Heterogeneous User-Centric Cluster Migration Improves the Connectivity-Handover Trade-Off in Vehicular Networks

Yan Lin, *Member, IEEE*, Zhengming Zhang, Yongming Huang, *Senior Member, IEEE*, Jun Li, *Senior Member, IEEE*, Feng Shu, *Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—User-centric (UC) clustering has recently emerged as a promising paradigm for enhancing the connectivity of mobile users by grouping an appropriate number of access points (APs), thus paving the way for seamlessly connected vehicular networks. However, for a high-velocity vehicular user, UC clustering may lead to overly frequent handovers (HOs), which increases the risk of throughput-reduction, call dropping and energy wastage. To mitigate this problem, we aim for reducing the HO overhead imposed on the heterogeneous UC (HUC) cluster migration process of vehicular networks. Specifically, we first conceive a novel hybrid HUC cluster migration strategy that adaptively switches between horizontal and vertical HOs for supporting both vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication. Then, a dynamic decision-making problem is formulated for balancing the benefits of HUC cluster migration and the total HO overhead, subject to realistic HUC clustering constraints. In the face of unknown vehicular mobility, we propose a sequential HUC cluster migration solution based on max-bipartite matching theory imposing a low complexity. As a design alternative, we also propose a holistic solution relying on model-free deep reinforcement learning (DRL). Finally, our numerical results reveal the superiority of the proposed cluster migration design in terms of striking a compelling trade-off between the per-user average data rate (PAR) and the number of HOs in different scenarios.**

***Index Terms*—User-centric clustering, vehicular networks, heterogeneous, handover, max-bipartite matching, deep reinforcement learning.**

I. Introduction

Vehicle-to-everything (V2X) communication has become one of the key enablers in enhancing comfort, safety and road traffic efficiency. With the rapid evolution of wireless communication, cellular V2X (C-V2X) technology together with next-generation networking (NGN) support vehicle-to-infrastructure (V2I), vehicle-to-vehicle (V2V) and vehicle-to-pedestrian (V2P) communications [1]–[4]. To meet the latency, capacity and reliability demands of V2X communication, a high data rate requirement must be guaranteed [5]–[9]. However, due to the drastic escalation of road traffic, vehicles should communicate with road unit sides (RSUs) and/or among themselves, which significantly alleviates the load of cellular base stations (BSs) [10]–[14].

As a popular candidate for enhancing connectivity, user-centric (UC) clustering has shown merits in terms of assisting the emerging NGNs by lending users as an empowered role [15] [16]. More explicitly, UC clustering is capable of seamlessly adapting to dynamic network topology fluctuation, hence still guaranteeing each user's data rate in the face of tele-traffic fluctuations. By grouping the most appropriate number of access points (APs), UC clusters are formed in the close proximity of users by exploiting their cooperation to support users at high data rate [17] [18]. In general, the selection of APs depends both on the network topology and on the data rate requirements, and their beneficial cooperation relies on the joint transmission technique. As a result, this architecture is eminently suitable for supporting substantial cell coverage expansion by eliminating any deleterious edge-effect, whilst controlling the latency and data rate, as well as balancing the tele-traffic in demanding communication scenarios [19]–[21]. In this light, the UC cluster supporting a specific vehicle has to accommodate its movements. Nevertheless, in vehicular networks, both RSUs and vehicles can play the role of APs, and vehicular APs (VAPs) can improve the reliability as a benefit of their close proximity. Based on the above characteristics, the UC clustering process conceived for supporting vehicular user equipment (VUE) constitutes a *heterogeneous UC (HUC) cluster migration* process, which is capable of substantially improving the connectivity probability of VUEs with the aid of both RSU cooperation and VAPs.

However, overly frequent HUC cluster migrations of VUEs lead to excessive handover (HO) overheads [22]–[25]. Specifically, the multiple-RSU association events result in an increase

This work was supported in part by the National Natural Science Foundation of China under Grants 62001225, 62071234, 61771244, 61727802 and 61872184, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20190454, in part by the Fundamental Research Funds for the Central Universities under Grant 30919011227, in part by the financial support of the Engineering and Physical Sciences Research Council projects EP/N004558/1, EP/P034284/1, EP/P034284/1, EP/P003990/1 (COALESCE), of the Royal Society's Global Challenges Research Fund Grant as well as of the European Research Council's Advanced Fellow Grant QuantCom. (*Corresponding authors: Feng Shu and Jun Li.*)

Y. Lin and J. Li are with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China. (Email: {yanlin, jun.li}@njust.edu.cn). Z. Zhang and Y. Huang are with the School of Information Science and Engineering, Southeast University, Nanjing 210096, China. (Email:{zmzhang,huangym}@seu.edu.cn). F. Shu is with the School of Information and Communication Engineering, Hainan University, Haikou 570228, China. (Email:shufeng0101@163.com). L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ, U.K. (Email: lh@ecs.soton.ac.uk).

in the number of reconfigured connections, while the VAP-VUE association may encounter more frequently HOs due to the high mobility of vehicles. The increased HO overheads may increase the risk of throughput-reduction and of call dropping, and increase the energy consumption [26] [27]. Fortunately, the HUC cluster migration process allows VUEs to perform smooth, seamless soft HOs, when they are associated with multiple RSUs. Additionally, as the VAPs move together with VUEs at a similar velocity along the road, the VAP-VUE association may maintain longer durations than the VAP-RSU association, thus providing an opportunity for reducing the frequency of HOs. Hence, the trade-off between the connectivity benefiting from HUC clustering and the HO overhead imposed has to be jointly considered in vehicular networks.

Although substantial attention has been dedicated to the HO problem of vehicular networks [28], most of the existing HO policies are cell-centric, based on the conventional single-AP association principle and hard HO. Moreover, the existing HO policies are typically based on two categories: *horizontal HO* and *vertical HO* [29] [30]. More explicitly, the horizontal HO takes place within the same access network, for maintaining the connections among either the VAPs or RSUs. By contrast, a vertical HO represents switching to a different access network, when a vehicle is roaming in an overlapping area of different multiple access networks, as in V2I and V2V communication. Hence we combine horizontal and vertical HOs into *hybrid HO* in the context of the above HUC cluster migration problem of vehicular networks. Nevertheless, this hybrid HO design encounters practical issues, including unknown vehicular mobility and association with multiple RSUs for supporting soft HOs. To the best of our knowledge, the hybrid HO problem of HUC clustering-based vehicular networks is still an open issue at the time of writing.

Against the above backdrop, in this paper, we solve this wide open HUC cluster migration problem of vehicular networks, and strike a compelling trade-off between the connectivity benefits of HUC clustering and the overhead imposed by the hybrid HO. Our decision-making problem focuses on striking an attractive trade-off between the sum data rate and the number of HOs within a finite time interval, while satisfying the minimum data rate requirement and the relevant association constraints per time slot (TS). Explicitly, the contributions of this paper are summarized as follows:

1) A novel HUC cluster migration framework is conceived for vehicular networks relying on hybrid HOs, which exploits the benefits of RSU cooperation and moving VAPs for improving the connectivity.
2) To strike the most appropriate trade-off between the connectivity and the HO overhead imposed, we formulate a decision-making optimization problem within a finite time interval, while satisfying the minimum data rate and the association constraints.
3) Based upon max-bipartite matching theory, a low-complexity sequential HUC cluster migration solution is developed in the face of unknown vehicular mobility. Then, a holistic solution is proposed for training the HUC cluster migration design with the aid of model-free deep reinforcement learning (DRL) [31].
4) Numerically, our results highlight the quantitative benefits of the proposed HUC cluster migration design over the state-of-the-art. It is shown that the DRL-aided holistic solution offers superior performance compared to the sequential one in terms of increasing the average data rate, whilst relying on the most appropriate number of HOs at the cost of a certain training overhead.

The rest of the paper is organized as follows: Section II introduces the related contributions, while Section III describes both the system model and our assumptions. Section IV formulates the design problem of HUC cluster migration. Then, Section V and Section VI detail the max-bipartite matching based sequential solution and the DRL-aided holistic solution, respectively. Our numerical results are provided in Section VII, and finally Section VIII concludes the paper.

II. Related Contributions

The HO problem of vehicular networks has received increasing research attention [28]. Most prior studies deal with the design of vertical HOs in heterogeneous vehicular networks consisting of cellular plus a range of other access techniques. For instance, Dwijaksara *et al.* [32] studied the HO problem of WiFi-aided vehicular networks, and a HO solution based on a real road topology was proposed. With the aim of maximizing the WiFi-based connection time, the proposed solution is capable of reducing the HO latency. Nevertheless, these researches have only considered cell-centric designs relying on a UE-AP association and the vertical HO, but they have neglected the benefits of HUC clustering as well as the trade-off between connectivity and hybrid HO overhead, which constitute the main focus of our paper.

Given the challenges of user mobility, numerous studies have aimed for balancing the mobility and connectivity in heterogeneous cellular networks. It has been shown in [33] and [34] that increasing the number of APs is beneficial for reducing the HO overhead and latency. As a further advance, Xu *et al.* [35] struck a trade-off between the effective capacity and the blocking probability. Based on convex optimization techniques, the proposed solution succeeded in reducing the blocking probability without widely reducing the effective capacity. The issue of optimizing the HO overhead is also a popular research topic in literature [36]–[38]. To seek the best HO target, the authors of [36] sought the most appropriate HO solution using a sophisticated analytical technique by ranking all available HO target options, which relies on the predicted user mobility. It has been shown that this solution is capable of mitigating frequent HOs as well as simultaneously enhancing the energy efficiency. Subsequently, Hasan *et al.* [37] conceived an algorithm for reducing the HO overhead by classifying the users as fast-moving or slow-moving users, where the fast-moving users remain connected to the macro BS, while the slow-moving users perform HO to the APs. They attained 79.56% of HO probability mitigation along with 10.82% network throughput increase. Moreover, in order to exploit AP cooperation, a network topology based HO solution was proposed in [38] for reducing the HO overhead, while maintaining a good average throughput. However, the solutions found in [36]–[38] are unable to seamlessly accommodate unknown mobility scenarios and only support HOs within conventional cellular networks. Moreover, [37] and [38] only analyzed the HO performance using a heuristic method, rather than solving the problem formulated by formally optimizing the HO performance. Hence we formally formulate the HO overhead optimization problem in this paper, which allows us to strike a feasible trade-off.

DRL has been widely exploited for solving diverse decision-making problems in wireless communication [43] [44], including the solution of HO problems, because DRL learns from experience and it is capable of finding a near-optimal or optimal policy even in the absence of knowing the environmental dynamics in advance. For instance, Wang *et al.* [39] developed an asynchronous multi-user DRL-aided HO scheme for reducing the HO overhead of a traditional cellular network, while guaranteeing a certain minimum network throughput. As a further advance, Ye *et al.* [40] adopted a deep deterministic policy gradient (DDPG) based solution for dealing with the AP on/off switching problem in heterogeneous cellular networks. Additionally, Zhao *et al.* of [41] utilized a double-deep Q-network for finding a near-optimal solution for jointly designing user association and resource allocation, which aims for maximizing the long-term network utility while ensuring the required signal to interference plus noise ratio (SINR). Khan *et al.* [42] investigated a vehicle-cell association problem and adopted an asynchronous actor-critic DRL algorithm for maximizing the time-averaged data rate per VUE, while ensuring the minimum data rate for all VUEs at a low signaling overhead, rather than HO overhead. Their asynchronous solution has assumed that all VUEs can perform training independently without any information exchange, and have not considered the association conflict encountered by multiple VUEs. However, none of these sophisticated DRL-aided HO solutions have taken the benefits of HUC clustering into account.

The comparisons between our contributions and the state-of-the-art are outlined at a glance in Table I, which allows the readers to capture their main differences.

III. HUC clustering based System Model

In this section, the notations used are introduced and then our HUC clustering based system model is presented.

A. Notations

Matrices and vectors are expressed in italic bold letter, and scalar variables are denoted by italic symbols. $|\mathcal{A}|$ denotes the cardinality of a set $\mathcal{A}$ and $|A|$ represents the absolute value of a scalar A. $\mathcal{A} \times \mathcal{A}'$ denotes the cartesian product of the sets $\mathcal{A}$ and $\mathcal{A}'$. $\mathbb{C}^{N\times M}$ represents the space of all $N \times M$ matrices having complex entries. $(\cdot)^T$ denotes the transpose of a matrix or a vector. The notations of the system model are listed in Table II.

B. Network Model

As illustrated in Fig. 1, we consider a downlink V2X network in an urban multi-lane freeway scenario, where a set $\mathcal{R}$ of R RSUs are uniformly distributed and a set of vehicles travel along the road in the same direction. The vehicles are classified into a pair of categories: VUEs and VAPs. The sets of VUEs and of VAPs are denoted by $\mathcal{U} = \{1, ..., U\}$ and by $\mathcal{A} = \{1, ..., A\}$, respectively. The system relies on equal-duration TSs with the index set of $\mathcal{T} = \{1, ..., T\}$, where the network topology and parameters remain unchanged for the duration of each TS. In Fig. 1, the moving trajectories of vehicles are depicted as the locations at the beginning of each TS.

