IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 17, 2022, accepted April 16, 2022, date of publication April 20, 2022, date of current version May 2, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3168986

Machine Learning Enables Radio Resource Allocation in the Downlink of Ultra-Low Latency Vehicular Networks

XINYUAN WANG[1], YINGZE WANG[1], QIMEI CUI[1], (Senior Member, IEEE), KWANG-CHENG CHEN[2], (Fellow, IEEE), AND WEI NI[3], (Senior Member, IEEE)

[1]National Engineering Laboratory for Mobile Network Technologies, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Department of Electrical Engineering, University of South Florida, Tampa, FL 33620, USA
[3]Digital Productivity and Services Flagship, Commonwealth Scientific and Industrial Research Organization (CSIRO), Marsfield, NSW 2122, Australia

Corresponding author: Qimei Cui (cuiqimei@bupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61971066, in part by the Joint Funds for Regional Innovation and Development of the National Natural Science Foundation of China under Grant U21A20449, in part by the Ministry of Education-China Mobile Research Foundation under Grant MCM 2020, and in part by the National Youth Top-notch Talent Support Program.

ABSTRACT Autonomous driving and intelligent transportation demand ultra-low latency and high reliability communication in future vehicular networks. Proactive wireless communication can facilitate minimal latency by open-loop communication, which discards traditional feedback control mechanisms. However, appropriate radio resource allocation in such proactive mobile networks has not been fully studied due to lacking channel state information (CSI) and the alleviation of multiple access interference (MAI) in multiple virtual cells. This paper aims to ensure the reliability of downlink communication by a novel radio resource allocation scheme in proactive vehicular networks with ultra-low latency. We regard data transmission success rate as the reliability indicator and propose a joint radio resource allocation model based on the ''generalized closed-loop'', where anchor node (AN) uses the radio resource utilization information (RRUI) from the vehicle in the immediate past uplink as a guide to assist resource allocation. Subsequently, we study the radio resource allocation model solution on the vehicle side and the network side respectively. On the vehicle side, vehicles use the local or global data transmission experience to select the radio resource with the best quality as the RRUI. On the network side, according to the latest RRUI of vehicle and resource occupancy information, deep reinforcement learning is proposed to make appropriate radio resource allocation decisions. Simulations demonstrate the effectiveness of the intelligent joint radio resource allocation scheme under the cooperation between vehicles and AN. When the resource load rate reaches 40%, the joint radio resource allocation scheme can achieve a data transmission success rate of more than 98%.

INDEX TERMS Proactive network, resource allocation, ultra-low latency, deep reinforcement learning, vehicular networks, 6G.

I. INTRODUCTION

Vehicle-to-Everything (V2X), including vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-pedestrian (V2P), vehicle-to-network/cloud (V2N/V2C) connections [1], emerges as an integral part of vehicular communication networks. The V2X supports the automation and intelligence of vehicles, promotes the safety and energy-efficiency of intelligent transportation systems [2]. The V2X applications which enhance road safety and autonomous driving in vehicular networks have extremely high requirements of latency and reliability [3]. Ultra-reliable and ultra-low latency communication is critical for V2X to carry out many artificial intelligence tasks [4].

Rich works have explored the realization and improvement of URLLC network performance from the following three aspects: (1) Some focus on enhancing existing technologies [5], such as HARQ [6]–[8], radio access [9], code [10], MIMO [11], grant-free [12]–[14], NOMA [15], short packet transmission [16]. (2) Some restructure the network architecture by introducing fog computing and distributed network technologies, then optimize the uplink

The associate editor coordinating the review of this manuscript and approving it for publication was Salekul Islam.

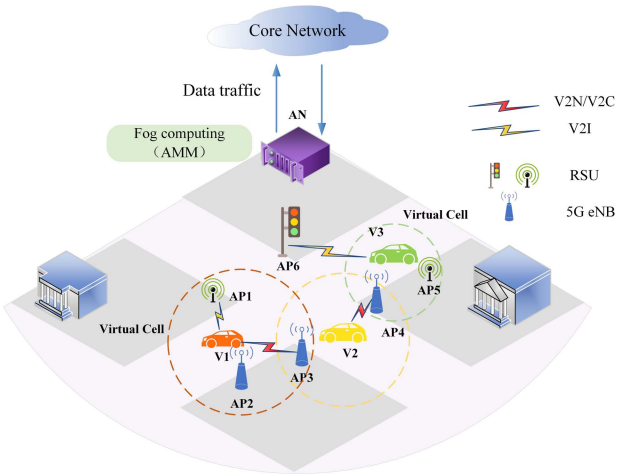


FIGURE 1. **The architecture of the proactive vehicular network. The APs refer to RSUs and 5G eNBs in the vehicular network.**

and downlink transmission, offloading, and online learning issues to reduce latency [17]–[19]. (3) Some consider advanced machine learning and artificial intelligence technologies to empower networks, concentrating on resource allocation [20], scheduling algorithms [21], link adaptation [22], and so on. The above studies are all based on traditional closed-loop networks. However, like vehicles, roadside units (RSUs), and large-scale smart devices access the wireless network, the control signaling overhead is huge. The high-speed movement of vehicles in the network will cause frequent switching with micro base stations, generate signaling storms, and greatly reduce the spectral utilization efficiency.

The proactive network based on open-loop communication novelly realizes ultra-reliable and low-latency vehicular networks [23]. It uses open-loop communication, and there is no complicated feedback mechanism and control signaling between the vehicle and the network, which greatly reduces the transmission latency [24]. As depicted in Fig. 1, the access network of the proactive vehicular network consists of multiple access points (APs) including RSUs and 5G eNBs, and anchor nodes (ANs) which are responsible for managing the APs. Once a vehicle is connected to the wireless network, it will actively select a couple of APs and radio resources for its services to form a virtual cell [25]. To guarantee the reliability of proactive multi-vehicle communication, at the physical layer, each vehicle communicates with multiple APs at the same time to form a multi-path wireless network and achieve macroscopic spatial diversity. Simultaneously, multi-user detection [26] and a specially designed open-loop error correction code [27] can further improve the reliability of the proactive vehicular network. At the network layer, ANs can effectively predict the next location of the vehicle through anticipatory mobility management (AMM) [28], thereby selecting high-quality APs for multipath transmission to improve the success rate of data transmission. However, since it is impossible to know the current network channel quality, the proactive vehicular network can only transmit data by selecting radio resources randomly. It will increase the probability of resource conflicts, leading to a sharp increase in MAI and affecting the correct reception of data. The proactive vehicular network urgently needs an intelligent radio resource allocation scheme to ensure the reliability of data transmission.

The authors of [29], [30] respectively use machine learning and random optimization to solve the problem of uplink resource allocation in proactive networks. Unfortunately, they are not suitable for downlink transmission resource allocation. The initiative of vehicles in the proactive vehicular network leads to different uplink and downlink resource management schemes. During downlink transmission, AN needs to use fog computing and AMM to centrally manage and allocate AP and radio resources. Contrarily, the uplink data transmission is much simpler. Each vehicle independently and actively selects radio resources without waiting for centralized resource management allocation and access control in AN. Currently, there is no effective resource management plan for the downlink. To innovate the downlink radio resource allocation, we must resolve the following challenges for the proactive vehicular network:

- Due to the lack of CSI in the proactive vehicular network without feedback mechanism, how to make reasonable radio resource allocation decisions in the unknown channel state environment is a difficult problem.
- Under the limitation of the ultra-low latency vehicular network, AN needs to make radio resource allocation decisions locally based on fog computing and requires an efficient radio resource allocation algorithm to make decisions quickly.

Existing research that studied resource management in traditional closed-loop low latency networks cannot solve the problem of proactive network downlink resource allocation. Slice resource reservation based on deep reinforcement learning is proposed to realize automated prediction and resource allocation in [31]. A V2V link selection algorithm based on greedy cells is designed in [32], which minimizes the total delivery delay by selecting specific V2V links and assigning appropriate channels. By considering communication factors and changes in vehicular platoon structure, a dynamic manager selection scheme based on joint resource allocation and coding rate optimization algorithms is proposed in [33]. An adaptive fuzzy logic strategy is developed in [34] to formulate rules for services to improve the system resource utilization. However, the method of resource reservation in [31] cannot solve the traffic explosion situation in the resource management of the proactive vehicular network. The researches of [32] and [33] have a single application scenario and cannot be extended to the proactive network. The study of [34] requires a feedback mechanism to provide information and cannot support the proactive network. Unfortunately, none of the existing resource management methods are suitable for downlink communication in the proactive vehicular network. We need to design a new and effective downlink

radio resource allocation scheme to optimize reliability under extremely low latency.

This paper focuses on the downlink transmission radio resource allocation problem in the proactive vehicular network, and aims to design an intelligent and effective downlink radio resource allocation scheme to break the dilemma of the blind selection of radio resources. The main contributions of this paper are summarized as follows:

- We propose a vehicle and AN cooperative radio resource allocation model based on a "generalized closed-loop" in the proactive vehicular network to optimize the success rate of data transmission. Due to lacking complete information about CSI, network management, and centralized coordination of transmissions, traditional radio resource allocation is not feasible anymore for proactive communications. In the process of "generalized closed-loop" downlink data transmission, the additional radio resource utilization information (RRUI) of the immediate past uplink transmission from the vehicle is to guide for AN to make radio resource allocation decisions.
- We make bidirectional optimization of the downlink radio resource allocation model from the vehicle side and the network side. On the vehicle side, we propose two RRUI generation strategies based on local experience (LE-RRUI) and global experience with AN assistance (AA-RRUI), so as to provide the best quality radio resource set to guide the radio resource allocation of downlink transmission. On the network side, a deep reinforcement learning based radio resource allocation algorithm (DRL-RRA) is proposed, which can quickly make reasonable radio resource allocation decisions with the assistance of the RRUI through offline training.
- Through the simulation of different joint schemes of the vehicle-side RRUI generation algorithm and the network-side radio resource allocation algorithm, we obtain the optimal joint radio resource allocation scheme. Simulations prove that under the resource load rate of 40%, AA-RRUI combined DRL-RRA can obtain a data transmission success rate of more than 98%.

