# Delay-Oriented Knowledge-Driven Resource Allocation in SAGIN-Based Vehicular Networks

Lei Huang*, Ruijin Sun*, Nan Cheng*, Yilong Hui* and Dandan Liang‡
*School of Telecommunications Engineering, Xidian University, Xi'an, China
‡Pengcheng Laboratory, Shenzhen, China
Email: 18010100111@stu.xidian.edu.cn, sunruijin@xidian.edu.cn, dr.nan.cheng@ieee.org, ylhui@xidian.edu.cn, liangdd@pcl.ac.cn

2023 IEEE Wireless Communications and Networking Conference (WCNC) | 978-1-6654-9122-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/WCNC55385.2023.10118888

***Abstract*—Space-air-ground integrated networks (SAGIN) have been envisioned as the promising and key network architecture for the 6G vehicular networks to provide seamless coverage for the connected vehicles. To access the most appropriate network quickly, this paper proposed a knowledge-driven network access approach, where the communication knowledge is explicitly integrated into neural networks, to deal with multiple tasks in SAGIN-based vehicular networks. Specifically, the formulated long-term network access problem is handled by asynchronous advantage actor-critic algorithm (A3C) in reinforcement learning. During this process, the space-time correlation knowledge is introduced to effectively reduce the action space in channel selection and the reward shaping exploiting the problem-specific communication and mathematical knowledge is adopted to solve the sparse reward problem in reinforcement learning. In addition, by modifying the sub-net learning rate of the A3C algorithm with experimental experience, this paper speeds up the network convergence speed by $1.5\%$. Numerical results also show that integrating knowledge into traditional deep reinforcement learning algorithm can improve the reward by $4\%$.**

***Index Terms*—Channel selection, knowledge-driven, network access, reward reshaping, space-air-ground integrated networks**

## I. Introduction

In the future sixth generation (6G) network, the vehicles are changing from the functional vehicles based on traditional electronic architecture to the mobile intelligent terminals that integrate many transformative technologies such as big data, cloud computing, artificial intelligence (AI) and the Internet of things (IoT). These connected vehicles may be deployed in the remote mountainous areas, deserts or forests, where the traditional communication systems rarely have stable communication coverage. Meanwhile, with the research of 6G, more complex communication scenarios are proposed, and the new communication services have more urgent requirements for seamless coverage. Traditional terrestrial networks cannot meet the new needs, while satellite and Unmanned Aerial Vehicle (UAV) communication networks are considered as promising complements to extend the terrestrial networks in order to suit for various scenarios. And satellite, UAV, and base station (BS) can complement each other, the integration of them, namely the space-air-ground integrated network (SAGIN), is proposed as a promising 6G wireless network, which is a promising solution to provide cost-effective, large-scale, and flexible wireless coverage and communication services [1]. SAGIN can also efficiently tackle the problems of network coverage and data transmission of the vehicular networks. At the same time, integrating SAGIN technology into these vehicular network can provide flexible and reliable services for vehicles by taking advantage of different networks [2] in SAGIN-based vehicular networks.

In SAGIN-based vehicular networks, resource scheduling is a research hotspot, which includes frequency, power, time, computing, sensing, storage resources and etc. The service transmission process in the vehicular network needs to consume a lot of network resources, while the network resources are limited and different services have different resource requirements, so it is necessary to efficiently schedule network resources according to the task requirements. And the network access decision is the key problem of resource scheduling. In traditional network access selection, two categories are mostly used, namely, selection methods driven by mathematical models such as optimization theory and game theory, and decision methods driven by data such as neural networks and deep reinforcement learning [3]. For example, by using a mathematical model-driven method, Guo *et al.* [4] formulated the joint route and access network selection problem as a semi-Markov decision process (SMDP) and derived a optimal algorithm to improve the throughput of the target vehicle. Besides, with a data-driven approach, Mchergui *et al.* [5] proposed a novel hybrid relay selection technique to perform the broadcasting task based on a reinforcement learning method to increase the success rate, save rebroadcasts, and reduce the delay in a grid map scenario with varied traffic densities.

