Dynamic Resource Orchestration for Service Capability Maximization in Fog-Enabled Connected Vehicle Networks

Duc-Nghia Vu, Nhu-Ngoc Dao, Woongsoo Na, and Sungrae Cho

***Abstract*—Technological advances in fog computing are precipitating an evolution in conventional vehicle networks to a new paradigm called fog-enabled connected vehicle networks (FCVNs). FCVNs provide communication efficiency for ensuring safe transportation through the massive Internet of vehicles. In FCVNs, massive vehicles tend to associate with roadside units and high power nodes, which act as fog nodes (FNs), when they have a good channel quality and/or popular contents. This circumstance may lead to a load imbalance among the FNs. This problem significantly decreases the resource utilization efficiency and service capability of the networks. In this paper, we propose a dynamic resource orchestration (DRO) scheme to harmonize resource allocation for connected vehicles by migrating the offloaded services among FNs. A graph-theoretic approach is utilized to transform the FCVN into a directed graph model, where the maximum resource reduction obtained by service migrations is considered the weight of the link between every two FNs. Subsequently, the maximum weight matching solution is used to determine optimal pairs of FNs for migrating services to maximize network resource utilization. Our simulation results reveal that the proposed DRO scheme achieves significant improvements in terms of service capability, throughput, and resource utilization efficiency as compared with existing algorithms.**

***Index Terms*—fog-enabled connected vehicle networks, resource orchestration, service capability, matching algorithm.**

I. Introduction

The rapid growth of communication technologies in the Internet of things paradigm has marked a milestone in the development of vehicle networks for smart transportation, wherein millions of vehicles may be connected and communicate over the Internet [1]. Advanced vehicle networks may support various applications that require complex data processing, high precision, and real-time responses (e.g., self-driving cars, navigation, and augmented reality assistants) for ensuring driving safety, traffic efficiency, and great convenience in transportation [2]. These applications are characterized by relatively high data communication, computation, and storage capacity requirements, which present great challenges to existing vehicle networks. In this context, fog-enabled connected vehicle networks (FCVNs) that integrate fog computing into conventional vehicle networks have emerged as a promising candidate to satisfy these requirements [3], [4]. Fig. 1 illustrates the architecture of FCVNs. In FCVNs, high power nodes (HPNs) are deployed to provide wide-area coverage and execute the control operations. The HPNs are interconnected via crosshaul links and coordinated by a central orchestrator, which interfaces between the HPNs and the cloud servers in the Internet. In contrast, roadside units (RSUs) are equipped with local caches, in which the interesting contents can be stored, as well as with computing processors for handling offloaded services, and are deployed mainly in close proximity to connected vehicles [5]. This allows the connected vehicles to access and request services promptly at a high-speed rate and very low transmission and computing latency. These RSUs are controlled by the HPNs. In the fog computing perspective, HPNs and RSUs are in general considered fog nodes (FNs).

The exploding number of smart vehicles that generate very large amounts of data each day has resulted in a significant increase in bandwidth consumption and competition, in the sense that a connected vehicle must compete against other devices for finite bandwidth [6]–[9]. To evaluate the resource utilization efficiency in FCVNs, the service capability, which represents the availability of the network for serving connected vehicles such that their diverse requirements are met, is considered a key criterion [10]. In other words, the service capability of a network is determined by the percentage of remaining resources relative to the total network resources.

Owing to radio resource constraints, efficient resource management for improving service capability is considered an emerging challenge in FCVNs [11]. For instance, some FNs tend to be heavily utilized by a very large number of connected vehicles when the FNs have a good signal-to-interference-plus-noise ratio (SINR), popular caching contents, or/and a high processing ability [12], [13]. This drives device association to apply either signal-aware (SA), content-aware (CoA), or capacity-aware (CaA) approaches. With the objective of providing a higher data rate and service quality, the device association in the SA approach prefers FNs that have a high signal quality [14]. Although this approach can achieve better network throughput and spectral efficiency, it also causes an imbalance among FNs, which can seriously degrade the

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2019R1A2C1090447) and the Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-00889, A study on core technology of 5G mobile communication using millimeter wave band).