This paper is organized as follows: Section II establishes a downlink radio resource allocation model based on "generalized closed-loop" and regards the long-term data transmission success rate as the optimization goal. Section III proposes two RRUI generation strategies on the vehicle side. Section IV proposes DRL-RRA on the network side and offers two benchmark solutions for comparison. Section V provides numerical results to validate the analysis and demonstrate the performance of the proposed radio resource allocation algorithm under the cooperation between vehicles and AN. Finally, conclusions are drawn in Section VI.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. THE ARCHITECTURE OF THE PROACTIVE VEHICULAR NETWORK

We consider a proactive vehicular network as shown in Fig. 1. The vehicular network includes V2I, and V2N communication modes [35]. The network includes the radio access network and the core network, radio access network consists of the APs and the ANs. Multiple ANs manage the access network with sufficient computing power and storage and are directly connected to the core network. Each AN manages APs within a certain range, and ANs' respective management areas do not overlap. "Proactive" in the proactive vehicular network describes the vehicle. When a vehicle accesses the wireless network, it can actively associate with the nearest APs and select radio resources to form a virtual cell and directly perform uplink communication without interacting with the AN in advance, so that the AN perceives the existence of the vehicle. Subsequently, when a downlink data task arrives from the core network or the infrastructure, AN allocates APs and radio resources for downlink transmission with the assistance of AMM. In the network, the distributions of vehicles and APs obey the homogeneous Poisson point process (PPP) with the density of λ_U and λ_B, respectively. Simultaneously, $\mathbb{B} = \{1, 2, \cdots, B\}$ and $\mathbb{U} = \{1, 2, \cdots, U\}$ collect APs and vehicles in the region. In order to enhance the readability of the paper, we summarize the abbreviations and notations respectively in Table 1 and Table 2.

TABLE 1. Glossary of abbreviations.

Abbreviation	Meaning
AN	Anchor Node
AMM	Anticipatory Mobility Management
AP	Access Point
CSI	Channel State Information
DQN	Deep Q-Network algorithm
DRL	Deep Reinforcement Learning
MAI	Multiple Access Interference
MCMF	Maximum Cost Maximum Flow problem
OFDMA	Orthogonal Frequency Division Multiple Access
PPP	Poisson Point Process
RB	Radio Block
RRA	Radio Resource Allocation
RRUI	Radio Resource Utilization Information
RSU	RoadSide Unit
RU	Radio Unit
SIR	Signal-to-Interference Ratio
SPFA	Shortest Path Faster Algorithm
V2X/V2V/V2I	Vehicle-to-Everything/Vehicle/Infrastructure
V2P/V2N/V2C	Vehicle-to-Pedestrian/Network/Cloud

B. RESOURCES IN THE SYSTEM

The resources in the proactive vehicular network are divided into radio resources and network resources. Orthogonal frequency division multiple access (OFDMA) is adapted for the physical layer transmission, where the radio resources are divided into multiple radio units (RUs), each containing a fixed number of subcarriers and symbols in the slot. The specific number of adjacent RUs are mapped into the link layer as the radio block (RB), which is defined as the basic scheduling element of the radio resources with the data transmission capacity of l. We denote the set of RBs by $\mathbb{J}$ and describe the quality of an RB by the data transmission

TABLE 2. Glossary of notations.

Notation	Description
$\lambda_U, U, \mathbb{U}$	Density, number and set of vehicles
$\lambda_B, B, \mathbb{B}$	Density, number and set of APs
$N_b, \mathbb{J}$	Number and set of RBs
N_s	Number of RUs
Π^j	The set of RUs that RBj mapped into the physical layer
l	The transmission capacity of every RB
$C_K^{t,u}$	The virtual cell of vehicle u at slot t
$V_K^{t,u}$	The set of K APs closest to vehicle u at slot t
$\Lambda^{t-\tau,u}$	The set of RBs with the highest communication quality (RRUI), which vehicle u reported to AN through its last uplink transmission at slot $t-\tau$
Φ^t, $\Phi^{t,u}$	The set of data tasks at slot t and one of the data task in it for vehicle u
$R^{t,u}$	Transmission rate requirement of $\Phi^{t,u}$
Ψ^t	The set of receiving vehicles for Φ^t at slot t
$n^{t,u}$, $\Gamma^{t,u}$	Number and set of RBs transmitted for data task $\Phi^{t,u}$
N_v	Number of RBs in RRUI
p	Transmit power of each RU in each AP
$h_b^{t,u}$	The complex channel coefficient of the transmission link from AP b to vehicle u at slot t
$g_b^{t,u}$	The small-scale fading experienced between AP b and vehicle u at slot t
$D_b^{t,u}$	The distance between AP b and vehicle u at slot t
$\gamma_{j,k}^{t,u}$	The SIR occupied RB j over the kth link in $V_K^{t,u}$
S^t	The RUs occupation matrix at slot t
C_s	Maximum number of RUs can be occupied by each AP in a slot
ρ	The long-term data transmission success rate
$O(t)$	Number of data tasks arriving at AN in slot t
$\delta(t,u)$	Whether data task $\Phi^{t,u}$ transmission is successful
γ_{th}	Minimum SIR threshold
$\mathbb{J}_l^{t-\tau,u}$	The communication quality of RBs based on vehicle u's local experience statistics at slot $t-\tau$
$\mathbb{J}_a^{t-\tau-\tau',u}$	The communication quality of RBs based on the AN global experience statistics at slot $t-\tau-\tau'$
ε_v	Probability of selecting RB randomly to form RRUI
ϖ	Probability of selecting high-quality RB from $\mathbb{J}_a^{t-\tau-\tau',u}$ to form RRUI
$M[i,n]$	The capacity of the edge from i to n in the network flow topology
$H[i,n]$	The cost of the edge from i to n in the network flow topology
$\mu_j^{t,u}$	Whether the transmission data $\Phi^{t,u}$ occupying the RB j can be successfully received
S^t,A^t,$R^t(s,a)$	State, action, reward of AN at slot t
β, φ	Learning rate and discount factor in DQN
ε_a	ε of $\varepsilon-greedy$ when AN selects action in DQN

success rate of the RB in a period of time. The partition of radio resources is illustrated in Fig. 2. This paper maps the physical layer radio resources into the link layer, AN makes allocation decisions in the link layer. The network resources refer to the APs managed by the AN.

The communication quality of the vehicle u is ensured by forming a virtual cell centered on itself, and we use $C_K^{t,u}$ to represent the virtual cell. The virtual cell $C_K^{t,u}$ includes the sets of network resource and radio resource that provide multipath communication for vehicle u, $C_K^{t,u} = (V_K^{t,u}, \Lambda^{t-\tau,u})$. Network resource $V_K^{t,u}$ corresponds to the set of K APs closest to the vehicle u (see Fig. 1). $\Lambda^{t-\tau,u}$ is RRUI, a set of RBs with

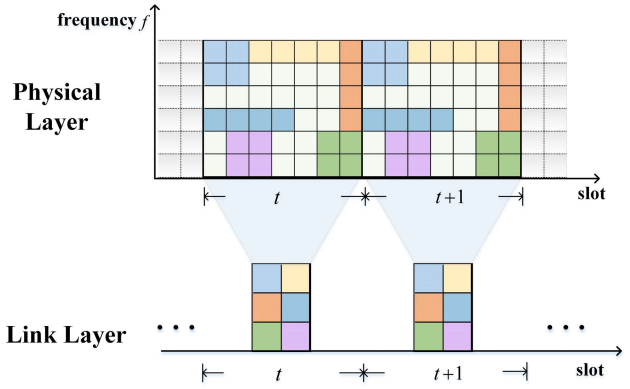


FIGURE 2. The logical relationship between the physical layer RU and the link layer RB.

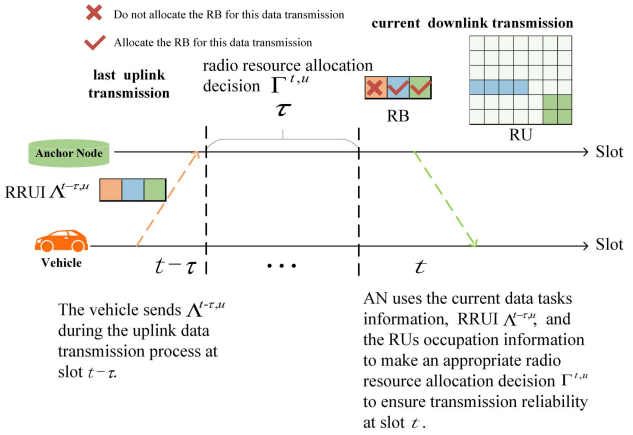


FIGURE 3. Downlink radio resource allocation scheme. RRUI is a set of RBs selected by vehicle u at slot $t-\tau$.

the highest communication quality obtained by the vehicle u through experience, which was reported to the AN at last uplink transmission time $t-\tau$.

C. RADIO RESOURCE ALLOCATION BASED ON "GENERALIZED CLOSED-LOOP"

Since the proactive vehicular network reduces the communication latency by discarding the feedback mechanism, the AN cannot directly or indirectly obtain the current wireless channel state. Through a more in-depth study in the timing operation process of the proactive vehicular network communication, we note a "generalized closed-loop" in a macroscopic and delayed manner (see Fig. 3). Real-time digital map updates and exchange of control information in autonomous driving will induce dense and frequent uplink and downlink communications between the vehicle and the AN [36], which suggests the possibility of using the "generalized closed-loop" framework for radio resource allocation.

