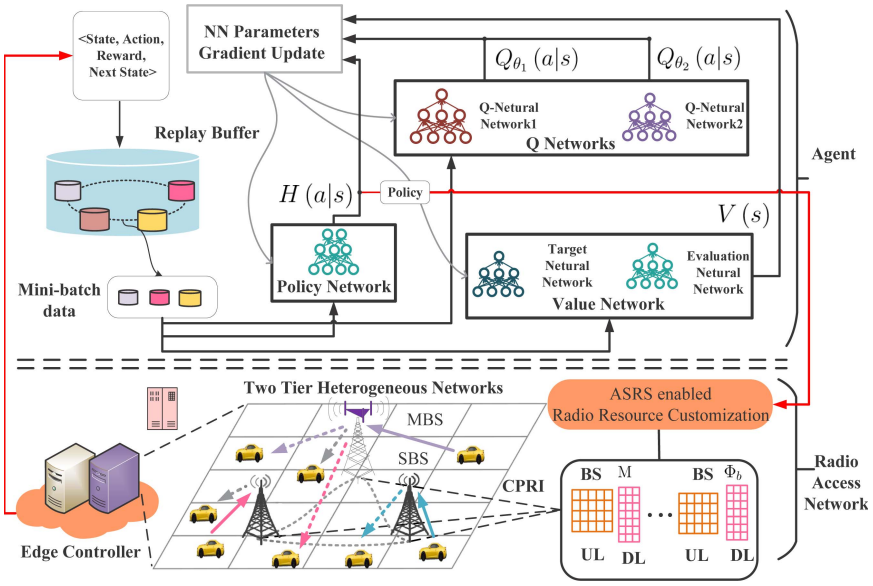


Fig. 2. The framework of two-tier RAN slicing for UL/DL decoupled cellular V2X.

expectation. Since $O(t)$ is modeled as a Poisson process with λ_d^f packet/s [19], it can be further expressed as:

$$\Lambda(\varpi) = \lambda_d^f \frac{e^{\varpi} - 1}{\varpi}. \tag{8}$$

The total transmission delay $\mathfrak{D}_{i,l}$ of a packet from the transmitter BS/vehicle-user l to the receiver BS/vehicle-user i exceeding the maximum delay constraint $\mathfrak{D}_{max}$ can be fitted as:

$$Pr\{\mathfrak{D}_{i,l} \geq \mathfrak{D}_{max}\} \approx e^{-\varpi p_{i,l} \mathfrak{D}_{max}} \leq \varepsilon, \tag{9}$$

where ε represents the delay violation probability, $p_{i,l} = \frac{c_{i,l}}{\mathrm{L}_d}$ is achievable rate from V2V user i to BS l (under the constraint of the number of packets transferred per second) and vice versa. The minimum realizable rate $p^{\min}$ is

$$p^{min} = -\frac{\log_{\varepsilon}}{\varpi \mathfrak{D}_{max}}. \tag{10}$$

According to the effective bandwidth theory, p^{min} should be equal to the effective bandwidth $\Lambda(\varpi)$ to ensure that the probability of delay violation does not exceed ε. So we can get $\varpi = \log\left(1 - \frac{\log \varepsilon}{\lambda_d^f \mathfrak{D}_{max}}\right)$. Then the minimum transmission rate of RAC-V2V communication can be derived as:

$$c^{min} = p^{min} \mathrm{L}_d = -\frac{\mathrm{L}_d \log_{\varepsilon}}{\mathfrak{D}_{max} \log\left(1 - \frac{\log_{\varepsilon}}{\lambda_d^f \mathfrak{D}_{max}}\right)}. \tag{11}$$

We consider an equal-length time-slotted system with the time as $\mathcal{T} = \{1, \cdots, T\}$. The t-time slot is set as hundreds of milliseconds. Hence, the vehicle locations do not change too much during this time slot. Additionally, the channel is assumed to be constant in each time period [33]. Therefore, the capacity of vehicle users and the minimum required transmission rate in (11) remains unchanged during the t-th time period.

IV. Two-Tier RAN Slicing Approach

A. MDP Model for RAN Slicing

The detailed process of learning, storing, and updating in our proposed two-tier slicing framework is illustrated in Fig. 2, which can be described as a Markov Decision Process (MDP). We denote the state, action, state change and return of the slice controller to get the fully observed MDP model. By dynamic interactions between slice controller and wireless network environment, we can derive the tuple: $(\mathbb{S}, \mathbb{A}, r, \mathfrak{P}, a)$, where $\mathbb{S}$ denotes the collection of states, $\mathbb{A}$ indicates a set of possible actions. $\mathfrak{P}$ is defined as the probability of state transition. In our environment, due to the dynamic and random nature of the state, it is impossible to know the probability of real-time state transition, so in the following, we adopt a model-free deep reinforcement learning algorithm to deal with the fully observed MDP problem. We denote $r_t(s_t, a_t)$ as the reward at a specific state s_t with taking action a_t, which will be further returned to the network slice controller.

State: We denote the cell state s_t as a tuple: $\{\Gamma_{t-1}, \Upsilon_t, r_{t-1}\}$, where Γ_t represents the bandwidth allocated to V2I and RAC-V2V slices at last time period $t-1$, Υ_t denotes the total bandwidth requirement of different slices at time period t, and r_{t-1} represents the reward last time period $t-1$.

Action: At a certain state, DRL agent performs an action of $a_t = \{\beta_{\mathrm{UL}}, \beta_{\mathrm{DL}}\}$, where β_{UL} and β_{DL} are restricted to $[0, 1]$.

Reward: The payoff for the state transfer is $r_t \in R(s, a)$. In the proposed scenario, at time period t, the reward including the utility function and the QoS of RAC-V2V cellular mode can be expressed as:

$$r_t(s, a) = \sum_i \sum_j \mathcal{A}\mathcal{P} + \sum_i \sum_j \mathcal{B}\mathcal{Q}. \tag{12}$$

We consider the dimensions of these matrices $\mathcal{A}$, $\mathcal{P}$, $\mathcal{B}$ and $\mathcal{Q}$ are the same, constituted by i rows and j columns. The

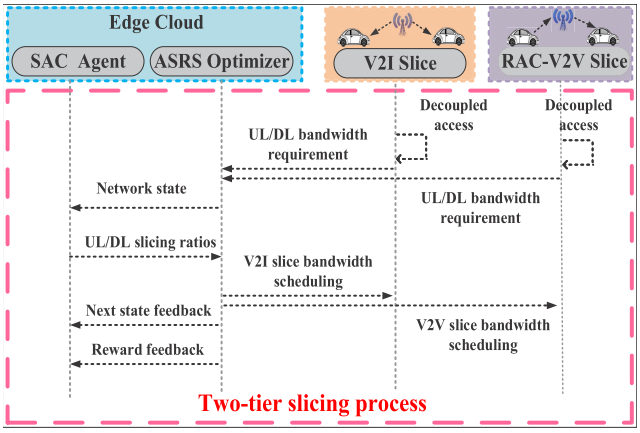


Fig. 3. Two-tier RAN slicing process.

index i represents the vehicles, while the index j represents the BSs. The matrices $\mathcal{A}$ and $\mathcal{B}$ denotes the UL/DL vehicle user's capacity with full bandwidth allocated to the associated BS, respectively. The matrices $\mathcal{P}$ and $\mathcal{Q}$ represents the UL/DL bandwidth slicing ratios to each user, respectively. As shown in Fig. 3, the two-tier bandwidth allocation diagram is described. Every vehicle user adopts UL/DL decoupled access procedure and reports its minimum rate requirement to each associated BS. Firstly, the edge cloud collected the rate requirement of different slices. Secondly, the policy network in SAC chooses an action (i.e., the UL/DL bandwidth allocation ratios) based on the network state. Thirdly, the ASRS is designed to allocate bandwidth to each user. When the allocation is completed, the rewards and new state will be put into the replay buffer.