D.-N. Vu is with Innowireless Inc., Seongnam, Republic of Korea.

N.-N. Dao is with the Department of Computer Science and Engineering, Sejong University, Seoul, Republic of Korea.

W. Na is with the Department of Computer Science and Engineering, Kongju National University, Cheonan, Republic of Korea.

S. Cho is with the School of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea.

(D.-N. Vu and N.-N. Dao are co-first authors.)

(Corresponding author: S. Cho.)

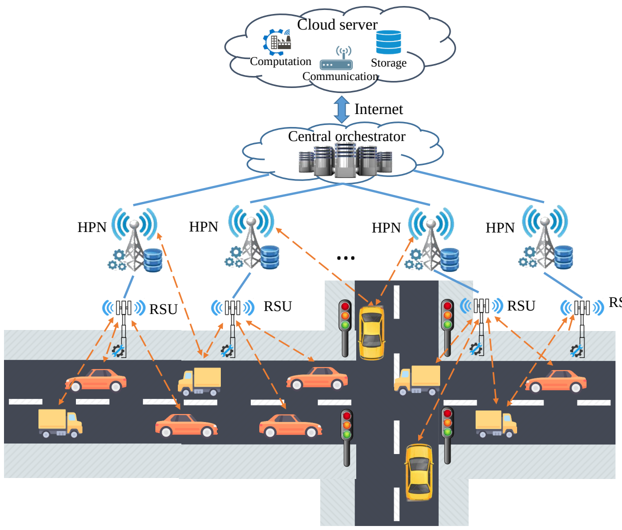

Fig. 1. Fog-enabled connected vehicle network architecture.

network service capability and availability. In contrast, the CoA approach focuses on directing users to associate with FNs that have their favorite contents [15]. This approach obtains a lower content-access latency and better service experiences. However, the resource utilization efficiency is not considered, resulting in a decrease in the service capability for massive numbers of connected vehicles. Finally, the CaA approach drives user association with respect to the resource availability of the FNs [16]. Although this approach addresses the resource imbalance and unfairness among FNs, it cannot ensure their service capability for massive numbers of connected vehicles owing to the low resource utilization efficiency [17].

In the aforementioned approaches, a device is associated with its preferred FNs without systematic orchestration; thus, these FNs may suffer from overload situations, whereas the other FNs are free. This circumstance leads to an imbalance in the resource utilization among FNs. Hence, this negatively affects the uplink services of connected vehicles. Because of the inefficient resource utilization, the uplink service quality of vehicles may not be guaranteed when connected to popular FNs (a.k.a the most preferred FNs).

To address these problems, we propose a dynamic resource orchestration (DRO) scheme to harmonize the resource scheduling among FNs for upstream offloading services to improve their service capability. On the basis of graph theory, the DRO scheme considers FNs as vertices, and the weight of the edge between two vertices is provided by the amount of maximum resource reduction when optimal service migration is conducted among the FNs. The optimal service migration in each pair of FNs for minimizing the resource utilization is obtained by using the steepest descent method. Finally, the *maximum weight matching* solution determines the optimal pairs of FNs for migrating connected vehicle services to achieve service capability maximization. The main contributions of this paper are summarized as follows.

- We present a graph model of FCVNs and formulate their characteristics that affect the network service capability and resource utilization efficiency.
- We propose the DRO scheme, which uses the *steepest gradient method* and *maximum weight matching solution* to harmonize resource allocation for connected vehicles by migrating the offloaded connected vehicle services among FNs.
- We describe rigorous simulations that demonstrate the effectiveness of the proposed DRO scheme. The evaluation results demonstrate that the proposed DRO scheme not only outperforms existing schemes but also achieves an approximation to the exact solution in terms of the service capability, throughput, and serviceability.

This paper is organized as follows. Section II summarizes the related work. Section III presents the problem statement and formulation. The DRO scheme for optimizing the network service capability is presented in Section IV. In Section V, we evaluate and analyze the effectiveness of the proposed DRO scheme. Finally, Section VI concludes the paper.