As shown in Fig. 3, when there is uplink data to be sent at slot $t-\tau$, the vehicle u selects N_v BRs as RRUI $\Lambda^{t-\tau,u}$, which according to a certain strategy (discussed in section III), and sends it to the AN with data together. The vehicle u then

monitors the RRUI to receive downlink data. At slot t, the AN intelligently allocates appropriate RBs in RRUI $\Lambda^{t-\tau,u}$ as $\Gamma^{t,u}$ for downlink transmission based on the latest network state. Note that the reason why AN only selects radio resources in RRUI $\Lambda^{t-\tau,u}$ is: 1) The proactive vehicular network has no feedback so that the vehicle cannot be informed of the transmission channel in time. 2) The complexity and latency of the resource allocation algorithm can be reduced by narrowing the radio resource selection space.

Since the network state changes rapidly, it is necessary to ensure the timeliness of RRUI. If the vehicle informs AN RRUI at time t_a, the information will remain unchanged during $min\{t'_a - t_a, \theta\}$ period, where t_a is the next uplink transmission time, and the vehicle will update RRUI at this time. When the vehicle has not updated the RRUI for a long time, it will be forced to update if the time limit θ is exceeded. The AN can realize cooperative intelligent downlink radio resource management with the interaction in "generalized closed-loop" communication composed of uplink and downlink.

D. DOWNLINK TRANSMIT MODEL

The number of data tasks is random in every slot. At slot t, the set of data tasks is Φ^t, where $\Phi^t = \left\{\Phi^{t,u} \middle| u \in \Psi^t\right\}$. Ψ^t is the collection of receiving vehicles for the data tasks at slot t. When a downlink data task $\Phi^{t,u}$ arrives at slot t, the AN allocates the K nearest APs to vehicle u to combine $V_K^{t,u}$ and make radio resource decision $\Gamma^{t,u}$ for transmission. The transmission model can be viewed as the MISO process. The APs in $V_K^{t,u}$ send the same packet synchronously to the vehicle u with each RB inside $\Gamma^{t,u}$.

The complex channel coefficients of the transmission link from AP b to vehicle u at slot t can be expressed as:

$$h_b^{t,u} = g_b^{t,u}\sqrt{\|D_b^{t,u}\|^{-\alpha}}, \tag{1}$$

where $g_b^{t,u}$ is the small-scale fading experienced between AP b and vehicle u, and obeys the Rayleigh distribution. $\|D_b^{t,u}\|^{-\alpha}$ is the path loss between AP b and vehicle u, where $D_b^{t,u}$ is the distance, and α is the path loss exponent ranging from 2 to 5 [37].

Assuming that there are N_s RUs at every slot and can be mapped into N_b RBs, as while each RB contains n_m RUs. And RB j is mapped into RUs as a set Π^j. We express the RUs occupation information as a matrix $S^t = \{0,1\}^{B\times N_s}$. $s_{b,m}^t = 1$ means that RU m is occupied by AP b to transmit data at slot t and $s_{b,m}^t = 0$ indicates the corresponding RU is not occupied by AP b.

Assume that every AP can only occupy at most C_s RUs in a slot for data transmission, and the transmit power of every RU is p. Therefore, the signal-to-interference ratio (SIR) $\gamma_{j,k}^{t,u}$ occupied RB j over the kth link in $V_K^{t,u}$ can be mapped into the physical layer, as given by:

$$\gamma_{j,k}^{t,u} = \frac{\left|h_k^t\right|^2 n_m p}{I_{j,k}^{t,u}}$$

$$= \frac{\left|h_k^t\right|^2 n_m p}{\sum_{b\in\mathbb{B}/V_K^{t,u},\quad m\in\Pi^j} s_{b,m}^t \left|h_b^{t,u}\right|^2 p}, \tag{2}$$

where h_k^t is the downlink channel coefficient between the kth selected AP and vehicle u, $h_b^{t,u}$ is the downlink channel coefficient between AP b and vehicle u. The noise is comparatively negligible in the presence of strong inter-virtual cell interference.

In this MISO-equivalent process, we assume that vehicle u adopts the selection combining strategy [38] to obtain the maximum SIR

$$\gamma_j^{t,u} = \max\left[\gamma_{j,1}^{t,u}, \gamma_{j,2}^{t,u}, \cdots, \gamma_{j,K}^{t,u}\right]. \tag{3}$$

Therefore, the downlink data task can be successfully received by vehicle u, under the following condition:

$$\min\{\gamma_j^{t,u} | j \in \Gamma^{t,u}\} \geqslant \gamma_{th}, \tag{4}$$

where γ_{th} is the SIR threshold that the data can be received correctly. Equation (4) means that every RB occupied by the data task can be received correctly.

E. PROBLEM STATEMENT

Since the data traffic in the vehicular network is not stable, the instant data transmission success rate cannot be used as an appropriate indicator to measure the performance of the entire network. We define the long-term data transmission success rate of the entire system as a downlink reliability indicator, denoted by ρ:

$$\rho = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{u\in\Psi^t}\delta(t,u)}{O(t)}, \tag{5}$$

where $\sum_{u\in\Psi^t}\delta(t,u)$ is the number of successful transmissions and $O(t)$ is the number of data tasks from the core network to the AN at slot t.

Therefore, we formulate the optimization problem of downlink data transmission success rate as:

$$\begin{aligned}
&\max_{\Gamma}\ \rho \\
&\text{s.t. } c1: \sum_{m=1}^{N_s} s_{b,m}^t \leq C_s, \quad b = 1, 2, \cdots, B \\
&\qquad c2: \delta(t,u) = \begin{cases} 1, & \min\{\gamma_j^{t,u} | j \in \Gamma^{t,u}\} \geq \gamma_{th}, \\ & n^{t,u} * l \geq R^{t,u}, u \in \Psi^t \\ 0, & else. \end{cases}
\end{aligned} \tag{6}$$

Γ is the set of radio resource allocation decisions for all slots from 1 to T. Condition $c1$ indicates that the maximum number of RUs that each AP occupies at every slot cannot exceed C_s. Condition $c2$ defines the requirements for successful data transmission in downlink communication. $n^{t,u}$ and $R^{t,u}$ are the number of RBs occupied for transmission and the rate requirement of $\Phi^{t,u}$, respectively.

When we fix the time variable, for each slot, (6) can be transformed into the famous traveling salesman problem,

Algorithm 1 Local Experience Based RRUI Generation Algorithm (LE-RRUI)

Input: the set of RBs $\mathbb{J}$, the communication quality of RBs based on the local experience statistics $\mathbb{J}_l^{t-\tau,u}$, ε_v of $\varepsilon-greedy$
Output: RRUI $\Lambda^{t-\tau,u}$

Initialization: Initialize $\Lambda^{t-\tau,u}=\emptyset$.
for $i=1:N_v$ **do**
  $x=$ random(0,1).
  if $x>\varepsilon_v$ **then**
    Vehicle selects the highest communication quality RB in $\mathbb{J}_l^{t-\tau,u}/\Lambda^{t-\tau,u}$ as RB i.
  else
    Vehicle randomly selects an RB from $\mathbb{J}/\Lambda^{t-\tau,u}$ as RB i.
  end if
  Add the chosen RB i in $\Lambda^{t-\tau,u}$.
end for

which is a typical NP-hard problem. Therefore (6) is an NP-hard problem. Since Γ is closely related to RRUI, in order to solve (6), we need to study and optimize from the generation of RRUI on the vehicle side and the radio resource allocation of AN on the network side.

III. RRUI GENERATION STRATEGY FOR VEHICLES

In the vehicular-helped downlink radio resource allocation model, since the radio resource decision $\Gamma^{t,u}$ of AN is generated in the RRUI, the generation strategy of the RRUI is significant to achieve the optimal resource allocation decision and improve the data transmission success rate of the overall system. Here we elaborate on the specific operations of RRUI $\Lambda^{t-\tau,u}$.

A. LOCAL EXPERIENCE BASED RRUI GENERATION STRATEGY

When vehicle u initially accesses the proactive network, it performs active network association by sensing nearby APs, and randomly selects radio resources for uplink transmission. Then in the communication process, vehicle u counts the number of successful transmissions of the system RBs within a period of time in local to measure the communication quality of RBs, and selects N_v RBs as RRUI $\Lambda^{t-\tau,u}$ to send to the AN. We use $\mathbb{J}_l^{t-\tau,u}$ to represent the communicating quality of the system RBs based on the local experience statistics of the vehicle u at slot $t-\tau$. In order to ensure the balance between exploration and exploitation, we assume that the vehicle u adopts a $\varepsilon-greedy$ policy when selecting RBs as RRUI. Algorithm 1 is the local experience based RRUI generation algorithm.

B. AN-ASSISTED RRUI GENERATION STRATEGY

There are certain limitations when the vehicle only utilizes local data transmission information to generate RRUI. Due to the high-speed changes of the network, for a single vehicle, the system RBs quality statistics based on local experience update slowly, the timeliness is not high enough and cannot fully represent the current radio resource quality. When we move the focus to the vehicle side of "generalized closed-loop" communication at slot $t-\tau$, AN can guide the vehicle to generate the RRUI with the global information of RBs' quality through the downlink transmission at slot $t-\tau-\tau'$. Fig. 4 is the AN-assisted RRUI generation process. We improve the RRUI generation strategy based on Algorithm 1. Assume that when selecting RBs to generate RRUI, the vehicle has a probability of ϖ to select the best quality RB among the global information of RB quality $\mathbb{J}_a^{t-\tau-\tau',u}$ provided by AN, and a probability of $1-\varpi-\varepsilon_v$ to select RB based on local experience. It should be noted that $1-\varpi-\varepsilon_v>0$, it is unreasonable to only use $\mathbb{J}_a^{t-\tau-\tau',u}$ sent by the AN to generate the RRUI. Because of a delay in $\mathbb{J}_a^{t-\tau-\tau',u}$, even the global information from the network cannot fully represent the current radio resource quality. On the other hand, the local experience of the vehicle is most suitable for the current location and environment, and has reference value. For detailed algorithm steps, please refer to Algorithm 2.