The goal of the slicing controller is to maximize the long-term reward $\sum_{t=1}^{T} \gamma^{t-1} r_t(s_t, a_t)$, which is represented as the discount summation of the reward over the time period $\mathcal{T}$. The discount factor γ is restrict to $[0, 1]$.

B. SAC for Allocating Spectrum to BSs

As a typical model-free algorithm, Q-learning performs poorly in processing continuous action situations, due to the fact that discretizing the actions will cause the exponential growth of action's number and some important information for solving the problem may be lost. Therefore, we propose a SAC algorithm based on the Actor-Critic framework. The adoption of SAC algorithm has the following three advantages: 1), based on policy iteration, it can deal with the problem of continuous action space; 2) the combined reward based on maximum entropy and state value can effectively explore more excellent actions, which gives a stronger exploration ability; 3) SAC overcomes the complexity brought by massive sampling and the sensitivity of reinforcement learning in hyperparameters (e.g., learning rates, exploration constants) [34].

The goal of traditional RL algorithms is to maximize long-term expected rewards $\sum_t \mathbb{E}_{(s_t,a_t)\sim\rho_\pi}[r(s_t, a_t)]$ according to a policy $\pi(a|s)$. However, in order to learn more effective movements, we add entropy to the long-term reward,

$$J(\pi) = \sum_{t=0}^{T} \mathbb{E}[r(s_t, a_t) + \alpha \mathcal{H}(\pi(\cdot|s_t))|\pi], \tag{13}$$

where the parameter α is the temperature index and represents the relative importance of the entropy compared to the reward. The larger the value of α, the more stochastic actions we want. Otherwise, the effective actions learned will be more specific. $\mathcal{H}(\pi(\cdot|s_t)) = -\log \pi(a|s)$ denotes an entropy for actions at state s. So Eq. (13) can be further described as,

$$J(\pi) = \sum_{t=0}^{T} \mathbb{E}[r(s_t, a_t) - \alpha \log \pi(\cdot|s)|\pi], \tag{14}$$

In the policy evaluation step of soft iteration, we wish to compute the value of a policy π according to the maximum entropy objective in Eq. (14). For a fixed policy, the discount soft Q-value can be computed iteratively, starting from any function Q and repeatedly applying a modified Bellman backup operator $\mathcal{F}^\pi$.

$$\mathcal{F}^\pi Q(s_t, a_t) \triangleq r(s_t, a_t) + \gamma \mathbb{E}[V(s_{t+1})]. \tag{15}$$

The soft value function can be obtained from Bellman backup operator, which is denoted as,

$$V(s_t) = \mathbb{E}[Q(s_t, a_t) - \alpha \log \pi(a_t|s_t)|\pi]. \tag{16}$$

Theorem 1: Given the policy π, when the action reward at any state is finite, the sequence Q^k can be converged to the soft Q-value, where $Q^{k+1} = \mathcal{F}^\pi Q^k$.

Proof: See Appendix B.1 in [34].

For the Actor-Critic algorithm in continuous action space, we run actor and critic alternately until convergence. Here, we use deep neutral network (DNN) to separately fit the Critic network and the Actor network, and then use the stochastic gradient descent method to update the parameters in the two networks alternately.

As shown in Fig. 2, there are 4 networks in the critic part, two value networks and two Q networks. Value networks are parameterized by $V(s) \approx V_\psi(s)$ and $V(s) \approx V_{\bar{\psi}}(s)$, i.e., evaluate network ψ and target network $\bar{\psi}$. The reason for using two value networks is to make the training more stable. The update of value network can be trained by minimizing the mean squared error (MSE) $L_V(\psi)$.

$$L_V(\psi) = \mathbb{E}\left[\frac{1}{2}\left(V_\psi(s_t) - \mathbb{E}\left[Q_\theta(s_t, a_t) - \alpha \log \pi_\phi(a_t|s_t)\right]\right)^2\right]. \tag{17}$$

The parameters needed for neural network update come from replay memory D, where the $\mathcal{D}$ is the distribution of previous sampled states and actions. The gradient of Eq. 17 can be estimated with an unbiased estimator,

$$\hat{\nabla}_\psi L_V(\psi) = \nabla_\psi V_\psi(s_t)\left(V_\psi(s_t) - Q_\theta(s_t, a_t) + \alpha \log \pi_\phi(a_t|s_t)\right). \tag{18}$$

The parameters of the evaluate network is updated by

$$\psi \leftarrow \psi - \vartheta_V \hat{\nabla}_\psi L_V(\psi), \tag{19}$$

where the parameter $\vartheta_V \geq 0$ is the learning rate of evaluate value network.

Leveraging the soft deep network parameter update method in double deep Q-learning (DDQN), we can derive target network update formula:

$$\bar{\psi} \leftarrow \psi + (1-\tau)\bar{\psi}, \tag{20}$$

where the parameter $\tau \in [0,1]$ is update weight of the target value network.

The two Q networks with parameters θ_1 and θ_2 are responsible for evaluating the state-action, which are illustrated in Fig. 2. The reason for using two Q networks is to select a lower Q value when updating the policy network to reduce overestimation, where the actions are sampled according to the current policy. The soft Q-function parameters can be trained to minimize the squared error,

$$L_Q(\theta) = \mathbb{E}\left[\frac{1}{2}\left(Q_\theta\left(s_t,a_t\right) - \hat{Q}\left(s_t,a_t\right)\right)^2\right], \tag{21}$$

with

$$\hat{Q}(s_t,a_t) = r(s_t,a_t) + \gamma \mathbb{E}\left[V_{\bar{\psi}}(s_{t+1})\right], \tag{22}$$

where $V_{\bar{\psi}}(s_{t+1})$ is the target state value in s_{t+1}. The MSE loss $L_Q(\theta)$ can be optimized with stochastic gradients:

$$\hat{\nabla}_\theta L_Q(\theta) = \nabla_\theta Q_\theta(s_t,a_t)\left(Q_\theta(s_t,a_t) - r(s_t,a_t) - \gamma V_{\bar{\psi}}(s_{t+1})\right). \tag{23}$$

From the above, we can update the Q network parameter θ_i, where $i \in \{1,2\}$, according to the following formula:

$$\theta_i \leftarrow \theta_i - \vartheta_Q \hat{\nabla}_{\theta_i} L_Q(\theta_i), \tag{24}$$

where the parameter $\vartheta_Q \geq 0$ is the learning rate of evaluate value network.