II. Related Works

As aforementioned in Section I, connected vehicle association is driven mostly by the SA, CoA, and CaA approaches. The application of each approach may achieve a better network performance for downlink services in terms of throughput and serviceability, as well as spectral efficiency. However, these approaches may lead to a load imbalance among FNs that may negatively affect the uplink services of the connected vehicles. This means that the optimal connection for a downlink may not be optimal for an uplink [18]. Recently, several researchers have aimed to resolve the load balancing problem with the objective of achieving the optimal resource allocation for the networks [19]–[21]. The resource scheduling algorithms in FCVNs can be classified based on their objectives including throughput, spectral efficiency, serviceability, and hybrid optimizations.

A. Throughput

According to these approaches, He *et al.* [22] proposed a semi-Markov decision process (SMDP) based resource allocation scheme to facilitate video streaming applications in heterogeneous cognitive vehicular networks. The scheme can boost the bandwidth utilization of the entire network to improve video streaming quality for vehicle users but cannot improve the network resource balancing, as well as serviceability. In contrast, Zheng *et al.* [23] proposed joint load balancing of the downlink and uplink for interference coordination in heterogeneous networks while considering different service classes. They presented the problem as mixed binary integer programming and provided a relaxed-rounding solution to their model. Although their proposed algorithm can obtain the desired load balance and user data throughput, it cannot improve the network service capability. Aiming at maximizing the downlink sum-rate, Vu *et al.* [24] proposed a strategy that uses the Hungarian method for resource scheduling in fog-enabled radio access networks.

B. Spectral efficiency

Although the aforementioned method can provide data rate optimization, it cannot obtain resource efficiency because of the non-exact solution. Meanwhile, a number of research studies have been conducted on spectral efficiency, which is a well-known typical metric for network evaluation. Several researchers have aimed at improving this feature by using optimization techniques [25], game theoretical models [26], etc., as indicated by the thorough survey in [12].

C. Serviceability

On the other hand, an adaptive resource balancing (ARB) scheme was proposed in [27] for serviceability maximization in fog radio access networks. By migrating the services among remote radio heads (RRHs), the algorithm can resolve the imbalance among cells and optimize network serviceability. However, resource utilization efficiency was not considered. Such single-objective approaches can achieve the best optimization in one perspective of the network. However, they do not consider other metrics that may decrease the network performance.

D. Hybrid optimization

In contrast, hybrid approaches are aimed at orchestrating the network resource to obtain a combined optimization for multiple metrics as mentioned above. For instance, to balance the load among cells in terms of user throughput and service capability, Xin *et al.* [28] proposed a joint user association and user scheduling for load balancing over the downlink of a wireless heterogeneous network, which they achieved by addressing a network-wide utility maximization problem. They approximated the nonconvex throughput achieved with user scheduling to a concave function and implemented a joint user association and user scheduling algorithm by exploiting a distributed convex optimization technique. Meanwhile, Cordeschi *et al.* [29] proposed a reliable adaptive resource management scheme for cognitive cloud vehicular networks to allow energy- and computing-limited car smartphones to utilize the available vehicle-to-infrastructure WiFi connections for performing traffic offloading to local or remote clouds. Their approach improved significantly the bandwidth and energy resource utilization in the network. Moreover, a number of joint algorithms were proposed in [30]–[32] aimed at obtaining load balancing, energy efficiency, and interference mitigation in heterogeneous cellular networks. Although these proposed algorithms significantly improve network throughput, serviceability, and energy efficiency, they cannot achieve resource utilization efficiency. In general, multiple-objective approaches can improve the network performance in terms of multiple metrics. However, the solving of these problems is complicated and obtains only an incomplete optimal solution.

Existing approaches have significantly contributed to improving the performance of FCVNs; however, most face drawbacks, such as high computational complexity, standalone metric optimization, and non-optimal solutions. In this paper, we propose a DRO scheme that can overcome the issues to obtain network service capability optimization, as well as to achieve improvements in terms of throughput and serviceability for resource allocation in FCVNs.