However, with the rapid increase of resource scheduling complexity of SAGIN-based vehicular network, it is difficult for the mathematical model methods or data-driven approaches to meet the updated needs, while integrating knowledge into these traditional approaches will hopefully optimize existing network structures or algorithm models. For instance, in a multi-radio multi-channel wireless mesh network (WMN) with wireless extenders (EXTs) whose locations change over time, Gačanin *et al.* [6] proposed a optimization algorithm by introducing the spectrum correlation knowledge to improve the timeliness and accuracy of the resource scheduling process. Additionally, in the fifth generation new radio (5G NR) downlink scheduler design, the base station optimized user scheduling and resource block allocation according to the channel state information (CSI) and queue length to maximize the number of resource blocks successfully received by

users. Based on expert knowledge, multi-head critic, reward reshaping and importance sampling were proposed in [7], which effectively enhanced the reliability of the network and accelerated the convergence of the network.

As vehicles become more intelligent, they can provide users with more infotainment services such as road safety, intelligent transportation, autonomous driving, in-car infotainment and so on, which are closely related to delay. Therefore, delay is a significant experimental index for the access network selection in vehicular networks [8], but the above knowledge-driven methods are rarely studied. And this paper adopts knowledge-driven vehicular access network selection method to minimize the delay of the vehicle access process in the SAGIN-based vehicular networks. The main contributions are summarized as follows:

- The paper designs a knowledge-driven access selection approach for SAGIN-based vehicular networks. This method solves the problem of vehicular service transmission in remote mountainous area, desert or other areas only covered by ground BSs.
- This paper introduces the space-time correlation knowledge to effectively avoid the influence of adjacent channel interference on vehicle access selection and designs a reward function by using the mathematics knowledge to solve the sparse reward problem.
- The algorithm can speed up the network convergence speed by setting different learning rates for each subnet of the asynchronous advantage actor-critic (A3C) algorithm and the access selection method can reduce the total convergence delay through integrating knowledge.

The reminder of the paper is organized as follows. In Section II, the access problem is formulated to minimize the total delay of vehicle task transmission. The knowledge-driven vehicle access algorithm is designed in Section III. Section IV presents the numerical results. Finally, conclusions are given in Section V.

## II. System Model And Problem Formulation

This section first introduces the system model. Then the communication rate model and the vehicle access delay model of vehicle access network are described.

### A. System Model

The network model is shown in Fig 1. The network deploys ground BSs, UAVs and LEOs to achieve full coverage of the SAGIN and provide stable transmission services for vehicles. Assuming there are $K$ vehicles, $n$ selectable access networks, including $N$ UAVs, $M$ BSs, and 1 LEO. $\mathcal{K}$ represents the set of the vehicles. Then the set of the available access networks is $\mathcal{N}$, which consists of the set of UAV $\{1, 2, ..., N\}$, the set of BS $\{1+N, 2+N, ..., M+N\}$, and the set of LEO $\{1+N+M\}$. Taking the processing capacity of UAVs, BSs, and LEO satellite into account, the vehicle needs to select the access network to reduce the service delay [9].

For each vehicle, a new task with different requirements will be generated in each time slot. Due to the limited computing capacity of the vehicle, it is necessary to transmit tasks to the cloud for processing. Since the data transfer rate may be lower than the data generation rate, there is a storage unit with limited memory capability to cache tasks. In this paper, all vehicles are assumed to be equipped with one antenna, namely only one task can be transmitted at a time for each vehicle. Meanwhile, different types of vehicle tasks have different latency requirements and are randomly generated. The services that are transmitted in this paper mainly include three types, namely, delay-sensitive task, rate-sensitive task, and simple task. The delay and rate requirements of different tasks are shown in Table I.

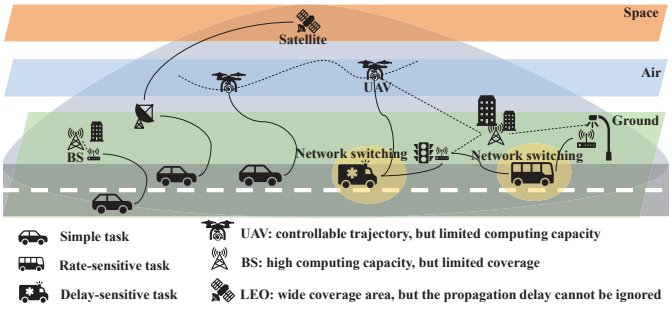


Fig. 1: The considered SAGIN-based vehicular networks.

When a new task is generated in a vehicle, the vehicle first needs to store the task locally in its storage unit. And all tasks in the storage unit are assumed to form a queue and the task that arrives first will be transferred first. A newly generated task will be dropped if the queue of tasks is too long that even transmitting the task with the highest rate, the processing delay is also higher than the delay requirement.