The policy network is responsible for outputting the Gaussian mean and variance of continuous actions. In the policy improvement step, for each state, we update the policy towards the exponential of the new Q-function, which is proved to lead to an improved policy with respect to the objective in Eq. (15) [35]. Also, we leverage KL-divergence to project the parameterized policy of Gaussian distribution. The entire soft strategy iteration algorithm alternates between soft strategy evaluation and soft strategy improvement. Finally, the policy parameters can be learned by directly minimizing the expected KL-divergence,

$$L_\pi(\phi) = \mathbb{E}\left[\mathrm{D}_{\mathrm{KL}}\left(\pi_\phi\left(\cdot|s_t\right)\middle|\middle|\frac{\exp\left(Q^{\pi_{\mathrm{old}}}\left(s_t,\cdot\right)\right)}{\mathrm{Z}^{\pi_{\mathrm{old}}}\left(s_t\right)}\right)\right]. \tag{25}$$

In our case, the target density is the Q-function, which is represented by a neural network. Therefore, it can be differentiated. And it is also convenient to apply the reparameterization trick instead, resulting in a lower variance estimator. The function $Z^{\pi_{\mathrm{old}}}$ is the normalized distribution, which has no contribution to the gradient and thus can be ignored. To that end, we reparameterize the policy using a neural network transformation,

$$a_t = f_\phi(\epsilon_t; s_t), \tag{26}$$

where ϵ_t is an input noise vector, sampled from Gaussian fixed distribution. Use reparameterization techniques, we can now rewrite the objective in Eq. (25) as:

$$L_\pi(\phi) = \mathbb{E}\left[\log \pi_\phi\left(f_\phi\left(\epsilon_t;s_t\right) - Q_\theta\left(s_t, f_\phi\left(\epsilon_t;s_t\right)\right)\right)\right], \tag{27}$$

where π_ϕ is defined implicitly in terms of f_ϕ, and we have noted that the function $Z^{\pi_{\mathrm{old}}}$ is independent of ϕ and thus can be ignored. We can approximate the gradient of Eq. (27) with

$$\begin{aligned}\hat{\nabla}_\phi L_\pi(\phi) &= \nabla_\phi \log \pi_\phi(a_t|s_t) \\ &\quad + \left(\nabla_{a_t}\log\pi_\phi(a_t,s_t) - \nabla_{a_t} Q\left(s_t,a_t\right)\nabla_\phi f_\phi(\epsilon_t;s_t)\right).\end{aligned} \tag{28}$$

To minimize $L_\pi(\psi)$, it is easy to get the policy network parameter update equation as:

$$\phi \leftarrow \phi - \vartheta_\pi \hat{\nabla} L_\pi(\phi), \tag{29}$$

where the parameter $\vartheta_\pi \geq 0$ is the learning rate of policy network.

Theorem 2: Given the policy $\pi \in \Pi$, the state-action value is finite, when we alternately update the critic network and the policy network, the policy π will converge to the optimal π^*, which can be expressed as $Q^{\pi^*}(s_t,a_t) \geq Q^{\pi}(s_t,a_t)\ \ \forall \pi \in \Pi$.

$Proof$: Detailed proof can be found in [34].

Algorithm 1 SAC-Based Slicing Strategy on Large Time Scales

Initialize parameters ψ, $\bar{\psi}$, θ, ϕ
for *time step* $t = 1, \cdots, T$ *in each training episode* **do**
  Initialize system state s_t;
  Choose slicing ratios$a_t = (\beta_{UL}, \beta_{DL})$ form the policy $a_t \sim \pi\left(\cdot|s_t,\phi\right)$;
  Perform action a_t and get the UL DL bandwidth for MBS and SBSs, respectively ;
  Perform Algorithm 2 to get the reward r_t;
  Observe the new state s_{t+1}, namely, the feedback tupe: $\{\Gamma_t, \Upsilon_{t+1}, r_t\}$;
  Store $\{s_t, a_t, r_t, s_{t+1}\}$ to the replay memory D;
for *each gradient step* **do**
  Sample $\{s_i, a_i, r_i, s_i\}$ from the D;
  Update the evaluate value network according to Eq. (19);
  Update the two Q networks, according to Eq. (24);
  Update the policy network according to Eq. (29);
  Update the target value network parameters according to Eq. (20);

Algorithm 1 presents the functionality of the agent part in Fig. 2, including parameters updating, action selecting, state storing, etc. The parameters ψ, $\bar{\psi}$, θ, ϕ need to be initialized before the algorithm starts. At each step of network slicing, the actor-network in SAC selects an action based on the current state. When the bandwidth of each BS is determined according to slicing ratios $(\beta_{\mathrm{UL}}, \beta_{\mathrm{DL}})$, Algorithm 2 will allocate the

bandwidth to each vehicle user, and then the edge controller obtains the network states of V2I and RAC-V2V Slice, and reward is sent to the replay buffer D, which is used to store experience. When the replay buffer reaches a certain amount, the network starts training and updates the network parameters.

C. ASRS for Allocating Spectrum to Vehicle Users

In terms of RAC-V2V communications, the UL and DL are independently associated with different BSs. However, since the communicating vehicles need to exchange safety-related information. By leveraging RAC-V2V, the roles of transmitter and receiver of RAC-V2V communications are constantly reversible to complete the interactive communications. Generally, UL and DL sessions of RAC-V2V cellular users are coupled to complete bidirectional information exchanges. To ensure stability and stringent delay bound requirement, UL/DL resource allocation must be considered simultaneously. The bidirectional safety-related data and signaling flow between RAC-V2V transmitter and receiver in UL/DL are general symmetrical traffic. Therefore, RAC-V2V communications require symmetric resource allocation in UL and DL. Therefore, we can formulate the QoS metric of RAC-V2V communication as $|R^{\mathrm{UL}} - R^{\mathrm{DL}}|$. The variable R^{UL} and R^{DL} represents the achievable rate in UL/DL, respectively. To ensure safe and timely communication, it is necessary to ensure that the UL/DL capacity of the RAC-V2V users is as the same as possible through synchronous resource allocation. Therefore, in the following, we will design the ASRS algorithm to minimize the RAC-V2V QoS metric. We use 1 to indicate that the vehicle user has established a link with the corresponding BS, and 0 indicates there is no link established with the BS. Therefore, we denote UL and DL association matrix as $\Re_{\mathrm{H}\times(\mathrm{B}+1)}$ and $\Im_{(\mathrm{H}+\mathrm{D})\times(\mathrm{B}+1)}$, respectively,

$$\Re = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix} \quad \Im = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

where the number of rows and columns of matrix $\Re$ and $\Im$ indicates the number of vehicles and BSs, respectively, $\Re_{i,j} \in 0, 1$ and $\Re \cdot e = e$, and $\Im_{i,j} \in 0, 1$ and $\Im \cdot e = e$. The unit vector $e = [1, \ldots, 1]^T$.

Here we also denote the UL and DL slicing ratios matrices, namely, $\mathcal{P}_{\mathrm{H}\times(\mathrm{B}+1)}$ and $\mathcal{Q}_{(\mathrm{H}+\mathrm{D})\times(\mathrm{B}+1)}$, which have the same order as $\Re$ and $\Im$, respectively. Besides, $\mathcal{P}_{i,j} \in (0, 1)$, $\mathcal{Q}_{i,j} \in (0, 1)$, which can be further represented as: $\Re \geq \mathcal{P}$; $\Im \geq \mathcal{Q}$. The sum of each column of matrix, i.e., $\sum_j \mathcal{P}_j$ and $\sum_j \mathcal{Q}_j$ are both equal to 1.