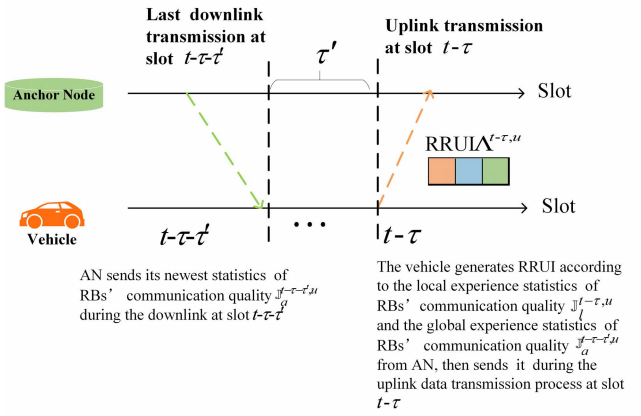


FIGURE 4. AN-assisted RRUI generation process.

The vehicle selects the RBs with good quality through the local and global RBs' communication quality information and sends it to the AN. After that, the vehicle receives downlink data by monitoring the RBs in the RRUI, and updates the local RBs quality statistics by the RBs occupation information in the downlink transmission.

IV. DOWNLINK RADIO RESOURCE ALLOCATION SOLUTIONS FOR AN

After the downlink data task arrives, the AN needs to make an appropriate radio resource allocation decision according to the RRUI provided by the target vehicle, so as to maximize the optimization objective in (6). To the best of our knowledge, there is no proper and mature solution for proactive network downlink radio resource allocation as a reference. In order to verify the effectiveness and pros and cons of DRL-RRA, this paper designs two benchmark solutions worthy of reference.

Algorithm 2 AN-Assisted RRUI Generation Algorithm (AA-RRUI)

Input: the set of RBs $\mathbb{J}$, the communication quality of RBs based on the local experience statistics $\mathbb{J}_l^{t-\tau,u}$, the communication quality of RBs based on the AN global experience statistics $\mathbb{J}_a^{t-\tau-\tau',u}$, ε_v of $\varepsilon-greedy$, probability ϖ of selecting high-quality RB from $\mathbb{J}_a^{t-\tau-\tau',u}$

Output: RRUI $\Lambda^{t-\tau,u}$

Initialization: Initialize $\Lambda^{t-\tau,u}=\emptyset$.
for $i=1:N_v$ **do**
  $x=$ random(0,1).
  if $x<\varepsilon_v$ **then**
    Vehicle randomly selects an RB from $\mathbb{J}/\Lambda^{t-\tau,u}$ as RB i.
  else if $x<\varepsilon_v+\varpi$ **then**
    Vehicle selects the highest communication quality RB in $\mathbb{J}_a^{t-\tau-\tau',u}/\Lambda^{t-\tau,u}$ as RB i.
  else
    Vehicle selects the highest communication quality RB in $\mathbb{J}_l^{t-\tau,u}/\Lambda^{t-\tau,u}$ as RB i.
  end if
  Add the chosen RB i in $\Lambda^{t-\tau,u}$.
end for

A. TWO BENCHMARK SOLUTIONS

Given the system scenario in (6), there exist two immediate realizations. One is the random selection scheme, that is, free selection with equal probability within the range of optional radio resources without considering any other factors. This scheme is intuitive and self-explanatory. If the gains of a radio resource allocation scheme to the system cannot even achieve the same performance as the random selection, the scheme is invalid or even damages the system.

Another realization is the heuristic algorithm which gets an approximate solution to the NP-hard problem. It is difficult to solve the highly dynamic random optimization problem in a variable range, so we divide the time into intervals and transform the dynamic problem into per-slot deterministic optimizations. The optimization goal is changed from ρ to the data transmission success rate in each slot, which is represented by:

$$\max_{\Gamma^{t_0}} \frac{\sum_{u\in\Psi^{t_0}}\delta(t_0,u)}{O(t_0)}, \tag{7}$$

where Γ^{t_0} is the set of radio resource allocation decisions for the data tasks at slot t_0 and $t_0\in\{1,2,\cdots,T\}$.

A data task can select multiple RBs, and each RB can only be selected by one data task at a slot. The problem of many-to-one radio resource matching in a fixed slot fits the network flow model [39]. The network flow topology diagram $G=(V,E,M,H)$ based on radio resource allocation is shown in Fig. 5. Where V is the set of vertices of the network graph, including virtual source and target points SO, TA, $O(t)$ downlink data task points, and N_b RB points. E is the set of directed edges in the network graph. The edge from i to n

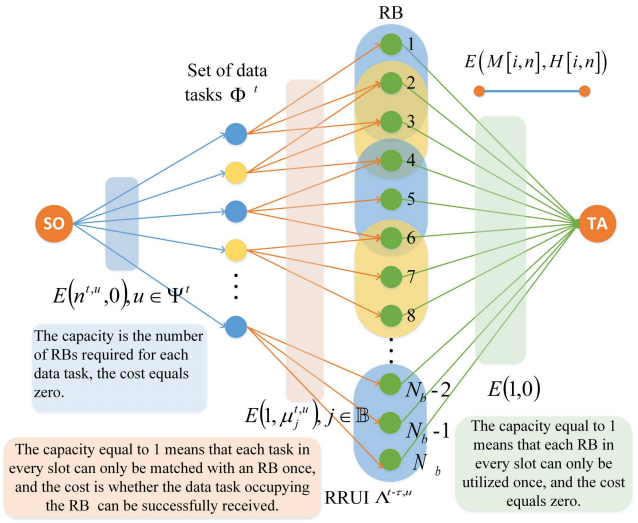


FIGURE 5. Design of network flow topology diagram based on radio resource allocation model.

has attributes $(M[i,n], H[i,n])$, which can be seen distinctly in Fig. 5. $M[i,n]$ is the capacity of the edge, and $H[i,n]$ is the cost per unit flow. In the network flow topology, the path selection from the data tasks to the system RBs is the process of radio resource allocation. The attributes setting of the edges between data tasks and the system RBs need to be specially explained. $M[i,n]=1\left(u\in\Psi^t\right)$ means that each task can only be matched with an RB once, and $H[i,n]=\mu_j^{t,u}\left(u\in\Psi^t, j=1,2,\cdots,N_b\right)$ is whether the transmission data occupying the RB j can be successfully received, where

$$\mu_j^{t,u}=\begin{cases}1, & \gamma_j^{t,u}\geq\gamma_{th}\\ 0, & else.\end{cases} \tag{8}$$

In order to maximize the number of successful data tasks per time slot, we need to set $H[i,n]=0$ for both the edges starting with SO and ending with TA. Finally, the optimization is treated as the maximum cost maximum flow problem (MCMF) to be solved by the shortest path faster algorithm (SPFA).

B. DOWNLINK RADIO RESOURCE ALLOCATION WITH DEEP REINFORCEMENT LEARNING

The radio resource allocation algorithm based on SPFA proposed in this paper has limitations in solving the NP-hard problem, it can only allocate radio resources in a single slot and obtain an approximation of the optimal solution of radio resource allocation. Reinforcement learning, as one of the paradigms and methodologies of machine learning, can utilize existing data in the proactive vehicular network to maximize specific goals in the process of interacting with the environment. This fits well with our dynamic optimization system and solves the radio resource allocation decision-making problem of highly dynamic systems.

1) THE BASIC MODEL OF REINFORCEMENT LEARNING

Consider the AN in the proactive vehicular network as an agent. For the downlink data task $\Phi^{t,u}$, regarding the RUs

occupation matrix S^t at slot t and RRUI $\Lambda^{t-\tau,u}$ at slot $t-\tau$, the AN will make an RB allocation decision and get a corresponding reward through the next uplink transmission. The whole process is a semi-Markov process, with the definitions of the state, action, and reward as follows.

a: STATE

We take the RUs occupation matrix S^t in the network as the environment state. One dimension of the matrix represents the set of all APs controlled by the AN, and the other dimension represents the set of all RBs in the network. Each element in the matrix is a 0-1 variable.

b: ACTION

The action is defined as $A^t = \left[a_1^t, a_2^t, \cdots, a_{N_b}^t\right]$, where $a_j^t = 1$ means RB j is utilized for data transmission, otherwise $a_j^t = 0$. There are 2^{N_b} actions that can be chosen at every step. As N_b increases, the action space will increase exponentially. For $\Phi^{t,u}$, $\Gamma^{t,u}$ can only be selected in RRUI $\Lambda^{t-\tau,u}$, resulting in a large number of unreasonable actions and affecting the learning performance. Therefore, we stipulate that at most N_v RBs can be selected for transmitting during each downlink transmission. In this way, the size of the action space is reduced from 2^{N_b} to $\sum_{n=0}^{N_v} \binom{N_b}{n}$.