TABLE I: Task Types And Properties

| Task type | Delay | Rate requirement |
|---|---|---|
| Delay-sensitive task | 30ms | 10Mbps |
| Rate-sensitive task | 2000ms | 10Gbps |
| Simple task | 3000ms | 1Mbps |

In this case, assuming the size of the arriving task of the $k$-th vehicle in the time slot $t$ is $\rho$ and the number of tasks is $A_k(t)$. In this paper, $U_k(t)$ is defined as the total amount of task data transmitted by the $k$-th vehicle at time slot $t$. At the same time, the amount of data stored locally in the $k$-th vehicle is set as the queue $Q_k(t)$ [10]. $\alpha_{k,n,f}(t)$ indicates that the $k$-th vehicle choose the $f$-th sub-channel of $n$-th access network at time slot $t$. $J$ is defined as the number of sub-channels with the set $\mathcal{J}$ [11].

### B. Communication Rate Model

*1) Vehicle-UAV Link:* When adopting the vehicle-UAV data transmission model, in consideration of the high-speed mobility of vehicles and UAVs will cause rapid changes in the channel state, this paper ignores small-scale fading and only considers large-scale channel fading [12]. So for the $k$-th vehicle, the path loss of the vehicle-UAV link can be calculated as

$$L_k(t) = 20\log\left(\frac{4\pi v_k(t)\sqrt{h_k^2(t)+r_k^2(t)}}{c}\right) + P_k^{\mathrm{LOS}}(t)\eta_k^{\mathrm{LOS}}(t) + (1-P_k^{\mathrm{LOS}}(t))\eta_k^{\mathrm{LOS}}(t), \quad (1)$$

where $r_k(t)$ represents the horizontal distance between the $k$-th vehicle and the UAV access network, $h_k(t)$ represents the flying height of the UAV, $v_k(t)$ represents the carrier frequency, $c$ represents the speed of light. $\eta_k^{\text{LOS}}(t)$ and $\eta_k^{\text{NLOS}}(t)$ represent the additional loss generated in the free-space path loss of line-of-sight (LOS) links and non-line-of-sight (NLOS) links of the $k$-th vehicle, respectively, meanwhile, both of them are variables determined by environmental information. And for the $k$-th vehicle, the probability of LOS in the vehicle-UAV link is

$$P_k^{\text{LOS}}(t) = \frac{1}{1 + a\exp\left\{-b\left[\arctan(\frac{h_k(t)}{r_k(t)}) - a\right]\right\}}, \quad (2)$$

where $a$ and $b$ are also variables determined by the environment. Therefore, the signal-to-noise ratio (SNR) of the vehicle-UAV link for the $k$-th vehicle, which chooses the $f$-th sub-channel of the $n$-th access network at time slot $t$, can be calculated as

$$\gamma_{k,n,f}(t) = \alpha_{k,n,f}(t)\frac{P_k^{\text{TX}}(t)10^{-\frac{L_k(t)}{10}}}{\delta^2 + \delta_{interf}}, n = 1, 2, ..., N, \quad (3)$$

where $P_k^{TX}(t)$ is the data transmission power of the $k$-th vehicle, and $\delta^2$ is the additive white Gaussian noise power. $\delta_{interf}$ represents the out-of-band power leakage caused by inter-carrier interference and in general, and the value of leakage is $1\%$ to $5\%$ of the transmitted power. Due to the fact that each sub-carrier of the access network can be only allocated to at most one vehicle, so for the $k$-th vehicle, we have

$$0 \leq \sum_{k\in\mathcal{K}} \alpha_{k,n,f}(t) \leq 1, \alpha_{k,n,f}(t) \in 0, 1, \forall f \in \mathcal{J}, t \geq 0. \quad (4)$$

*2) Vehicle-BS Link:* When the $k$-th vehicle selects the $f$-th sub-channel of the $n$-th access network in the time slot $t$, the SNR of the vehicle-BS link is

$$\gamma_{k,n,f}(t) = \alpha_{k,n,f}(t)\frac{P_k^{\text{TX}}(t)G_{k,n,f}(t)}{\delta^2 + \delta_{interf}}, \quad n = N+1, N+2, ..., N+M, \quad (5)$$

where $G_{k,n,f}(t)$ is the uplink channel gain of the vehicle-BS link in the time slot $t$.