TABLE I
KEY NOTATION DESCRIPTION

Notation	Description
N	Number of FNs
R	Radius of FN coverage area
λ	Mean arrival rate of connected vehicles
μ	Mean departure rate of connected vehicles
C_i	Capacity of the ith FN (unit: RB)
$O_i(t)$	Current occupied capacity of the ith FN assigned to connected vehicles at time slot t
$A_i(t)$	Remaining capacity of the ith FN at time slot t
$S_i(t)$, $S(t)$	Current service capability of the ith FN and the network at time slot t, respectively
p_i	Probability that a connected vehicle associates with the ith FN
$D_i^{in}(t)$, $D_i^{out}(t)$	Set of connected vehicles arriving and departing to/from the ith FN at time slot t, respectively
$D_{ij}^{cov}(t)$	Set of connected vehicles located in the coverage area of the ith and jth FNs at timeslot t
$D_i(t)$	Set of connected vehicles associated in the ith FN at time slot t
$D_{ij}^{fea}(t)$	Set of connected vehicles that is feasible for service migration between the ith FN and the jth FN at timeslot t
b_{ij}	Number of resource blocks required from the ith FN to satisfy the data rate r_j of the jth connected vehicle
$d_{ij}^{*}(t)$	Optimal set of connected vehicles for service migration from FN i to FN j at time slot t
$W_{ij}(t)$	Weight that is obtained when service migrations of the connected vehicles between FN i and FN j at time slot t are conducted

III. PROBLEM STATEMENT

We consider the FCVN system model in terms of communication and spectral resource management, as depicted in Fig. 1. The FNs are managed by a central orchestrator and geographically distributed according to the local traffic density. Without loss of generality, the arrival rate and departure rate of the connected vehicles to/from the network are assumed to follow a Poisson process with a mean value λ and an exponential process with a mean value μ, respectively. A list of the key notations used in this paper is provided in Table I.

We assume that the jth connected vehicle issues a request for services with the required data rate r_j to the FNs. Therefore, the number of resource blocks (RBs) b_{ij} that the ith FN must assign to the jth connected vehicle [33] is derived as

$$b_{ij} = \left\lceil \frac{r_j}{\Delta f \log_2(1+\mathrm{SINR}_{ij})} \right\rceil, \tag{1}$$

where $b_{ij} \in \mathbb{N}$, Δf is the bandwidth that one RB utilizes during 1 ms (i.e., 180 KHz [34]) and SINR_{ij} is the signal-to-interference-plus-noise ratio (SINR) on the data channel between the ith FN and the jth connected vehicle [35].

The set of connected vehicles arriving at the ith FN at time slot t is defined as $D_i^{in}(t)$. Let $D_i^{out}(t)$ be a set of connected vehicles departing from the ith FN at timeslot t, as their offloaded workloads have been successfully executed and returned by the computing entities. In this circumstance, we denote the mean departure rate of the connected vehicles from the network by μ. It is readily observed that

$$\mathbb{E}\left[\sum_{i=0}^{N}|D_i^{out}(t)|\right] = \mu, \tag{2}$$

where N is the number of FNs in the network. Because of the resource constraints, the FNs can serve only a limited number of connected vehicles at timeslot t, defined as $D_i(t)$. The set $D_i(t)$ can be derived as

$$D_i(t) = D_i(t-1) \cup D_i^{in}(t) \backslash D_i^{out}(t). \tag{3}$$

Accordingly, we can obtain the remaining RBs $A_i(t)$ of the ith FN after it has been occupied by connected vehicles as

$$A_i(t) = C_i - O_i(t) = C_i - \sum_{j=0}^{|D_i(t)|} b_{ij}, \tag{4}$$

where C_i and $O_i(t)$ are the capacity and occupied RB of the ith FN at timeslot t, respectively. Following the definition of service capability in Section I, the service capability of the ith FN at timeslot t is given by