It should be particularly noted that when the utilized RB is selected and mapped into the physical layer RUs for transmission, some of the RUs may have been occupied. At this time, the RUs should be discarded, and the operation is called dropout in this algorithm.

c: REWARD

We use whether a downlink task is successfully transmitted as a reward. The reward can be obtained from the subsequent uplink communication of the vehicle. When there is downlink data to be sent, according to the system state and related actions, the reward can be designated as:

$$R^t(s,a) = \begin{cases} 1, & \sum_{m=1}^{N_s} s_{b,m}^t \le C_s, \min\{\gamma_j^{t,u} | j \in \Gamma^{t,u}\} \ge \gamma_{th} \\ 0, & \sum_{m=1}^{N_s} s_{b,m}^t \le C_s, \min\{\gamma_j^{t,u} | j \in \Gamma^{t,u}\} < \gamma_{th} \\ -1, & \sum_{m=1}^{N_s} s_{b,m}^t > C_s, \end{cases} \quad (9)$$

where $s = S^t$,$a = A^t$. At slot t, if the downlink transmission SIR requirement ($min\left\{\gamma_j^{t,u} \middle| j \in \Gamma^{t,u}\right\} \ge \gamma_{th}$) and the AP's occupying number of RUs requirement ($\sum_{m=1}^{N_s} s_{b,m}^t \le C_s$) are met, the data transmission is successful and $reward = 1$. If the SIR requirement cannot be met, the data transmission fails and $reward = 0$. If the AP's occupying number of RUs requirement cannot be met, the action selection is unreasonable and a certain penalty $reward = -1$ is required.

The problem that the reinforcement learning model needs to solve is to find the optimal policy π^*:

$$\pi^* = \arg\max_{\pi} \rho. \quad (10)$$

2) DEEP Q-NETWORK STRATEGY

The Q-learning method has been increasingly exploited for solving the reinforcement learning problem but shows infeasibility for numerous state-action scenarios. When the state-action pair is sufficiently large, traversing all the samples stored in a Q-table at each step is challenging. To overcome the drawbacks of Q-learning, the Deep Q-network algorithm (DQN) is used in this paper.

At step t, AN makes action $A^t = \pi\left(S^t\right)$ through policy π under the current state S^t. State-action value function $Q^\pi\left(S^t, A^t\right)$ is the expected return and can be expressed as

$$Q^\pi(s,a) = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \varphi^k R^{t+k}\left(s = S^t, a = A^t\right)\right], \quad (11)$$

where $\varphi \in [0,1]$ is the discount factor to balance current reward and long-term reward. In the Q-learning algorithm, the evaluation of $Q(s,a)$ is denoted as:

$$Q(s,a) \leftarrow Q(s,a) + \beta\left(R^t(s,a) + \varphi \max_{a'} Q\left(s', a'\right) - Q(s,a)\right), \quad (12)$$

where $\beta \in (0,1]$ is the learning rate. $Q(s,a)$ is recorded in Q-table and AN selects the largest $Q(s,a)$ value as action.

In DQN strategy, in order to deal with large state and action spaces, it uses neural networks to estimate $Q(s,a)$. We define evaluated Q-network $Q(s,a;\omega)$ and target Q-network $Q\left(s,a;\omega'\right)$, and the weights ω of the evaluated network $Q(s,a;\omega)$ are updated according to the target value y_t:

$$y_t = R^t + \varphi \max_{a'} Q\left(s', a'; \omega'\right). \quad (13)$$

The ω' is updated periodically by copying ω, which can remove correlations in the observation sequence.

In order to further improve the stability of agent learning, DQN introduces an experience replay mechanism. The agent stores the experience $e_t = (S^t, A^t, R^t, done, S^{t+1})$ of each step in the experience replay memory buffer, and randomly selects a set of experience samples from it, then trains the network weights through gradient descent to minimize the relevant objective loss function of y in (13).

In the following, we will introduce the training and testing operations of the proposed DRL-based radio resource allocation algorithm on the network side.

3) TRAINING AND TEST

The DRL-based radio resource allocation scheme on the network side has two stages: training and testing. The training stage trains the Q-network by simulating the generation of

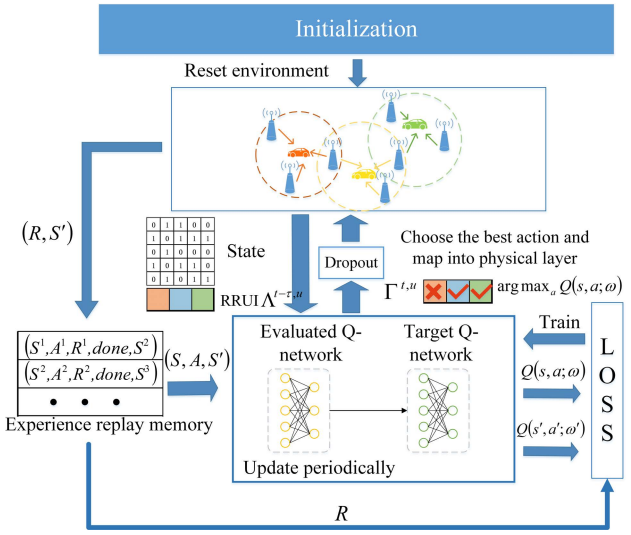


FIGURE 6. DRL-based radio resource allocation framework.

downlink data flow and the interaction between the vehicle and the AN. In the testing phase, the AN first loads the trained Q network parameters ω and ω' and initializes the experience replay memory buffer, and then interacts with the environment. Actions made by the AN will be chosen based on the output of the Q-network with loaded parameters, and states will be generated depending on its local observations [40]. Fig. 6 and Algorithm 3 provide the framework and algorithm procedure of DRL-based radio resource allocation algorithm, respectively.

It should be noted that the main parameters of our proposed DRL-based radio resource allocation model, such as state dimension, action dimension, etc., are only related to the resources (number of APs and RBs) managed by the AN in the system. It is independent of other environmental variables such as the number of vehicles and the arrival of downlink data tasks. In the real environment, the number of APs managed by each AN and the number of radio resources are fixed, which makes this scheme have good adaptability and can be quickly deployed in other ANs in the edge network.

4) LATENCY ANALYSIS OF PROACTIVE NETWORK RADIO RESOURCE ALLOCATION SCHEME BASED ON DRL

How to implement an effective radio resource allocation scheme in the low-latency proactive vehicle network is our concern. In this paper, we optimize the latency from the following aspects:

- Fog computing is used for distributed radio resource allocation on the AN, which reduces the scale and latency of the solution compared to the centralized radio resource management in the core network.
- Aiming at the convergence problem of DQN, we further de-redundancy by deleting illegal actions when setting actions to speed up the convergence of the algorithm.
- Considering that the computational complexity is related to the structure of the Q network in the real deployment, the redundant number of hidden layers of the neural network will increase the computational latency. Therefore, we use two fully connected (FC) layers as the hidden layers of the neural network, which can reduce the computational latency while ensuring fitting accuracy. The specific settings of the neural network are shown in Fig. 7.
- The parameters used by the radio resource allocation scheme based on DRL in the model application are already trained offline and can be directly put into use online in the real environment.

Algorithm 3 DRL-Based Radio Resource Allocation Algorithm (DRL-RRA)

Initialization:
Initialize evaluated Q-network and target Q-network with parameters ω and ω'.
for episode$= 1 : M$ **do**
  Initialize the proactive vehicular network environment.
  for $t = 1 : T$ **do**
    AN receives the set of downlink data tasks Φ^t.
    for $i = 1 : O(t)$ **do**
      AN gets the downlink data task $\Phi^{t,u}$.
      AN senses the current environment state S^i.
      AN makes action A^i according to $\Phi^{t,u}$ and S^i based on $\varepsilon - greedy$ policy.
      AN obtains reward R^i and next state S^{i+1}.
      AN stores $(S^i, A^i, R^i, done, S^{i+1})$ in the experience replay memory.
      Sample random minibatch of experiences $(S^k, A^k, R^k, done, S^{k+1})$ from the experience replay memory.
      if episode terminates at step $k+1$ **then**
        $y_k = R^k$,
      else
        $y_k = R^k + \varphi \max_{A'} Q\left(S^{k+1}, A'; \omega'\right)$.
      end if
      Train evaluated Q-network to minimize $L(\omega)$.
      Every P steps, update target Q-network.
      $S^k \leftarrow S^{k+1}$
    end for
  end for
end for

V. SIMULATIONS

A. SYSTEM SETTINGS

To obtain the numerical results, we use a CPU-based server with 3.70 GHz Intel Core i9-10900k processor and 64 GB RAM, and the software environment is Python 3.7.6 with Tensorflow 1.13.0 and Hmmlearn 0.2.7.

The arrival rate process of data flow in the proactive vehicular network conforms to certain spatio-temporal

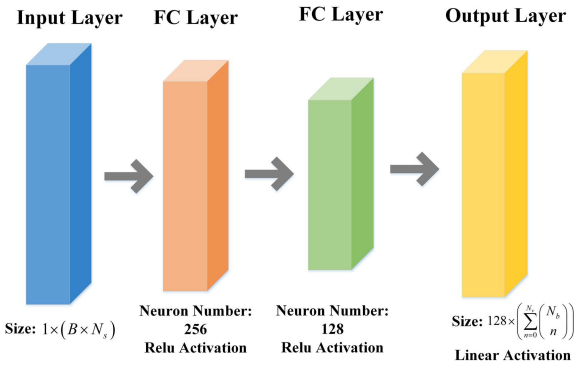


FIGURE 7. The neural network structure of Q network.

characteristics. It can be described by a new structured time series model based on the hidden Markov model [41]. Assume that the network's arrival number of data tasks is divided into three states: trough, mid-term, and peak, corresponding to Y_1, Y_2, and Y_3 data tasks, respectively. Then the number of data tasks in each slot obeys $\lambda^t = (\pi_d, F, Z)$, where π_d is the initial state probability matrix, F is the hidden state transition probability matrix, and Z is the observation state transition probability matrix. We use Hmmlearn 0.2.7 to simulate the arrival of data flow.