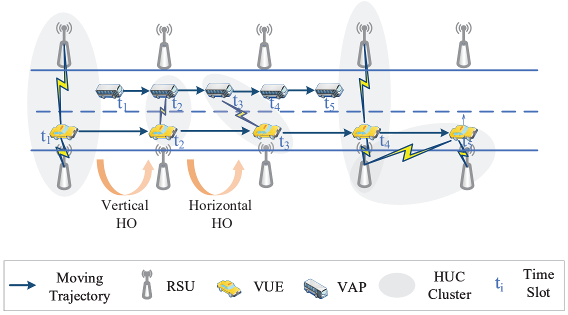


Fig. 1: An example of a HUC clustering V2X network.

C. Vehicle Mobility Model

The network's performance critically depends on the modeling of mobility, albeit it is generally stochastic and unknown. In practice, the velocities of vehicles at adjacent TSs are correlated due to the physical laws of motion. More explicitly, a vehicle's current velocity depends upon the previous velocity. In our work, the velocity of vehicles is modeled as a Gauss-Markov stochastic process [45]. Specifically, when an initial velocity $v_{i,0}$ is assigned to vehicle i, the velocity $v_{i,t}$ of vehicle i at TS t is calculated based upon the velocity $v_{i,t-1}$ at TS $t-1$, the asymptotic velocity and a random variable, given by

$$v_{i,t} = \alpha_i v_{i,t-1} + (1-\alpha_i)\bar{v}_i + \bar{\sigma}_i \sqrt{(1-{\alpha_i}^2)}n. \tag{1}$$

Herein, $\bar{v}_i$ and $\bar{\sigma}_i$ are the corresponding asymptotic mean and standard deviation of vehicle i's velocity. The parameter $\alpha_i \in [0, 1]$ denotes the memory-depth of past velocities determining the temporal correlation in movements of vehicle i. Note that as α_i tends to 1, the current velocity of vehicle i becomes more dependent on the previous velocity. Moreover, n is an uncorrelated random Gaussian process with zero mean and variance σ_n^2.

D. HUC Cluster Migration Model

In order to achieve seamless connections in vehicular networks, the UC clustering architecture [15] [16] is adopted, where each VUE can be associated with a set of RSUs in its close proximity both for improving the data rate and for receiving the latest road traffic information. For supporting both V2I and V2V communication, we define the HUC cluster, which may either be a limited number of RSUs or a single VAP. In the example of Fig. 1, the HUC clusters consist of RSUs at TS t_1, t_4 and t_5, while the VUE is supported by the VAP at TS t_2 and t_3.

TABLE I: Related Contributions

	[32] -2018	[33] -2018	[34] -2019	[35] -2019	[36] -2019	[37] -2019	[38] -2018	[39] -2018	[40] -2020	[41] -2019	[42] -2019	**this paper**
cellular network		✓	✓	✓	✓	✓	✓	✓	✓	✓		
vehicular network	✓										✓	✓
user mobility	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
multiple-association		✓	✓				✓				✓	✓
data rate				✓			✓		✓	✓	✓	✓
HO overhead	✓				✓	✓	✓	✓				✓
vertical HO	✓	✓	✓			✓						✓
horizontal HO	✓			✓	✓	✓	✓	✓			✓	✓
HO performance analysis		✓	✓									
heuristic HO solution	✓					✓	✓					✓
DRL-aided HO solution								✓	✓	✓	✓	✓

TABLE II: Notation Definitions

$\mathcal{R}$	RSU set of $\{1, ..., R\}$
$\mathcal{A}$	VAP set of $\{1, ..., A\}$
$\mathcal{U}$	VUE set of $\{1, ..., U\}$
$\mathcal{T}$	TS set of $\{1, ..., T\}$
$v_{i,t}$	the velocity of vehicle i at TS t
$\bar{S}_{\text{max}}$	the maximum number of RSUs associated with each VUE
$\bar{d}_t^{\text{R}}$	the coverage distance threshold for RSU association
$\bar{d}_t^{\text{A}}$	the coverage distance threshold for VAP association
$\mathcal{C}_{i,t}$	the HUC cluster supporting VUE i at TS t
p_c	the transmit power of transmitter $c \in \mathcal{R} \cup \mathcal{A}$
$h_{c,i,t}$	the channel gains spanning from transmitter c to VUE i at TS t
$e_{i,t}$	the number of HOs, i.e. the number of connections changing their A/P status when VUE i moves at TS t
$R_{i,t}$	the achievable data rate of VUE i at TS t
$\bar{R}_i$	the minimum data rate requirement of VUE i at each TS

Based upon the above fact, the HUC cluster migration strategy is adopted by supporting hybrid HOs, which adaptively switches between the horizontal and vertical HOs. An illustration of HUC cluster migration is depicted in the example of Fig. 1, wherein it has the options of horizontal and vertical HOs at each TS. For instance, the VUE selects the horizontal HO when it moves from t_2 to t_3 and from t_4 to t_5. By contrast, the components of its HUC cluster switch between RSUs and VAP from t_1 to t_2 and from t_3 to t_4, thus resulting in vertical HOs.

We set the maximum number of RSUs associated with each VUE as $\bar{S}_{\text{max}}$, and set the coverage distance threshold as $\bar{d}_t^{\text{R}}$ for RSU association and as $\bar{d}_t^{\text{A}}$ for VAP association, respectively. We let $\mathcal{C}_{i,t}$ denote the HUC cluster supporting VUE $i \in \mathcal{U}$ at TS t, and $|\mathcal{C}_{i,t}|$ represent the size of the HUC cluster for VUE i at TS t. Thus, the HUC cluster $\mathcal{C}_{i,t}$ obeys $\mathcal{C}_{i,t} \subset \mathcal{R} \cup \mathcal{A}$ as well as $|\mathcal{C}_{i,t}| \leq \bar{S}_{\text{max}}$ due to the association restriction. Our assumption is that the HUC cluster migration of each VUE takes place right at the beginning of each TS. In contrast to the traditional hard and horizontal HO relying on a single connection, HUC cluster migration takes place, when any one of the links from different access networks supporting a VUE changes its status from active (A) to passive (P) or vice versa. Hence, the HO overhead imposed can be represented by the specific number of connections changing their A/P status, when VUE i moves at TS t, given by

$$e_{i,t} = \max\{|\mathcal{C}_{i,t-1}|, |\mathcal{C}_{i,t}|\} - |\mathcal{C}_{i,t-1} \cap \mathcal{C}_{i,t}|. \tag{2}$$

Herein, the first item denotes the maximum number of connections for VUE i during HO at TS t, whilst the second item is the specific number of the same connections. Notably, due to $|\mathcal{C}_{i,t}| \leq \bar{S}_{\text{max}}$, the HO overhead per TS shall not exceed $\bar{S}_{\text{max}}$.

E. Wireless Communication Model

For simplicity, all RSUs and vehicles are assumed to be equipped with a single antenna. We assume that the real-time locations of both the vehicles and the RSUs are known relying on pre-installed sensors and positioning technology. In order to mitigate the inter-vehicle interference, multiple orthogonal resource blocks are employed to support the vehicles.

For VUE i, we let p_c denote the transmit power of its transmitter c and let $h_{c,i,t}$ denote the channel gains of the link spanning from transmitter c to VUE i at TS t. In this model, we only consider the small-scale fading and path loss of V2V communication and of V2I communication, but neglect the effects of shadow fading. Accordingly, the achievable data rate of VUE i at TS t is represented by

$$R_{i,t} = \log_2(1 + \frac{\sum_{c \in \mathcal{C}_{i,t}} p_c h_{c,i,t}}{\sigma^2}), \tag{3}$$

where σ^2 is the variance of the additive Gaussian white noise (AWGN).

IV. Problem Formulation

Based upon the above HUC cluster migration strategy, we have to strike a trade-off between the total hybrid HO overhead and the quantitative connectivity benefits of HUC clustering. Firstly, let us define the variables $\boldsymbol{X}_{i,t} = \{x_{i,j,t}\}_{j \in \mathcal{R} \cup \mathcal{A}}$ representing the HUC clustering policy of VUE i at TS t. Explicitly, $x_{i,j,t} = 1$ indicates that transmitter j belongs to the HUC cluster $\mathcal{C}_{i,t}$ of VUE i at TS t, i.e.

$$\mathcal{C}_{i,t} = \{j | x_{i,j,t} = 1, j \in \mathcal{R} \cup \mathcal{A}\}. \tag{4}$$

Next, the constraints of the optimization problem are formulated. Due to the above definition, the HUC clustering policy should obey

$$\mathrm{C1}:\ x_{i,j,t}=\{0,1\},\forall i\in\mathcal{U},\ \forall j\in\mathcal{R}\cup\mathcal{A},\forall t\in\mathcal{T}. \tag{5}$$

When the HUC cluster is a set of RSUs at a TS, the VUE should be associated with no more than $\bar{S}_{\max}$ RSUs, thus we have

$$\mathrm{C2}:\ \sum_{j\in\mathcal{R}}x_{i,j,t}\le\bar{S}_{\max},\ \forall i\in\mathcal{U},\ \forall t\in\mathcal{T}. \tag{6}$$

Similarly, as a VAP plays a role of the transmitter at a TS, the VUE can only connect with a single VAP, whilst each VAP can serve at most one VUE simultaneously. In this case, the HUC clustering constraints become

$$\mathrm{C3}:\ \sum_{j\in\mathcal{A}}x_{i,j,t}\le 1,\ \forall i\in\mathcal{U},\ \forall t\in\mathcal{T} \tag{7}$$

and

$$\mathrm{C4}:\ \sum_{i\in\mathcal{U}}x_{i,j,t}\le 1,\ \forall j\in\mathcal{A},\ \forall t\in\mathcal{T}. \tag{8}$$

Moreover, we assume that the RSUs and the VAP cannot serve the VUE simultaneously. Accordingly, the HUC clustering policy is subjected to

$$\mathrm{C5}:\ \sum_{j\in\mathcal{A}}x_{i,j,t}=0,\text{if}\sum_{j\in\mathcal{R}}x_{i,j,t}\ge 1,\forall i\in\mathcal{U},\ \forall t\in\mathcal{T} \tag{9}$$

and

$$\mathrm{C6}:\ \sum_{j\in\mathcal{R}}x_{i,j,t}=0,\text{if}\sum_{j\in\mathcal{A}}x_{i,j,t}=1,\forall i\in\mathcal{U},\ \forall t\in\mathcal{T}. \tag{10}$$

Additionally, in order to guarantee a seamless connection, the data rate at each TS has to reach

$$\mathrm{C7}:\ R_{i,t}\ge\bar{R}_i,\ \forall i\in\mathcal{U},\ \forall t\in\mathcal{T}. \tag{11}$$

Herein, $\bar{R}_i$ denotes the minimum data rate required by VUE i at each TS.

In order to formulate our optimization problem, the connectivity benefits of HUC clustering are quantified in terms of the average data rate attained over all TSs, whilst the hybrid HO overhead is characterized by the number of HOs over all TSs. Ideally, we should only allow the number of HOs to increase, if the connectivity probability sufficiently increased. Hence we have to strike a trade-off. Let us continue by defining a normalized utility function for VUE i at TS t for striking a trade-off, which is formulated as

$$E_{i,t}(\boldsymbol{X}_{i,t},\boldsymbol{X}_{i,t-1})=\kappa\frac{R_{i,t}(\boldsymbol{X}_{i,t})}{\bar{R}_i}-(1-\kappa)\frac{e_{i,t}(\boldsymbol{X}_{i,t},\boldsymbol{X}_{i,t-1})}{\bar{S}_{\max}}, \tag{12}$$

where $\kappa\in[0,1]$ quantifies the weighting factor of the connectivity benefits, while $(1-\kappa)$ is the weighting factor of the HO overhead. Explicitly, the first term represents the relative data rate contribution of VUE i during TS t, while the second term is the normalized HO-rate, i.e. the ratio of the actual number of connections changing their A/P status to the maximum one, when VUE i moves at TS t. It should be noted that $E_{i,t}$ is a function of both the current HUC clustering policy $\boldsymbol{X}_{i,t}$ at TS t and the previous HUC clustering policy $\boldsymbol{X}_{i,t-1}$ at TS $t-1$, because the HUC cluster $C_{i,t}$ is a function of $\boldsymbol{X}_{i,t}$ according to (4).

Given the above constraints and the normalized objective function (OF) of (12), our optimization problem can be cast as

$$(\mathbf{P}):\max\ \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{U}}E_{i,t}(\boldsymbol{X}_{i,t},\boldsymbol{X}_{i,t-1}) \tag{13a}$$

$$\text{s.t.}\ \ \text{C1-C7}. \tag{13b}$$

It can be observed that $(\mathbf{P})$ is an NP-hard binary integer programming problem, which has no exact solution in polynomial time. More explicitly, the exhaustive search has an excessive computational complexity, which is related to the densities of RSUs, VAPs and VUEs as well as the finite number of TSs. Furthermore, the unknown stochastic vehicular mobility makes it infeasible to derive the optimal solution without an explicit model of the environmental dynamics.