*3) Vehicle-Satellite Link:* For the $k$-th vehicle on the $f$-th sub-channel in the $n$-th access network at time slot $t$, the SNR of the vehicle-to-satellite link can be calculated as follows,

$$\gamma_{k,n,f}(t) = \alpha_{k,n,f}(t)\frac{P_k^{\text{TX}}((t)G_{G,n,f}(t)G_{rsd,f}(t)L_{k,fn,f}(t)L_{k,rn,f}(t)}{\delta^2 + \delta_{interf}}, \quad n = N+M+1, \quad (6)$$

where $G_{G,n,f}(t)$ is the uplink channel gain of the vehicle-LEO link in the time slot $t$, $G_{rsd,f}(t)$ is the gain of the LEO receiving antenna, $L_{k,fn,f}(t)$ and $L_{k,rn,f}(t)$ are the free space loss and the atmospheric attenuation between the LEO and the ground BS [13], respectively.

### C. Vehicle Access Delay Model

Delay is an important data indicator in this paper. The total delay in the network access process consists of three parts: queue delay, computation delay and uplink transmission delay.

*1) Queue Delay:* During the driving process of the vehicle, it is necessary to store the services for transmission, so there will be many tasks waiting to be processed in the local storage unit of the vehicle. According to the little theorem, the queue delay is equal to the ratio of the average queue waiting amount of the task data to the average arrival rate. From this, the queue delay in the local storage unit of the $k$-th vehicle is

$$\tau_k^Q(t) = \frac{Q_k(t)}{\tilde{A}_k(t)}, \quad (7)$$

where $\tilde{A}_k(t)$ is the average arrival rate of the $k$-th vehicle buffer data.

*2) Computation Delay:* Computation delay is equal to the ratio of the processed task size to the computing capacity of the vehicle. Define $\lambda$ as the computational complexity, that is, the number of CPU cycles required to process 1-bit service data, so that the computation delay of the $k$-th vehicle processing the task in the time slot $t$ is

$$\tau_k^{com}(t) = \frac{U_k(t)\lambda}{\tilde{fc}_k(t)}, \quad (8)$$

where $\tilde{fc}_k(t)$ is the central processing unit (CPU) frequency of the $k$-th vehicle in the time slot $t$.

*3) Uplink Transmission Delay:* The uplink transmission delay depends on the size of the transmitted task, the rate of the uplink transmission as well as the available transmission resources. The vehicle transmits the vehicular task to the UAV or BS through the orthogonal sub-channel. Assuming that the channel bandwidth is $B_0$, according to the Shannon formula, the maximum achievable data transmission rate of the $k$-th vehicle on sub-carrier $f$ in access network $n$ in the time slot $t$, the sum rate of the $k$-th vehicle and the amount of task data that can be transmitted are respectively expressed as

$$R_{k,n,f}(t) = \alpha_{k,n,f}(t)B_0\log(1 + \gamma_{k,n,t}), \quad (9)$$

$$R_k(t) = \sum_{n\in\mathcal{N}} R_{k,n}(t) = \sum_{n\in\mathcal{N}}\sum_{f\in\mathcal{J}} R_{k,n,f}(t), \quad (10)$$

$$u_k^{tra}(t) = \min\left\{Q_k(t) + \rho A_k(t), R_k(t)\tau\right\}, \quad (11)$$

where $\tau$ is the slot length. Then the transmission delay of the $k$-th vehicle in the time slot $t$ is

$$\tau_k^{tra}(t) = \frac{U_k(t)}{R_k(t)} = \min\left\{\frac{Q_k(t) + \rho A_k(t)}{R_k(t)}, \tau\right\}. \quad (12)$$

The uplink transmission delay of the satellite consists of two parts, which are the delay from the uplink transmission of the vehicle to the ground BS and the ground BS to the satellite. The data transmission rate and delay of the ground BS for the $k$-th vehicle can be respectively calculated as

$$R_k^{prog}(t) = \sum_{n\in\mathcal{N}}\sum_{f\in\mathcal{J}} \alpha_{k,n,f}(t)B_c\log(1 + \gamma_{k,n,f}(t)), \quad n \in \{N+M+1\}, \quad (13)$$

$$\tau_k^{prog}(t) = \frac{U_k(t)}{R_k^{prog}(t)}, \tag{14}$$

where $B_c$ is the bandwidth of the ground BS-satellite link and $\tau_k^{prog}(t)$ is the delay of the uplink transmission from the ground BS to the satellite. The achievable data transmission rate of the satellite link in the time slot $t$ is the bandwidth of the satellite link. Therefore, at time slot $t$, the total delay for the $k$-th vehicle to choose access network for service transmission is

$$\tau_k(t) = \begin{cases} \tau_k^Q(t) + \tau_k^{tra}(t) \\ +\tau_k^{com}(t), \quad n \in \{1, ..., N, N+1, ..., N+M\}, \\ \tau_k^Q(t) + \tau_k^{tra}(t) + \tau_k^{com}(t) \\ +\tau_k^{prog}(t), \quad n \in \{N+M+1\}, \end{cases} \tag{15}$$