In the network, the basic parameters used in our system simulation are shown in Table 3. In the urban environment, there are 14 APs and 64 vehicles. The channel band is 10 MHz [35]. Meanwhile, a vehicle randomly selects one direction out of eight directions to move a certain distance at every slot in the environment.

B. RESULTS AND ANALYSIS

On the vehicle side, we propose two RRUI generation strategies based on local experience and AN assistance, respectively. On the network side, after receiving the RRUI, the AN can make resource allocation decisions through three resource allocation algorithms: DRL-RRA, SPFA-based radio resource allocation algorithm (SPFA-RRA), and random-based radio resource allocation algorithm (R-RRA). We first analyze the effectiveness of the two RRUI generation algorithms. In order to verify the effectiveness of LE-RRUI and AA-RRUI, we compare them with the random RRUI generation strategy. Fig. 8 shows the performance of the six radio resource allocation algorithms under the condition of $N_v = 4$, $\varpi = 0.3$, and resource load rate $= 40\%$, where the resource load rate of each slot is the ratio of the occupied resource number in S^t to the total number of resources and used to measure the density of downlink data flow in the network. When AN chooses R-RRA, the data transmission success rate under LE-RRUI and AA-RRUI is 1.6 and 2 times that of the RRUI random generation algorithm respectively, and the data transmission success rate of AA-RRUI is 10.7% higher than that of LE-RRUI. After fixing the RRUI generation algorithm, the superiority of DRL-RRA algorithm on the network side is reflected. When RRUI is randomly generated, the data transmission success rate of DRL-RRA is 1.4 and 2.4 times that of SPFA-RRA and R-RRA, respectively. AA-RRUI combined with DRL-RRA can achieve a data transmission success rate of 99.3%. Fig. 9 shows that DRL-RRA has good convergence.

TABLE 3. Basic simulation parameters.

Wireless Environment Parameter	Value
Environment length	300 m
Environment width	200 m
Number of RBs N_b	16
Number of RUs N_s	64
Number of RBs in RRUI N_v	4
Maximum RUs can be occupied by each AP C_s	40
Vehicle density λ_U	$0.002 / m^2$
AP density λ_B	$0.0005 / m^2$
Vehicle speed v	72 km/h
Transmit power of each RU in each AP p	1 W
Path loss factor α	3
Minimum SIR threshold γ_{th}	4.5 dB
Transmission rate requirement $R^{t,u}$	[40,60] bit/s
Probability of selecting RB randomly to form RRUI ε_v	0.1
Probability of selecting high-quality RB from $\mathbb{J}_a^{t-\tau-\tau',u}$ to form RRUI ϖ	0.4
Training Parameter	**Value**
Number of training episodes	1500
Number of slots in an episode	100
Experience replay memory capacity	5000
Size of minibatch	32
Number of hidden layers	2
Number of neurons in each hidden layer	256,128
Active function of each hidden layer	ReLU
Learning rate β	0.01
Discount factor φ	0.9
ε_a of $\varepsilon - greedy$ policy	Decrease linearly from 1 to 0.0001 with the ε_a decay rate of 10^{-4}

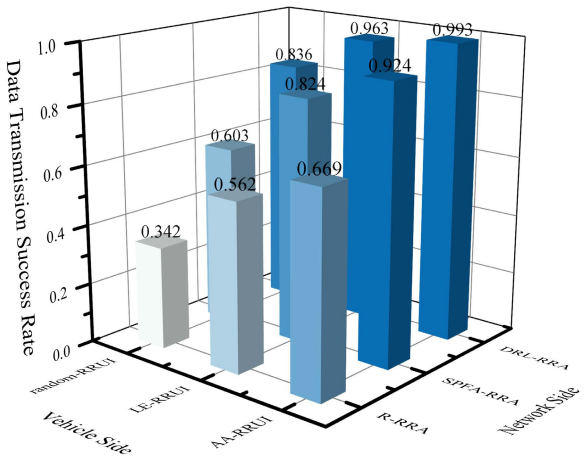


FIGURE 8. The performance of the six downlink joint radio resource allocation algorithms under the condition of $N_v = 4$, $\varpi = 0.3$, and resource load rate = 40%.

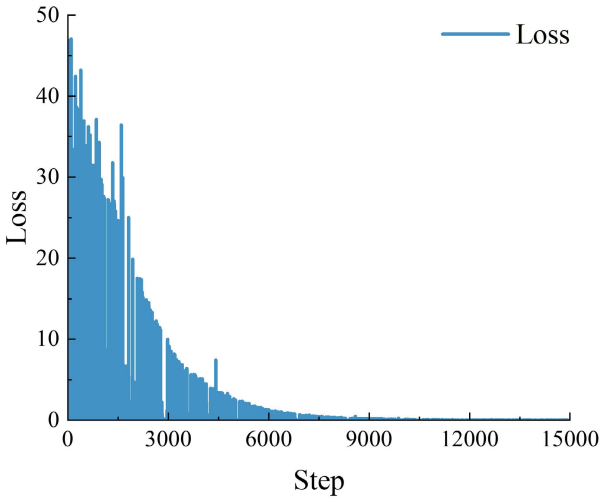


FIGURE 9. The loss of DRL-RRA.

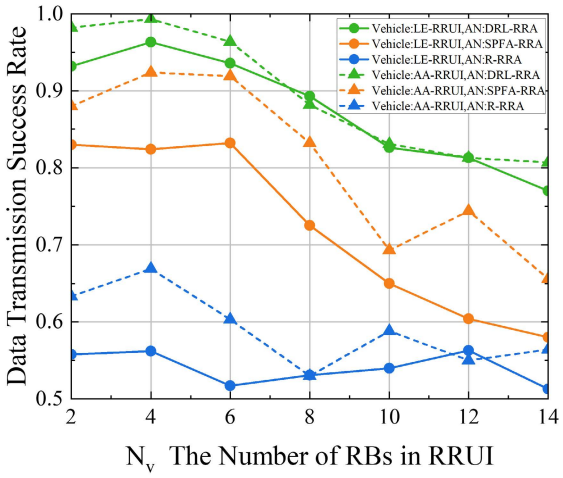


FIGURE 10. The effect of N_V (the number of RBs in the RRUI) on the performance of the downlink resource allocation algorithms when $\varpi = 0.3$ and resource load rate = 40%.

Next, we discuss the parameters of the vehicle-side RRUI generation algorithm. Fig. 10 describes the effect of N_v (the number of RBs in the RRUI) on the performance of the downlink resource allocation algorithms when $\varpi = 0.3$ and resource load rate = 40%. The optimal value of N_v is in the range of 2 to 6. If N_v continues to increase, the data transmission success rates of the six radio resource allocation algorithms all drop significantly. When N_v is close to the total number of RBs in the system, it is meaningless for the vehicle to send RRUI to guide the AN for resource allocation. The DRL-RRA and SPFA-RRA algorithms on the network side are sensitive to the size of N_v. In particular, the size of N_v in RRUI directly affects the size of the action space in DRL-RRA. If N_v is too large, the action space becomes larger and affects the learning performance and convergence of DQN. Fig. 11 shows the effect of ϖ (the probability of selecting RB from global information about RB quality sent by AN)

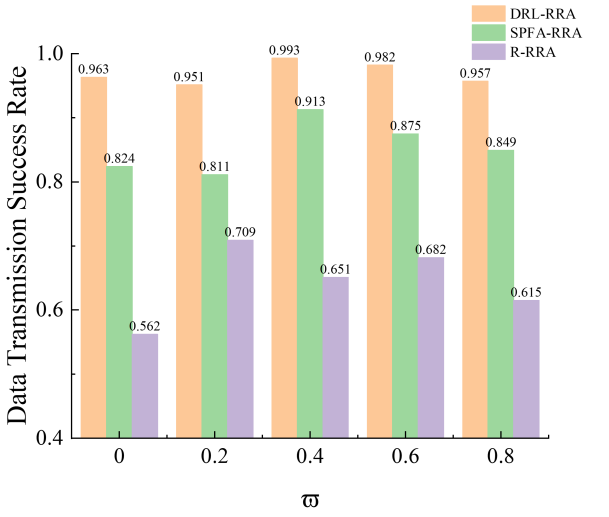


FIGURE 11. The effect of ϖ in AA-RRUI on the performance of the three radio resource allocation algorithms when $N_V = 4$ and resource load rate = 40%.

in AA-RRUI on the performance of the algorithms when $N_v = 4$ and resource load rate = 40%. $\varpi = 0$ means that the vehicle adopts LE-RRUI. The performance of AA-RRUI is higher than that of LE-RRUI, which is because the number of data transmissions of the vehicle itself in a fixed period is much lower than the number of data transmissions of the entire network, estimating the RBs' quality has a certain deviation. However, since the local experience is more suitable for the location and radio environment of the vehicle at the moment, the local data transmission experience cannot be directly discarded. Due to the learning ability of DRL-RRA, with the increase of ϖ, the data transmission success rate of DRL-RRA only increases slightly, and the data success rate of DRL-RRA can reach 99.3% when $\varpi = 0.4$. The performance fluctuations of SPFA-RRA and R-RRA affected by ϖ are 10% and 14%, respectively. DRL-RRA, SPFA-RRA, and R-RRA have higher data transmission success rates when ϖ is in the range of 0.2 - 0.6.

Since the downlink data task must be sent in the next slot once it reaches the AN, the increase of the downlink data traffic leads to the increase of the resource load rate. Fig. 12 shows the data transmission success rates of six radio resource allocation algorithms under different resource load rates. R-RRA is most sensitive to the resource load situation. Since its radio resource selection is completely random, once the occupied resource in the network increase, the interference during downlink transmission will increase. When the resource load rate in the vehicular network does not exceed 40%, both DRL-RRA and SPFA-RRA have high data transmission success rates, and the AA-RRUI and DRL-RRA joint radio resource allocation algorithm can achieve a data transmission success rate of more than 90%. Unfortunately, as the average resource load rate further increases, the performance of both algorithms degrades significantly.