Based on the above analysis, we know that the bandwidth allocation of each BS is given by Algorithm 1. Hence, in the second tier, the goal is to minimize the QoS metric of RAC-V2V communications. The objective function and corresponding constraints are formulated as follows:

$$\underset{\{\mathcal{P}^*,\ \mathcal{Q}^*\}}{\text{Minimize}} \quad \sum_i^N \left| \sum_j \mathcal{A}_{i,j}\mathcal{P}_{i,j} - \sum_j \mathcal{B}_{i,j}\mathcal{Q}_{i,j} \right| \tag{30a}$$

$$\text{s.t.} \quad 0 \leq \mathcal{P}_{i,j} \leq 1,\ \mathcal{P}^T \cdot e = e, \mathcal{P}_{i,j} \leq \Re_{i,j}, \tag{30b}$$

$$0 \leq \mathcal{Q}_{i,j} \leq 1,\ \mathcal{Q}^T \cdot e = e, \mathcal{Q}_{i,j} \leq \Im_{i,j}, \tag{30c}$$

$$\sum_j A_{i,j}\mathcal{P}_{i,j} \geq c_{min},\ i \in \{1, \cdots, \mathrm{H}\}, \tag{30d}$$

$$\sum_j \mathcal{B}_{i,j}\mathcal{Q}_{i,j} \geq c_{min},\ i \in \{1, \cdots, \mathrm{H}\}, \tag{30e}$$

$$\sum_j \mathcal{B}_{i,j}\mathcal{Q}_{i,j} \geq \lambda_h \mathrm{L}_h,\ i \in \{\mathrm{H}+1, \cdots, \mathrm{H}+\mathrm{D}\}. \tag{30f}$$

The solution of the optimization problem in (30) is the bandwidth allocation in both UL and DL while satisfying the minimum rate requirements of V2I and RAC-V2V users. The objective function is an absolute value function. In addition, the absolute value function is NP-Hard and not smooth [26], traditional gradient-descent based methods are difficult to solve it. The objective function in Eq. (30) for the optimization problem is not convex. Therefore, we let Γ be equivalent to the objective function, and then the original optimization problem can be equivalently transformed into,

$$\underset{\{\mathcal{P}^*,\ \mathcal{Q}^*\}}{\text{Minimize}} \quad \Gamma \tag{31a}$$

$$s.t.\ 0 \leq \mathcal{P}_{i,j} \leq 1,\ \mathcal{P}^T \cdot e = e, \mathcal{P}_{i,j} \leq \Re_{i,j}, \tag{31b}$$

$$0 \leq \mathcal{Q}_{i,j} \leq 1,\ \mathcal{Q}^T \cdot e = e, \mathcal{Q}_{i,j} \leq \Im_{i,j}, \tag{31c}$$

$$\sum_j \mathcal{A}_{i,j}\mathcal{P}_{i,j} \geq c_{min},\ i \in \{1, \cdots, \mathrm{H}\}, \tag{31d}$$

$$\sum_j \mathcal{B}_{i,j}\mathcal{Q}_{i,j} \geq c_{min},\ i \in \{1, \cdots, \mathrm{H}\}, \tag{31e}$$

$$\sum_j \mathcal{B}_{i,j}\mathcal{Q}_{i,j} \geq \lambda_h \mathrm{L}_h,\ i \in \{\mathrm{H}+1, \cdots, \mathrm{H}+\mathrm{D}\}, \tag{31f}$$

$$\Gamma \geq \sum_i^N \left[\sum_j \mathcal{A}_{i,j}\mathcal{P}_{i,j} - \sum_j \mathcal{B}_{i,j}\mathcal{Q}_{i,j} \right], \tag{31g}$$

$$\Gamma \geq -\sum_i^N \left[\sum_j \mathcal{A}_{i,j}\mathcal{P}_{i,j} - \sum_j \mathcal{B}_{i,j}\mathcal{Q}_{i,j} \right], \tag{31h}$$

$$\Gamma \geq 0. \tag{31i}$$

However, this is still not a standard convex problem because the inequality constraints $(32g)$, $(32h)$ are dual non-convex (i.e., one of the two constraints must be non-convex). By leveraging the block coordinated descent algorithm [36], we designed ASRS, namely, fixed matrix $\mathcal{P}$ to get the optimal $\mathcal{Q}^*$, then we can utilize $\mathcal{Q}^*$ to calculate the optimal $\mathcal{P}^*$. In each iteration loop of ASRS, we need to fix one variable to solve another variable, so that the problem (31) can be transformed into two linear programming problems. In each iteration, first, we fix the matrix $\mathcal{P}$ to solve for the optimal matrix $\mathcal{Q}$, so the subproblem is described as,

$$\underset{\{\mathcal{Q}^*\}}{\text{Minimize}} \quad \Gamma$$
$$\text{s.t.} \quad (31b), (31d) - (31h). \tag{32}$$

After the above problem (32) is solved, we get the optimal matrix $\mathcal{Q}^*$. Due to the characteristics of cyclic optimization, in the next iteration, we use the optimal $\mathcal{Q}^*$ to solve the optimal $\mathcal{P}^*$. So another sub-problem (33) is expressed as,

$$\begin{aligned} &\underset{\{\mathcal{P}^*\}}{\text{Minimize}} \quad \Gamma \\ &\text{s.t.} \quad (31a), (31c), (31e) - (31h). \end{aligned} \tag{33}$$

In this way, the original non-convex optimization problem is split into two sub-optimization problems. Through the loop calculation, each iteration needs to solve two sub-problems, which is illustrated below:

$$\begin{aligned} &\underbrace{\{\mathcal{P}^0, \mathcal{Q}^0\} \rightarrow \Gamma^0}_{\text{initialization}} \rightarrow \cdots \rightarrow \underbrace{\mathcal{Q}^{n-1} \rightarrow \{\mathcal{P}^{n-1}, \Gamma^{n-1}\}}_{(n-1)\text{ th iteration}} \rightarrow \cdots \\ &\rightarrow \underbrace{\mathcal{P}^n \rightarrow \{\mathcal{Q}^n, \Gamma^n\}}_{(n)\text{ th iteration}} \rightarrow \cdots \rightarrow \underbrace{\{\mathcal{P}^*, \mathcal{Q}^*\} \rightarrow \Gamma^*}_{\text{optimal value}}. \end{aligned} \tag{34}$$

Before the loop of the ASRS algorithm, we need to find an initial bandwidth allocation matrices $\mathcal{P}^0, \mathcal{Q}^0$. Therefore, we need to calculate the minimum communication rate for each vehicle user, and the minimum bandwidth requirement for each uplink user can be represented as:

$$\mathcal{P}_{i,j}^{\min} = \begin{cases} \dfrac{c_{\min}}{\log(1+\text{SNR}^{i,j})} & j = 1, \Re_{i,j} \neq 0; \\ \dfrac{c_{\min}}{\log(1+\text{SINR}^{i,j})} & j \neq 1, \Re_{i,j} \neq 0. \end{cases} \tag{35}$$

