This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3163430, IEEE Transactions on Vehicular Technology

Energy-Efficient Resource Allocation for UAV-Assisted Vehicular Networks with Spectrum Sharing

Weijing Qi, Qingyang Song, *Senior Member, IEEE*, Lei Guo, and Abbas Jamalipour, *Fellow, IEEE*

***Abstract*—Vehicular networks are envisioned to deliver data transmission services ubiquitously, especially in the upcoming autonomous driving era. Accordingly, the high data traffic load poses a heavy burden to the terrestrial network infrastructure. Unmanned Aerial Vehicles (UAVs) show enormous potential to assist vehicular networks in providing services. In a dense UAV-assisted vehicular network with a large number of users, spectrum sharing is leveraged for alleviating the spectrum scarcity. However, the increasing data traffic still leads to the UAV energy consumption problem. This paper considers a specific network scenario where a UAV transmits its cached content files to vehicular users over UAV-to-vehicle (U2V) links while vehicle-to-vehicle (V2V) links reuse the U2V spectrum for safety-critical message exchanges. To improve the UAV's energy efficiency while guaranteeing users quality of service (QoS), we jointly optimize content placement, spectrum allocation, co-channel link pairing, and power control, which are the key factors affecting energy efficiency and QoS. The joint optimization problem is formulated as a mixed-integer nonlinear programming (MINLP) problem, which is solved by combining Hungarian and DDQN. We perform system performance evaluations, demonstrating that our approach can not only improve the UAV's energy efficiency while satisfying the users' QoS requirements but also increase the timeliness of making decisions.**

***Index Terms*—UAV, vehicular networks, spectrum sharing, resource allocation, reinforcement learning**

I. Introduction

As a typical application of Internet of Things (IoT), intelligent transportation attracts broad attention both from academia and industry and develops fast. Advanced communication technologies such as Cellular Vehicle-to-Everything (C-V2X) and Dedicated Short Range Communication (DSRC) can achieve a vehicular network, helping vehicles to interact seamlessly with other vehicles, road infrastructures, pedestrians, and the Internet. With the flourishing of advanced vehicular applications especially in the upcoming autonomous driving era, vehicular networks are envisioned to deliver high-bandwidth services ubiquitously. The heavy traffic burden poses a great challenge to the terrestrial network infrastructure which is expected to guarantee the quality of vehicles' services [1]. This situation will be worse on busy roads and during rush hours. In addition, a sudden disaster may cause the terrestrial communication system breakdown. Unmanned Aerial Vehicles (UAVs) have good covering ability, disaster tolerance capability, and flexible deployment advantages, showing enormous potential in alleviating the data traffic burden of terrestrial infrastructures and improving network invulnerability. Therefore, UAVs become an attractive complementary to the terrestrial infrastructure in vehicular networks.

Recently, UAV-assisted vehicular networks where UAVs play the roles of aerial base stations, relays, or edge computing/caching servers have been extensively studied [2]. In some works [3–7] that take UAVs as relays or base stations, the researchers have primarily focused on how to improve system performance in terms of throughput, packet delivery ratio, and end-to-end delay. The authors in [4] took UAVs as flying base stations responsible for routing incoming vehicular data. Through analyzing the vehicle-to-UAV (V2U) path availability including the connectivity probability and connection time, the vehicle connectivity is improved. Detection of malicious vehicles is also done during the routing process in [5]. Furthermore, the UAV-assisted computing offloading problem is studied in [8, 9], where UAVs are playing a remarkable role in vehicular computation offloading. The system energy cost and the number of offloaded tasks have been taken into consideration. Moreover, by caching large-size files (e.g., high-resolution map, video, etc.) at the edge of the network, UAV caching has been proved enormous potential in reducing the traffic burden of backhaul links and the content delivery latency significantly [10–12]. The authors in [10] studied a joint caching and trajectory optimization problem aiming at maximizing the overall network throughput. While the objectives in [11] and [12] are to minimize the transmission delay and maximize the number of served vehicles, respectively.

Despite their extensive applications, UAVs typically have limited onboard energy due to their relatively low payload, significantly affecting the system performance. Some work on reducing UAVs' energy consumption [13] has been done. Take a scenario where a UAV plays as an aerial base station that transmits content files for vehicular users within its communication range. To cope with the conflict between the massive data transmission demand and low-quality UAV-to-

This work was supported by the National Key Research and Development Program of China under Grant 2018YFE0206800, the National Natural Science Foundation of China under Grant 62025105, the Chongqing Municipal Education Commission under Grant CXQT21019, and the Natural Science Foundation of Chongqing under Grant cstc2020jcyj-msxmX0918.

W. Qi, Q. Song, L. Guo are with the Institute of Intelligent Communication and Network Security, School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, China.

A. Jamalipour is with the School of Electrical and Information Engineering, The University of Sydney, Sydney, Australia.

The corresponding author is Q. Song: songqy@cqupt.edu.cn

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3163430, IEEE Transactions on Vehicular Technology

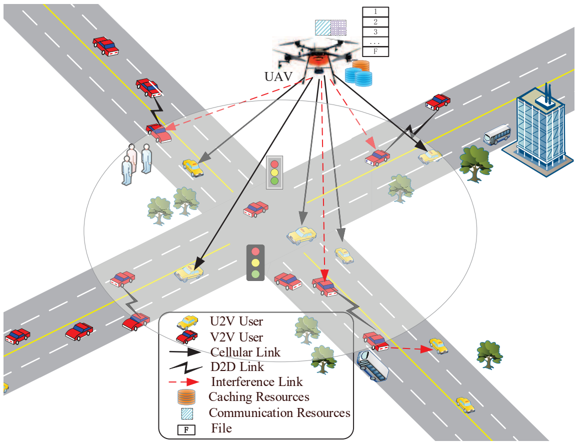


Fig. 1. Illustration of UAV-assisted vehicular network with spectrum sharing.

vehicle (U2V) channel conditions caused by vehicles' high-speed movements, UAVs always have to increase the transmission power for gaining high data rates. This causes severe interference and high energy consumption. How to take full advantage of energy is usually evaluated by energy efficiency, defined as the ratio of total throughput to total power consumption. Hence energy efficiency becomes a more important metric than energy consumption in UAV-assisted vehicular networks. How to maximize energy efficiency while fullfilling the diverse QoS requirements of users is an urgent problem to be solved. The optimal resource allocation such as spectrum and transmission power allocation for U2V communications has been proved effective to solve this problem [14–16]. In addition, since the UAV has limited storage resources, a good content placement policy with a high cache-hit rate is also beneficial to the system throughput improvement and consequently to the UAV's energy efficiency improvement [10, 12]. Traditional resource allocation schemes are based on heuristic algorithms. For example, the resource allocation approach in [17] considers the large-scale fading effect and applies a heuristic solution for solving the power allocation. However, these algorithms' complexity is usually high, and it takes a long time to find an optimal solution. They are obviously not suitable for a highly dynamic vehicular network which needs real-time decisions. Deep reinforcement learning (DRL) is proved to be effective for solving energy problem in vehicular networks. The authors in [18] proposed a DRL framework to minimize UAVs average energy consumption. While in [19], the authors introduced a DRL algorithm named asynchronous advantage actor-critic (A3C) to minimize service execution delay, reduce energy consumption and maximize the system throughput.

Another increasingly prominent problem in UAV-assisted vehicular networks is spectrum scarcity with the exponentially growing number of vehicles and data explosion. Effective spectrum sharing can improve spectrum efficiency [20, 21]. Hence the network scenario we focus on is extended to a system supporting spectrum sharing. Specifically, as shown in Fig. 1, the system contains two types of users: yellow vehicles requiring high-bandwidth content services served by U2V links; red vehicles requiring high-reliability information-sharing services served by V2V links. A UAV preloads popular content files so that the users in its coverage can download high-bandwidth content files from the UAV. Meanwhile, each V2V link does not occupy an additional frequency spectrum but reuses the frequency spectrum of a U2V link for improving the spectrum efficiency. However, most current related works focus on U2V links without spectrum sharing. UAV-assisted vehicular networks with spectrum sharing still face the following challenges, which are seldom studied:

1) Co-channel interference between U2V and V2V links caused by spectrum sharing must be carefully managed to satisfy diverse quality of service (QoS) requirements.
2) Spectrum sharing involves a Co-channe link pairing problem between U2V and V2V links. The matching must be jointly optimized with spectrum and power allocation from a UAV energy efficiency perspective.
3) In such a highly dynamic vehicular network, optimization decisions should be made in real time.