The network access decision in each time slot can seriously affect the service transmission delay. The optimization objective studied in this paper is to minimize the total delay of the vehicle, and the optimization factor is the selection set of the access network. The problem can be expressed as Equation (16a-c), with P1 denoting the optimization objective,

$$P1: \min_{\{\alpha_{k,n,f}(t)\}} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{K} \sum_{k=1}^{K} \tau_k(t), \tag{16a}$$

$$s.t. \quad C1: 0 \le \sum_{k\in\mathcal{K}} \alpha_{k,n,f}(t) \le 1, \alpha_{k,n,f}(t) \in \{0,1\}, \forall f \in \mathcal{J}, \tag{16b}$$

$$C2: R_k(t) \ge R_{\min}^m, \quad \forall k \in \mathcal{K}, \quad \forall m \in \{\text{type1}, \text{type2}, \text{type3}\}. \tag{16c}$$

where $R_k(t)$ represents the data transmission rate of the $k$-th vehicle at time slot t. And $m$ is expressed as the task type, type1, type2 and type3 mean the minimum data transmission rate of the delay-sensitive task, the rate-sensitive task and simple task, respectively. And (16a) is the objective function that minimizes the time-average delay of all collected tasks of $K$ vehicles over $T$ epochs. (16b) shows that each sub-channel of the access network can be only allocated to at most one vehicle and (16c) constrains the minimum data transfer rate of the $k$-th vehicle.

## III. Knowledge-Driven Vehicle Access Algorithm Design

The real model of the considered SAGIN-based vehicular network needs to be approximated gradually by learning and the corresponding model is unknown. As a consequence, the deep reinforcement algorithm are supposed to be an effective method to solve the vehicle access problem in this paper. This section firstly introduces the deep reinforcement algorithm and then fine-tunes the learning rate of the A3C algorithm subnets. Besides, we analyze the problems if A3C algorithm is implemented directly and propose to integrate knowledge to solve them.

### A. Deep Reinforcement Learning Framework

Problem P1 can be reformulated as a Markov decision process (MDP) , with three components $\langle s(t), a(t), r\rangle$, where $s(t)$, $a(t)$, and $r$ denote system state space, action space, and reward function, respectively. For the aforementioned problem, the state, action, and reward in the MDP model are formulated as follows.

*1) State:* $s(t)$ is the state space of the heterogeneous network at time slot $t$, including channel state and the type of vehicle transmission business.

*2) Action:* $a(t)$=$\{\alpha_{k,n,f}(t)\}$ is the action space of heterogeneous networks at time slot $t$, which specifically refers to access network selection and sub-carrier selection.

*3) Reformulated reward:* $r(t)$ is related to the total delay of the $k$-th vehicle $\tau_k(t)$ and the data transmission sum rate $R_k(t)$. The reward value function is

$$r(t) = \sum_{k=1}^{K} w_1 \tau_k(t) + w_2 R_k(t), \tag{17}$$

where the weights $w_1$ and $w_2$ are proposed to describe the emphasis of different service types, which represents the weight of delay and the weight of data transmission rate, respectively. For the delay-sensitive task, $w_1 = -1$, $w_2 = 0.1$, for the rate-sensitive task, $w_1 = -0.1$, $w_2 = 1$ and for the simple task, $w_1 = -0.1$, $w_2 = 0.1$. Set different values to increase the difference in weights.

### B. Asynchronous Advantage Actor-Critic Algorithm

Considering the complexity and the numerous configuration parameters of the SAGIN-based vehicular networks, it is very suitable to use A3C algorithm in vehicle access selection network to improve efficiency and reduce convergence delay. So, this paper utilizes an A3C algorithm to address vehicle access network selection problem in SAGIN. In the A3C algorithm, each sub-network will upload the learning results to the global network after interacting with the environment, and the global network will distribute the shared parameters to each sub-network, and the sub-network will continue to interact with the environment after obtaining the updated parameters.