Similarly, the minimum bandwidth requirement for the downlink users $\mathcal{Q}_{i,j}^{\min}$ is represented as:

$$\mathcal{Q}_{i,j}^{\min} = \begin{cases} \dfrac{c_{\min}}{\log(1+\text{SNR}^{i,j})} & j = 1, i \leq \text{H}, \Im_{i,j} \neq 0; \\ \dfrac{\lambda_h \text{L}_h}{\log(1+\text{SNR}^{i,j})} & j = 1, \text{H} \leq i \leq \text{H+D}, \Im_{i,j} \neq 0; \\ \dfrac{c_{\min}}{\log(1+\text{SINR}^{i,j})} & j \neq 1, i \leq \text{H}, \Im_{i,j} \neq 0; \\ \dfrac{\lambda_h \text{L}_h}{\log(1+\text{SINR}^{i,j})} & j \neq 1, \text{H} \leq i \leq \text{H+D}, \Im_{i,j} \neq 0. \end{cases} \tag{36}$$

Therefore, we can get the minimum uplink bandwidth requirement for the j th BS $b_i^{\text{UL}} = \sum_i \mathcal{P}_{i,j}$. Similarly, the minimum downlink bandwidth requirement for the j th BS is $b_i^{\text{DL}} = \sum_i \mathcal{Q}_{i,j}$.

The ASRS slicing algorithm is described in Algorithm 2. It is responsible for timely resource customization. We design a heuristic initial bandwidth allocation method, which can find suitable iterative initial test matrices $\mathcal{P}^0, \mathcal{Q}^0$. In the process of cyclic iteration, we first fix a matrix $\mathcal{P}$ and find the optimal objective function Γ and the optimal matrix $\mathcal{Q}$, then utilize the optimal matrix $\mathcal{Q}$ to find the optimal matrix $\mathcal{P}$ and the optimal objective function Γ.

Convergence Analysis: Base on the results presented in the previous two subsections, we propose an iteration algorithm ASRS for problem (31) by the block coordinated decent method, also known as the alternating optimization method. Specifically, the entire optimization variables in original problem (31) are partitioned into two blocks, i.e., $\{\mathcal{P}, \mathcal{Q}\}$. Then

Algorithm 2 The ASRS Slicing Algorithm

Input Parameters $\{\beta^{\text{UL}}, \beta^{DL}\}$ from Algorithm 1, maximum iteration times $\hbar$, and the tolerance $\omega \leq 0$.
Output Optimal bandwidth slicing ratio matrix $\{\mathcal{P}^*, \mathcal{Q}^*\}$, the optimal value Γ^* and reward r.
Calculate $\mathcal{P}_{i,j}^{\min}$ and $\mathcal{Q}_{i,j}^{\min}$.
for $j \in \{1, \cdots, 1+B\}$ **do**
  If $\Re_{i,j} \neq 0$,
    $\mathcal{P}_{i,j}^0 = \frac{\text{UL bandwidth of BS } j - \sum_i \mathcal{P}_{i,j}^{\min}}{\sum_i \Re_{i,j}} + \mathcal{P}_{i,j}^{\min}$,
  If $\Im_{i,j} \neq 0$,
    $\mathcal{Q}_{i,j}^0 = \frac{\text{DL bandwidth of BS } j - \sum_i \mathcal{Q}_{i,j}^{\min}}{\sum_i \Im_{i,j}} + \mathcal{Q}_{i,j}^{\min}$,
Get the initialized matrices $\{\mathcal{P}^0, \mathcal{Q}^0\}$.
Initialize iteration variable $n = 0$.
while $\left|\Gamma^n - \Gamma^{n-1}\right| \geq \omega$ **do**
  Given uplink slicing ratio matrix $\mathcal{P}^n$, obtain the downlink slicing ratio matrix $\mathcal{Q}^n$ and the objective function Γ^n according to problem (32),
  Calculate matrix $\mathcal{P}^n$ and update Γ^n by leveraging the above matrix $\mathcal{Q}^n$ according to problem (33),
Return: Optimal slicing ratio matrix $\{\mathcal{P}^*, \mathcal{Q}^*\}$, the optimal objective value Γ^* and the reward r according to Eq. (12).

TABLE II
DETAILED SIMULATION PARAMETERS

Parameters	Value	Parameters	Value
num V2I users	30	num V2V users	10
P_M	40 dBm	P_{Φ_b}	23 dBm
$Q_{i,M}$	15 dBm	Q_{i,Φ_b}	15 dBm
α_M, α_k	3	vehicle speed	72 km/h
W^{UL}	20 MHz	W^{DL}	20 MHz
D_{max}	10 ms	ε	1 ms
λ_h	20 packet/s	λ_d	5 packet/s
L_h	9000 bit	L_d	2000 bit

the uplink spectrum bandwidth scheduling ratios $\mathcal{P}$ and downlink spectrum scheduling ratios $\mathcal{Q}$ are alternately optimized, by solving problem (32) and (33) correspondingly, while keeping another block of variables fixed. Furthermore, the obtained solution in each iteration is used as the input of the next iteration. The details of this algorithm are summarized in Algorithm 2. It is worth pointing out that in the classical block coordinate descent method, the sub-problem for updating each block of variables is required to be solved exactly with optimality in each iteration to guarantee the convergence [37].

V. PERFORMANCE EVALUATION

Our simulation implementation is divided into two parts, e.g., the SAC reinforcement learning is implemented in Python 3.6.10 and PyTorch 1.4.0, while the optimization algorithm of ASRS is based on CVXPY 1.0.31. In terms of hardware, we utilize the 1.60GHz intel(R)Core(TM) i5 CPU with 256GB memory, and the RAM is 8GB 2133MHz.The important modeling parameters are listed in Table II.

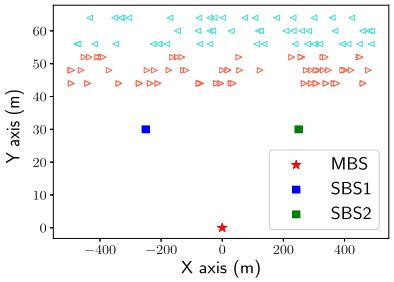


Fig. 4. The two-way six-lane free-way. There is a macro BS and two small BSs.

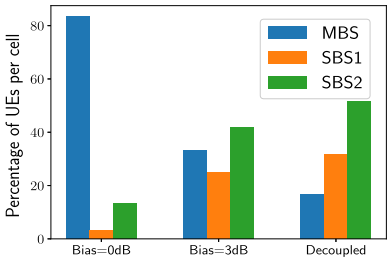


Fig. 5. Percentage of users accessing each BS when the bias = 0 dB, 3 dB and decoupled access.

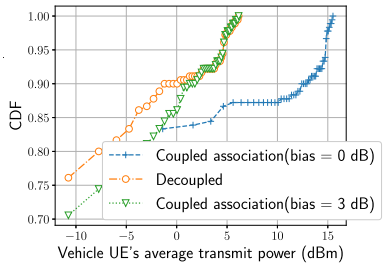


Fig. 6. The CDF of vehicle transmission power when the bias = 0 dB, 3 dB and decoupled access.