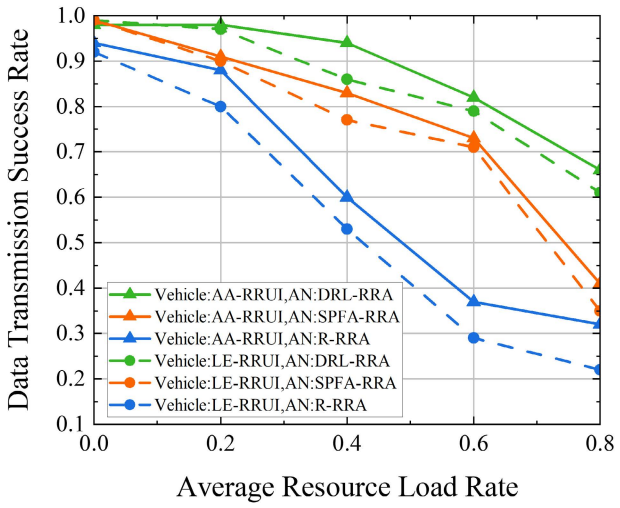


FIGURE 12. The data transmission success rates of six joint radio resource allocation algorithms under different resource load rates.

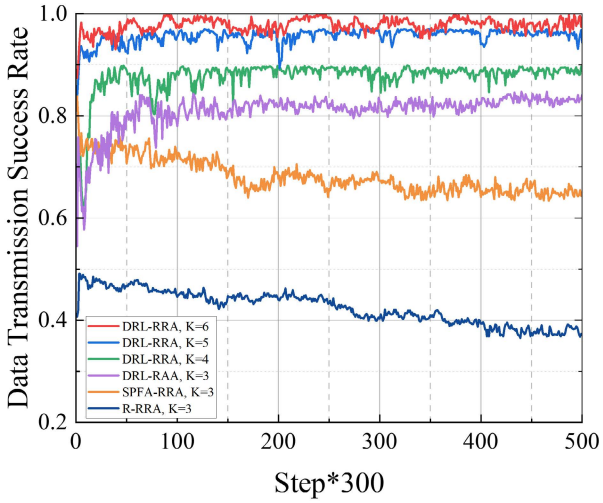


FIGURE 13. The relationship between the number of APs K in a virtual cell and the data transmission success rate.

When the resource load rate reaches 80%, the performance of DRL-RRA and SPFA-RRA drops to around 80% and 62%, respectively.

When the downlink data flow in the proactive vehicular network is dense, appropriately increasing the network resources (number of APs) in each vehicle's virtual cell is an intuitive and effective method to improve the reliability of data transmission. As long as the SIR of one link is higher than the SIR threshold when a vehicle receives data, the data transmission is successful. Fig. 13 shows the relationship between the number of APs K in a virtual cell and the data transmission success rate when the resource load rate reaches 60%. Assuming that the AA-RRUI algorithm is used on the vehicle side, by increasing the number of APs in the virtual cell, the data transmission success rate of the DRL-RRA algorithm can be improved. In the process of increasing the value of K from 3 to 5, the success rate of data transmission increases significantly. When K is increased to 6, the data transmission success rate of DRL-RRA can reach 98%. It indicates K to be a critical design factor for high traffic load.

VI. CONCLUSION

This paper proposed an intelligent radio resources allocation scheme in an ultra-low latency vehicular network. Aim to break the status quo of blind radio resource selection in AN due to the open-loop communication without CSI feedback, we constructed a downlink radio resource allocation framework based on the "generalized closed-loop" with the guide of RRUI from the vehicle's immediate past uplink transmission. Subsequently, we took the long-term success rate of data transmission as a reliability indicator, and established a downlink radio resource allocation model based on cooperation between vehicles and AN to optimize the success rate of data transmission. On the vehicle side, we proposed two RRUI generation strategies, LE-RRUI and AA-RRUI, to select high communication quality RBs from local experience and global experience as RRUI. On the network side, we proposed an intelligent radio resource allocation algorithm based on deep reinforcement learning (that is DRL-RRA). The simulation verified the effectiveness of the joint radio resource allocation algorithm under the cooperation between the vehicle side and the network side.

We noted that the performance of the radio resource allocation scheme based on the joint vehicle and AN was limited when the resource occupancy rate exceeded 60%. Increasing the number of APs in a virtual cell could improve the success rate of data transmission, but at the same time, it would further increase the resource load rate. This paper proposes further research direction. We study the radio resource allocation after fixing the number of network resources (APs) in the virtual cell. If each downlink data task can be flexibly allocated from two dimensions of network resources and radio resources, the success rate of data transmission can be further improved. The difficulty in the direction is that the resource allocation model will be more complex, and the realization of more fine-grained resource allocation under extremely low latency communication has high requirements on algorithm performance.

REFERENCES

[1] A. Alalewi, I. Dayoub, and S. Cherkaoui, "On 5G-V2X use cases and enabling technologies: A comprehensive survey," *IEEE Access*, vol. 9, pp. 107710–107737, 2021.

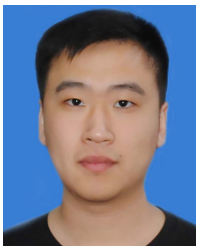
[2] Z. Ali, S. Lagen, L. Giupponi, and R. Rouil, "3GPP NR V2X mode 2: Overview, models and system-level evaluation," *IEEE Access*, vol. 9, pp. 89554–89579, 2021.

[3] C. R. Storck and F. Duarte-Figueiredo, "A survey of 5G technology evolution, standards, and infrastructure associated with vehicle-to-everything communications by internet of vehicles," *IEEE Access*, vol. 8, pp. 117593–117614, 2020.

[4] E. Ko, K.-C. Chen, and S.-Y. Lien, "Collaborative partially-observable reinforcement learning using wireless communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2021, pp. 1–6.