Therefore, for the network system in Fig. 1 described above, we propose a resource (including storage, spectrum, and power) allocation scheme in this paper, aiming to maximize the UAV's energy efficiency while satisfying vehicles' QoS requirements. Taking the interference between U2V and V2V links caused by spectrum sharing into account, resource allocation is formulated as an optimization problem, in which the objective is to maximize the long-term energy efficiency of the UAV. Based on the scenario, the problem subjects to the orthogonality between resource blocks (RBs), the UAV's maximum transmission power, and the users' QoS requirements. Containing numerous binary variables, the problem here is a mixed-integer nonlinear programming (MINLP) problem, which is challenging to solve. Therefore, we decompose it into two sub-problems: spectrum allocation and power allocation. And then, we explore the power of combining classical optimization and reinforcement learning to solve these two sub-problems. Specifically, we can obtain the optimal RB allocation solution and the link pairing result by implementing two rounds of the Hungarian algorithm. After that, a Double Deep Q-learning Network (DDQN)-based algorithm is used to realize the optimal power allocation. We perform extensive simulations, which show that our approach significantly outperforms other existing algorithms that address the same problem. The contributions of our work can be summarized as follows:

1) We propose an energy-efficient resource allocation strategy in the UAV-assisted vehicular network with spectrum sharing. To our knowledge, this is the first study on optimization of content placement, spectrum allocation, co-channel link pairing, and power control, which are all factors affecting energy efficiency and QoS.
2) The joint optimization problem is formulated as a mixed-integer nonlinear programming (MINLP) problem. Through solving the problem by combining Hungarian and DDQN, improved UAV energy efficiency is achieved. Taking the efficient Hungarian algorithm as an initial step can reduce the state and action space of DDQN, thus improving the convergence speed of DDQN.

3) We perform extensive simulations in different settings, and also compare our proposed approach with traditional methods. Results demonstrate its effectiveness in increasing the timeliness in real-time decisions, improving the UAV's energy efficiency, and adapting to rapid environmental changes in a way that satisfies users' QoS requirements.

The rest of this paper is organized as follows. We present the system model and formulate the energy efficiency optimization problem in Section II. In Section III, we decompose the problem into two sub-problems and propose a combined resource allocation algorithm. The simulation results are given in Section IV. Finally, we conclude our work in Section V.

II. System Model and Problem Formulation

This section introduces the UAV-assisted vehicular network system with spectrum sharing, where the network architecture and communication model are introduced. In this system, to improve the UAV's energy efficiency for QoS guarantee, we optimize the allocation of the UAV's caching, spectrum, and power resources. Note that the caching resource optimization is done by selecting a batch of files that yield the best system performance to be cached at the UAV. The spectrum resource optimization includes a U2V spectrum allocation problem and a link paring problem between the U2V and V2V links that use the same spectrum. We then formulate the resource allocation as an optimization problem with the objective of maximizing the long-term energy efficiency under diverse QoS constraints. Finally, we propose a content placement and updating strategy so that the problem can be simplified.

A. Network Architecture

We consider a UAV-assisted vehicular network with spectrum sharing as shown in Fig. 1. In this network scenario, the UAV, working as an aerial base station, is assumed to be nearly static, hovering at a fixed altitude. U2V and V2V connections are supported, and the V2V links reuse the spectrum of the U2V links.We imagine K vehicle users denoted as U2V-UEs require high-capacity U2V links, and M pairs of vehicle users denoted as V2V-UEs perform V2V data exchanges. Let $\mathcal{K} = \{1, 2, ..., K\}$ and $\mathcal{M} = \{1, 2, ..., M\}$ represent the U2V-UE set and V2V-UE pair set, respectively. Assume that there are a total of F content files $\mathcal{F} = \{1, 2, ..., F\}$ in the content server that users are interested in. The size of the f^{th} file is C_f bits. The UAV with a limited caching capacity of C_U bits is preloaded with content files $\tilde{\mathcal{F}} \in \mathcal{F}$. We denote the caching state of the f^{th} file as $s_f \in \{0, 1\}$. Specifically, $s_f = 1$ if $f \in \tilde{\mathcal{F}}$ holds, and $s_f = 0$, otherwise. We define $\beta_{k,f} \in \{0, 1\}$ which represents whether the k_{th} U2V-UE sends a request to the UAV for the f_{th} file ($\beta_{k,f} = 1$) or not ($\beta_{k,f} = 0$). Note that a U2V-UE obtains its requested file from the UAV directly if it is cached in the UAV, and otherwise, it has to request the file from a base station. In time-division dual (TDD)-orthogonal frequency-division multiplexing (OFDM) 5G systems, transmissions are scheduled in the group of 12 subcarriers in the frequency domain, known as a new radio (NR) RB. The RB bandwidth is not fixed and depends on sub-carrier spacing. We assume that the UAV can assign one RB to at most one user and $\mathcal{N} = \{1, 2, ..., N\}$ is the set of RBs in the system. To improve spectrum utilization efficiency, V2V-UEs reuse the downlink spectrum orthogonally allocated for U2V-UEs. The maximum allowed transmission power of the UAV when sending data to the k^{th} U2V-UE is P_k^{max}.

TABLE I
Notation list

Symbol	Definition
$\mathcal{K}$	Set of U2V-UEs in the network.
$\mathcal{M}$	Set of V2V-UEs in the network.
$\mathcal{F}$	Set of content files that users are interested in.
$\tilde{\mathcal{F}}$	Set of content files preloaded by the UAV.
$\mathcal{T}$	Set of time slots.
C_f	Size of the f_{th} file in bits.
C_U	UAV's caching capacity in bits.
s_f	Binary parameter denoting whether the f_{th} file is in the UAV.
$\beta_{k,f}$	Binary parameter denoting whether the k_{th} U2V-UE sends a request to the UAV for the f_{th} file.
$\mathcal{N}$	Set of available RBs in the network.
$g_{k,n}$	Channel gain between the UAV and the k_{th} U2V-UE on RB n.
$h_{k,n}$	Fast fading component of $g_{k,n}$.
$g_{m,k}$	Channel gain between the m_{th} pair of V2V-UEs which reuses the k_{th} V2V-UE's RB.
$\widetilde{g}_{m,k}$	Interference channel gain from the m_{th} V2V-UE to the k_{th} U2V-UE.
$\widetilde{g}_{k,m}$	Interference channel gain from the UAV to the m_{th} V2V-UE receiver on the RB of the k_{th} U2V-UE.
$\gamma_{k,n}$	Received instantaneous downlink SINR at the k^{th} U2V-UE on RB n.
$p_{k,n}$	Transmission power of the k^{th} U2V-UE on RB n.
$p_{m,k}$	Transmission power of the m^{th} V2V-UE sharing the k^{th} U2V-UE's RB.
$\rho_{m,k}$	An end-to-end path from node s_k to the RSU nearby.
$\delta_{k,n}$	Binary variable indicating whether the UAV allocates RB n to the k^{th} U2V-UE ($\delta_{k,n} = 1$) or not ($\delta_{k,n} = 0$).
γ_k	Received instantaneous downlink SINR at the k^{th} U2V-UE.
$\gamma_{m,k}$	SINR of the m^{th} V2V-UE sharing the same spetrum of the k^{th} U2V-UE.
γ_m	SINR of the m^{th} V2V-UE.
$D_{k,f}^T$	Transmisison delay when the k^{th} U2V-UE requests the f^{th} file which is cached in the UAV.
$D_{k,f}^F$	Fronthaul delay for the k^{th} U2V-UE requesting the f^{th} file which is not cached in the UAV.
$D_{k,f}$	Delay of the k^{th} U2V-UE requesting for the f^{th} file.
Q_k	The k^{th} U2V-UE's request queue length at time slot t.
A	Constant request arrival rate per time slot.
b_f	Cache placement control action.
P_k^{total}	Total power consumption of the k^{th} U2V-UE.
P^C	Circuit power transmitting data, incurred by signal processing and active circuit blocks.
R_k^{min}	Minimum transmission capacity requirement of the k^{th} U2V-UE.
γ_m^0	Minimum SINR needed by the m^{th} V2V-UE to establish a reliable link.
Pr^0	Tolerant outage probability of V2V links.

The network system works over a time window, which is divided into discrete time slots denoted by $t \in \mathcal{T} = \{0, 1, 2, ..., T-1\}$. According to its resource allocation strategy, the UAV allocates RBs and power for vehicles that send data transmission requests at the beginning of each time slot.

B. Communicaiton Model

Taking the channel fading into account, we describe the channel gain between the UAV and the k_{th} U2V-UE on RB

n as follows:

$$g_{k,n} = |h_{k,n}|^2 \alpha_{k,n} \tag{1}$$

where $h_{k,n}$ is the fast fading component and $\alpha_{k,n}$ captures path loss and shadow fading. The channel gain $g_{m,k}$ between the m_{th} pair of V2V-UEs which reuses the k_{th} U2V-UE's RB is similarly defined. We define $\widetilde{g}_{m,k}$ as the interference channel gain from the m_{th} V2V-UE to the k_{th} U2V-UE caused by sharing the same RB. We make $\widetilde{g}_{k,m}$ as the interference channel gain from the UAV to the m_{th} V2V-UE receiver on the RB of the k_{th} U2V-UE.