$$S_i(t) = \frac{A_i(t)}{C_i}. \tag{5}$$

Similarly, we obtain the service capability of the network at timeslot t as

$$S(t) = \frac{\sum_{i=0}^{N} A_i(t)}{\sum_{i=0}^{N} C_i}. \tag{6}$$

Under the popular contents and signal intensity effects, interesting FNs may attract a massive amount of incoming connected vehicles. The number of connected vehicles $|D_i(t)|$ arriving at these FNs rapidly increases in the cumulative time. This leads to the FNs reaching overcapacity and becoming unable to provide the service capability because of resource limitations. However, the new incoming connected vehicles that do not have sufficient competitive capability to be served by the best FNs are served by the FNs that have poorer conditions (e.g., a low SINR). This means that these connected vehicles will consume a larger number of the RBs b_{ij} to satisfy their desired requirements (e.g., data rate and latency). This problem leads to a decrease in both the resource utilization efficiency and the service capability of the network.

With the objective of alleviating the burden on the interesting FNs by balancing resources among the FNs, as well as optimizing the resource utilization efficiency, we propose the DRO scheme, which conducts service migration among FNs to achieve service capability maximization. We observe that the services of a connected vehicle can be migrated between two FNs only if the connected vehicle is located in the overlapped coverage and served by those two FNs.

The service migration among FNs for minimizing the RB occupation in resource-constrained FCVNs is presented as a matching problem. In a directed graph model of FNs, the matching weight between the FNs is determined by the maximum reduction of the occupied RBs when the services of the connected vehicles are moved between the FNs. A dynamic algorithm based on the steepest gradient method for finding the matching weight between two FNs is proposed. Finally, we address the maximum weight matching algorithm to achieve the optimal service migrations among the FNs, which results in resource utilization efficiency optimization, as well as in service capability maximization.

IV. Dynamic Resource Orchestration

A. Matching Problem

We consider a graph model $G(V, E)$ of N FNs, where V and E are the set of vertices of FNs and edges among FNs, respectively. Let $D_{ij}^{fea}(t)$ be a set of connected vehicles for which services can feasibly me migrated between the ith and jth FNs at timeslot t. The edge between FN i and FN j is feasible if these FNs have at least one connected vehicle in the overlapped coverage served by these FNs. This means that the set of connected vehicles for which services can feasibly be migrated is $D_{ij}^{fea}(t) \neq \emptyset$. Matching in a graph is defined by *a set of edges without common vertices*. Accordingly, maximum weight matching is a matching that has the maximum weight sum of the edges [36]. The maximum weight matching problem for maximizing the reduction in RB occupation when the services of connected vehicles are migrated among FNs can be formulated as

$$\max \sum_{i=1}^{N}\sum_{j=1}^{N} x_{ij} W_{ij}(t) \tag{7}$$

$$\text{s.t. } x_{ij} \in \{0,1\}, \tag{8}$$

$$x_{ij} = x_{ji} \in \{0,1\}, \forall i, j = 1, 2, \ldots, N, \tag{9}$$

$$\sum_{i=1}^{N} x_{ij} \leq 1, \ \forall j = 1, 2, \ldots, N, \tag{10}$$

$$\sum_{j=1}^{N} x_{ij} \leq 1, \ \forall i = 1, 2, \ldots, N, \tag{11}$$

where the indicator x_{ij} is given by

$$x_{ij} \triangleq \begin{cases} 1 & \text{if the services of connected vehicles are migrated} \\ & \text{between } i\text{th and } j\text{th FNs;} \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

and W_{ij} is the matching weight that can be determined by the maximum occupied RB reduction when service migration of the connected vehicles between the ith and jth FNs at timeslot t is conducted. The optimal service migration between two FNs for finding the matching weight W_{ij} can be formulated as an *integer programming* problem, which is addressed in Section III.B. By solving the problem (7), we can minimize the RB occupation, as well as optimize the resource utilization efficiency, thereby maximizing the service capability.