[5] D. Feng, L. Lai, J. Luo, Y. Zhong, C. Zheng, and K. Ying, "Ultra-reliable and low-latency communications: Applications, opportunities and challenges," *Sci. China Inf. Sci.*, vol. 64, no. 2, pp. 1–12, Jan. 2021.
[6] F. E. Airod, H. Chafnaji, and H. Yanikomeroglu, "HARQ in full-duplex relay-assisted transmissions for URLLC," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 409–422, 2021.
[7] H. Jang, J. Kim, W. Yoo, and J.-M. Chung, "URLLC mode optimal resource allocation to support HARQ in 5G wireless networks," *IEEE Access*, vol. 8, pp. 126797–126804, 2020.
[8] E. Dosti, M. Shehab, H. Alves, and M. Latva-Aho, "Ultra reliable communication via optimum power allocation for HARQ retransmission schemes," *IEEE Access*, vol. 8, pp. 89768–89781, 2020.
[9] A. Hossain, Z. Pan, M. Saito, J. Liu, and S. Shimamoto, "Multiband massive channel random access in ultra-reliable low-latency communication," *IEEE Access*, vol. 8, pp. 81492–81505, 2020.
[10] D. Xu and P. Ren, "Quantum learning based nonrandom superimposed coding for secure wireless access in 5G URLLC," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2429–2444, 2021.
[11] J. Zeng, T. Lv, R. P. Liu, X. Su, Y. J. Guo, and N. C. Beaulieu, "Enabling ultrareliable and low-latency communications under shadow fading by massive MU-MIMO," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 234–246, Jan. 2020.
[12] Y. Liu, Y. Deng, M. Elkashlan, A. Nallanathan, and G. K. Karagiannidis, "Analyzing grant-free access for URLLC service," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 741–755, Mar. 2021.
[13] T. H. Jacobsen, R. Abreu, G. Berardinelli, K. I. Pedersen, I. Z. Kovács, and P. Mogensen, "Multi-cell reception for uplink grant-free ultra-reliable low-latency communications," *IEEE Access*, vol. 7, pp. 80208–80218, 2019.
[14] R. Qi, X. Chi, L. Zhao, and W. Yang, "Martingales-based ALOHA-type grant-free access algorithms for multi-channel networks with mMTC/URLLC terminals co-existence," *IEEE Access*, vol. 8, pp. 37608–37620, 2020.
[15] F. Salehi, N. Neda, M.-H. Majidi, and H. Ahmadi, "Cooperative NOMA-based user pairing for URLLC: A max–min fairness approach," *IEEE Syst. J.*, early access, Oct. 13, 2021, doi: 10.1109/JSYST.2021.3116112.
[16] J. Khan and L. Jacob, "Availability maximization framework for CoMP enabled URLLC with short packets," *IEEE Netw. Lett.*, vol. 2, no. 1, pp. 1–4, Mar. 2020.
[17] B. Kharel, O. L. A. López, N. H. Mahmood, H. Alves, and M. Latva-Aho, "Fog-RAN enabled multi-connectivity and multi-cell scheduling framework for ultra-reliable low latency communication," *IEEE Access*, vol. 10, pp. 7059–7072, 2022.
[18] Q. Cui, J. Zhang, X. Zhang, K.-C. Chen, X. Tao, and P. Zhang, "Online anticipatory proactive network association in mobile edge computing for IoT," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4519–4534, Jul. 2020.
[19] Q. Cui, Z. Gong, W. Ni, Y. Hou, X. Chen, X. Tao, and P. Zhang, "Stochastic online learning for mobile edge computing: Learning from changes," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 63–69, Mar. 2019.
[20] G. Tan, H. Zhang, and S. Zhou, "Resource allocation in MEC-enabled vehicular networks: A deep reinforcement learning approach," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Jul. 2020, pp. 406–411.
[21] A. Elgabli, H. Khan, M. Krouka, and M. Bennis, "Reinforcement learning based scheduling algorithm for optimizing age of information in ultra reliable low latency networks," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2019, pp. 1–6.
[22] S. Praveen, J. Khan, and L. Jacob, "Reinforcement learning based link adaptation in 5G URLLC," in *Proc. 8th Int. Conf. Smart Comput. Commun. (ICSCC)*, Jul. 2021, pp. 159–163.
[23] K.-C. Chen, T. Zhang, R. D. Gitlin, and G. Fettweis, "Ultra-low latency mobile networking," *IEEE Netw.*, vol. 33, no. 2, pp. 181–187, Mar./Apr. 2019.
[24] S.-Y. Lien, S.-C. Hung, K.-C. Chen, and Y.-C. Liang, "Ultra-low-latency ubiquitous connections in heterogeneous cloud radio access networks," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 22–31, Jun. 2015.
[25] S.-C. Hung, H. Hsu, S.-M. Cheng, Q. Cui, and K.-C. Chen, "Delay guaranteed network association for mobile machines in heterogeneous cloud radio access network," *IEEE Trans. Mobile Comput.*, vol. 17, no. 12, pp. 2744–2760, Dec. 2018.
[26] C.-H. Zeng and K.-C. Chen, "Downlink multiuser detection in the virtual cell-based ultra-low latency vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4651–4666, May 2019.
[27] I. W. Lai, C. H. Lee, K. C. Chen, and E. Biglieri, "Path-permutation codes for end-to-end transmission in ad hoc cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3309–3321, Jun. 2015.
[28] C.-Y. Lin, K.-C. Chen, D. Wickramasuriya, S.-Y. Lien, and R. D. Gitlin, "Anticipatory mobility management by big data analytics for ultra-low latency mobile networking," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–7.
[29] Y. Wang, Q. Cui, and K.-C. Chen, "Machine learning enables predictive resource recommendation for minimal latency mobile networking," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2021, pp. 1363–1369.
[30] Y. Zhang, L. Zhao, G. Zheng, X. Chu, Z. Ding, and K.-C. Chen, "Resource allocation for open-loop ultra-reliable and low-latency uplink communications in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2590–2604, Mar. 2021.
[31] M. Wen, T. Hai, L. Zhang, J. Hao, G. Zhao, Z. Zhen, Y. Zhao, and L. Feng, "Deep reinforcement learning-based resource reservation method for power emergency Internet-of-Things slice," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Jun. 2021, pp. 63–67.
[32] F. Abbas, P. Fan, and Z. Khan, "A novel low-latency V2V resource allocation scheme based on cellular V2X communications," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2185–2197, Jun. 2019.
[33] Z. Dong, X. Zhu, Y. Jiang, and H. Zeng, "Manager selection and resource allocation for 5G-V2X platoon systems with finite blocklength," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–6.
[34] M. Zhang, Y. Dou, P. H. J. Chong, H. C. B. Chan, and B.-C. Seet, "Fuzzy logic-based resource allocation algorithm for V2X communications in 5G cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2501–2513, Aug. 2021.
[35] *Study on NR Vehicle-to-Everything (V2X)*, document TR 38.885, Version 16.0.0, 3GPP, Mar. 2019.
[36] Z. Sheng, A. Pressas, V. Ocheri, F. Ali, R. Rudd, and M. Nekovee, "Intelligent 5G vehicular networks: An integration of DSRC and mmWave communications," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2018, pp. 571–576.
[37] A. Chockalingam, M. Zorzi, L. B. Milstein, and P. Venkataram, "Performance of a wireless access protocol on correlated Rayleigh-fading channels with capture," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 644–655, May 1998.
[38] F. D. P. Calmon and M. Yacoub, "MRCS—Selecting maximal ratio combined signals: A practical hybrid diversity combining scheme," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3425–3429, Jul. 2009.
[39] D. P. Williamson, *Network Flow Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2019.
[40] Y. Lin, Z. Zhang, Y. Huang, J. Li, F. Shu, and L. Hanzo, "Heterogeneous user-centric cluster migration improves the connectivity-handover trade-off in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16027–16043, Dec. 2020.
[41] Y. Xie, S.-Z. Yu, S. Tang, and X. Huang, "A two-layer hidden Markov model for the arrival process of web traffic," in *Proc. IEEE 19th Annu. Int. Symp. Modeling, Anal., Simulation Comput. Telecommun. Syst.*, Jul. 2011, pp. 469–471.

XINYUAN WANG received the B.S. degree in communication engineering from the Xi'an University of Posts and Telecommunications, China, in 2019. She is currently pursuing the M.S. degree in information and communications engineering with the Beijing University of Posts and Telecommunications, Beijing, China. Her research interests include graph theory, reinforcement learning, and their applications to resource allocation in ultra-reliable and low-latency networks.



YINGZE WANG received the B.S. degree in communication engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2015. He is currently pursuing the Ph.D. degree in information and communications engineering with the Beijing University of Posts and Telecommunications, Beijing, China. His current research interests include the proactive communication theory, intelligent wireless networks, and the applications of reinforcement learning within resource management.

QIMEI CUI (Senior Member, IEEE) received the B.E. and M.S. degrees in electronic engineering from Hunan University, Changsha, China, in 2000 and 2003, respectively, and the Ph.D. degree in information and communications engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2006. In 2016, she was a Visiting Professor at the Department of Electronic Engineering, University of Notre Dame, IN, USA. Since 2014, she has been a Full Professor with the School of Information and Communication Engineering, BUPT. Her research interests include 5G/B5G wireless communications, mobile computing, and the IoT. She was a recipient of the Best Paper Award from the IEEE International Symposium on Communications and Information Technologies (ISCIT) 2012, the IEEE Wireless Communications and Networking Conference (WCNC) 2014, the APCC 2018, the WCSP 2019, the Honorable Mention Demo Award from the ACM MobiCom 2009, and the Young Scientist Award from the URSI GASS 2014. She served as a Technical Program Committee Member of several international conferences, such as the IEEE International Conference on Communications (ICC), the IEEE Wireless Communications and Networking Conference (WCNC), the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), the IEEE International Conference on Communications (ICCC), the WCSP 2013, and the IEEE International Symposium on Communications and Information Technologies (ISCIT) 2012. She is the Editor of the *Science China Information Sciences* and a Guest Editor of the *EURASIP Journal on Wireless Communications and Networking* and the *International Journal of Distributed Sensor Networks*.

KWANG-CHENG CHEN (Fellow, IEEE) received the B.S. degree from National Taiwan University, in 1983, and the M.S. and Ph.D. degrees from the University of Maryland, College Park, MD, USA, in 1987 and 1989, respectively, all in electrical engineering. From 1987 to 1998, he worked with SSE, COMSAT, IBM Thomas J. Watson Research Center, and National Tsing Hua University. From 1998 to 2016, he was a Distinguished Professor at National Taiwan University, Taipei, Taiwan, also served as the Director for the Graduate Institute of Communication Engineering, the Director for the Communication Research Center, and the Associate Dean for Academic Affairs, College of Electrical Engineering and Computer Science, from 2009 to 2015. Since 2016, he has been a Professor of electrical engineering with the University of South Florida, Tampa, FL, USA. His recent research interests include wireless networks, multi-robot systems, the IoT and CPS, social networks and data analytics, and cybersecurity. He is actively involving in the organization of various IEEE conferences as the General/TPC Chair/Co-Chair, and has served in editorships for a few IEEE journals. He also actively participates in and has contributed essential technology to various IEEE 802, Bluetooth, LTE and LTE-A, 5G-NR, and ITU-T FG ML5G wireless standards. He was a recipient of a number of awards, including the 2011 IEEE Communications Society (COMSOC) Wireless Communications Technical Committee (WTC) Recognition Award, the 2014 IEEE Jack Neubauer Memorial Award, and the 2014 IEEE Communications Society (COMSOC) AP Outstanding Paper Award.

WEI NI (Senior Member, IEEE) received the B.E. and Ph.D. degrees in electronic engineering from Fudan University, Shanghai, China, in 2000 and 2005, respectively. He was a Postdoctoral Research Fellow at Shanghai Jiao Tong University, from 2005 to 2008, a Research Scientist and the Deputy Project Manager of the Bell Laboratories R&I Center, Alcatel/Alcatel-Lucent, from 2005 to 2008, and a Senior Researcher at the Devices Research and Development Department, Nokia, from 2008 to 2009. He is currently the Group Leader and the Principal Research Scientist of the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia. He is also an Adjunct Professor with the University of Technology Sydney, and an Honorary Professor with Macquarie University. His research interests include optimization, game theory, and graph theory, as well as their applications to networks and security. He is a Program Committee Member of the International Conference on Communications and Networking, China, in 2014, and a TPC Member of the IEEE International Conference on Communications 2014, the International Conference on Culture Collections 2015, EICE 2014, and the Wireless Communications and Networking Conference 2010. He served as the Student Travel Grant Chair for the International Symposium on Wireless Personal Multimedia Communications 2014. He has been the Chair of the IEEE New South Wales (NSW) Vehicular Technology Society (VTS) Chapter, since 2020, and the Vice-Chair of the IEEE NSW VTS Chapter, since 2019. He is the Secretary of the Chapter, from 2015 to 2018, the Track Chair of VTC-Spring 2017, the Track Co-Chair of IEEE Conference on Vehicular Technology (VTC)-Spring 2016, and the Publication Chair of BodyNet 2015. He has been the Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, since 2018.