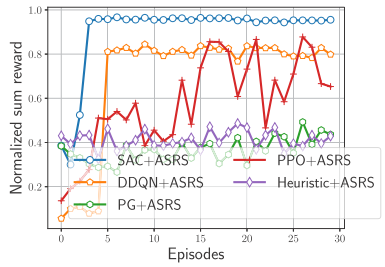


Fig. 7. The learning effect of SAC compared with PPO, DDQN, PG, and Heuristic algorithm.

A. Evaluation of Decoupled Access

We first evaluate the performance of decoupled access. As shown in Fig. 4, the simulation scenario is set up according to 3GPP TR 36.885 [5]. We consider the classic 1000-meters-long road with 6 lines and 2 directions. In this scenario, there is a heterogeneous cellular network with an MBS and two SBSs. The MBS is 40 meters away from the road, and the MBS is set to the origin of coordinates. Also, the two SBSs are located close to the road (set the distance as 10 meters away from the road), and the coordinates of the two SBSs are (−250m, 30m) and (250m, 30m). The vehicles are located on the roads according to the spatial Poisson process and the vehicle density is determined by the vehicle speed [30]. The H V2I users and D pairs of RAC-V2V users are randomly selected from the generated vehicles.

Fig. 5 shows the performance comparison of the decoupled access, the traditional access (bias = 0 dB), and the bias-access, which is commonly used in LTE (e.g., bias = 3 dB means that only when the transmit power of macro is 3 dB higher than the small BSs that vehicles choose to access macro BS). However, the decoupled access performs better in offloading UL traffic to small BSs in contrast to traditional access and bias manner. This is meaningful to macro BS which suffers the heavy burden in both UL and DL.

Fig. 6 shows the cumulative distribution function (CDF) curve of transmit power for vehicle users using decoupled, traditional access (bias = 0 dB) and the bias access (bias = 3 dB). Here we use channel inversion uplink power compensation for vehicle users, and the minimum received power is −106dBm. Obviously, it can be seen that transmission power CDF of decoupled access is located on the left, which means that the transmit power is lower compared with traditional and bias access. It is worth mentioning that the CDF curve of decoupled access is not significantly better than that of bias access because the distance between the vehicles on the Y-axis is relatively small.

B. Evaluation of Two-Tier Slicing Algorithm

Since the distribution of vehicle flow is stable in a certain period time [11], it is reasonable to assume that the number of vehicles that conduct V2I and RAC-V2V communications in each time slot of a period time is subject to Poisson distribution. Here we assume that the mean of Poisson process of the V2I and the RAC-V2V communications are 30 and 10 pairs, respectively.

The simulation of the proposed two-tier slicing framework is shown in detail below. First, in terms of the parameters of SAC neural networks, we chose the Adam optimizer, the learning rate of NN is set to 1e-4 (including 2 value networks, 2 Q-networks, and 1 policy network), and the discount factor γ is set to 0.999. Besides, the size of the replay buffer is 1000, the hidden layer of all neural networks is set to 3 layers, each layer has 128 neurons. At last, we chose Relu as the activation function of the neural networks.

In Fig. 7, we train 30 episodes, each episode has 100 steps. The X-axis Fig. 7 represents the value of the normalized capacity in each episode through the training process. Further, we give the normalized capacity curve through the training process of different DRL algorithms, including off-policy discrete action-space double deep Q-learning (DDQN), on-policy continuous action-space policy gradient (PG), and on-policy actor-critic continuous action-space PPO algorithm. We can see the excellent performance of SAC in improving the network capacity and better stability in contrast to other listed DRL algorithms.

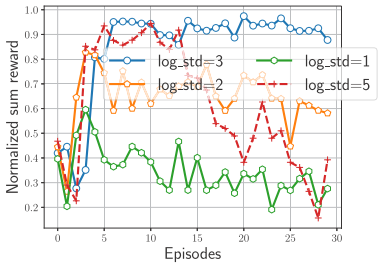


Fig. 8. The learning effect between different action variance when the clipped variance is 1, 2, 3, and 5.

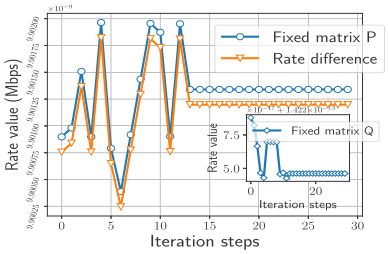


Fig. 9. The sum rate-difference with iteration steps of each sub-problem in proposed ASRS algorithm.

The outputs of the policy network in SAC are the mean and variance of Gaussian distribution. The actions defined as UL/DL slicing ratios are sampled according to the mean and variance of the Gaussian distribution obtained from the policy network. In the actual simulation, we limit the maximum and minimum variance values of the policy network. Since the slicing ratios are limited to $(0, 1)$, the maximum and minimum variance values will have a huge impact on the sampled actions. So a proper variance is crucial. Here we compare the effects of four cases of variances, namely, logarithmic standard (logstd) on the total normalized reward in each episode. As can be seen from Fig. 8, when the variance is equal to 3, the SAC algorithm can effectively converge to a higher reward and has more stability. When the variance is large (i.e., logstd $= 5$), it can quickly learn the optimal action. However, because the variance is too large, it is easy to cause the sampling actions to move away from the optimal slicing ratios, resulting in performance degradation. When the variance is small (i.e., logstd $= 1$), because of the small change of action, it is easy to cause the action to fall into the local optimal, to obtain a relatively poor learning effect.

In Fig. 9, we draw the iteration process of the proposed ASRS algorithm, including the blue, yellow line in the main picture. The blue line in the main-picture represents the iterative process of subproblem 32. The blue line in the sub-picture shows the iterative process of subproblem 33. Moreover, the yellow line represents the value of the objective function value of subproblem 31. Fig. 9 shows that the ASRS algorithm has converged when iterating at about the 15th step. It is worth mentioning that after one optimization cycle, the value of the objective function is reduced to 1e-9.

We also compare the ASRS algorithm with traditional resources allocation algorithms, such as the round-robin (RR) and best channel quality indicator (BCQI). Fig. 10 shows the CDF curve of the total uplink and downlink rate when using the ASRS, RR, and BCQI algorithms. It can be seen that the UL/DL rates of ASRS are almost identical. However, from the distribution, we can see that the proposed ASRS algorithm sacrifices the total throughput of the network while obtaining a very low rate difference.

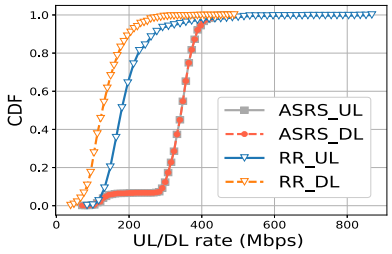


Fig. 10. CDF of UL/DL sum rate of ASRS compared to RR bandwidth allocation method.

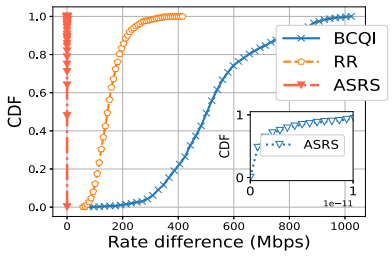


Fig. 11. The CDF comparison of UL/DL rate difference using ASRS, RR, and BCQI algorithm.