Again, the trade-off utility function of (12) depends only upon the present policy as well as upon the previous policy, but it is independent of any earlier policy. Inspired by this, we can decouple the problem $(\mathbf{P})$ into a series of sequential subproblems to be solved for single TS, or exploit its Markovian nature for constructing a Markov Decision Process (MDP). In what follows, we will first propose a low-complexity sequential HUC cluster migration solution in the face of unknown vehicular mobility, by decoupling our problem $(\mathbf{P})$ into sequential decision-making subproblems with the aid of a bipartite matching framework. Next, in order to adapt to unknown dynamic environments, we attempt to adopt a model-free DRL algorithm for designing a holistic HUC cluster migration solution after designing a MDP problem.

V. Max-Bipartite Matching-aided Sequential HUC cluster migration solution

Due to the mobility of vehicles, the HUC cluster migration problem consists of sequential decision-making subproblems at each TS. For reducing the computational complexity imposed, we find the optimal solution for each sequential subproblem without knowing the vehicles' mobility model as a low-complexity alternative solution for solving problem $(\mathbf{P})$. Motivated by this, in this section we propose a low-complexity sequential HUC cluster migration solution by comparing the solution of exhaustive search on all possible VUE-RSU combinations to that of all possible VUE-VAP association combinations using a max-bipartite matching framework. Following the algorithm's description, we analyze the complexity and the optimality of the proposed solution.

A. Solving Sequential Subproblem

In order to decouple problem $(\mathbf{P})$ into sequential decision-making subproblems, our optimization problem becomes that of the optimal solution at each TS based upon the results at the previous TS. Bearing in mind that the minimum data rate constraint C7 is in a non-closed form function of $\boldsymbol{X}_{i,t}$ due to (4), we relax it by ensuring that each VUE can be served by at least one transmitter at each TS, namely by substituting it with $\sum_{j\in\mathcal{R}\cup\mathcal{A}}x_{i,j,t}\ge 1$ for $\forall i\in\mathcal{U}$ and $\forall t\in\mathcal{T}$. Then, given $\{\boldsymbol{X}_{i,t-1}\}_{i\in\mathcal{U}}$, the optimization subproblem $(\mathbf{P_t})$ at TS t is given by

$$(\mathbf{P_t}):\ \max \sum_{i\in\mathcal{U}} E_{i,t}(\boldsymbol{X}_{i,t}) \tag{14a}$$
$$\text{s.t. } x_{i,j,t}=\{0,1\},\ \forall i\in\mathcal{U},\ \forall j\in\mathcal{R}\cup\mathcal{A}, \tag{14b}$$
$$\sum_{j\in\mathcal{R}} x_{i,j,t}\le \bar{S}_{\max},\ \forall i\in\mathcal{U}, \tag{14c}$$
$$\sum_{j\in\mathcal{A}} x_{i,j,t}\le 1,\ \forall i\in\mathcal{U}, \tag{14d}$$
$$\sum_{i\in\mathcal{U}} x_{i,j,t}\le 1,\ \forall j\in\mathcal{A}, \tag{14e}$$
$$\sum_{j\in\mathcal{A}} x_{i,j,t}=0,\ \text{if} \sum_{j\in\mathcal{R}} x_{i,j,t}\ge 1,\ \forall i\in\mathcal{U}, \tag{14f}$$
$$\sum_{j\in\mathcal{R}} x_{i,j,t}=0,\ \text{if} \sum_{j\in\mathcal{A}} x_{i,j,t}= 1,\ \forall i\in\mathcal{U}, \tag{14g}$$
$$\sum_{j\in\mathcal{R}\cup\mathcal{A}} x_{i,j,t}\ge 1,\ \forall i\in\mathcal{U}. \tag{14h}$$

As such, this problem corresponds to a many-to-many bipartite matching problem of classic bipartite graph theory, where some nodes are allowed to connect with multiple counterparts. But this kind of many-to-many max-bipartite matching problem formulated under specific constraints cannot be solved directly by the existing max-bipartite matching techniques. More explicitly, classic max-bipartite matching methods, such as the Kuhn-Munkres algorithm, are used for solving the personnel-assignment problem (page 32 in [46]) that "chooses a set of n independent[1] elements of the matrix so that the sum of these elements is maximum". This means that the Kuhn-Munkres algorithm can only allow each node to connect with at most one counterpart without any additional constraint. However, inspired by constraints (14f) and (14g), the optimal solution can be obtained by comparing the independently found solutions based upon the set of $\mathcal{R}$ and that of $\mathcal{A}$, respectively. Accordingly, we first conceive the subproblem of finding an optimal RSU set as follows:

$$(\mathbf{P^1_t}):\ \max \sum_{i\in\mathcal{U}} E_{i,t}(\boldsymbol{X}_{i,t}) \tag{15a}$$
$$\text{s.t. } x_{i,j,t}=\{0,1\},\forall i\in\mathcal{U},\forall j\in\mathcal{R}\cup\mathcal{A}, \tag{15b}$$
$$\sum_{j\in\mathcal{R}} x_{i,j,t}\le \bar{S}_{\max},\ \forall i\in\mathcal{U}, \tag{15c}$$
$$\sum_{j\in\mathcal{R}} x_{i,j,t}\ge 1,\ \forall i\in\mathcal{U}, \tag{15d}$$
$$\sum_{j\in\mathcal{A}} x_{i,j,t}= 0,\ \forall i\in\mathcal{U}. \tag{15e}$$

Let $\boldsymbol{X}^{\mathrm{R}}_{i,t}$ denote the optimal solution of problem $(\mathbf{P^1_t})$ and $\mathcal{R}^*_{i,t}$ represent the corresponding optimal RSU set for VUE i at TS t. Having acquired the results of $\{\mathcal{R}^*_{i,t}\}_{\forall i\in\mathcal{U}}$, we then add it as an alternative HUC cluster for each VUE at TS t and compare it to other VUE-VAP associations. Consequently, problem $(\mathbf{P_t})$ becomes a classic bipartite matching problem of finding the optimal solution for all VUEs based upon all available VUE-VAP associations and the optimal VUE-RSU associations solved by $(\mathbf{P^1_t})$.

To be more specific, we first define a set $\mathcal{A}'=\{1,..,A,A+1,...,A+U\}$ denoting all the associations for all VUEs based upon $\{\mathcal{R}^*_{i,t}\}_{\forall i\in\mathcal{U}}$. Then, we introduce an auxiliary vector $\boldsymbol{X}'_{i,t}=\{x'_{i,\iota,t}\}_{\iota\in\mathcal{A}'}$ for VUE i at TS t, whose elements indicate its association either to a VAP or to the optimal RSU set. More explicitly, let ι^* denote the index satisfying $x'_{i,\iota,t}=1$. Then we have

$$\mathcal{C}_{i,t}=\begin{cases}\{\iota^*\},\ \text{if } \iota^*\in\mathcal{A},\\ \mathcal{R}^*_{i,t},\ \text{if } \iota^*=[\mathcal{A}'\setminus\mathcal{A}]_i,\\ \emptyset,\ \text{otherwise},\end{cases} \tag{16}$$

where $[\mathcal{A}'\setminus\mathcal{A}]_i$ denotes the i-th selection in the set $\mathcal{A}'$, except for $\mathcal{A}$, which exactly corresponds to the optimal RSU set for VUE i. Herein, the first case represents that VUE i is associated with the ι^*-th VAP at TS t, whilst the second case indicates that VUE i selects its optimal RSU set as its HUC cluster at TS t. Let us consider Fig. 2 as an example, where $U=3$ and $A=2$. At a TS, three VUEs share two VAPs, while having their optimal RSU set. Each VUE can select VAP 1, VAP 2 or its optimal RSU set as its HUC cluster.

As a consequence, on the basis of the solution of $(\mathbf{P}^1_t)$, problem $(\mathbf{P_t})$ can be reformulated as problem $(\mathbf{P}^2_t)$, given by

$$(\mathbf{P^2_t}):\ \max \sum_{i\in\mathcal{U}} E_{i,t}(\boldsymbol{X}'_{i,t}) \tag{17a}$$
$$\text{s.t. } x'_{i,j,t}=\{0,1\},\forall i\in\mathcal{U},\forall j\in\mathcal{A}', \tag{17b}$$
$$\sum_{j\in\mathcal{A}'} x'_{i,j,t}=1,\forall i\in\mathcal{U}, \tag{17c}$$
$$\sum_{i\in\mathcal{U}} x'_{i,j,t}\le 1,\ \forall j\in\mathcal{A}'. \tag{17d}$$

More explicitly, the following Theorem formalizes the relationship between problem $(\mathbf{P}^1_t)$, problem $(\mathbf{P}^2_t)$ and problem $(\mathbf{P_t})$.

Theorem 1: The optimal solution of $(\mathbf{P}^2_t)$ is the optimal solution of problem $(\mathbf{P_t})$ on the basis of the solution of $(\mathbf{P}^1_t)$.

Proof The proof is given in Appendix A. □

In what follows, we will elaborate on the solutions of $(\mathbf{P^1_t})$ and $(\mathbf{P^2_t})$.

1) Stage I - Solving $(\mathbf{P^1_t})$*:* Since $(\mathbf{P^1_t})$ represents the binary integer programming problem of finding an optimal RSU set, its optimal solution can be obtained by exhaustive search, which has a computational complexity related to the number of available RSUs and of VUEs. Moreover, $(\mathbf{P^1_t})$ can also be further decoupled into the A independent subproblems of each VUE. Thus, we may opt for searching no more than $\bar{S}_{\max}$ available RSUs, which maximizes each OF $E_{i,t}$ for VUE $\forall i\in\mathcal{U}$ independently.

[1] In [46], a set of elements of a matrix are defined as "independent" provided that no two of them lie in the same row or column of the matrix simultaneously.

2) Stage II - Solving $(\mathbf{P_t^2})$*:* In essence, problem $(\mathbf{P_t^2})$ corresponds to finding a permutation for one-to-one associations. However, according to the constraint (17d), the VUEs may compete for becoming associated with the same VAP. Hence, the solutions of all VUEs are coupled to avoid such an association conflict. Given that each VUE can only be associated with at most one VAP, because it has to satisfy both constraints (17b) and (17c), this problem is constituted by a max-bipartite matching problem. In order to solve it, we adopt the Kuhn-Munkres algorithm [46] [47] for finding the optimal VUE-VAP association.

B. Algorithm Description

Before describing our proposed algorithm, we introduce the following definitions of max-bipartite matching [48] [49]. First, let a graph be denoted by $\mathcal{G} = (\mathcal{K}, \mathcal{E})$, where $\mathcal{K}$ is the vertex set and $\mathcal{E}$ is the edge set. Then, formally, a *bipartite matching* is stated as follows:

Definition 1: A graph $\mathcal{G} = (\mathcal{K}, \mathcal{E})$ is *bipartite* if there exist partitions $\mathcal{Y}$ and $\mathcal{Z}$ that satisfy $\mathcal{K} = \mathcal{Y} \cup \mathcal{Z}$ with $\mathcal{Y} \cap \mathcal{Z} = \emptyset$ and $\mathcal{E} \subseteq \mathcal{Y} \times \mathcal{Z}$.

Definition 2: A *bipartite matching* is a subset $\mathcal{M} \subseteq \mathcal{K}$ for the bipartite graph $\mathcal{G} = (\mathcal{K}, \mathcal{E})$, so that at most one edge in $\mathcal{M}$ is incident upon $\forall k \in \mathcal{K}$. The weight of $\mathcal{M}$ is defined as the sum of the weights of all edges in $\mathcal{M}$.

Based upon the above fundamental definitions, a *max-bipartite matching* is stated as follows:

Definition 3: A *max-bipartite matching* is a matching $\mathcal{M}$ from $\mathcal{Y}$ to $\mathcal{Z}$ having a maximum weight.