We assume the channel state information (CSI) of vehicle-to-UAV (V2U) links can be estimated at the UAV. Due to the channel reciprocity in TDD-OFDM systems, the CSI of downlinks can be seen as the same as that of uplinks. That is, $g_{k,n}$ and $\widetilde{g}_{m,k}$ are known. At the same time, the CSI of vehicular links is fed back to the UAV with a delay. The channel variation over the feedback delay is modeled as follows [22]:

$$h = \varepsilon \hat{h} + e \tag{2}$$

where $\hat{h}$ and h are the channel fast fading component in the previous (before the feedback delay) and current time slot, e is distributed according to $\mathcal{CN}(0, 1-\varepsilon^2)$. The channel correlation between the two consecutive time slots is reflected in ε, which is given by $\varepsilon = J_0(2\pi f_d T)$. $J_0(\cdot)$ is the zero-order Bessel function where the maximum Doppler frequency $f_d = v f_c / c$ with the speed of light $c = 3 \times 10^8$ m/s, carrier frequency f_c and vehicle speed v.

The received instantaneous signal to interference and noise ratio (SINR) at the k^{th} U2V-UE on RB n is given as follows:

$$\begin{aligned}\gamma_{k,n} &= \frac{p_{k,n} g_{k,n}}{\sum_{m=1}^{M} \rho_{m,k} p_{m,k} \widetilde{g}_{m,k} + N_0 B} \\ &= \frac{p_{k,n} |h_{k,n}|^2 \alpha_{k,n}}{\sum_{m=1}^{M} \rho_{m,k} p_{m,k} |\widetilde{h}_{m,k}|^2 \widetilde{\alpha}_{m,k} + N_0 B}\end{aligned} \tag{3}$$

where $p_{k,n}$ and $p_{m,k}$ are the transmission power of the k^{th} U2V-UE on RB n and the m^{th} V2V-UE sharing the k^{th} U2V-UE's RB, respectively. $\rho_{m,k}$ indicates whether the m^{th} V2V-UE reuses the spetrum of the k^{th} U2V-UE ($\rho_{m,k} = 1$) or not ($\rho_{m,k} = 1$). N_0 is the power spectral density of the additive white Gaussian noise with zero mean and variance σ^2. The bandwidth of each RB is B.

We define a binary variable $\delta_{k,n}$ to indicate whether the UAV allocates RB n to the k^{th} U2V-UE ($\delta_{k,n} = 1$) or not ($\delta_{k,n} = 0$). The received instantaneous downlink SINR at the k^{th} U2V-UE is given as follows:

$$\gamma_k = \frac{\sum_{n=1}^{N} \delta_{k,n} p_{k,n} |h_{k,n}|^2 \alpha_{k,n}}{\sum_{m=1}^{M} \rho_{m,k} p_{m,k} |\widetilde{h}_{m,k}|^2 \widetilde{\alpha}_{m,k} + N_0 B} \tag{4}$$

The SINR of the m^{th} V2V-UE sharing the same spetrum of the k^{th} U2V-UE is given as follows:

$$\begin{aligned}\gamma_{m,k} &= \frac{p_{m,k} g_{m,k}}{\sum_{n=1}^{N} \delta_{k,n} p_{k,n} \sum_{k=1}^{K} \rho_{m,k} \widetilde{g}_{k,m} + N_0 B} \\ &= \frac{p_{m,k} \alpha_{m,k} (\varepsilon_m^2 |\hat{h}_{m,k}|^2 + |e_m|^2)}{\sum_{n=1}^{N} \sum_{k=1}^{K} \delta_{k,n} \rho_{m,k} p_{k,n} \alpha_{k,m} (\varepsilon_k^2 |\hat{h}_{k,m}|^2 + |e_k|^2) + N_0 B}\end{aligned} \tag{5}$$

Given the association between a single U2V-UE and a V2V-UE pair, the SINR of the m^{th} V2V-UE is expressed as follows:

$$\gamma_m = \frac{\sum_{k=1}^{K} \delta_{k,n} p_{m,k} \alpha_{m,k} (\varepsilon_m^2 |\hat{h}_{m,k}|^2 + |e_m|^2)}{\sum_{n=1}^{N} \sum_{k=1}^{K} \delta_{k,n} \rho_{m,k} p_{k,n} \alpha_{k,m} (\varepsilon_k^2 |\hat{h}_{k,m}|^2 + |e_k|^2) + N_0 B} \tag{6}$$

The achievable transmission data rate of the k^{th} U2V-UE at time slot t is calculated by the Shannon equation, as follows:

$$R_k(t) = B log_2(1 + \gamma_k(t)) \tag{7}$$

where $\gamma_k(t)$ is the received SINR at the k^{th} U2V-UE at time slot t.

If the k^{th} U2V-UE requests the f^{th} file which is cached in the UAV at time slot t, the transmisison delay is caculated as follows:

$$D_{k,f}^{T}(t) = \frac{C_f}{R_k(t)} \tag{8}$$

where C_f is the size of the f^{th} file.

If the file is not cached in the UAV, the user has to fetch the file from the associcated base station through a fronthaul link. Besides the transmission delay, we need to consider an added fronthaul delay, which is usually related to the average link distance and traffic load in practice. For simplification, we assume the fronthaul delay as a fixed value D_F. The fronthaul delay of the k^{th} U2V-UE requesting the f^{th} file at time slot t can be caculated as:

$$D_{k,f}^{F}(t) = (1 - s_{k,f}) D_F \tag{9}$$

Consequently, the delay for the k^{th} U2V-UE requesting for the f^{th} file at time slot t can be written as:

$$D_{k,f}(t) = \frac{C_f}{R_k(t)} + (1 - s_{k,f}) D_F \tag{10}$$

C. Problem Formulation

The energy efficiency of the UAV for transmitting data to the k^{th} U2V-UE is:

$$\eta_k^{ee} = \frac{R_k}{P_k^{total}} = \frac{B log(1 + \gamma_k)}{\sum_{n=1}^{N} \delta_{k,n} (p_{k,n} + P^C)} \tag{11}$$

where P_k^{total} is the total power consumption of the k^{th} U2V-UE, which includes the transmission power $p_{k,n}$ and circuit power P^C.

Considering the caching state of the UAV, the UAV's energy efficiency of the whole system at time slot t can be calculated by the following formulation:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3163430, IEEE Transactions on Vehicular Technology

$$\eta^{ee}(t) = \frac{\sum_{k=1}^{K} \beta_{k,f}(t) s_f B log_2(1+\gamma_k(t))}{\sum_{k=1}^{K} \beta_{k,f}(t) s_f \sum_{n=1}^{N} \delta_{k,n}(t)(p_{k,n}(t)+P^C)} \tag{12}$$

We formulate an optimization problem to maximize the long-term energy efficiency by jointly optimizing the spectrum and power resource allocation for vehicle users and content placement at the UAV. The orthogonality of OFDMA, maximum energy consumption of the UAV, and diverse QoS requirements of different users, i.e., high data rates of U2V-UEs and high reliabilities of V2V-UEs are considered as constraints. Let $S=\{s_f,\forall f\}$, $\Delta=\{\delta_{k,n}(t),\forall k,n\}$, $X=\{\rho_{m,k}(t),\forall m,k\}$, $P=\{p_{k,n}(t),\forall k,n\},\{p_{m,k}(t),\forall m,k\}$ be the file caching state matrix in the UAV, RB allocation matrix for U2V-UEs, spetrum reusing association matrix, and UAV transmission power matrix, respectively. The problem can be formulated as:

$$(P1)$$

$$\max_{\{S,\Delta,X,P\}} \frac{1}{T}\sum_{t=0}^{T-1} \eta^{ee}(t) \tag{13}$$

$$s.t.\quad R_k(t) \geq R_k^{min}, \forall k \in K \tag{13a}$$

$$Pr\{\gamma_m(t) \geq \gamma_m^0\} \leq Pr^0, \forall m \in M \tag{13b}$$

$$\frac{1}{K}\sum_{k=1}^{K} D_{k,f}(t) \geq D^{thr} \tag{13c}$$

$$0 \leq p_{k,n}(t) \leq P_k^{max}, \forall n \in N, \forall k \in K \tag{13d}$$

$$0 \leq p_{m,k}(t) \leq P_m^{max}, \forall m \in M, k \in K \tag{13e}$$

$$\sum_{f=1}^{F} s_f C_f \leq C_U, \tag{13f}$$

$$\sum_{k=1}^{K} \rho_{k,m}(t) \leq 1, \forall m \in M \tag{13g}$$

$$\sum_{m=1}^{M} \rho_{k,m}(t) \leq 1, \forall k \in K \tag{13h}$$

$$\sum_{k=1}^{K} \delta_{k,n}(t) \leq 1, \forall n \in N \tag{13i}$$

$$\rho_{m,k}(t) \in \{0,1\}, \forall m \in M, \forall k \in K \tag{13j}$$

$$\delta_{k,n}(t) \in \{0,1\}, \forall k \in K, \forall n \in N \tag{13k}$$

$$s_f \in \{0,1\}, \forall f \in F \tag{13l}$$

where R_k^{min} is the minimum transmission capacity requirement of the k^{th} U2V-UE. γ_m^0 is the minimum SINR needed by the m^{th} V2V-UE to establish a reliable link and Pr^0 is the tolerant outage probability. $Pr\{\cdot\}$ represents the probability of the input. D^{thr} is the threshold value of the average delay of all U2V-UEs. A relative high cache hit rate can be guranteed through tuning this parameter. P_k^{max} is the maximum transmission power of the UAV when sending data to the k^{th} U2V-UE and P_m^{max} is the maximum transmission power of the m^{th} V2V-UE sender. C_U is the caching capacity of the UAV.