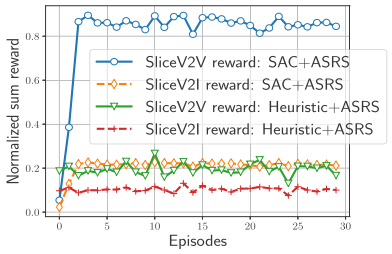


Fig. 12. The comparison of sum rate of SAC and heuristic algorithm with ASRS on different slice.

Fig. 11 shows the CDF of the UL/DL rate difference of ASRS, RR, BCQI, respectively. It can be seen that ASRS has stronger robustness and better performance than RR and BCQI. Almost every iteration, The RAC-V2V QoS metric can be reduced to a minimum value.

The two logical networks, namely the V2I slice, and RAC-V2V slice are shown in Fig. 12, we compare the proposed two-tier slicing algorithm with the heuristic algorithm. In different slices (i.e., V2I slices and RAC-V2V slices), it can be seen that the normalized reward brought by our proposed slicing algorithm is always higher than that of the heuristic with ASRS algorithm in the V2I slice and RAC-V2V slice.

Fig. 13 shows one episode of the stable training. It can be seen that, in one episode, the reward value of our two-tier slicing algorithm is higher than that of the SAC+RR

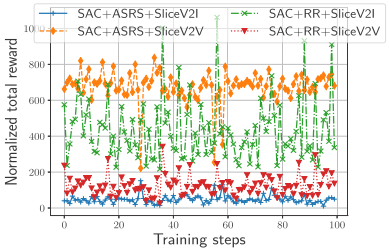


Fig. 13. The comparison of sum capacity of SAC+ASRS and SAC+RR in different slice.

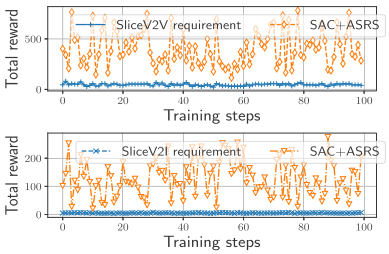


Fig. 14. The comparison of SAC+ASRS algorithm and slice capacity requirements on UL/DL.

algorithm. Specifically, the sum capacity of the RAC-V2V slice is significantly higher than that of V2I slice, and the reason is that we need more capacity/bandwidth to ensure low-latency communications of RAC-V2V. In Fig. 14, we can see that our two-tier slicing algorithm can cope with the dynamic changes in different QoS requirements. More specifically, the capacity of V2I and RAC-V2V slice is always higher than the rate demand of the corresponding slices.

VI. Conclusion

In this paper, we have investigated the feasibility of uplink/downlink decoupled RAN framework for cellular V2X communications, which can play a vital role in offloading the traffic from MBS to SBSs and decreasing uplink transmission power of vehicle users. Under the framework, we have detailedly analyzed the characteristics of relay-assisted cellular-V2V communications and innovatively proposed the RAC-V2V QoS metric. To fulfill the different QoS requirements of V2I and RAC-V2V slice, we have proposed a two-tier RAN slicing solutions including the reinforcement learning SAC algorithm on the first tier and ASRS derived from the block descent algorithm on the second tier. Compared with other algorithms, the performances of SAC algorithm have been verified to be significantly better than their other counterparts. In addition, the ASRS algorithm has been proved to quickly converge to a minimum value, and it has been theoretically proved to reach a stable point. In the future, we will jointly consider the traffic load forecasting and network slicing issues in UL/DL decoupled 5G and beyond cellular V2X networks.

References

[1] W. Liu, S. Jin, M. Matthaiou, and X.-H. You, "Transmission scheme and performance analysis of multi-cell decoupled heterogeneous networks," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4423–4436, Jul. 2020.

[2] S. Zhou, T. Zhao, Z. Niu, and S. Zhou, "Software-defined hyper-cellular architecture for green and elastic wireless access," *IEEE Commun. Mag.*, vol. 54, no. 1, pp. 12–19, Jan. 2016.

[3] Z. Sattar, J. V. C. Evangelista, G. Kaddoum, and N. Batani, "Spectral efficiency analysis of the decoupled access for downlink and uplink in two-tier network," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4871–4883, May 2019.

[4] F. Boccardi *et al.*, "Why to decouple the uplink and downlink in cellular networks and how to do it," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 110–117, Mar. 2016.

[5] *Technical Specication Group Radio Access Network: Study LTE-Based V2X Services: (Release 14)*, document 3GPP TR 36.885, 3GPP, 2016.

[6] X. Peng, H. Zhou, B. Qian, K. Yu, F. Lyu, and W. Xu, "Enabling security-aware D2D spectrum resource sharing for connected autonomous vehicles," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3799–3811, May 2020.

[7] K. A. Hafeez, L. Zhao, B. Ma, and J. W. Mark, "Performance analysis and enhancement of the DSRC for VANET's safety applications," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3069–3083, Sep. 2013.

[8] B. Qian *et al.*, "Leveraging dynamic Stackelberg pricing game for multi-mode spectrum sharing in 5G-VANET," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 6374–6387, Jun. 2020.

[9] H. Zhou, W. Xu, J. Chen, and W. Wang, "Evolutionary V2X technologies toward the internet of vehicles: Challenges and opportunities," *Proc. IEEE*, vol. 108, no. 2, pp. 308–323, Feb. 2020.

[10] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 94–100, May 2017.

[11] H. Zhou *et al.*, "WhiteFi infostation: Engineering vehicular media streaming with geolocation database," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2260–2274, Aug. 2016.

[12] X. Shen *et al.*, "AI-assisted network-slicing based next-generation wireless networks," *IEEE Open J. Veh. Technol.*, vol. 1, pp. 45–66, 2020.

[13] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, Jul. 2003.

[14] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Downlink and uplink decoupling: A disruptive architectural design for 5G networks," in *Proc. IEEE Global Commun. Conf.*, Dec. 2014, pp. 1798–1803.

[15] M. A. Lema, E. Pardo, O. Galinina, S. Andreev, and M. Dohler, "Flexible dual-connectivity spectrum aggregation for decoupled uplink and downlink access in 5G heterogeneous systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 2851–2865, Nov. 2016.

[16] M. Bacha, Y. Wu, and B. Clerckx, "Downlink and uplink decoupling in two-tier heterogeneous networks with multi-antenna base stations," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2760–2775, May 2017.

[17] M. Chen, W. Saad, and C. Yin, "Echo state networks for self-organizing resource allocation in LTE-U with uplink–downlink decoupling," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 3–16, Jan. 2017.

[18] R. Su *et al.*, "Resource allocation for network slicing in 5G telecommunication networks: A survey of principles and models," *IEEE Netw.*, vol. 33, no. 6, pp. 172–179, Nov./Dec. 2019.

[19] Q. Ye, W. Zhuang, S. Zhang, A.-L. Jin, X. Shen, and X. Li, "Dynamic radio resource slicing for a two-tier heterogeneous wireless network," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9896–9910, Oct. 2018.