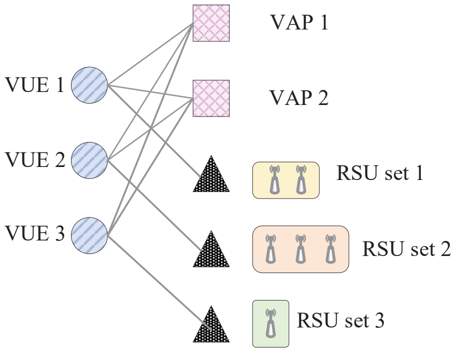
To solve this max-bipartite matching problem of $(\mathbf{P_t^2})$ by finding the matching having the maximum sum of weights of all edges in $\mathcal{G}$, first the graph $\mathcal{G}$ and its partitions have to be defined. Let graph $\mathcal{G} = (\mathcal{U} \cup \mathcal{A}', \mathcal{E})$, where the VUE set $\mathcal{U}$ and the union set $\mathcal{A}'$ are its partitions. Herein, $\mathcal{E}$ is the corresponding edge set that represents the association between VUEs and VAPs or RSUs. Then, let the weight $w(i,k)$ of the edge $(i,k) \in \mathcal{E}$ denote the trade-off utility function when VUE $i \in \mathcal{U}$ connects node $k \in \mathcal{A}'$, formulated as

$$w(i,k) = \begin{cases} E_{i,t}(\boldsymbol{X}_{i,t}^{\mathrm{A}}), & \text{if } k \in \mathcal{A}, \\ E_{i,t}(\boldsymbol{X}_{i,t}^{\mathrm{R}}), & \text{if } k = [\mathcal{A}' \setminus \mathcal{A}]_i, \\ 0, & \text{otherwise}, \end{cases} \tag{18}$$

where $\boldsymbol{X}_{i,t}^{\mathrm{A}} = \boldsymbol{X}_{i,t}|_{(x_{i,k,t}=1, x_{i,j,t}=0, \forall j \neq k)}$ is the utility function when VUE i connects to VAP k at TS t, and $\boldsymbol{X}_{i,t}^{\mathrm{R}}$ is the optimal solution of problem $(\mathbf{P_t^1})$. For those invalid connections, we add virtual vertices and edges with zero weight to construct a complete weighted graph. Mathematically, finding a max-bipartite matching can be formulated as the problem of

$$\max_{\mathcal{M}:\mathcal{U} \to \mathcal{A}'} \sum_{i \in \mathcal{U}} \sum_{k \in \mathcal{A}'} w(i,k). \tag{19}$$

In the example of Fig. 2, the weights between VUEs and VAPs $[w(1,1)$, $w(1,2)$, $w(2,1)$, $w(2,2)$, $w(3,1)$ and $w(3,2)]$ are obtained by assuming that their connections exist, respectively. By contrast, the weights $w(1,3)$, $w(2,4)$ and $w(3,5)$ represent the maximum trade-off utility function (12) for each VUE, when it connects to its optimal RSU set.



	VAP 1	VAP 2	RSU set 1	RSU set 2	RSU set 3
VUE 1	$w(1,1)$	$w(1,2)$	$w(1,3)$	0	0
VUE 2	$w(2,1)$	$w(2,2)$	0	$w(2,4)$	0
VUE 3	$w(3,1)$	$w(3,2)$	0	0	$w(3,5)$

Fig. 2: An illustration of the Proposed Max-Bipartite Matching Model Construction.

Based upon the weight of edges, a function termed as *feasible vertex labeling* is introduced as follows:

Definition 4: A *feasible vertex labeling* in $\mathcal{G}$ is a real-valued function l on $\mathcal{U} \cup \mathcal{A}'$, so that for all $i \in \mathcal{U}$ and $k \in \mathcal{A}'$,

$$l(i) + l(k) \geq w(i,k). \tag{20}$$

In this paper, we adopt a simple feasible vertex labeling which relies on simply labelling a vertex with the largest weight associated with all edges leading to the vertex, namely $l(i) = \max_{k \in \mathcal{A}'} w(i,k)$ for $i \in \mathcal{U}$ and $l(k) = 0$ for $k \in \mathcal{A}'$. In this case, a graph $\mathcal{G}_l = (\mathcal{U} \cup \mathcal{A}', \mathcal{E}_l)$ is termed as the *equality subgraph* for a given l, where we have

$$\mathcal{E}_l = \{(i,k) | l(i) + l(k) = w(i,k)\}. \tag{21}$$

Based on the above construction, the detailed Kuhn-Munkres algorithm starts with an initial feasible vertex labeling l on one side of the graph having the maximum weights. Then, an initial maximum matching $\mathcal{M}$ is formed by finding the matching having the maximum sum of weights of all edges in $\mathcal{G}_l$. If the matching is complete, the algorithm will be terminated, giving the maximum weights. Otherwise, it begins iterating by attempting to find a larger matching by using the *augmenting alternating paths* technique and updating its labeling to find the corresponding maximum assignment. Through increasing the size of $\mathcal{M}$ in each iteration, the above procedure will eventually terminate due to the limited number of vertices in $\mathcal{G}$, and a maximum-weighted bipartite matching will be found.

In a nutshell, the proposed max-bipartite matching-based sequential HUC cluster migration solution is presented in Algorithm 1 (ALG1), where the stages I and II are repeatedly and sequentially performed at each TS.

C. Algorithm Analysis

In the following, we analyze the optimality and the computational complexity of the proposed sequential HUC cluster migration solution.

Algorithm 1 Proposed Max-Bipartite Matching-Based Sequential HUC Cluster Migration Solution

```
1:  Initialize: $\boldsymbol{X}^{\mathrm{R}}_{i,0} = \mathbf{0}^{(R+A)\times 1}$, $\boldsymbol{X}^{\mathrm{A}}_{i,0} = \mathbf{0}^{(R+A)\times 1}$ $(\forall i \in \mathcal{U})$;
2:  for all t=1,..,T do
3:      Stage I – Solving ($\mathbf{P^1_t}$):
4:      for all i=1,...,U do
5:          Find the optimal VUE-RSU association solution $\boldsymbol{X}^{\mathrm{R}}_{i,t} = \arg\max_{||\boldsymbol{X}_{i,t}||_1 \le \bar{S}_{\max}} E_{i,t}(\boldsymbol{X}_{i,t})$ by exhaustive search and calculating $E_{i,t}(\boldsymbol{X}_{i,t}) = \kappa \frac{R_{i,t}(\boldsymbol{X}_{i,t})}{\bar{R}_i} - (1-\kappa)\frac{e_{i,t}(\boldsymbol{X}_{i,t}, \boldsymbol{X}_{i,t-1})}{\bar{S}_{\max}}$ ;
6:      end for
7:      Calculate all the associations $\mathcal{A}'$ based upon $\boldsymbol{X}^{\mathrm{R}}_{i,t}$;
8:      Stage II – Solving ($\mathbf{P^2_t}$):
9:      Construct graph $\mathcal{G} = (\mathcal{U} \cup \mathcal{A}', \mathcal{E})$ and calculate weight $w(i,k)$ for $i \in \mathcal{U}$ and $k \in \mathcal{A}'$:
            $$w(i,k) = \begin{cases} E_{i,t}(\boldsymbol{X}^{\mathrm{A}}_{i,t}), \text{ if } k \in \mathcal{A}, \\ E_{i,t}(\boldsymbol{X}^{\mathrm{R}}_{i,t}), \text{ if } k = [\mathcal{A}' \setminus \mathcal{A}]_i, \\ 0, \text{ otherwise.} \end{cases}$$
10:     Initialize: $l(i) = \max_{k \in \mathcal{A}'} w(i,k)$ for $i \in \mathcal{U}$ and $l(k) = 0$ for $k \in \mathcal{A}'$;
11:     Determine equality subgraph $\mathcal{G}_l = (\mathcal{U} \cup \mathcal{A}', \mathcal{E}_l)$ such that
            $$\mathcal{E}_l = \{(i,k) | l(i) + l(k) = w(i,k)\},$$
        and choose an initial maximum matching $\mathcal{M}$ in $\mathcal{G}_l$;
12:     while matching $\mathcal{M}$ is NOT complete for $\mathcal{G}$ do
13:         Find an unmatched VUE $i \in \mathcal{U}$; set $\mathcal{S} = \{i\}$ and $\mathcal{P} = \emptyset$;
14:         Set $\mathcal{J}_l(\mathcal{S}) = \{k | \forall i \in \mathcal{S} : (i,k) \in \mathcal{E}_l\}$.
15:         if $\mathcal{J}_l(\mathcal{S}) = \mathcal{P}$ then
16:             Find $\alpha_l = \min_{i \in \mathcal{S}, k \in \mathcal{A}' \setminus \mathcal{P}} \{l(i) + l(k) - w(i,k)\}$ and construct
                $$l'(\kappa) = \begin{cases} l(\kappa) - \alpha_l, \forall \kappa \in \mathcal{S}, \\ l(\kappa) + \alpha_l, \forall \kappa \in \mathcal{P}, \\ l(\kappa), \text{otherwise.} \end{cases}$$
17:             Replace $l$ by $l'$ and $G_l$ by $G_{l'}$.
18:         else
19:             Select $k$ from $\mathcal{J}_l(\mathcal{S}) \setminus \mathcal{P}$;
20:             if $k$ is unmatched then
21:                 Augmenting $\mathcal{M}$ with path $(i,k)$; Go to step 12;
22:             else
23:                 Find the matched VUE $i' \in \mathcal{U}$ with node $k$ and extend the alternating path by updating $\mathcal{S} = \mathcal{S} \cup \{i'\}$ and $\mathcal{P} = \mathcal{P} \cup \{k\}$. Go to step 14.
24:             end if
25:         end if
26:     end while
27:     Calculate $\boldsymbol{X}_{i,t}$ according to $\mathcal{M}$.
28: end for
29: Output: $\{\boldsymbol{X}_{i,t}\}_{\forall i, \forall t}$
```

1) Optimality: In max-bipartite matching theory, the Kuhn-Munkres algorithm is the optimal solution that achieves the maximum weight, as long as its adopted labeling function l is feasible and its output M is a complete matching [47]. Based upon this property, the optimality of our solution is presented in the following proposition:

Proposition 1: The complete matching of problem ($\mathbf{P^2_t}$) is the optimal solution of problem ($\mathbf{P_t}$).

Remark 1: Proposition 1 manifests the optimality of ALG1 for solving problem ($\mathbf{P_t}$) when the solution of ($\mathbf{P_{t-1}}$) is given, but ALG1 may not be the optimal solution of problem ($\mathbf{P}$).

2) Computational Complexity: Let us now discuss the computational complexity of ALG1. As far as Stage I is concerned, there are $\sum_{\tau=1}^{\bar{S}_i} \binom{L_i}{\tau}$ possible VUE-VAP association combinations for VUE i, where L_i is the number of candidate RSUs within the coverage distance threshold $\bar{d}^{\mathrm{R}}_t$ and the actual number of RSUs associated with VUE i is at most $\bar{S}_i = \min\{\bar{S}_{\max}, L_i\}$. Next, in terms of Stage II, the computational complexity of the Kuhn-Munkres algorithm is related to the number of vertices in the constructed graph, which is given by $\mathcal{O}\left((2U+A)^3\right)$. To sum up, the overall sequential HUC cluster migration solution has a computational complexity of $\mathcal{O}\left(T \cdot \left[\sum_{i=1}^{U} \sum_{\tau=1}^{\bar{S}_i} \binom{L_i}{\tau} + (2U+A)^3\right]\right)$, which increases polynomially.

VI. DRL-aided Holistic HUC cluster migration Solution

Although the sequential HUC cluster migration process does indeed generate a potential solution at a moderate complexity in the face of unknown vehicular mobility, the overall optimality of such a dynamic decision-making solution cannot be guaranteed. As an emerging innovative method of finding a good policy for model-free MDP problems, DRL relies on deep neural networks (DNNs) invoked for sophisticated mappings between the input and the desired output based upon a large amount of training data, which eventually yields a beneficial mapping from the state space to the action space. The advantage of DRL is that the agent can learn from experience via interaction with the environment, even in the absence of knowing the environmental dynamics in advance. Once the policy becomes sufficiently well trained, it can be employed by the agent in the same environment during the test stage, or even be generalized for the agent in different environments. In this section, we resort to an alternative policy optimization based model-free DRL method – namely to the DDPG algorithm of [50] for solving our problem, which benefits from the advantages of being an actor-critic deterministic policy gradient algorithm, as well as Deep Q Network (DQN) [31]. Compared to the most representative DQN method, which can only handle the problems having a low-dimensional discrete action space, the DDPG method converges much faster than DQN and also supports high-dimensional state-action spaces. The underlying reason is that DQN requires exhaustive evaluation of the Q-function of all possible actions at each step, whilst DDPG is capable of generating a deterministic action from the policy network.

To better understand the theoretical basis of DDPG, in the following, we will first introduce the relevant fundamentals and then present our MDP design for the proposed DRL-aided holistic HUC cluster migration solution.

A. Fundamentals of DDPG

1) MDP: A MDP is formalized by a 5-tuple $(\mathcal{S}, \mathcal{A}, p, r, \gamma)$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $p : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ is the transition probability, $r : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ is the reward function and $\gamma \in (0, 1)$ is the discount factor. In general, the return of agent from a state at TS t is a discounted future cumulative reward from TS t, formulated as

$$G_t = \sum_{l=0}^{\infty} \gamma^l r_{t+l+1}, \tag{22}$$

Herein, r_{t+l+1} denotes the reward at TS $t+l+1$. The essence of DRL is that of finding a policy capable of maximizing the expectation of this discounted future cumulative reward [51].