Constraints (13a) and (13b) represent the minimum transmission data rate requirement of the U2V-UEs and reliability requirement of the V2V-UEs, respectively. The evaluation of the tolerant outage probability considers the CSI feedback delay. Constraint (13c) ensures that the average delay of all U2V-UEs is not larger than the threshold. A reasonable threshold promotes a good content placement in which popular files are cached at the UAV. Note that different vehicular users may have different transmission requirements. In our scenario containing U2V and V2V links, the large transmission power of UAV and V2V-UE transmitters will cause serious interference and excessive system energy consumption. Constraints (13d) and (13e) enforce the maximum allowed transmission power of the UAV and each V2V-UE, respectively. Constraint (13f) denotes the file caching constraint of the UAV. Constraints (13g) and (13h) together with (13j) ensure our consumption that one U2V link's RB can only be reused by one V2V link and one V2V link can only access one U2V link's RB. Constraint (13i) together with (13k) is consistent with our assumption that one RB can only be allocated to one U2V-UE within the service range of a UAV, which meets the orthogonality constraints of OFDMA. Constraint (13l) represents that there is only one state for a file, that is, whether it is cached at the UAV or not.

D. Problem Simplification

Problem (P1) contains a storage optimization and a communication resource optimization, where the former is done every time window with T time slots and the latter is done every time slot. We propose a popularity-based content placement strategy that is executed at the beginning of a time window, in order to obtain the file caching state matrix S and hence simplify Problem (P1).

Each U2V-UE has a queue buffer to store requests. Let $Q_k(t)$ denote the k^{th} U2V-UE's request queue length at time slot t, and the queue is updated as:

$$Q_k(t+1) = max(Q_k(t) - \beta_{k,f} s_f \mathbf{1}(R_k(t) \geq R_k^{min}), 0) + A \tag{14}$$

where $\mathbf{1}(\cdot)$ is an indicator funciton and it values 1 when $\cdot$ is true. A is the constant request arrival rate per time slot, under the assumption of deterministic periodic arrivals.

The cached file library of the UAV can be updated at regular intervals using technologies such as millimeter-wave beamforming. The cache update strategy is to replace the F_l least popular content files in the UAV's storage with the F_m most popular content files. The sizes of F_l and F_m are determined according to constraint $13(f)$. Instantaneous popularity at time slot t is computed according to the cumulative sum of instantaneous length of the request queues $Q_f = \sum_{t=1}^{T}\sum_{k=1}^{K} Q_{k,f}(t)$. The parameters $B = \arg\max_{\mathcal{B}\in\mathcal{F},|\mathcal{B}|=F_m} \sum_{f\in\mathcal{B}} Q_f$ and $S = \arg\min_{\mathcal{S}\in\mathcal{F},|\mathcal{S}|=F_m} \sum_{f\in\mathcal{S}} Q_f$ are denoted as the sets of data files which should be charged and discharged from the UAV respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3163430, IEEE Transactions on Vehicular Technology

We perform the content flie placement b_f as follows:

$$b_f = \begin{cases} 1, & \text{if } f \in \mathcal{B} \text{ and } f \notin \tilde{\mathcal{F}} \\ -1, & \text{if } f \in \mathcal{S} \text{ and } f \in \tilde{\mathcal{F}} \\ 0, & \text{if } f \in \mathcal{B} \text{ and } f \in \tilde{\mathcal{F}} \end{cases} \tag{15}$$

where 1 means adding the f^{th} content file into the UAV, -1 means deleting the f^{th} file from the UAV, and 0 no change. Finally, the state of the f^{th} file cached in the UAV has the following dynamic in the new time window: $s_f(t_w+1) = s_f(t_w) + b_f$.

So far, the variable matrixs in Problem (P1) are Δ, X and P. And constraint $13(f)$ is removed since it has been already guaranteed by the sizes of F_l and F_m. Since the aim of constraint $13(c)$ is to promote a content placement in which popular content files are cached at the UAV, consistent with the idea of our popularity-based content placement strategy, we also remove it for simplification. Problem (P1) is turned into Problem (P2) as follows:

$$(P2)$$

$$\max_{\{\Delta,X,P\}} \frac{1}{T}\sum_{t=0}^{T-1} \eta^{ee}(t) \tag{16}$$

$$s.t. \quad (13a), (13b), (13d), (13e), (13g) - (13k)$$

III. Combined Solution Based on Classical Optimization and Reinforcement Learning

In this section, we will focus on obtaining the solution to Problem (P2). Containing discrete binary variables $\delta_{k,n}$ and $\rho_{m,k}$ for spectrum allocation and continuous variables $p_{k,n}$ and $p_{m,k}$ for power control, Problem (P2) is a MINLP problem, a typical Non-deterministic Polynomial Hard (NP-hard) problem. Due to the large quantities of content files, vehicles, RBs, and power levels in this network, Problem (P2) cannot be solved in polynomial time. A straightforward method to obtain the optimal solution is decomposing the problem into multiple sub-problems and conducting exhaustive searches. However, in such a highly dynamic vehicular network, optimization decisions should be made in real time.

A. Motivation for Reinforcement Learning Combining Classical Optimization

Reinforcement Learning (RL), in which intelligent agents improve their behavior policy based on the knowledge obtained from interactions with the environment, is an effective technique enabling intelligent decision making in dynamic scenarios. Deep Q Network (DQN) approximates the Q value in RL through the Deep Network framework. As an improvement over DQN, Double Deep Q Network (DDQN) solves the overestimating issue in the training process, improving training stability and convergence speed. Therefore, to deal with the large-scale state and action space caused by rapid environmental changes, we suggest using classical optimization to provide a good initialization state for DDQN and thus improve DDQNs convergence speed. Specifically, we decompose Problem (P2) into two sub-problems, i.e., RB allocation and power allocation sub-problems. Based on the strict restriction that one RB can only be occupied by one vehicular user, Hungarian can solve the RB allocation sub-problem. After obtaining the RB allocation result, power allocation based on DDQN is performed.

B. RB allocation based on Hungarian

1) RB allocation for U2V-UEs: Suppose that the UAV uses the same transmission power $p^*_{k,n}$ when sending data to each U2V-UE and each V2V-UE sender use the same transmission power $p^*_{m,k}$. We ignore the interference between the U2V links and the V2V links for the moment. The optimization problem (P2) is simplified to (P3), as shown below:

$$(P3)$$

$$\max_{\{\Delta\}} \frac{1}{T}\sum_{t=0}^{T-1} \frac{\sum_{k=1}^{K} Blog_2(1+\frac{\sum_{n=1}^{N}\delta_{k,n}(t)p^*_{k,n}(t)|h_{k,n}(t)|^2\alpha_{k,n}(t)}{N_0 B})}{\sum_{k=1}^{K}\sum_{n=1}^{N}\delta_{k,n}(t)(p^*_{k,n}(t)+P^C)} \tag{17}$$

$$s.t.$$

$$Blog_2(1+\frac{\sum_{n=1}^{N}\delta_{k,n}(t)p^*_{k,n}(t)|h_{k,n}(t)|^2\alpha_{k,n}(t)}{N_0 B}) \geq R_k^{min}, \quad \forall k \in K \tag{17a}$$

$$Pr\{\frac{\sum_{k=1}^{K}\delta_{k,n}(t)p^*_{m,k}(t)\alpha_{m,k}(t)(\varepsilon_m^2|h_{m,k}(t)|^2+|e_m|^2)}{N_0 B} \geq \gamma_m^0\} \leq Pr^0, \forall m \in M \tag{17b}$$

$$(13i), (13k)$$

Problem (P3) is a maximization problem that the traditional Hungarian cannot directly solve. We transform the efficiency matrix to one that Hungarian can solve. The steps of solving the RB allocation problem for U2V-UEs are described below:

Step 1: Construct an efficiency matrix $\mathcal{C}_0$, where element $C_{k,n}$ in the k^{th} row and the n^{th} column is the energy efficiency obtained by the k^{th} U2V-UE on RB n.