[20] V. Sciancalepore, K. Samdanis, X. Costa-Perez, D. Bega, M. Gramaglia, and A. Banchs, "Mobile traffic forecasting for maximizing 5G network slicing resource utilization," in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, May 2017, pp. 1–9.

[21] Y. Wei, F. R. Yu, M. Song, and Z. Han, "Joint optimization of caching, computing, and radio resources for fog-enabled IoT using natural actor–critic deep reinforcement learning," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2061–2073, Apr. 2019.

[22] K. Xiong, S. Leng, J. Hu, X. Chen, and K. Yang, "Smart network slicing for vehicular fog-RANs," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3075–3085, Apr. 2019.

[23] S. Zhang, W. Quan, J. Li, W. Shi, P. Yang, and X. Shen, "Air-ground integrated vehicular network slicing with content pushing and caching," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 2114–2127, Sep. 2018.

[24] H. Khan, S. Samarakoon, and M. Bennis, "Enhancing video streaming in vehicular networks via resource slicing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 3513–3522, Apr. 2020.

[25] F. Wang, J. Xu, and S. Cui, "Optimal energy allocation and task offloading policy for wireless powered mobile edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2443–2459, Apr. 2020.

[26] A. Aijaz, "Hap—SliceR: A radio resource slicing framework for 5G networks with haptic communications," *IEEE Syst. J.*, vol. 12, no. 3, pp. 2285–2296, Sep. 2018.
[27] H. D. R. Albonda and J. Pérez-Romero, "An efficient RAN slicing strategy for a heterogeneous network with eMBB and V2X services," *IEEE Access*, vol. 7, pp. 44771–44782, 2019.
[28] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1889–1897.
[29] F. Fu, Y. Kang, Z. Zhang, F. R. Yu, and T. Wu, "Soft actor–critic DRL for live transcoding and streaming in vehicular fog-computing-enabled IoV," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1308–1321, Feb. 2021.
[30] L. Liang, G. Y. Li, and W. Xu, "Resource allocation for D2D-enabled vehicular communications," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 3186–3197, Jul. 2017.
[31] C. Huang, R. Zhang, and S. Cui, "Optimal power allocation for outage probability minimization in fading channels with energy harvesting constraints," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 1074–1087, Feb. 2014.
[32] J. Tang and X. Zhang, "Cross-layer resource allocation over wireless relay networks for quality of service provisioning," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 645–656, May 2007.
[33] C. Guo, L. Liang, and G. Y. Li, "Resource allocation for low-latency vehicular communications: An effective capacity perspective," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 905–917, Apr. 2019.
[34] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," 2018, *arXiv:1801.01290*. [Online]. Available: https://arxiv.org/abs/1801.01290
[35] T. Haarnoja, H. Tang, P. Abbeel, and S. Levine, "Reinforcement learning with deep energy-based policies," 2017, *arXiv:1702.08165*. [Online]. Available: https://arxiv.org/abs/1702.08165
[36] Q. Shi, H. Sun, S. Lu, M. Hong, and M. Razaviyayn, "Inexact block coordinate descent methods for symmetric nonnegative matrix factorization," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5995–6008, Nov. 2017.
[37] M. Hong, M. Razaviyayn, Z.-Q. Luo, and J.-S. Pang, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 57–77, Jan. 2016.

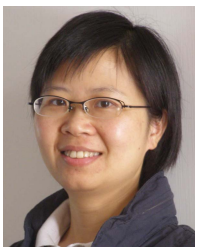
Kai Yu (Graduate Student Member, IEEE) received the B.S. degree in detection, guidance, and control technology from the University of Electronic Science and Technology of China, Chengdu, China, in 2019. He is currently pursuing the Ph.D. degree with the School of Electronic Science and Engineering, Nanjing University, China. His research interests include resource allocation, machine learning for wireless communications, and heterogeneous networks.

Haibo Zhou (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. From 2014 to 2017, he was a Post-Doctoral Fellow with the Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo. He is currently an Associate Professor with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. His research interests include resource management and protocol design in vehicular *ad-hoc* networks, cognitive networks, and space-air-ground integrated networks. He served as a TPC Member of many IEEE conferences, including GLOBECOM, ICC, and VTC. He was a recipient of the 2019 IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award. He served as the Invited Track Co-Chair of ICCC'2019 and VTC-Fall'2020. He also served as an Associate Editor for IEEE Comsoc technically co-sponsored journal: the *Journal of Communications and Information Networks* (JCIN) from April 2017 to March 2019 and a Guest Editor for the *IEEE Communications Magazine* in 2016, the *International Journal of Distributed Sensor Networks* (Hindawi) in 2017, and *IET Communications* in 2017. He is currently an Associate Editor of IEEE INTERNET OF THINGS JOURNAL, *IEEE Network*, and IEEE WIRELESS COMMUNICATIONS LETTER.

Zhixuan Tang (Member, IEEE) received the B.S. degree in communication engineering from Nanjing University, Nanjing, China, in 2020, where she is currently pursuing the M.S. degree with the School of Electronic Science and Engineering. She mainly focuses on the dynamic resource management and networking optimization in the field of emerging wireless networks.

Xuemin (Sherman) Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His researches focus on network resource management, wireless network security, the Internet of Things, 5G and beyond, and vehicular networks.

He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, a Chinese Academy of Engineering Foreign Member, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society. He received 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, the Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) in 2019, the James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, the Joseph LoCicero Award in 2015 and Education Award in 2017 from the IEEE Communications Society (ComSoc), and the Technical Recognition Award from Wireless Communications Technical Committee in 2019 and AHSN Technical Committee in 2013. He has also received the Excellent Graduate Supervision Award from the University of Waterloo in 2006 and the Premier's Research Excellence Award (PREA) from the Province of Ontario, in 2003. He served as the Technical Program Committee Chair/Co-Chair of IEEE Globecom'16, IEEE Infocom'14, IEEE VTC'10 Fall, and IEEE Globecom'07, and the Chair of the IEEE ComSoc Technical Committee on Wireless Communications. He was the Vice President of Technical and Educational Activities, the Vice President of Publications, the Member-at-Large on the Board of Governors, the Chair of the Distinguished Lecturer Selection Committee, and a member of IEEE Fellow Selection Committee of the ComSoc. He is the President Elect of the IEEE ComSoc. He served as the Editor-in-Chief of the IEEE INTERNET OF THINGS JOURNAL, *IEEE Network*, and *IET Communications*.



Fen Hou (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, Canada, in 2008. She is currently an Associate Professor with the State Key Laboratory of IoT for Smart City and the Guangdong-Hong Kong-Macao Joint Laboratory for Smart Cities, Department of Electrical and Computer Engineering, University of Macau. Her research interests include resource allocation intelligent computing networks, mechanism design, and optimal user behavior in crowd sensing networks. She was a co-recipient of the IEEE Globecom Best Paper Award in 2010, the Distinguished Service Award at the IEEE MMTC in 2011, and IEEE VTC-Fall Best Student Paper Award in 2021. She served as the TPC Chair and Co-Chair of several IEEE conferences, such as ICCS 2014, INFOCOM 2014, ICCC 2015, ICC 2016, and ICCC 2021. She currently serves as an Associate Editor for *IET Communications*.