2) Value function: Let $\pi : \mathcal{S} \times \mathcal{A} \to [0, 1]$ be a stochastic policy. The value function of the policy π at state s is defined as the expected discounted future cumulative reward from state s:

$$v^{\pi}(s) = \mathbb{E}[G_t | s_t = s]. \tag{23}$$

Similarly, the action-value function $Q^{\pi}(s, a)$ is defined as the expected discounted future cumulative reward from state s when the agent takes action a:

$$Q^{\pi}(s, a) = \mathbb{E}[G_t | s_t = s, a_t = a]. \tag{24}$$

3) Policy Gradient: In practice, when seeking the optimal policy becomes infeasible for large state spaces, the policy gradient is used for evaluating the performance of the policy by parameterizations. More explicitly, let $\pi^{\theta}(a|s)$ denote the policy at state s, when taking action a using the parameter vector θ. Accordingly, the probability of taking action a_t based on θ is

$$\pi_{\theta}(a|s) = P\{a_t = a | s_t = s, \theta_t = \theta\}. \tag{25}$$

Let $J(\theta)$ be the OF of policy optimization-based algorithms, and the policy gradient be given by

$$\nabla_{\theta} J(\theta) = \mathbb{E}_{\pi_{\theta}}[\nabla_{\theta} \log \pi_{\theta}(a|s) Q^{\pi_{\theta}}(s, a)]. \tag{26}$$

4) DDPG: Essentially, DDPG consists of an actor network and a critic network, which learns the policy and estimates the Q-function, respectively. Thus, a Q network and a policy network in DDPG are defined as follows:

- Q network: Having said that, the Q-function describes the expected reward after taking an action in the current state, when following a policy π. As the policy is deterministic, the Bellman equation reflecting the recursive nature of the Q-function can be described by

  $$Q(s_t, a_t) = \mathbb{E}_{s_{t+1}}[r_t + \gamma Q(s_{t+1}, a_{t+1})]. \tag{27}$$

  In the case of a high-dimensional action or state space, the Q-function of (27) can be approximated by a DNN having the weights of $\{\theta^Q\}$ as a Q-network $Q(s, a|\theta^Q)$. Once $\{\theta^Q\}$ is determined, $Q(s, a)$ will represent the outputs of the DNN. Then, the Q-network plays a role of a critic function by appraising the benefits of the action, and the function approximators $\{\theta^Q\}$ are optimized based on minimizing the loss function, represented as

  $$L(\theta^Q) = \mathbb{E}_{s_t, a_t, r_t}[(\zeta_t - Q(s_t, a_t|\theta^Q))^2], \tag{28}$$

  where

  $$\zeta_t = r_t + \gamma Q(s_{t+1}, \pi(s_{t+1})|\theta^Q). \tag{29}$$

- Policy network: As an actor network, a policy network in DDPG outputs the deterministic action, given the current state. Similarly, the policy network $\pi(s|\theta^{\pi})$ can also be approximated by a DNN in conjunction with the weights $\{\theta^{\pi}\}$. According to Theorem 1 in [52], the policy is optimized by following the policy gradient, given by

  $$\begin{aligned} \nabla_{\theta^{\pi}} J &\approx \mathbb{E}_{s_t}[\nabla_{\theta^{\pi}} Q(s, a|\theta^Q)|_{s=s_t, a=\pi(s_t|\theta^{\pi})}] \\ &= \mathbb{E}_{s_t}[\nabla_a Q(s, a|\theta^Q)|_{s=s_t, a=\pi(s_t)} \nabla_{\theta^{\pi}} \pi(s|\theta^{\pi})|_{s=s_t}]. \end{aligned} \tag{30}$$

In order to improve the stability of learning, a pair of separate target networks has to be established as a copy of the actor and critic networks, namely $Q'(s, a|\theta^{Q'})$ and $\pi'(s|\theta^{\pi'})$, respectively. For the sake of distinction, the original networks are termed as online Q-networks and online policy networks, respectively, since both target networks are updated by arranging for them to slowly track the networks learned, which is formulated as:

$$\theta^{Q'} \leftarrow \tau \theta^Q + (1 - \tau)\theta^{Q'}, \tag{31}$$

$$\theta^{\pi'} \leftarrow \tau \theta^{\pi} + (1 - \tau)\theta^{\pi'}, \tag{32}$$

where $\tau \in [0, 1]$ is the tracking parameter.

Furthermore, the replay buffer strategy of [50] is employed in our DDPG to avoid having correlated samples. At each step, the actor and the critic are updated by uniformly sampling a minibatch $\{(s_t, a_t, r_t, s_{t+1})\}_{t=1}^{N}$ of N transitions from this buffer. As a result, the sampled policy gradient can be represented by

$$\nabla_{\theta^{\pi}}^{\text{sampled}} J \approx \frac{1}{N} \sum_{t=1}^{N} \nabla_a Q(s, a|\theta^Q)|_{s=s_t, a=\pi(s_t)} \nabla_{\theta^{\pi}} \pi(s|\theta^{\pi})|s_t. \tag{33}$$

Additionally, an exploration policy π' is constructed by adding noise sampled from a Ornstein-Uhlenbeck (OU) process of $\mathcal{U}^{\text{OU}}$ to our actor policy as follows

$$\pi'(s_t) = \pi(s_t|\theta^{\pi}) + \mathcal{U}^{\text{OU}}. \tag{34}$$

In a nutshell, the DDPG method exploits both the off-policy data and the Bellman equation to learn the DQN, and then uses the DQN to learn the policy via another DNN, which eventually concurrently learns a Q-function and a policy. The framework of our DDPG is depicted in Fig. 3 at a glance. It has been shown in [50] that the DDPG method is capable of learning beneficial policies using its straightforward actor-critic architecture.

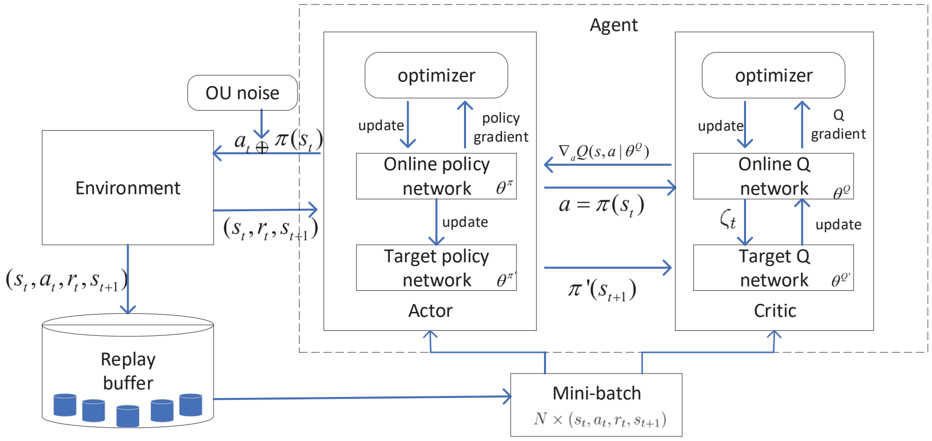

Agent
OU noise
$a_t \oplus \pi(s_t)$
Environment
(s_t, r_t, s_{t+1})
(s_t, a_t, r_t, s_{t+1})
Replay buffer
optimizer
update
policy gradient
Online policy network θ^π
update
Target policy network $\theta^{\pi'}$
Actor
$\nabla_a Q(s,a|\theta^Q)$
$a = \pi(s_t)$
$\pi'(s_{t+1})$
optimizer
update
Q gradient
Online Q network θ^Q
ζ_t
update
Target Q network $\theta^{Q'}$
Critic
Mini-batch
$N \times (s_t, a_t, r_t, s_{t+1})$

Fig. 3: The framework of the DDPG algorithm.

B. MDP design

Clearly, our problem is actually a dynamic decision-making problem. As a result of its unknown state transition probability, the problem constitutes a model-free MDP defined as follows:

- *Agent:* All VUEs are jointly considered as the agent.
- *Action Space* $\mathcal{A}$: At TS t, the action of VUE i is defined by $a_{i,t} = \{x_{i,j,t}\}$. Hence, the system's action at TS t can be represented as $a_t = \{a_{i,t}\}$. Given the constraint of having a limited maximum number of the associated RSUs, the dimension of the action space at each TS is given by $2^{U(\bar{S}_{\max}+A)}$.
- *State Space* $\mathcal{S}$: At each TS, the state of the environment for the agent includes the locations of all VUEs, that of each of their observable RSUs, the locations of all VAPs, and their corresponding HUC clustering status (the locations of the associated transmitters) in the previous TS. The state exhibits the Markovian property due to the assumption of having a Gauss-Markov vehicular mobility model. To satisfy constraint C2, we assume that each VUE only observes at most $\bar{S}_{\max}$ closest RSUs within $\bar{d}_t^{\mathrm{R}}$ at each TS. Additionally, for satisfying constraint C3 and C4, we define a variable ξ_t to represent the collision status of all VAPs at TS t. Explicitly, since there are more than one connections to any VAP at each TS, a collision takes place and we have $\xi_t = 1$. Let $\mathcal{L}_{i,t}^{\mathrm{VUE}}$, $\mathcal{L}_{i,t}^{\mathrm{RSU}}$, $\mathcal{L}_{i,t}^{\mathrm{HUC}}$ and $\mathcal{L}_t^{\mathrm{VAP}}$ denote the location set of VUE i, that of its observable RSUs, that of its previous HUC clustering (the associated RSUs or VAP at the previous TS), and of all VAPs at TS t, respectively. Thus, the system state at TS t becomes

$$s_t = \left[\{\mathcal{L}_{i,t}^{\mathrm{VUE}}\}_i, \{\mathcal{L}_{i,t}^{\mathrm{RSU}}\}_i, \{\mathcal{L}_{i,t-1}^{\mathrm{HUC}}\}_i, \mathcal{L}_t^{\mathrm{VAP}}, \xi_t\right]. \quad (35)$$

- *Immediate Reward*: In our scenario, the immediate reward r_t is determined by the trade-off utility function of (12). To facilitate the learning process, we add a penalty term ρ_t when C5-C7 are not satisfied and consider the per-user average trade-off performance , represented as

$$r_t = \frac{1}{U}\sum_{i\in\mathcal{U}} E_{i,t} + \rho_t. \quad (36)$$

In what follows, we will elaborate on the training and testing stages of our DDPG-based holistic HUC cluster migration solution.

C. Training and Testing

Algorithm 2 Training Stage for the DDPG-based Holistic HUC Cluster Migration Solution

```
Randomly initialize online networks $Q(s,a|\theta^Q)$ and $\pi(s|\theta^\pi)$;
Initialize target networks $Q'(s,a|\theta^{Q'})$ and $\pi'(s|\theta^{\pi'})$ with $\theta^{Q'} \leftarrow \theta^Q$ and $\theta^{\pi'} \leftarrow \theta^\pi$;
Initialize replay buffer $B$;
for all episode=1,..,M do
    Reset simulation parameters for HUC cluster migration vehicular networks environment;
    Initialize $\mathcal{U}^{\mathrm{OU}}$;
    Generate $s_1$ based on the generated initial velocity of all vehicles;
    for all t=1,...,T do
        Select action $a_t$ according to $\pi'(s_t) = \pi(s_t|\theta^\pi) + \mathcal{U}^{\mathrm{OU}}$ and 'binarize' $a_t$ by comparing to 0.5;
        Execute $a_t$ and observe $r_t$, $s_{t+1}$;
        Store the tuple $(s_t, a_t, r_t, s_{t+1})$ in $B$;
        Sample a random minibatch of $N$ tuples $\{(s_t, a_t, r_t, s_{t+1})\}_{t=1}^N$ from $B$;
        Set $\zeta_t = r_t + \gamma Q(s_{t+1}, \pi(s_{t+1})|\theta^Q)$;
        Update the critic network by minimizing $L(\theta^Q) = \mathbb{E}_{s_t,a_t,r_t}[(\zeta_t - Q(s_t, a_t|\theta^Q))^2]$;
        Update the actor network by using $\nabla_{\theta^\pi}^{\mathrm{sampled}} J \approx \frac{1}{N}\sum_{t=1}^N \nabla_a Q(s,a|\theta^Q)|_{s=s_t,a=\pi(s_t)} \nabla_{\theta^\pi}\pi(s|\theta^\pi)|s_t$;
        Update the target network:
            $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$
            $\theta^{\pi'} \leftarrow \tau\theta^\pi + (1-\tau)\theta^{\pi'}$
    end for
end for
```

There are two stages including training and testing for the DDPG based HUC cluster migration solution. The training stage is used for generating training data by the simulated environment, which mimics the interaction of the VUEs with the vehicular networks. In the testing stage, the agent will first load its network parameters (θ^{Q}, $\theta^{Q'}$, θ^{π}, $\theta^{\pi'}$) and then reset its replay buffer by interacting with a randomly initialized environment. The action will be selected according to the output of the actor network with loading parameters, while the state will be yielded relying on its local observation.

The detailed training stage is illustrated in Algorithm 2 (ALG2). The initial state s_1 is based upon the observation according to the initial policy, where the VUEs associate with the nearest $\bar{S}_{\max}$ observable RSUs. Then, the policy is gradually improved by invoking the updated actor and critic networks. It is worthy noting that since the action of the DDPG outputs a continuous value, we can 'binarize' the action by comparing the continuous output to an appropriately selected threshold (we set it as 0.5 in this paper). In this way, once the policy becomes sufficiently well trained, it can be employed for VUEs in the same environment during the test stage for dynamically selecting its HUC cluster, which eventually completes the HUC cluster migration in the face of unknown vehicular mobility by judiciously balancing the HO overhead against the connectivity benefits of HUC clustering in terms of the average data rate.