Step 2: Calculate the transmission capacity $C_{k,n}$ obtained by the k^{th} U2V-UE on RB n. If the constraint (17a) is not satisfied, the corresponding energy efficiency value $C_{k,n} = 0$ and we get the filtered matrix $\mathcal{C}_1$.

Step 3: Find the largest element in matrix $\mathcal{C}_1$, and subtract each element in the matrix from the largest element. The new matrix is denoted as $\mathcal{C}_2$. Therefore, the energy efficiency maximization problem is transformed into a minimization one, which can be solved by Hungarian.

Step 4: For each row in matrix $\mathcal{C}_2$, subtract the minimum element from each element in that row; for each column, perform the same action, ensuring that there are zero elements in each row and each column finally. The optimal solution will not be affected. The obtained matrix is denoted as $\mathcal{C}_3$.

Step 5: Cover all zeros in matrix $\mathcal{C}_3$ by a minimum number of horizontal or vertical lines. If the number of the required lines is less than the matrix dimension, go to Step 6; otherwise, go to Step 8.

Step 6: Find the smallest element that is not covered by the line in Step 3. Subtract the value of it from all uncovered elements and add it to all elements covered twice. Additional zeros can be created.

Step 7: Repeat steps 5 and 6 until the number of lines covering zeros is not less than the matrix dimension.

Step 8: Obtain the optimal solution. Find m zero elements in different rows and columns of matrix C_3, which correspond to the RB allocation result. Zero element $C_{k,n}$ corresponds to allocating RB n to vehicle user k, i.e., $\delta_{k,n} = 1$.

2) Link pairing: After we obtain the optimal RB allocation for U2V-UEs, we need to pair V2V links and U2V links that share the same spectrum. Given the optimal RB allocation $\delta^*_{k,n}$ and power allocation $p^*_{k,n}$ for U2V-UEs, the value of η_k^{ee} is only related to C_k. Therefore, to maximize η_k^{ee} is equal to maximize γ_k, which has a positive correlation with C_k. Considering the interference from the m^{th} V2V-UE pair, the received instantaneous downlink SINR at the k^{th} U2V-UE is given as follows:

$$\gamma_{k,m} = \frac{\sum_{n=1}^{N} \delta_{k,n} p_{k,n} |h_{k,n}|^2 \alpha_{k,n}}{\rho_{m,k} p_{m,k} |\widetilde{h}_{m,k}|^2 \widetilde{\alpha}_{m,k} + N_0 B} \tag{18}$$

Through checking all possible combinations of the reuse pairs, problem (P2) can be converted to

$$\begin{aligned} &(P4) \\ &\max_{\{X\}} \frac{1}{T} \sum_{t=0}^{T-1} \sum_{m=1}^{m=M} \sum_{k=1}^{k=K} \rho_{m,k}(t) \gamma_{k,m}(t) \qquad (19) \\ &s.t. \qquad (13g), (13h), (13j) \end{aligned}$$

which can be solved by Hungarian as above since it is a maximum weight bipartite matching problem.

C. Power allocation based on DDQN

Since the state transition probability and the expected reward are unknown in real communication environments, we choose the model-free DDQN algorithm to solve the power allocation problem. Given the RB allocation result obtained by the Hungarian algorithm, Problem (P2) is then converted to (P5), as follows:

$$\begin{aligned} &(P5) \\ &\max_{\{P\}} \frac{1}{T} \sum_{t=0}^{T-1} \eta^{ee}(\delta_{k,n} = \delta^*_{k,n}, \rho_{m,k} = \rho^*_{m,k}) \qquad (20) \\ &s.t. \\ &R_k(\delta_{k,n} = \delta^*_{k,n}, \rho_{m,k} = \rho^*_{m,k})(t) \geq R_k^{min}, \forall k \in K \qquad (20a) \\ &Pr\{\gamma_m(\delta_{k,n} = \delta^*_{k,n}, \rho_{m,k} = \rho^*_{m,k})(t) \geq \gamma_m^0\} \leq Pr^0, \forall m \in M \qquad (20b) \\ &(13a), (13b), (13c), (13d), (13e) \end{aligned}$$

where $\delta^*_{k,n}$ and $\rho^*_{m,k}$ is the RB allocation result obtained from Section III-B.

In this sub-section, our goal is to achieve the optimal power allocation for each vehicle in the network. Problem (P5) is modeled as an Markov Decision Process (MDP), defined by $\{\mathcal{S}, \mathcal{A}, \mathcal{R}\}$.

1) The state $s_t \in \mathcal{S}$ is defined as the communication channel conditions of each vehicle on its allocated RB at time slot t. It is collected at the UAV at each time slot.

$$s_{k,t} = \{G_{k,t}, I_{k,t}\} \tag{21}$$

where $G_{k,t}$ and $I_{k,t}$ are the channel power gain and interference channel gain obtained by the k_{th} U2V-UE at time slot t, respectively. The value of $s_{k,t}$ is known at each time slot. Similarly, $G_{m,t}$ and $I_{m,t}$ are those obtained by the m_{th} V2V-UE.

$$s_{m,t} = \{G_{m,t}, I_{m,t}\} \tag{22}$$

Hence the network state at the k^{th} time slot is expressed as follows:

$$s_t = \{s_{k,t}\} \cup \{s_{m,t}\}, \forall k \in \mathcal{K}, \forall k \in \mathcal{M} \tag{23}$$

The state set $\mathcal{S} = \{s_t\}, \forall t \in \mathcal{T}$.

2) The action $a_t \in \mathcal{A}$ taken by the agent is defined as to allocate the transmission power for the UAV sending data to each U2V-UE and for each V2V-UE transmitter sending data to the corresponding V2V-UE receiver. Since the core of DDQN is Q learning, we use discrete power levels to reduce the huge action space brought by continuous power values. Assuming there are L_k and L_m discrete power levels for U2V and V2V links, respectively. The action undertaken by the UAV for sending data to the k^{th} U2V-UE is as shown in equ. (29):

$$a_{k,t} \in \{0, \frac{P_k^{max}}{L_k - 1}, \frac{2P_k^{max}}{L_k - 1}, ..., P_k^{max}\} \tag{24}$$

Similarly, the action undertaken by the m^{th} V2V-UE transmitter is as shown in equ. (30):

$$a_{m,t} \in \{0, \frac{P_m^{max}}{L_m - 1}, \frac{2P_m^{max}}{L_m - 1}, ..., P_m^{max}\} \tag{25}$$

At each time slot t, the agent takes actions for a single U2V-UE and a V2V-UE sender which share the same spetrum simultaneously. The action a_t is expressed as equ. (31):

$$a_t = \{pair(a_{k,t}, a_{m,t}) | \rho_{m,k} = 1\}, \forall m \in \mathcal{K} \tag{26}$$

The action set $\mathcal{A} = \{a_t\}, \forall t \in \mathcal{T}$.

3) The reward $r_t \in \mathcal{R}$ captures the expected saved energy efficiency when taking action a_t under state s_t, which can also reflect the vehicles' satisfaction degree with the current action. It is necessary to seek a reward function conducive to achieving our optimization goal, i.e., maximizing the energy efficiency of the UAV. We make those actions leading to the greater energy efficiency to obtain the higher corresponding reward. Besides, constraints $(13a)$ and $(13b)$ need to be considered. To ensure user fairness, we add a penalty to the communication channel which cannot meet the minimum communication requirements of the user. Therefore, the defined reward function includes three parts, one is the contribution to the energy efficiency, and the other two are the penalties when the transmission rate and link reliability cannot

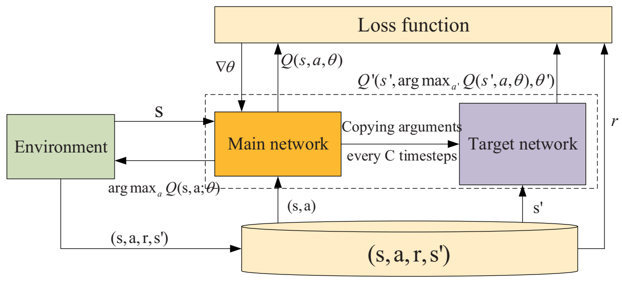


Fig. 2. DDQN structure for resource allocation in UAV-assisted vehicular networks.

satisfy users' demands. The reward function is as shown in equ. (27):

$$r^t = \omega_1 \eta^{ee} - \omega_2 \sum_{k \in \mathcal{K}} \xi_(k,n) R_k \tag{27}$$

where r_t is the reward at time t; ω_1 and ω_2 are the weights of contribution and penalty, respectively. $\xi_(k)$ is a binary variable and its value is taken as 1 only when the transmission rate cannot satisfy user k' demands.

Over time slot t, the agent performs decision making, state observation, and reward obtaining in a time slot. Denote π: $S \rightarrow A$ as the policy function, Under which, the power allocation decision at time slot $t+1$ will be performed according to $a_{t+1} = \pi(s_t)$. Each UAV uses a system reward to measure the energy efficiency increment of taking a power allocation action.