Although the A3C algorithm is a good match for this network, as to be shown as follows, there are still some issues to be addressed in the straightforward implementation. In the vehicle access selection network, each access network has multiple channels for vehicles to select. Between adjacent channels of the access network, partial power will scatter in the signal transmission process, which is related to the transmit power that cannot be too small to be ignored. Therefore, this paper must consider the influence of adjacent sub-channel interference on the vehicle network access selection process.

Besides, in the early stage of training, due to the small number of samples that the agent can obtain, it will cause the problem of difficult training and inaccurate results. And interacting with the physical network directly using the unreliable results runs the risk of degrading network performance. Meanwhile, in the process of sample acquisition, the agent needs to continuously interact with the environment, which consumes a lot of general computing and storage resources in terms of time and security. As a consequence, it is necessary to solve the reward sparsity in reinforcement learning.

### C. Knowledge Integration

To address the issues in the training phase of the straightforward implementation, this paper proposes the knowledge-driven vehicle access algorithm. Specifically, the integrated knowledge includes 1) the space-time correlation knowledge 2) the mathematics knowledge. With the help of these knowledge, the improved algorithm can not only solve the problems caused by only applying A3C algorithm but also improve the performance of the network.

TABLE II: KNOWLEDGE INTRODUCTION

| Approach | Knowledge | Advantages | Integration | Problems |
|---|---|---|---|---|
| Channel choice | Space-time correlation knowledge | Reduce action space | Improve learning algorithm | Inter-channel interference |
| Reward reshape | Mathematics knowledge | Accelerate covergence speed | Improve learning algorithm | Sparse reward |

*1) Channel Choice:* In order to reduce the interference of adjacent channels, this paper proposes a method of channel selection by using the relevant knowledge of communication. When the vehicle selects the channel, it will select the channel whose adjacent channels are all idle, until all the channels that meet the requirements are selected. There will be a random channel selected for access among the idle channels where the adjacent channel interferes with the adjacent channel.

*2) Reward Reshape:* The reward design and learning method adopted in this paper is to artificially set the reward function. In the original algorithm, if the task fails to complete the transmission in this time slot, even if it successfully transmits part of the service in this time slot, the reward it gets is still 0, and only when the transmission completes the whole task will there be a positive reward. Exploring optimization strategies by delaying rewards consumes a lot of time. To address the reward sparsity problem, this paper applies reward shaping in generating non-zero immediate rewards per time slot. Referring [7], this paper also defines a potential function $\Phi\left(d_k(t)\right)$, and the agent will get a reward for falling from a high potential to a low potential. The specific function is

$$r_k^{'}(t) = r_k(t) - \Phi\left(\tau_k(t)\right) + \gamma\Phi\left(\tau_k(t+1)\right), \qquad (18)$$

where $\tau_k(t)$ is the delay of the $k$-th vehicle at time slot $t$ and the value of $\gamma$ in this paper is 0.9. The potential-based function $\Phi\left(\tau_k(t)\right)$, a linear function, can be expressed as

$$\Phi\left(\tau_k(t+1)\right) = \frac{\Phi_{\max} - \Phi_{\min}}{\tau_{\min}}\tau_k(t) + \Phi_{\min}, \qquad (19)$$

where $\tau_{\min}$ is a minimum delay bound, if the delay is smaller than it, the reward of the system is zero. And the $\Phi_{\max}$ and $\Phi_{\min}$ are artificially designed parameters.

To summarize, the knowledge-driven A3C algorithm is outlined as Algorithm 1.

## IV. NUMERICAL RESULTS

In this section, extensive numerical results are conducted to evaluate the proposed knowledge-driven vehicle access approch. Specifically, this section first elaborates on the parameter settings. Afterward, the performance evaluation of the proposed method is carried out.

**Algorithm 1** Knowledge-Driven A3C

Set the global shared parameter vectors $\theta$ and $\theta_v$.
Set the specific parameter vectors $\hat{\theta}$ and $\hat{\theta_v}$.
Initialize the parameters of the NNs, e.g. the learning rate of the sub-nets.
Initialize the global counter $N = 0$.
Set the network to an initial state.
**for** $N<N_{max}$ **do**
    Reset the gradients: $\mathrm{d}\theta \leftarrow 0$ and $\mathrm{d}\theta_v \leftarrow 0$.
    Update the parameter vectors $\hat{\theta} = \theta$ and $\hat{\theta_v} = \theta_v$.
    Initialize the thread step counter $n \leftarrow 0$ and $n_0 \leftarrow 0$.
    **for** $k \leq K$ **do**
        Get the state $s(t)$.
        Calculate the SNR of the link from (1)-(6), where $\delta_{interf}$ depends on the number of the chosen adjacent channels.
        Choose action $a_t$ from policy network $\pi(a_n|s_n;\hat{\theta})$ and execute the action.
        Calculate the reshaped delay $\tau(t)_k^{'}$ from (19).
    **end for**
    Receive the reward $r(t)$ from (17).
    Update counter $n \leftarrow n+1$, $N \leftarrow N+1$.
    Update vectors $\theta \leftarrow \mathrm{d}\theta$ and $\theta_v \leftarrow \mathrm{d}\theta_v$.
**end for**