D. Computational Complexity

The computational complexity of the DDPG approach imposed during both the training and testing stages mainly depends on the actor and the critic networks, where the number of floating point operations (FLOPs) is widely adopted as the complexity metric. Let us assume that a three-layer fully connected DNN is adopted for both the actor and the critic network, where D_1, D_2, D_3 are the number of neurons in the three hidden layers. When a state is the input of the actor network during each TS, its computational complexity can be shown to be on the order of $\mathcal{O}(D_1 + D_1D_2 + D_2D_3 + D_3)$ in terms of the number of FLOPs due to the fact that the output is a deterministic action. Subsequently, the number of FLOPs required for the critic network during each TS can be represented as $\mathcal{O}(D_1+D_1D_2+D_2D_3+D_3)$, when the current state and the executed action are the inputs. By contrast, the computational complexity of the DQN during each TS requires $\mathcal{O}(D_1 + D_1D_2 + D_2D_3 + D_3 2^{U(\bar{S}_{\max}+A)})$ FLOPs, when a three-layer fully connected DNN is adopted as well. Therefore, the DDPG method adopted significantly reduces the computational complexity in our problem.

To reflect on the applicability of our solution in realistic networks, we also compare the overall computation time of the DDPG method and of the DQN method in the context of our HUC clustering design. We tested it on our computer using an Intel Core i9-9920X processor and 32 GB RAM for 1000 training episodes and 2 VUEs. The computation time is 410.351482 seconds for DDPG and 1372.472 seconds for DQN, respectively. Hence, the DDPG method is more attractive for realistic networks.

VII. Numerical Evaluation

In this section, we characterize the performance of two proposed HUC cluster migration solutions by our numerical simulations. We focus our attention on a V2X network wherein the RSUs are located uniformly along both sides of the road at the same interval. The default system parameters are listed in TABLE III. Both the VUEs and VAPs are travelling along the different lanes of the road with a random mean velocity from the interval of $[6\text{m/s}, 12\text{m/s}]$ by default. Their initial velocity is set to be the same as the mean velocity. Moreover, the VUEs are located in the center of a lane in sequence, such as their x-coordinates are $[0, 1, 2, 3, 4]$ in meters for 5 VUEs. By contrast, the VAPs are randomly located in the center of another lane along the road. Additionally, the numerical results are averaged over 100 episodes.

As far as ALG2 is concerned, the DNNs invoked for the actor and for the critic rely on a three-layer fully connected neural network, where both the number of the neurons in the hidden layers is 30. The activation function 'Relu' and 'Tanh' [53] are used in the hidden layer and the output layer of the actor DNN, respectively, whilst the activation function 'Relu' and 'Linear' are used in the hidden layer and the output layer of the critic DNN, respectively. We adopt the adaptive moment estimation method (Adam) optimizer [54] for training. The other default training parameters are also listed in TABLE III.

TABLE III: Simulation Parameters

System Parameters	
Length of road L	1 km
Width of road W	$3.75 \times 2 = 7.5$ m
Number of RSUs R	20
Number of VAPs A	5
Number of VUEs U	5
Coverage distance threshold $\bar{d}_t^{\rm R}$	200 m
Coverage distance threshold $\bar{d}_t^{\rm A}$	50 m
Maximum number of associated RSUs $\bar{S}_{\max}$	4
Minimum data rate constraint	15 b/s/Hz
Transmit power of RSUs p_0	30 dBm
Transmit power of VAPs p_1	30 dBm
Noise power density (5 dB figure)	-174 dBm/Hz
Subcarrier bandwidth	180 kHz
Vehicle mobility model	$\alpha_i = 0.1,\ \bar{\sigma}_i = 0.1,$ $\bar{v}_i = [6$ m/s, 12 m/s$],$ $v_{i,0} = [6$ m/s, 12 m/s$],$ $\sigma_n^2 = 1$
Fast fading	Rayleigh fading $\mathcal{CN}(0,1)$
Path loss model	[11]
Weighting factor w	0.5

Training Parameters	
Learning rate of actor network	0.001
Learning rate of critic network	0.002
Noise standard deviation of $\mathcal{U}^{\rm OU}$	2 in a decay rate of .9995
Buffer capacity B	10 000
Discount factor γ	0.9
Size of minibatch N	32
Tracking parameter τ	0.01
Penalty term ρ_t	-1

For comparison, we consider the following two benchmarkers:

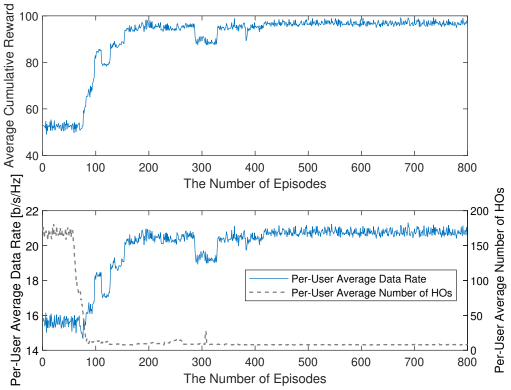


Fig. 4: Convergence of the proposed ALG2 based on simulations and the parameters of Table III.

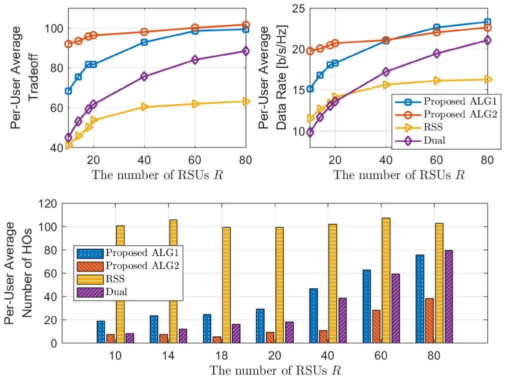


Fig. 5: Per-user average performance versus the number of RSUs R.

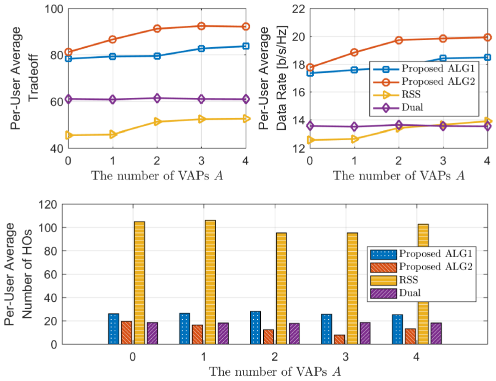


Fig. 6: Per-user average performance versus the number of VAPs A.

1) The received signal strength benchmarker (legend as 'RSS'): The VUE will hand over to an alternative RSU/VAP with the highest RSS among all the alternatives, when it detects that the RSS of the alternative RSU/VAP is also higher than that of the current connection at the beginning of each TS. This benchmarker has been widely used for comparison in terms of HO problems [55] [56].
2) The dual connection benchmarker (legend as 'Dual') [38]: The VUE can be served by its closest and second closest RSUs simultaneously. At the beginning of each TS, the VUE will connect to the common RSU between the current serving RSU set and the alternative RSU set, provided that it exists, and it will disconnect from the other RSU. Otherwise, the VUE will associate with the two closest RSUs.

A. *Convergence of the proposed ALG2*

Let us now first show the convergence of our proposed holistic solution during the training stage in Fig. 4. Firstly, it can be seen that the average cumulative reward of our proposed ALG2 converges at about 400 episodes, which demonstrates the efficiency of our proposed solution. Secondly, we observe that the convergence of the average cumulative reward follows a similar trend to that of the corresponding per-user average data rate (PAR). This is due to the fact that the specific selection of HUC clustering substantially affects the fluctuations of the PAR, which implicitly reflects the significance of the data rate versus HO-rate trade-off. Moreover, it is worth mentioning that the the proposed ALG2 requires non-negligible training overhead, relying on substantial computational capability and storage capacity for keeping track of both vehicles and RSUs, as well as imposing a high information signalling overhead.

B. *Impact of the number of RSUs and VAPs*

Fig. 5 illustrates the per-user average performance for different number of RSUs. First of all, we can clearly see from the left subfigure of Fig. 5, that the per-user average trade-off utility function (PAT) can be increased by adding more RSUs for all solutions. However, when the number of RSUs becomes sufficiently high, the PAT saturates. This phenomenon is due to the fact that the PAR of all solutions increases with the number of RSUs at different rates as a result of having an increased number of nearby RSUs, which can be observed from the right subfigure of Fig. 5. Meanwhile, VUEs may encounter more frequent HOs in an effort to increase their PAR performance, as seen from the lower subfigure of Fig. 5.

The proposed ALG2 outperforms the benchmarkers, especially at a lower number of RSUs ($R = [10, 40]$ at the left of the upper subfigure). Moreover, the proposed ALG1 is capable of approaching the best performance of the proposed ALG2 for $R \geq 60$. This phenomenon indicates that ALG2 has explicit benefits in our HUC clustering design at a lower number of RSUs, whilst ALG1 attains an improved performance by increasing the RSU density. Additionally, by comparing the per-user average number of HOs in the lower subfigure of Fig. 5, the proposed ALG2 reduces the frequency of HOs at least by 50% compared to all the other solutions, when $R \geq 14$. This verifies its superiority in terms of striking a compelling connectivity versus HO-rate trade-off.

Fig. 6 shows the per-user average performance with regard to various number of VAPs. Firstly, observe at the left of the upper subfigure, that the PAT can be increased by adding more VAPs for all solutions, except for the VAP-agnostic 'Dual' solution. The underlying reason is that their PAR (at

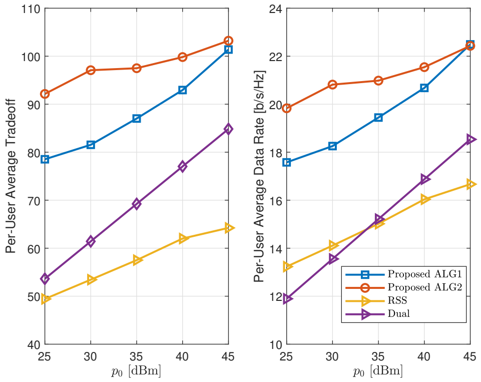


Fig. 7: PAT and PAR versus the transmit power of RSUs p_0.

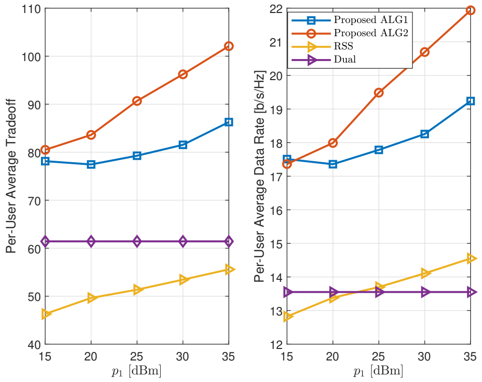


Fig. 8: PAT and PAR versus the transmit power of VAPs p_1.

the right of the upper subfigure) increases with the number of VAPs, whilst there is a modest fluctuation in terms of the per-user average number of HOs (the lower subfigure). After accumulating the PAR over all TSs, the PAT is also increased, obeying a similar trend. This trend exhibits explicit benefits for increasing the number of VAPs in our HUC clustering design. Additionally, it is worth noting that the well-trained ALG2 proposed exhibits superior performance over the other solutions in terms of all the above average performance metrics. This phenomenon reflects that as a benefit of HUC clustering, the DRL-aided solution is capable of striking a compelling connectivity versus HO-rate trade-off, provided that a certain training overhead is allowed. Furthermore, by comparing the PAT and PAR performance, we can conclude that our HUC clustering design achieves at least 30% higher PAT and 25% higher PAR than the benchmarkers.

In terms of the per-user average number of HOs seen in the lower subfigure of Fig. 6, we further observe that the proposed ALG2 achieves the best performance, while the proposed ALG1 only achieves a modest performance, as $A \geq 1$. In this case, the VAPs are allowed for data transmission and may maintain longer association durations than VUE-RSU, thus providing more opportunities for reducing the frequency of HOs. When comparing the results of $A = 0$ and of $A \geq 1$, it can be also seen that the proposed ALG2 is capable of making better use of VAPs for reducing the HO-rate, which verifies again the superiority of the DRL approach.

C. Impact of the transmit power of RSUs and VAPs

The performance is further investigated in Fig. 7 for different transmit powers of the RSUs. First of all, we observe from the figure that all the solutions exhibit an increasing trend upon increasing the transmit power of RSUs p_0 in terms of PAT, since the increased p_0 contributes to the PAR. Next, it can be seen that the proposed ALG1 and ALG2 solutions achieve higher PAT than the benchmarkers as a benefit of our HUC clustering design. Furthermore, observe that as p_0 grows, the PAT gap between ALG1 and ALG2 is gradually reduced, whilst the PAT gap between the 'Dual' solution and the 'RSS' solution widens. This behavior is deemed to be due to the benefits of multiple RSU associations, which is explicitly affected by their transmitted power. Hence, the sequential HUC clustering solution constitutes an attractive design alternative to strike a compelling trade-off without an excessive training overhead at high p_0.

Next, the performance versus the transmit power of the VAPs is investigated in Fig. 8. We first observe from the figure that as the power p_1 of VAPs increases, the superiority of ALG2 in terms of both its PAT and PAR becomes more prominent. The underlying reason for this is that the increased power of VAPs enhances the connectivity benefits brought about by VAPs. For the same reason, the PAT of ALG1 and the 'RSS' solution is also gradually increased when p_1 is increased, which can be clearly seen from Fig. 8. Moreover, it is worth mentioning that the PAT of the 'Dual' solution does not affect the transmit power of VAPs, since it only supports the RSU-association.