Based on the above definition, the power allocation task can be realized based on the DDQN algorithm. In the DDQN algorithm, the performance of the current power allocation policy π is evaluated by an action-value function, which calculates the long-term reward brought by a certain power allocation action in a certain state. The cumulative discount reward function is defined as equ. (28):

$$R^t = \sum_{\tau=0}^{\infty} \gamma^\tau r^{t+\tau+1} \tag{28}$$

where r represents a reward, and a discount factor $\gamma \in (0,1)$ is used to weight the importance of the current and future rewards.

The Q function $Q(s,a)$ corresponding to action a in state s under a policy π is shown in equ. (29):

$$Q(s,a;\theta) = E_\pi[R^t | s^t = s, a = a^t] \tag{29}$$

where θ represents the DDQN network parameter, and $E[\cdot]$ is the expectation operator.

The approximate Q value is obtained through neural network estimation. In the DDQN algorithm, the action corresponding to the maximum Q value in the current Q network is found firstly, which is then used to get the target Q value. The network's estimated reward for the action is as shown in equ. (30):

$$y^t = r^t + \gamma Q^{'}(s^{t+1}, arg \max_{a^{'}} Q(s^{t+1,a;\theta}); \theta^{'}) \tag{30}$$

where y^t is the optimal Q value.

Algorithm 1 Hungarian and DDQN-Based Resource Allocation Algorithm

Input: number of RB N, number of U2V-UEs K, number of V2V-UEs M, fading parameters, noise power N_0, inherent circuit power of UAV P^C, maximum transmission power of UAV P_k^{max} , number of power levels L_m, weights of contribution and penalty ω_1, ω_2, update interval of main network C .

Output: RB and power levels assigned by the UAV to each vehicle user, i.e., $\{\delta_{k,n}\}$, $\{\rho_{k,n}\}$ and $\{p_{k.n}\}$, $\{p_{m,k}\}$.

```
Initialization: random Q network weight θ, targets Q network weight θ' = θ, initial state s_1;
Execute Hungarian to obtain initial RB allocation results;
Initialize state s_1;
for all t ∈ T do
    Select action a_t according to the ϵ greedy strategy;
    Execute action a_t and obtain reward r_t;
    Re-allocate RB based on Hungarian;
    Generate new state s_{t+1};
    Store data item (s_t, a_t, r_t, s_{t+1}) in the database;
    randomly select a data item (s_j, a_j, r_j, s_{j+1}) from the database;
    if j ≠ T − 1 then
        y^j = r^j + γQ'(s^{j+1}, arg max_{a'} Q(s^{j+1,a;θ}); θ');
    else
        y_j=r_j;
    end if
    Obtain loss value L(θ) = E[(y^j − Q(s^j, a^j; θ))^2];
    if step mod C==0 then
        θ'=θ;
    end if
end for
Return {δ_{k,n}}, {ρ_{k,n}} and {p_{k.n}}, {p_{m,k}}.
```

The loss function is defined in equ. (31):

$$L(\theta) = E[(y^t - Q(s^t, a^t; \theta))^2] \tag{31}$$

We use the stochastic gradient descent method to train the neural network parameters, and the optimal parameter θ is finally obtained. The parameter θ is updated as follows:

$$\theta^{t+1} = \theta^t + \eta(y^t - Q(s^t, a^t; \theta^t)) \nabla Q(s^t, a^t; \theta^t) \tag{32}$$

where η is the learning rate.

The DDQN structure is shown in Fig. 2. At each step, the UAV stores the current state, power allocation action, network reward, and the next state as an experience item in the database. The neural network randomly selects a part of the experience for training. The loss function is used to evaluate the training performance. The weights of the target network and the main network are updated through the neural network's backpropagation. In the training process, the power allocation action with the largest Q value is finally selected.

The pseudocode of the resource allocation algorithm based on Hungarian and DDQN is shown in Algorithm 1.

Since the vehicular network environment is highly dynamic, applying the data-driven DDQN into vehicular networks re-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3163430, IEEE Transactions on Vehicular Technology

quires many computing and storage resources. Considering the network resources are limited, a two-step training framework is adopted. First, DDQN is used in the simulated wireless communication system for offline pre-training. Transfer learning is then adopted for further minor adjustments to real scenes, reducing fully online training pressure.

IV. Simulation Results

We build a reinforcement learning framework based on Tensorflow, and simulate our resource allocation method using Python language. In this section, we introduce the simulation settings first and then we compare our proposed Hungarian and DDQN-based resource allocation method against the random baseline, the classical Q-learning-based, the DQN-based [23] and the DDQN-based [24] methods. Finally, we analyze the simulation results, in order to verify the effectiveness of our proposed scheme in terms of timeliness in real-time decisions, UAV's energy efficiency, and users' average delay.

A. Simulation settings

The DDQN is composed of 3 fully connected hidden layers, which contain 500, 250, and 120 neurons respectively. The Rectified Linear Unit (ReLU) is used as the activation function, and the RMSProp optimizer is used to update the network parameters with a learning rate of 0.01. The training exploration rate drops from 0.9 to 0.01, and then remains unchanged. The simulation setting is shown in Table II.

TABLE II
SIMULATION SETTING.

Variables	Values
Number of RBs N	$4 \sim 40$
Number of U2V-UEs K	$4 \sim 40$
Number of V2V-UEs M	$4 \sim 40$
Radio range of the UAV	$300\ m$
Noise power N_0	$-114\ dBm$
Circuit power of the UAV P^C	$5\ dBm$
Number of power levels L_m	5
Maximum transmission power of the UAV P_k^{max}	$17(23)\ dBm$
CSI feedback period of vehicle	$1\ ms$
Weights on contribution and penality ω_1, ω_2	$0.8, 0.2$
Learning rate η	0.1
Initial decay rate	0.9
SINR threshold γ_m^0	$20\ dB$

B. Simulation Analysis

1) Selection of contribution and penality weights: The contribution weight ω_1 and the penalty weight ω_2 in DDQN greatly influence the algorithm's performance. When the contribution weight is large, the algorithm prioritizes energy efficiency over vehicle users' minimum transmission data rate requirements. The opposite happens if the penalty weight is larger. As shown in Fig. 3, the larger ω_1 is, the larger energy efficiency is obtained. We set $\omega_1 = 0.8$, $\omega_2 = 0.2$ in this paper.

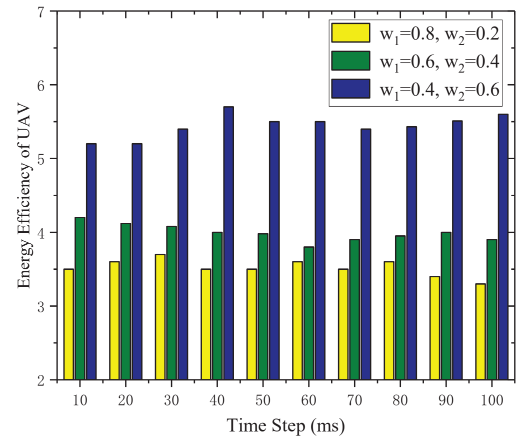


Fig. 3. Comparision of UAV's energy efficiency under different values of ω_1 and ω_2.

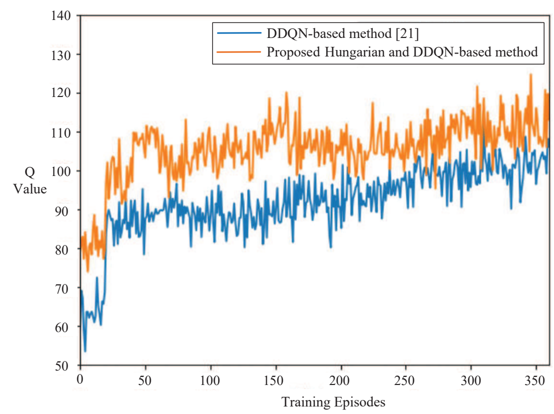


Fig. 4. Comparision of Q value changes.

2) Timeliness in real-time decisions: In highly dynamic vehicular networks, optimization decisions should be made in real time. We compare our Hungarian and DDQN-based resource allocation method with the DDQN-based one to verify the effectiveness of introducing Hungarian. Fig. 4 shows the Q value changes in the iterative process of these two algorithms. It can be seen that, by introducing Hungarian, our resource allocation method can obtain a higher Q value much faster than the DDQN-based method; that is, the network receives a higher reward value faster. The real-time performance of our method is superior to that of the DDQN-based one. Specifically, the Q value gains 100 within 25 iterations in our method while it takes more than 250 iterations in the DDQN-based one. The reason is that the RB allocation result obtained by Hungarian provides a better initial state of DDQN, reducing the search space dimension.