### A. Parameter Settings

In the experiments, a straight road with a length of 60km is considered, and 4 BSs, 2 UAVs, as well as 1 LEO are interspersed on the entire road. The UAVs locate at 20km and 50km of the road respectively, flying at a constant speed and a constant altitude. When one UAV flying two kilometers to the end, a new one will immediately take off from the initial position to ensure the stable communication function of the UAV network. And other experiment parameters are listed in Table III.

TABLE III: PARAMETER SETTINGS

| Parameters | Value | Parameters | Value |
|---|---|---|---|
| Number of RB | 6 | BS coverage radius | 500m |
| UAV coverage radius | 2km | LEO coverage radius | 30km |
| Vehicle speed | 36km/h | RB bandwidth | 150K |
| UAV speed | 72km/h | UAV fight altitude | 100m |
| Vehicle transmission power | 20dBm | Vehicle channel bandwidth | 1MHz |
| LEO channel bandwidth | 20MHz | AWGN power | -114dBm |

### B. Performance of the Network

*1) The Impact of Modifying the Learning Rate:* The initial learning rate of the network is set to $0.0001$. It can be seen from Fig. 2 that the workers in the sub-network can finally achieve convergence regardless of whether they follow the same learning rate or different learning rates, but the workers whose learning rate decreases according to the exponential of $0.9$ have the fastest convergence speed.

*2) The Impact of Knowledge Integration:* In the experiment, the vehicle access process adopts three ways to improve the learning method: reward reshaping, channel selection and changing the learning rate and the final result is shown in Fig.

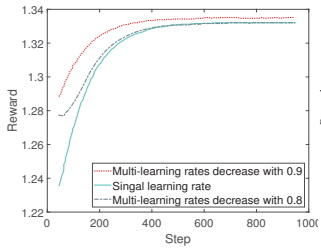


Fig. 2: Channel choice numerical results.

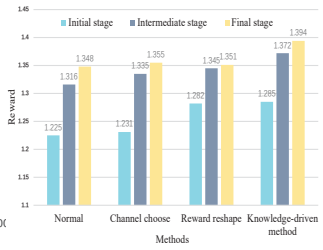


Fig. 3: Knowledge integration experiment results.

3. The learned reward value of the knowledge-driven training algorithm is the highest. And the results show that integrating knowledge can effectively reduce about $4\%$ convergence delay in the vehicle access process.

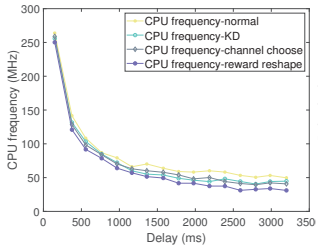


Fig. 4: Results of the relationship between different CPU frequencies and delay.

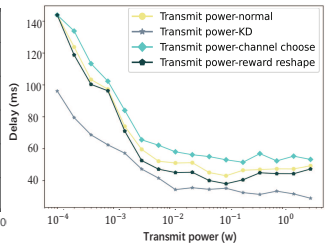


Fig. 5: Results of relation between different transmit power and delay.

*3) The Impact of Different CPU Frequencies:* Fig. 4 shows the relationship between different CPU frequencies and rewards. As can be seen from the figure, the delay decreases with the increase of CPU frequency. But latency does not improve indefinitely with increasing CPU frequency. In this simulation scenario, when the user's computing frequency rises to 2.4GHz, the delay curve tends to be stable and there is no improvement. This is mainly because the total system delay is composed of queuing delay, computation delay and transmission delay. Increasing the computation frequency can only reduce the computation delay, but has no effect on queuing delay and transmission delay. Therefore, when the computation frequency is high enough, the other two kinds of delay dominate the total delay, and the delay does not change with the increase of CPU frequency.

*4) The Impact of Different Transmit Power:* Fig. 5 illustrates the relationship between different transmit power and reward. It can be seen from the figure that the delay decreases with the increase of the transmit power. However, increasing the transmit power does not affect the calculation delay and transmission delay, so when the transmit power increases to a certain extent, the total delay tends to converge and does not change with the increase of the transmit power.