D. Impact of the data rate and load constraints

The PAT performance is further investigated in Fig. 9 versus the minimum required data rate constraint for different $\bar{S}_{\max}$. The first point to observe is that the PAT is reduced as the data rate constraint increases. This is because the data rate constraint substantially reduces the data rate contribution in the normalized trade-off utility function. Secondly, it is clearly observed that ALG2 outperforms other solutions, regardless of the data rate constraint, which reflects the superiority of the DRL approach.

Moreover, we observe that all the solutions using $\bar{S}_{\max} = 4$ outperform their counterpart associated with $\bar{S}_{\max} = 2$ apart from the 'Dual' solution, when the minimum data rate constraint is $15 - 30$ b/s/Hz. This is because it is possible to increase the PAT by relying on RSU cooperation using the increased number of associated RSUs, subject to an appropriate minimum data rate constraint. However, as the minimum data rate constraint is 10 b/s/Hz, the PAT of ALG2 associated with $\bar{S}_{\max} = 2$ is higher than that with $\bar{S}_{\max} = 4$. This phenomenon indicates that at a low minimum data rate, the DRL approach may learn a policy with high HO overhead as $\bar{S}_{\max}$ increases, since the contribution of the increased HO overhead to the normalized reward exceeds that of the increased data rate. By contrast, for different $\bar{S}_{\max}$, we observe that only the 'Dual' solution using $\bar{S}_{\max} = 2$ achieves better average cumulative reward than that with $\bar{S}_{\max} = 4$, regardless of the minimum

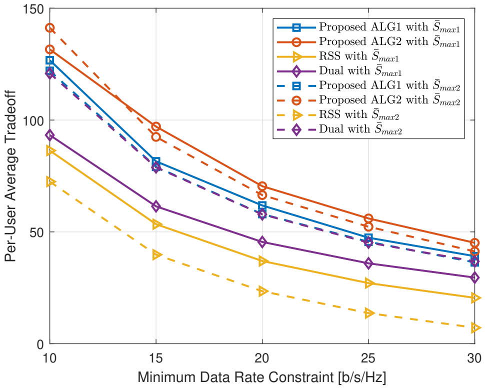


Fig. 9: PAT versus the minimum data rate constraint for $\bar{S}_{\max 1} = 4$ and $\bar{S}_{\max 2} = 2$.

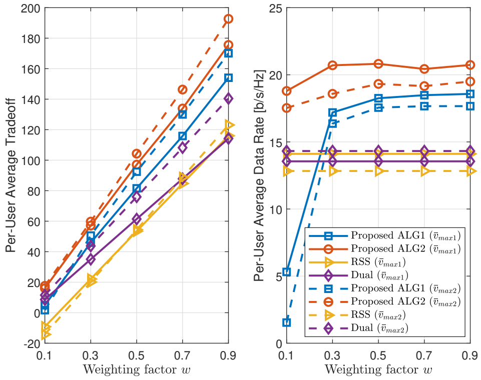


Fig. 10: PAT and PAR versus the weighting factor w for $\bar{v}_{\max 1} = 12$ m/s and $\bar{v}_{\max 2} = 8$ m/s.

data rate constraint. The reason behind this trend is that the 'Dual' solution allows VUEs to select a common RSU as the alternative rather than to select a closest RSU. Hence, as $\bar{S}_{\max}$ decreases, there are more opportunities to select a closer RSU for VUE by observing at most $\bar{S}_{\max}$ closest RSUs.

E. Impact of the weighting factor

In this subsection, the performance versus the weighting factor is investigated. First of all, it can be seen from the left of Fig. 10 that increasing the weighting factor has a beneficial effect on the PAT. This implies that the contribution of the PAR in the normalized utility function is higher than that of the HO overheads. When aiming for maximizing the PAT, the PAR of the proposed ALG1 and ALG2 solutions is also increased when w increases, as shown in the right of Fig. 10, which is in line with our expectation. In the second place, the proposed ALG2 outperforms all other solutions in terms of both PAT and PAR, regardless of the weighting factor. By contrast, the proposed ALG1 exhibits superior performance to both benchmarkers, when a high weighting factor is applied to the connectivity benefits. Furthermore, it is worth noting that reducing the maximum velocity of VUEs and VAPs results in increasing the number of TSs required for travelling a certain distance along the road. Hence the PAT performance of $\bar{v}_{\max 2}$ is higher than that of $\bar{v}_{\max 1}$.

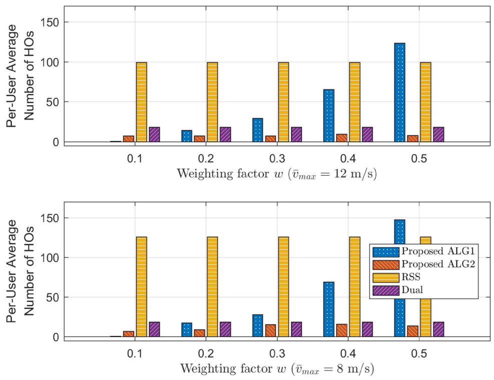


Fig. 11: Per-user average number of HOs versus the weighting factor w for $\bar{v}_{\max 1} = 12$ m/s and $\bar{v}_{\max 2} = 8$ m/s.

As a further step, the per-user average number of HOs versus the weighting factor is studied. Observe from Fig. 11 that as w increases, the number of HOs of the proposed ALG1 solution is gradually increased for both $\bar{v}_{\max 1}$ and $\bar{v}_{\max 2}$, due to using an increased number of RSUs for cooperation. By contrast, the number of HOs of the proposed ALG2 exhibits slight fluctuations upon increasing w relying on DRL. Since the association process in two benchmarkers is independent of the weighting factor, their number of HOs remains constant. Therefore, the results of Fig. 10 and Fig. 11 can offer us some insights into how we should select the weighting factor.

VIII. Conclusions

Incorporating UC clustering into vehicular networks is a promising technique of enhancing connectivity versus the HO-rate trade-off. A novel HUC clustering framework was conceived relying on both RSU cooperation and V2V communication, which is capable of supporting hybrid HOs. In the face of unknown vehicular mobility, we proposed a pair of efficient solutions for solving our HUC cluster migration problem, with the aid of max-bipartite matching and the powerful DRL approach, respectively. Our numerical results have shown that the two proposed HUC cluster migration designs achieve a more beneficial trade-off than the benchmarkers considered, and demonstrated the superiority of the DRL-aided solution, albeit at the cost of a certain training overhead. The results contribute to a better understanding of the benefits of UC clustering in mobile scenarios, which will improve the connectivity-HO tradeoff by at least 30%. Our future research will include 1) the consideration of real-world traffic demands and dynamics; 2) the deployment of a multi-agent DRL framework; 3) the dual function of communicating and computing for both RSUs and APs; 4) the most ambitious, but promising Pareto-Optimization of multi-component OFs.

Appendix

A. Proof of Theorem 1

Let $\Xi = [\boldsymbol{\xi}_1^T, ..., \boldsymbol{\xi}_U^T]^T \in \mathbb{C}^{U\times(A+U)}$ and $\Xi^* \in \mathbb{C}^{U\times(A+U)}$ be an arbitrarily feasible solution and the optimal solution of problem $(\mathbf{P_t^2})$, respectively, wherein the i-the column vector

$\boldsymbol{\xi}_i^T \in \mathbb{C}^{(A+U)\times 1}$ is the corresponding association variable for VUE i. For the sake of simplicity, we define F_i as the trade-off utility function for VUE i and $F(\Xi) = \sum_i F_i(\boldsymbol{\xi}_i)$. Thus, we firstly have

$$F(\Xi^*) = \max_{\Xi} F(\Xi) \geq \sum_{i\in\mathcal{U}} F_i(\boldsymbol{\xi}_i). \tag{37}$$

Considering the independence of the VUE-VAP association and VUE-RSU association, we classify VUEs into a pair of categories, namely VUEs connecting to VAP ($\mathcal{U}^A$) and VUEs connecting to RSU ($\mathcal{U}^R$). Then, we arrive at

$$\sum_{i\in\mathcal{U}} F_i(\boldsymbol{\xi}_i) = \sum_{i\in\mathcal{U}^A} F_i(\boldsymbol{\xi}_i) + \sum_{i\in\mathcal{U}^R} F_i(\boldsymbol{\xi}_i). \tag{38}$$

Assuming that $\boldsymbol{\eta} = [\boldsymbol{\eta}_1^T, ..., \boldsymbol{\eta}_U^T]^T$ is an arbitrarily feasible solution we have for a RSU set, namely solving problem ($\mathbf{P_t^1}$), and $\boldsymbol{\eta}_i^T$ denotes the corresponding association variable for VUE i, thus

$$F_i(\boldsymbol{\xi}_i) = \max_{\{\boldsymbol{\eta}_i\}} F_i(\boldsymbol{\eta}_i), \forall i \in \mathcal{U}^R. \tag{39}$$

Hence, upon substituting (39) into (38), we have

$$\sum_{i\in\mathcal{U}} F_i(\boldsymbol{\xi}_i) \geq \sum_{i\in\mathcal{U}^A} F_i(\boldsymbol{\xi}_i) + \sum_{i\in\mathcal{U}^R} F_i(\boldsymbol{\eta}_i). \tag{40}$$

According to the formulation of ($\mathbf{P_t^1}$), ($\mathbf{P_t^2}$) and ($\mathbf{P_t}$), it is clearly found that both the solution of ($\mathbf{P_t^1}$) and ($\mathbf{P_t^2}$) is one of the feasible solutions of ($\mathbf{P_t}$). In other words, Ξ and $\boldsymbol{\eta}$ are also the solutions of problem ($\mathbf{P_t}$). Additionally, $\mathcal{U}^A$ and $\mathcal{U}^R$ can be any kind of classification. Again, owing to the independence of the VUE-VAP association and VUE-RSU association, we have

$$\sum_{i\in\mathcal{U}^A} F_i(\boldsymbol{\xi}_i) + \sum_{i\in\mathcal{U}^R} F_i(\boldsymbol{\eta}_i) = F(\Xi'), \tag{41}$$

where Ξ' denotes an arbitrarily feasible solution of problem ($\mathbf{P_t}$).

Finally, by combining (37), (40) and (41), we conclude that Ξ^* is also the optimal solution of problem ($\mathbf{P_t}$).

References

[1] L. Liang, H. Peng, G. Y. Li, and X. Shen, "Vehicular communications: A physical layer perspective," *IEEE Trans. on Veh. Technol.*, vol. 66, pp. 10647–10659, Dec. 2017.

[2] S. Zhang, J. Chen, F. Lyu, N. Cheng, W. Shi, and X. Shen, "Vehicular communication networks in the automated driving era," *IEEE Commun. Mag.*, vol. 56, pp. 26–32, Sep. 2018.

[3] H. Peng, Le Liang, X. Shen, and G. Y. Li, "Vehicular communications: A network layer perspective," *IEEE Trans. on Veh. Technol.*, vol. 68, pp. 1064–1078, Feb 2019.

[4] F. Jameel, S. Wyne, S. J. Nawaz, and Z. Chang, "Propagation channels for mmwave vehicular communications: State-of-the-art and future research directions," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 144–150, 2019.

[5] J. Chen, G. Mao, C. Li, W. Liang, and D. Zhang, "Capacity of cooperative vehicular networks with infrastructure support: Multiuser case," *IEEE Trans. on Veh. Technol.*, vol. 67, no. 2, pp. 1546–1560, 2018.

[6] J. Yang, M. Ding, G. Mao, Z. Lin, D. Zhang, and T. H. Luan, "Optimal base station antenna downtilt in downlink cellular networks," *IEEE Trans. on Wireless Commun.*, vol. 18, no. 3, pp. 1779–1791, 2019.

[7] W. Viriyasitavat, M. Boban, H.-M. Tsai, and A. Vasilakos, "Vehicular Communications: Survey and Challenges of Channel and Propagation Models," *IEEE Veh. Technol Mag.*, vol. 10, pp. 55–66, June 2015.

[8] H. Yang, K. Zhang, K. Zheng, and Y. Qian, "Joint frame design and resource allocation for ultra-reliable and low-latency vehicular networks," *IEEE Trans.s on Wireless Commun.*, vol. 19, no. 5, pp. 3607–3622, 2020.

[9] X. Ge, "Ultra-reliable low-latency communications in autonomous vehicular networks," *IEEE Trans. on Veh. Technol.*, vol. 68, no. 5, pp. 5005–5016, 2019.

[10] K. Zheng, Q. Zheng, P. Chatzimisios, W. Xiang, and Y. Zhou, "Heterogeneous vehicular networking: A survey on architecture, challenges, and solutions," *IEEE Commun. Surveys & Tutorials*, vol. 17, pp. 2377–2396, Fourthquarter 2015.

[11] L. Liang, H. Ye, and G. Y. Li, "Spectrum sharing in vehicular networks based on multi-agent reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 37, pp. 2282–2292, Oct. 2019.