B. Optimal Service Migration

The optimal service migration of connected vehicles between the ith and jth FNs is that which achieves the maximum reduction in RB occupation while guaranteeing the quality of service of the connected vehicles. We define the set of connected vehicles located in the overlapped coverage area of the ith and jth FNs as $D_{ij}^{cov}(t)$. As previously mentioned, it is observed that the services of connected vehicles can be migrated between two FNs if these connected vehicles are served and located in the overlapped coverage area of the two FNs. The set of connected vehicles $D_{ij}^{fea}(t)$ for which services can feasibly be migrated between the ith and jth FNs can be determined as

$$D_{ij}^{fea} = \left(D_i(t) \cup D_j(t)\right) \cap D_{ij}^{cov}(t), \tag{13}$$

where $D_i(t)$ and $D_j(t)$ are the sets of connected vehicles that are served by the ith and jth FNs at timeslot t, respectively. The RBs occupied by the connected vehicles in set $D_{ij}^{fea}(t)$ before the service $O_{ij}(t)$ is migrated are derived as

$$O_{ij}(t) = \sum_{k=1}^{|D_i(t) \cap D_{ij}^{fea}(t)|} b_{ik} + \sum_{k=1}^{|D_j(t) \cap D_{ij}^{fea}(t)|} b_{jk}. \tag{14}$$

Let $D_i^*(t)$ and $D_j^*(t)$ be the sets of connected vehicles associated in the ith and jth FNs after optimal service migration has been conducted. Thus, the RBs occupied by these connected vehicles in set $D_{ij}^{fea}(t)$ after service migration $O_{ij}^*(t)$ can be obtained by

$$O_{ij}^*(t) = \sum_{k=1}^{|D_i^*(t) \cap D_{ij}^{fea}(t)|} b_{ik} + \sum_{k=1}^{|D_j^*(t) \cap D_{ij}^{fea}(t)|} b_{jk}. \tag{15}$$

Accordingly, the weight matching $W_{ij}(t)$ is the marginally occupied RBs after the services between are moved the ith and jth FNs, given as

$$W_{ij}(t) = O_{ij}(t) - O_{ij}^*(t). \tag{16}$$

Similarly, the optimal set of connected vehicles for service migrations from the ith FN to the jth FN and its converse are defined as $d_{ij}^*(t)$ and $d_{ji}^*(t)$, respectively, derived as

$$\begin{cases} d_{ij}^*(t) = D_i(t) \backslash D_i^*(t), \\ d_{ji}^*(t) = D_j(t) \backslash D_j^*(t). \end{cases} \tag{17}$$

With the aim to of maximizing the resource utilization efficiency, we migrate the services of connected vehicles in the overlapped coverage area of the ith and jth FNs to minimize the number of RBs occupied by the connected vehicles. The optimal service migration between FNs i and j can be formulated as

$$\min \sum_{m \in \{i,j\}} \sum_{n=1}^{|D_{ij}^{fea}(t)|} b_{mn} x_{mn} \tag{18}$$

$$\text{s.t.} \sum_{n=1}^{|D_{ij}^{fea}(t)|} b_{mn} x_{mn} \leq \delta_m, \forall m \in \{i,j\}, \tag{19}$$

$$\sum_{m \in \{i,j\}} x_{mn} = 1, \forall n = 1, 2, \ldots, |D_{ij}^{fea}(t)|, \tag{20}$$

$$x_{mn} \in \{0,1\}, \forall m \in \{i,j\}, \forall n = 1, 2, \ldots, |D_{ij}^{fea}(t)|, \tag{21}$$

where the indicator x_{mn} is given by

$$x_{mn} \triangleq \begin{cases} 1 & \text{if the services of the } n\text{-}th \text{ connected vehicle are} \\ & \text{migrated to the } m\text{-}th \text{ FN;} \\ 0 & \text{otherwise,} \end{cases} \tag{22}$$

and

$$\begin{cases} \delta_i = A_i(t) + \sum_{k=1}^{|D_i(t) \cap D_{ij}^{fea}(t)|} b_{ik}, \\ \delta_j = A_j(t) + \sum_{k=1}^{|D_j(t) \cap D_{ij}^{fea}(t)|} b_{jk}. \end{cases} \tag{23}$$