3) Energy efficiency comparison: The introduction of Hungarian can improve timeliness in real-time decisions. We then verify the effectiveness of utilizing DDQN. Hence we compare our method with the random, the classical Q-learning-based and the DQN-based methods. In the random method, the UAV randomly allocates RBs and power levels to vehicle users. The energy efficiency values under the four methods are as shown in Fig. 5. It can be seen that, compared with the other three methods, our proposed method contributes higher energy efficiency. This is because it can thoroughly learn the rela-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3163430, IEEE Transactions on Vehicular Technology

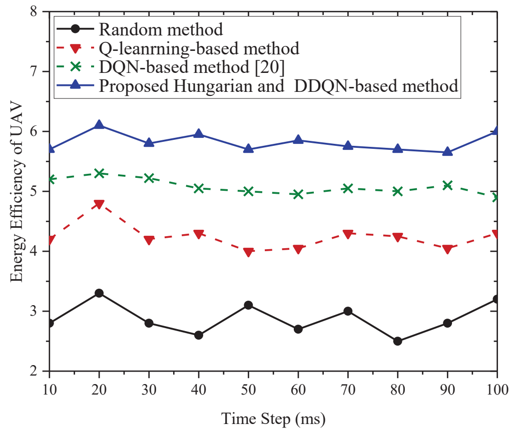


Fig. 5. Comparision of UAV's energy efficiency.

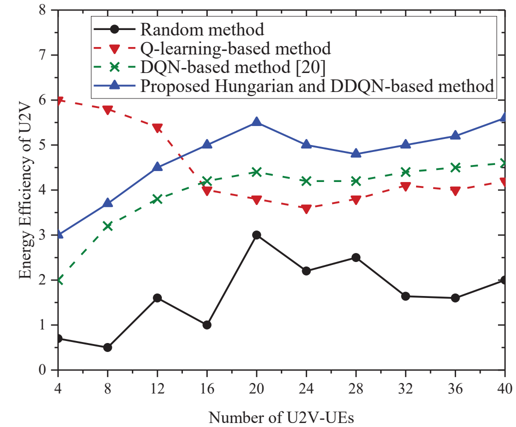


Fig. 6. Comparision of UAV's energy efficiency under different numbers of vehicles.

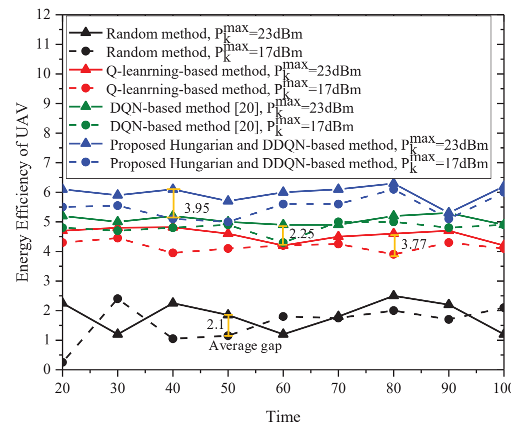


Fig. 7. Comparision of UAV's energy efficiency under different maximum transmission power.

tionship between environmental states and resource allocation actions, making real-time decisions that are more conducive to improving the UAV's energy efficiency. In the random allocation method where the network's energy efficiency is the lowest, the UAV randomly allocates RB and power to vehicle users without considering the real-time environment and users' demands. Hence, the network's energy efficiency is the lowest. In the DQN-based resource allocation method, the overestimation of Q values may make the agent take a sub-optimal action which yields a relatively low energy efficiency. Similar to the DQN-based method, the Q-learning-based one suffers the overestimation of Q values caused by high space dimensionality. Additionally, performing the Q-learning method creates a large computing and time overhead, making the Q-learning-based resource allocation method worse than those based on DDQN and DQN.

The number of U2V-UEs within a UAV's coverage affects the vehicles' channel conditions and will further affect the optimal allocation of RBs and power. Fig. 6 shows the comparison result of the energy efficiency obtained by the above four methods with different numbers of U2V-UEs. We can see that with a small number of vehicles in the system, the joint allocation based on Q-learning performs better than those based on DQN and DDQN. The reason is that in the small-scall network, Q-learning runs faster and more accurately with less data to learn. As the number of vehicles increases, the Q-learning-based method has to built and maintain a huge Q-table with high state and action space dimensionality. The data explosion results in a performance degradation accordingly. Both the DQN-based and the DDQN-based methods predict Q values using neural networks instead of computing Q values accurately. They can fast catch the environmental changes and take actions in real time. In addition, the DDQN-based method can effectively avoid the overfitting problem to gain higher energy efficiency for the UAV. The random allocation method without considering the real-time environmental state and users' demands achieves the lowest energy efficiency.

Energy efficiency is defined as the ratio of data transmission rate to power consumption. Hence, the maximum transmission power of UAVs directly affects the energy efficiency performance. Fig. 7 shows the comparison results of the energy efficiency obtained by the four methods under different maximum transmission power for U2V-UEs. It can be seen that the energy efficiency value of the our proposed allocation method is greater than the other three methods in most cases. In addition, for our proposed method, the greater the maximum transmission power, the greater the energy efficiency obtained. The average gap of gained energy efficiency between the allocation with 23 dBm power and that with 17 dBm is 3.95, which is the largest, indicating that our proposed method can make best use of transmission power gain to improve energy efficiency as much as possible. Unfortunately, the transmission power gain is often/occasionally wasted in the other three methods, which is illustrated by the red intersecting points and the black/green coincident points.

4) Average delay performance: Due to the caching capability of the UAV, the U2V-UEs can get cached files directly from the UAV without the fronthaul delay. Fig. 8 approves that the average delay will reduce as the UAV's caching capacity enhances. However, the decreasing trend is not apparent when the total number of files is much larger than the capacity.

In addition, we plot the delay performance of our proposed resource allocation scheme with cached content library updating and the traditional scheme without content updating under different file request arrival rates in Fig. 9. The average delay of U2V-UEs increases with the arrival rate. When the arrival

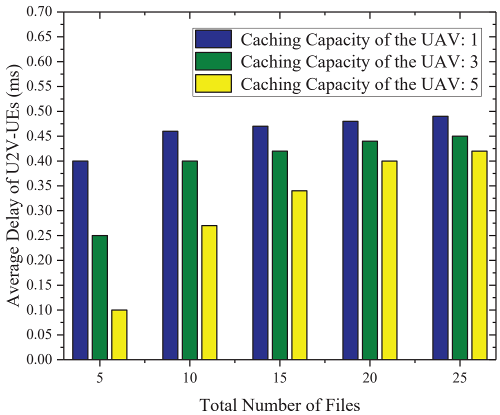


Fig. 8. Comparision of U2V-UEs' average delay with different caching capacity of the UAV.

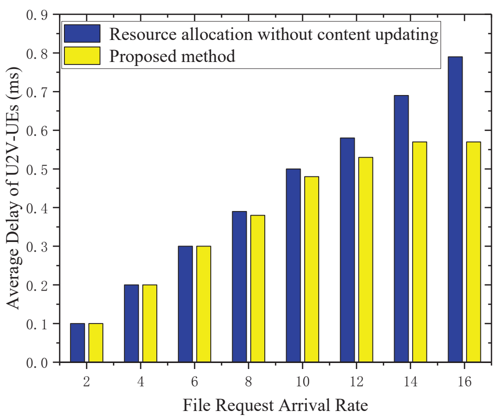


Fig. 9. Comparision of U2V-UEs' average delay under different file request arrival rate.

rate is low, the delay performance of the resource allocation scheme without content updating is likely to be the same as our proposed scheme. However, when the arrival rate becomes larger, our scheme's delay performance is superior to the one without content updating. This is because the latter solution cannot adaptively capture the content popularity and request queue dynamics. In contrast, our proposed solution can help the UAV cache the most popular files to handle the high arrival rate caused by the frequent requests of popular content files.

V. Conclusion

In this paper, we have studied the resource allocation problem for UAV-assisted vehicular networks with spectrum sharing. The objective of the optimization problem is to maximize the UAVs long-term energy efficiency. Focusing on the scenario where U2V and V2V communications are supported simultaneously, we have performed a joint optimization of content placement, spectrum allocation, link pairing, and power control for UAV-assisted vehicular networks. To solve the complex issue, we have proposed a content placement strategy for problem simplification. Classical optimization and intelligent optimization have been combined to decide the optimal spectrum and power allocation with fast convergence speed. We have finally conducted simulations to evaluate our proposed resource allocation method. The simulation results have demonstrated that our approach can significantly increase the timeliness in real-time decisions and improve the UAV's energy efficiency while satisfying the users' QoS requirements.

In our future work, a more dynamic scenario where a UAV flies along an optimized trajectory will be considered. As a result, the trade-off between the UAVs service range and service time will be a challenge. The analysis and prediction of traffic streams along roads may be explored to help design the UAVs trajectory, in order to serve more vehicles.