[12] J. Wang, C. Jiang, K. Zhang, T. Q. S. Quek, Y. Ren, and L. Hanzo, "Vehicular sensing networks in a smart city: Principles, technologies and applications," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 122–132, 2018.

[13] P. Duan, G. Mao, W. Liang, and D. Zhang, "A unified spatio-temporal model for short-term traffic flow prediction," *IEEE Trans. on Intell. Transp. Syst.*, vol. 20, no. 9, pp. 3212–3223, 2019.

[14] X. Ge, H. Cheng, G. Mao, Y. Yang, and S. Tu, "Vehicular communications for 5G cooperative small-cell networks," *IEEE Trans. on Veh. Technol.*, vol. 65, no. 10, pp. 7882–7894, 2016.

[15] S. Chen, F. Qin, B. Hu, X. Li, and Z. Chen, "User-centric ultra-dense networks for 5G: Challenges, methodologies, and directions," *IEEE Wireless Commun.*, vol. 23, pp. 78–85, Apr. 2016.

[16] Y. Lin, R. Zhang, L. Yang, C. Li, and L. Hanzo, "User-centric clustering for designing ultradense networks: Architecture, objective functions, and design guidelines," *IEEE Veh. Technol. Mag.*, vol. 14, pp. 107–114, Sep. 2019.

[17] Y. Lin, R. Zhang, C. Li, L. Yang, and L. Hanzo, "Graph-based joint user-centric overlapped clustering and resource allocation in ultra dense networks," *IEEE Trans. on Veh. Technol.*, vol. 67, pp. 4440–4453, May 2018.

[18] R. Zhang, J. Wang, Z. Wang, Z. Xu, C. Zhao, and L. Hanzo, "Visible light communications in heterogeneous networks: Paving the way for user-centric design," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 8–16, 2015.

[19] T. Sahin, M. Klugel, C. Zhou, and W. Kellerer, "Virtual cells for 5G V2X communications," *IEEE Commun. Standards Mag.*, vol. 2, pp. 22–28, Mar. 2018.

[20] L. Wang, T. Han, Q. Li, J. Yan, X. Liu, and D. Deng, "Cell-less communications in 5G vehicular networks based on vehicle-installed access points," *IEEE Commun. Mag.*, vol. 24, pp. 64–71, Dec 2017.

[21] L. Ding, Y. Wang, P. Wu, L. Li, and J. Zhang, "Kinematic information aided user-centric 5G vehicular networks in support of cooperative perception for automated driving," *IEEE Access*, vol. 7, pp. 40195–40209, 2019.

[22] T. Bilen, B. Canberk, and K. R. Chowdhury, "Handover management in software-defined ultra-dense 5G networks," *IEEE Network*, vol. 31, no. 4, pp. 49–55, 2017.

[23] H. Song, X. Fang, and L. Yan, "Handover scheme for 5G C/U plane split heterogeneous network in high-speed railway," *IEEE Trans. on Veh. Technol.*, vol. 63, pp. 4633–4646, Nov. 2014.

[24] H. Zhang, W. Huang, and Y. Liu, "Handover probability analysis of anchor-based multi-connectivity in 5G user-centric network," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 396–399, 2019.

[25] X. Xu, Z. Sun, X. Dai, T. Svensson, and X. Tao, "Modeling and analyzing the cross-tier handover in heterogeneous networks," *IEEE Trans.s on Wireless Commun.*, vol. 16, no. 12, pp. 7859–7869, 2017.

[26] Y. Li, B. Cao, and C. Wang, "Handover schemes in heterogeneous LTE networks: challenges and opportunities," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 112–117, 2016.

[27] L. Yan, H. Ding, L. Zhang, J. Liu, X. Fang, Y. Fang, M. Xiao, and X. Huang, "Machine learning-based handovers for Sub-6 GHz and mmWave integrated vehicular networks," *IEEE Trans. on Wireless Commun.*, vol. 18, pp. 4873–4885, Oct 2019.

[28] J. M. Marquez-Barja, H. Ahmadi, S. M. Tornell, C. T. Calafate, J. Cano, P. Manzoni, and L. A. DaSilva, "Breaking the vehicular wireless communications barriers: Vertical handover techniques for heterogeneous networks," *IEEE Trans. on Veh. Technol.*, vol. 64, pp. 5878–5890, Dec 2015.

[29] A. Ahmed, L. M. Boulahia, and D. Gaiti, "Enabling vertical handover decisions in heterogeneous wireless networks: A state-of-the-art and a classification," *IEEE Commun. Surveys & Tutorials*, vol. 16, no. 2, pp. 776–811, 2014.

[30] K. Abboud, H. A. Omar, and W. Zhuang, "Interworking of DSRC and cellular network technologies for V2X communications: A survey," *IEEE Trans. on Veh. Technol.*, vol. 65, pp. 9457–9470, Dec 2016.

[31] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, p. 529–533, 2015.

[32] M. H. Dwijaksara, W. S. Jeon, and D. G. Jeong, "A graph-based handover scheduling for heterogeneous vehicular networks," *IEEE Access*, vol. 6, pp. 53722–53735, 2018.

[33] B. Yang, X. Yang, X. Ge, and Q. Li, "Coverage and handover analysis of ultra-dense millimeter-wave networks with control and user plane separation architecture," *IEEE Access*, vol. 6, pp. 54739–54750, 2018.

[34] R. Arshad, H. Elsawy, L. Lampe, and M. J. Hossain, "Handover rate characterization in 3D ultra-dense heterogeneous networks," *IEEE Trans. on Veh. Technol.*, vol. 68, pp. 10340–10345, Oct 2019.

[35] X. Xu, X. Tang, Z. Sun, X. Tao, and P. Zhang, "Delay-oriented cross-tier handover optimization in ultra-dense heterogeneous networks," *IEEE Access*, vol. 7, pp. 21769–21776, 2019.

[36] M. Alhabo, L. Zhang, and N. Nawaz, "GRA-based handover for dense small cells heterogeneous networks," *IET Commun.*, vol. 13, pp. 1928–1935, 2019.

[37] M. M. Hasan, S. Kwon, and S. Oh, "Frequent-handover mitigation in ultra-dense heterogeneous networks," *IEEE Trans. on Veh. Technol.*, vol. 68, pp. 1035–1040, Jan 2019.

[38] E. Demarchou, C. Psomas, and I. Krikidis, "Mobility management in ultra-dense networks: Handover skipping techniques," *IEEE Access*, vol. 6, pp. 11921–11930, 2018.

[39] Z. Wang, L. Li, Y. Xu, H. Tian, and S. Cui, "Handover control in wireless systems via asynchronous multiuser deep reinforcement learning," *IEEE Internet of Things J.*, vol. 5, no. 6, pp. 4296–4307, 2018.

[40] J. Ye and Y. J. Zhang, "Drag: Deep reinforcement learning based base station activation in heterogeneous networks," *IEEE Trans. on Mobile Comput.*, early access, 2020 (DOI: 10.1109/TMC.2019.2922602).

[41] N. Zhao, Y. Liang, D. Niyato, Y. Pei, M. Wu, and Y. Jiang, "Deep reinforcement learning for user association and resource allocation in heterogeneous cellular networks," *IEEE Trans. on Wireless Commun.*, vol. 18, pp. 5141–5152, Nov 2019.

[42] H. Khan, A. Elgabli, S. Samarakoon, M. Bennis, and C. S. Hong, "Reinforcement learning-based vehicle-cell association algorithm for highly mobile millimeter wave communication," *IEEE Trans. on Cogn. Commun. and Netw.*, vol. 5, pp. 1073–1085, Dec 2019.

[43] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, 2017.

[44] J. Wang, C. Jiang, H. Zhang, Y. Ren, K. Chen, and L. Hanzo, "Thirty years of machine learning: The road to pareto-optimal wireless networks," *IEEE Commun. Surveys & Tutorials*, early access, 2020 (DOI:10.1109/COMST.2020.2965856).

[45] S. Batabyal and P. Bhaumik, "Mobility models, traces and impact of mobility on opportunistic routing algorithms: A survey," *IEEE Commun. Surveys & Tutorials*, vol. 17, no. 3, pp. 1679–1707, 2015.

[46] J. Munkres, "Algorithms for the assignment and transportation problems," *J. Soc. Ind. Appl. Math.*, vol. 5, no. 1, pp. 32–38, 1957.

[47] H. W. Kuhn, "The Hungarian method for the assignment problem," *Nav. Res. Logist. Quart.*, vol. 2, pp. 83–97, Mar. 1955.

[48] A.E.Roth and M.Sotomayor, *Two Sided Matching: A study in Game-Theoretic Modeling and Analysis*. Cambridge, UK: Cambridge University Press, 1991.

[49] S. Bayat, Y. Li, L. Song, and Z. Han, "Matching theory: Applications in wireless communication," *IEEE Signal Process. Mag.*, vol. 33, pp. 103–122, Nov 2016.

[50] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv:1509.02971*, 2015.

[51] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[52] D. Sliver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," *International Conf. on Machine Learning, Beijing*, pp. 387–395, 2014.

[53] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[54] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, arXiv:1412.6980.

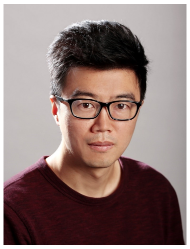
[55] B. Yang, X. Wang, and Z. Qian, "A multi-armed bandit model based vertical handoff algorithm for heterogeneous wireless networks," *IEEE Commun. Letter*, vol. 22, no. 10, pp. 2116–2119, 2018.

[56] X. Zhang, Y. Xie, Y. Cui, Q. Cui, and X. Tao, "Multi-slot coverage probability and SINR-based handover rate analysis for mobile user in hetnet," *IEEE Access*, vol. 6, pp. 17868–17879, 2018.

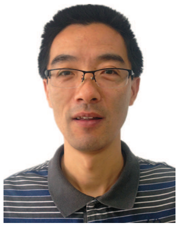
Yan Lin (M'16) received the M.S. and Ph.D. degree from Southeast University, China, in 2013 and 2018, respectively. She is currently a lecturer in the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, China, from 2018. She visited Southampton Wireless Group in Southampton University, U.K. from Oct. 2016 to Oct. 2017. Her current research interests include vehicular networks, mobile edge computing and reinforcement learning for resource allocation in wireless communication.

Zhengming Zhang received the B.S. in electronic information science and technology from Nanjing Agricultural University, Nanjing, China, in 2016. Since September 2016, he is pursuing his PhD degree in information and communication engineering with the School of Information Science and Engineering. His current research interests include wireless big data, machine learning, 5G mobile networks, UAV aided communication and resource management.

Yongming Huang (M'10-SM'16) received the B.S. and M.S. degrees from Nanjing University, China, in 2000 and 2003, respectively. In 2007 he received the Ph.D. degree in electrical engineering from Southeast University, China. Since March 2007 he has been a faculty in the School of Information Science and Engineering, Southeast University, China, where he is currently a full professor. During 2008- 2009, Dr. Huang was visiting the Signal Processing Lab, Electrical Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden. His current research interests include MIMO wireless communications, cooperative wireless communications and millimeter wave wireless communications.



Jun Li (M'09-SM'16) received Ph. D degree in Electronic Engineering from Shanghai Jiao Tong University, Shanghai, P. R. China in 2009. From January 2009 to June 2009, he worked in the Department of Research and Innovation, Alcatel Lucent Shanghai Bell as a Research Scientist. From June 2009 to April 2012, he was a Postdoctoral Fellow at the School of Electrical Engineering and Telecommunications, the University of New South Wales, Australia. From April 2012 to June 2015, he was a Research Fellow at the School of Electrical Engineering, the University of Sydney, Australia. From June 2015 to now, he is a Professor at the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China. He was a visiting professor at Princeton University from 2018 to 2019. His research interests include network information theory, game theory, distributed intelligence, multiple agent reinforcement learning, and their applications in ultra-dense wireless networks, mobile edge computing, network privacy and security, and industrial Internet of things.



Feng Shu received the Ph.D. degree from Southeast University, Nanjing, China, in 2002. Since 2020, he is a full professor and a Supervisor of the Ph.D. students with the School of Information and Communication Engineering, Hainan University, Haikou, China. From 2005 to 2020, he was School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, where he was promoted to a full Professor and also a Supervisor of the Ph.D. students in 2013. From 2009 to 2010, he held a visiting postdoctoral position with The University of Texas at Dallas. He is also with Fujian Agriculture and Forestry University and awarded with Mingjian Scholar Chair Professor in Fujian Province. His research interests include wireless networks, wireless location, and array signal processing.

Lajos Hanzo (F'08) (http://www-mobile.ecs.soton.ac.uk) FREng, FIET, Fellow of EURASIP, D.Sc., received his degree in electronics in 1976 and his doctorate in 1983. He holds honorary doctorates from the Technical University of Budapest (2009) and the University of Edinburgh (2015). He is a foreign member of the Hungarian Academy of Sciences and a former Editor-in-Chief of IEEE Press. He has served several terms as a Governor of both IEEE ComSoc and VTS. He has coauthored 19 John Wiley, IEEE Press books and 1900+ contributions at IEEE Xplore.