The constraint (19) ensures that the RBs occupied by the connected vehicles when services are moved do not exceed the capacity of the FNs. By addressing the problem in (18), we obtain the matching weight $W_{ij}(t)$, optimal set $d_{ij}^*(t)$, and $d_{ji}^*(t)$ of the connected vehicles for the optimal service migration between the ith and jth FNs. We define

$$\begin{aligned} X_m &= \left\{ (x_{mn})_{n=1}^{|D_{ij}^{fea}(t)|} \middle| (19) \text{ and } (21) \text{ are satisfied} \right\} \\ &= \left\{ (x_{mn}^k)_{n=1}^{|D_{ij}^{fea}(t)|} \middle| k = 1, 2 \ldots, K_m \right\}, \end{aligned} \tag{24}$$

where K_m is the maximum number of elements in X_m and $(x_{mn}^k)_{n=1}^{|D_{ij}^{fea}(t)}$ is the k-th element of the set X_m. The certain values of x_{mn}, $n = 1, 2, \ldots, |D_{ij}^{fea}(t)|$ fulfill the constraints in (19) and (21) only if

$$x_{mn} = \sum_{k=1}^{K_m} y_m^k x_{mn}^k, n = 1, 2, \ldots, |D_{ij}^{fea}(t)|, \tag{25}$$

where $\sum_{k=1}^{K_m} y_m^k = 1$ and $y_m^k \in \{0,1\}$. Then, the problem in (18) is equivalent to

$$\begin{aligned} \min & \sum_{m \in \{i,j\}} \sum_{n=1}^{|D_{ij}^{fea}(t)|} \sum_{k=1}^{K_m} b_{mn} y_m^k x_{mn}^k \\ &= \sum_{m \in \{i,j\}} \sum_{k=1}^{K_m} \left(\sum_{n=1}^{|D_{ij}^{fea}(t)|} b_{mn} x_{mn}^k \right) y_m^k \end{aligned} \tag{26}$$

$$\text{s.t.} \sum_{m \in \{i,j\}} \sum_{k=1}^{K_m} y_m^k x_{mn}^k = 1, \forall n = 1, 2, \ldots, |D_{ij}^{fea}(t)|, \tag{27}$$

$$\sum_{k=1}^{K_m} y_m^k = 1, \forall m \in \{i,j\}, \tag{28}$$

$$y_m^k \in \{0,1\}, \forall m \in \{i,j\}, k = 1, 2, \ldots, K_m. \tag{29}$$

Let u_n and w_m be the dual variables associated with the constraints in (27) and (28), respectively. The linear programming dual problem of the continuous relaxation of the equivalent problem in (26) can be obtained by

$$\max \sum_{m \in \{i,j\}} w_m + \sum_{n=1}^{|D_{ij}^{fea}(t)|} u_n \tag{30}$$

$$\begin{aligned} \text{s.t.} & \sum_{n=1}^{|D_{ij}^{fea}(t)|} x_{mn}^k u_n + w_m \leq \sum_{n=1}^{|D_{ij}^{fea}(t)|} b_{mn} x_{mn}^k, \\ & \forall m \in \{i,j\}, k = 1, 2, \ldots, K_m. \end{aligned} \tag{31}$$

Moreover, it is observed that the constraint in (31) is equivalent to $w_m \leq \sum_{n=1}^{|D_{ij}^{fea}(t)|}(b_{mn}-u_n)x_{mn}^k$. This means that $w_m \leq \min\sum_{n=1}^{|D_{ij}^{fea}(t)|}(b_{mn}-u_n)x_{mn} = \max\sum_{n=1}^{|D_{ij}^{fea}(t)|}(u_n-b_{mn})x_{mn}$. Thus, the dual problem can be represented as

$$\max \sum_{n=1}^{|D_{ij}^{fea}(t)|} u_n + \sum_{m\in\{i,j\}} \begin{cases} \max\sum_{n=1}^{|D_{ij}^{fea}(t)|}(u_n-b_{mn})x_{mn} \\ \text{s.t.} \sum_{n=1}^{|D_{ij}^{fea}(t)|} b_{mn}x_{mn} \leq \delta_m, \\ x_{mn}\in\{0,1\}, n=1,2,\ldots,|D_{ij}^{fea}(t)|. \end{cases} \tag{32}$$