References

[1] F. Tang, Y. Kawamoto, N. Kato, and J. Liu, "Future intelligent and secure vehicular network toward 6G: Machine-learning approaches," *Proceedings of the IEEE*, vol. 108, no. 2, pp. 292–307, 2019.

[2] Z. Niu, X. S. Shen, Q. Zhang, and Y. Tang, "Space-air-ground integrated vehicular network for connected and automated vehicles: Challenges and solutions," *Intelligent and Converged Networks*, vol. 1, no. 2, pp. 142–169, 2020.

[3] H. Abualola, H. Otrok, H. Barada, M. Al-Qutayri, and Y. Al-Hammadi, "Matching game theoretical model for stable relay selection in a UAV-assisted Internet of vehicles," *Vehicular Communications*, vol. 27, p. 100290, 2021.

[4] M. Khabbaz, C. Assi, and S. Sharafeddine, "Multihop V2U path availability analysis in UAV-assisted vehicular networks," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10 745–10 754, 2021.

[5] H. Fatemidokht, M. K. Rafsanjani, B. B. Gupta, and C.-H. Hsu, "Efficient and secure routing protocol based on artificial intelligence algorithms with UAV-assisted for vehicular Ad Hoc networks in intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4757–4769, 2021.

[6] O. Bouachir, M. Aloqaily, I. A. Ridhawi, O. Alfandi, and H. B. Salameh, "UAV-assisted vehicular communication for densely crowded environments," in *IEEE/IFIP Network Operations and Management Symposium*, 2020, pp. 1–4.

[7] Y. He, D. Zhai, Y. Jiang, and R. Zhang, "Relay selection for UAV-assisted urban vehicular ad hoc networks," *IEEE Wireless Communications Letters*, vol. 9, no. 9, pp. 1379–1383, 2020.

[8] L. Zhao, K. Yang, Z. Tan, X. Li, S. Sharma, and Z. Liu, "A novel cost optimization strategy for SDN-enabled UAV-assisted vehicular computation offloading," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3664–3674, 2021.

[9] H. Peng and X. Shen, "Multi-agent reinforcement learning based resource management in MEC- and UAV-assisted vehicular networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 131–141, 2021.

[10] H. Wu, F. Lyu, C. Zhou, J. Chen, L. Wang, and X. Shen, "Optimal UAV caching and trajectory in aerial-assisted vehicular networks: A learning-based approach," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 12, pp. 2783–2797, 2020.

[11] Z. Su, M. Dai, Q. Xu, R. Li, and H. Zhang, "UAV enabled content distribution for Internet of connected vehicles in 5G heterogeneous networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 5091–5102, 2021.

[12] A. Al-Hilo, M. Samir, C. Assi, S. Sharafeddine, and D. Ebrahimi, "A cooperative approach for content caching and delivery in UAV-assisted vehicular networks," *Vehicular Communications*, p. 100391, 2021.

[13] M. Samir, D. Ebrahimi, C. Assi, S. Sharafeddine, and A. Ghrayeb, "Leveraging UAVs for coverage in cell-free vehicular networks: A deep reinforcement learning approach," *IEEE Transactions on Mobile Computing*, vol. 20, no. 9, pp. 2835–2847, 2021.

[14] S. Zhang, J. Zhou, D. Tian, Z. Sheng, X. Duan, and V. C. M. Leung, "Robust cooperative communication optimization for multi-UAV-aided vehicular networks," *IEEE Wireless Communications Letters*, vol. 10, no. 4, pp. 780–784, 2021.

[15] M. Samir, D. Ebrahimi, C. Assi, S. Sharafeddine, and A. Ghrayeb, "Trajectory planning of multiple dronecells in vehicular networks: A reinforcement learning approach," *IEEE Networking Letters*, vol. 2, no. 1, pp. 14–18, 2020.

[16] H. Bian, H. Dai, and L. Yang, "Throughput and energy efficiency maximization for UAV-assisted vehicular networks," *Physical Communication*, vol. 42, p. 101136, 2020.

[17] W. Sun, E. G. Ström, F. Brännström, K. C. Sou, and Y. Sui, "Radio resource management for D2D-based V2V communication," *IEEE*

Transactions on Vehicular Technology, vol. 65, no. 8, pp. 6636–6650, 2015.

[18] O. S. Oubbati, M. Atiquzzaman, A. Baz, H. Alhakami, and J. Ben-Othman, "Dispatch of UAVs for urban vehicular networks: A deep reinforcement learning approach," *IEEE Transactions on Vehicular Technology*, 2021.

[19] Y. Ren, X. Chen, S. Guo, S. Guo, and A. Xiong, "Blockchain-based VEC network trust management: A DRL algorithm for vehicular service offloading and migration," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 8148–8160, 2021.

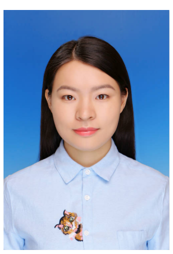
[20] B. Hamdaoui, B. Khalfi, and N. Zorba, "Dynamic spectrum sharing in the age of millimeter-wave spectrum access," *IEEE Network*, vol. 34, no. 5, pp. 164–170, 2020.

[21] F. Tang, Z. M. Fadlullah, N. Kato, F. Ono, and R. Miura, "AC-POCA: Anticoordination game based partially overlapping channels assignment in combined UAV and D2D-based networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1672–1683, 2017.

[22] T. Kim, D. J. Love, and B. Clerckx, "Does frequent low resolution feedback outperform infrequent high resolution feedback for multiple antenna beamforming systems?" *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1654–1669, 2010.

[23] H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep reinforcement learning based resource allocation for V2V communications," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3163–3173, 2019.

[24] D. Zhao, H. Qin, B. Song, Y. Zhang, X. Du, and M. Guizani, "A reinforcement learning method for joint mode selection and power adaptation in the V2V communication network in 5G," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 452–463, 2020.



Weijing Qi received the Ph.D. degree in communication and information system from the Northeastern University, Shenyang, China, in 2019. She is currently a Lecturer in Communication and Information System with the Chongqing University of Posts and Telecommunications, Chongqing, China. Her research interests include resource management in vehicular networks, edge computing, and machine learning-based optimizaiton.

Qingyang Song (S'05-M'07-SM'14) received the Ph.D. degree in telecommunications engineering from the University of Sydney, Australia, in 2007. From 2007 to 2018, she was with Northeastern University, China. She joined Chongqing University of Posts and Telecommunications in 2018, where she is currently a Professor. She has authored more than 80 papers in major journals and international conferences. Her current research interests include cooperative resource management, edge computing, vehicular ad hoc network (VANET), mobile caching, and simultaneous wireless information and power transfer (SWIPT). She serves on the editorial boards of two journals, including Editor for IEEE Transactions on Vehicular Technology and Technical Editor for Digital Communications and Networks.

Lei Guo received the Ph.D. degree in communication and information system from the University of Electronic Science and Technology of China, Chengdu, China, in 2006. He is currently a Full Professor of Communication and Information System with the Chongqing University of Posts and Telecommunications, Chongqing, China. He has authored or coauthored more than 200 technical papers in international journals and conferences. He is an Editor for several international journals. His research interests include communication networks, optical communications, and wireless communications.

Abbas Jamalipour (S'86-M'91-SM'00-F'07) received the Ph.D. degree in Electrical Engineering from Nagoya University, Nagoya, Japan in 1996. He holds the positions of Professor of Ubiquitous Mobile Networking with the University of Sydney and since January 2022, the Editor-in-Chief of the IEEE Transactions on Vehicular Technology. He has authored nine technical books, eleven book chapters, over 550 technical papers, and five patents, all in the area of wireless communications and networking. Prof. Jamalipour is a recipient of the number of prestigious awards, such as the 2019 IEEE ComSoc Distinguished Technical Achievement Award in Green Communications, the 2016 IEEE ComSoc Distinguished Technical Achievement Award in Communications Switching and Routing, the 2010 IEEE ComSoc Harold Sobol Award, the 2006 IEEE ComSoc Best Tutorial Paper Award, as well as over 15 Best Paper Awards. He was the President of the IEEE Vehicular Technology Society (2020-2021). Previously, he held the positions of the Executive Vice-President and the Editor-in-Chief of VTS Mobile World and has been an elected member of the Board of Governors of the IEEE Vehicular Technology Society since 2014. He was the Editor-in-Chief IEEE WIRELESS COMMUNICATIONS, the Vice President-Conferences, and a member of Board of Governors of the IEEE Communications Society. He sits on the Editorial Board of the IEEE ACCESS and several other journals and is a member of Advisory Board of IEEE Internet of Things Journal. He has been the General Chair or Technical Program Chair for several prestigious conferences, including IEEE ICC, GLOBECOM, WCNC, and PIMRC. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), the Institute of Electrical, Information, and Communication Engineers (IEICE), and the Institution of Engineers Australia, an ACM Professional Member, and an IEEE Distinguished Speaker.