## V. Conclusion

This paper has focused on how to minimize the delay of vehicle service transmission service in the access selection of vehicles in SAGIN-based vehicular networks. In detail, an A3C-based knowledge-driven vehicle access approach has been presented to optimize the access decision of the network to improve the network performance by integrating the space-time correlation knowledge and mathematics knowledge. Numerical results have shown that the proposed knowledge-driven approach outperformed the original method without integrated knowledge in terms of access delay and network convergence.

## Acknowledgment

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1807700, in part by the National Natural Science Foundation of China (NSFC) under Grant 62201414, 62071356 and 62201310, and in part by the Chongqing Key Laboratory of Mobile Communications Technology under Grant cqupt-mct-202202.

## References

[1] N. Cheng, W. Quan, W. Shi, H. Wu, Q. Ye, H. Zhou, W. Zhuang, X. Shen, and B. Bai, "A comprehensive simulation platform for space-air-ground integrated network," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 178–185, Feb. 2020.

[2] B. Cao, J. Zhang, X. Liu, Z. Sun, W. Cao, R. M. Nowak, and Z. Lv, "Edge–cloud resource scheduling in space–air–ground-integrated networks for internet of vehicles," *IEEE Internet Things J.*, vol. 9, no. 8, pp. 5765–5772, Mar. 2021.

[3] C. She, C. Sun, Z. Gu, Y. Li, C. Yang, H. V. Poor, and B. Vucetic, "A tutorial on ultrareliable and low-latency communications in 6G: Integrating domain knowledge into deep learning," *in Proc. IEEE Inst Electr Electron Eng*, vol. 109, no. 3, pp. 204–246, Mar. 2021.

[4] X. Guo, M. H. Omar, K. M. Zaini, G. Liang, M. Lin, and Z. Gan, "Multiattribute access selection algorithm for heterogeneous wireless networks based on fuzzy network attribute values," *IEEE Access*, vol. 10, pp. 74 071–74 081, May 2022.

[5] A. Mchergui and T. Moulahi, "A novel deep reinforcement learning based relay selection for broadcasting in vehicular Ad Hoc networks," *IEEE Access*, vol. 10, pp. 112–121, Dec. 2022.

[6] H. Gačanin, E. Perenda, S. Karunaratne, and R. Atawia, "Self-optimization of wireless systems with knowledge management: An artificial intelligence approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 9682–9697, Jun. 2019.

[7] Z. Gu, C. She, W. Hardjawana, S. Lumb, D. McKechnie, T. Essery, and B. Vucetic, "Knowledge-assisted deep reinforcement learning in 5G scheduler design: From theoretical framework to implementation," *IEEE J. Select. Areas Commun.*, vol. 39, no. 7, pp. 2014–2028, Sep. 2020.

[8] R. Sun, Y. Wang, L. Lyu, N. Cheng, S. Zhang, T. Yang, and X. Shen, "Delay-oriented caching strategies in D2D mobile networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8529–8541, May 2020.

[9] Y. Che, F. Lin, and J. Liu, "Deep reinforcement learning in M2M communication for resource scheduling," in *in Proc. 2021 World Conf. Comput. Commun. Technol. (WCCCT)*, Jan. 2021, pp. 97–100.

[10] H. Liao, Z. Zhou, W. Kong, Y. Chen, X. Wang, Z. Wang, and S. Al Otaibi, "Learning-based intent-aware task offloading for air-ground integrated vehicular edge computing," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 8, pp. 5127–5139, Oct. 2020.

[11] M. Qin, W. Wu, Q. Yang, R. Zhang, N. Cheng, H. Zhou, R. R. Rao, and X. Shen, "Green-oriented dynamic resource-on-demand strategy for multi-RAT wireless networks powered by heterogeneous energy sources," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5547–5560, Aug. 2020.

[12] N. Cheng, F. Lyu, W. Quan, C. Zhou, H. He, W. Shi, and X. Shen, "Space/aerial-assisted computing offloading for IoT applications: A learning-based approach," *IEEE J. Select. Areas Commun.*, vol. 37, no. 5, pp. 1117–1129, Mar. 2019.

[13] D. Zhou, M. Sheng, Y. Wang, J. Li, and Z. Han, "Machine learning-based resource allocation in satellite networks supporting internet of remote things," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 10, pp. 6606–6621, Apr. 2021.