Further, the problem in (32) is equivalent to finding $\varphi^* = \max_{u\geq 0}\varphi(u)$, where $\varphi(u)$ is given by

$$\begin{aligned} \max f(x) = &-\sum_{m\in\{i,j\}}\sum_{n=1}^{|D_{ij}^{fea}(t)|} b_{mn}x_{mn} \\ &+ \sum_{n=1}^{|D_{ij}^{fea}(t)|} u_n\left(1+\sum_{m\in\{i,j\}} x_{mn}\right) \\ \text{s.t.} &\sum_{n=1}^{|D_{ij}^{fea}(t)|} b_{mn}x_{mn} \leq \delta_m, \forall m\in\{i,j\}, \\ &x_{mn}\in\{0,1\}, \forall m\in\{i,j\}, n=1,2,\ldots,|D_{ij}^{fea}(t)|, \end{aligned} \tag{33}$$

which is the Lagrangian dual problem. It is found that the problem (33) can be represented as two independent *Knapsack* problems [37], wherein $m=i$ and $m=j$, respectively. Here, $\varphi(u)$ is transformed to be

$$\begin{aligned} \max f(x) &= \sum_{m\in\{i,j\}}\sum_{n=1}^{|D_{ij}^{fea}(t)|}(u_n-b_{mn})\,x_{mn} + \sum_{n=1}^{|D_{ij}^{fea}(t)|} u_n \\ &= \sum_{n=1}^{|D_{ij}^{fea}(t)|}(u_n-b_{in})\,x_{in} + \sum_{n=1}^{|D_{ij}^{fea}(t)|}\left(u_n-b_{jn}\right)x_{jn} \\ &\quad + \sum_{n=1}^{|D_{ij}^{fea}(t)|} u_n \\ \text{s.t.} &\sum_{n=1}^{|D_{ij}^{fea}(t)|} b_{in}x_{in} \leq \delta_i, \\ &\sum_{n=1}^{|D_{ij}^{fea}(t)|} b_{jn}x_{jn} \leq \delta_j, \\ &x_{in}\in\{0,1\}, \forall n=1,2,\ldots,|D_{ij}^{fea}(t)|, \\ &x_{jn}\in\{0,1\}, \forall n=1,2,\ldots,|D_{ij}^{fea}(t)|. \end{aligned} \tag{34}$$

It is observed that the problem in (18) can derive the exact solution using the branch and bound (BAB) method [38]. However, this method is relatively complicated because of the computational complexity, $O(n2^n)$. In this paper, we propose a dynamic heuristic scheme based on the steepest descent method to obtain rapidly the near optimal solution for the problem in (34), and thus the solution for the problem in (18). First, the proposed algorithm solves the two Knapsack problems in (32) with the given u to find the maximum value of $f(x)$. Thereafter, the steepest descent method [39] is used to find the optimal u^* to maximize $\varphi(u)$. A problem arises when we address the two Knapsack problems separately, in that we can find two different sets of connected vehicles defined by x_i^* and x_j^*. Then, it cannot be ensured that the constraint $\sum_{m\in\{i,j\}} x_{mn} = 1, n=1,2,\ldots,|D_{ij}^{fea}(t)|$. This means that we cannot ensure that a service of connected vehicles is assigned to an FN. This leads to a solution of the Knapsack problem that is infeasible. In order to guarantee the feasibility of the solution after the two Knapsack problems have been solved, we propose an approximation algorithm that comprises solving KP1 and KP2:

$$\text{KP1: } \max \sum_{n=1}^{|D_{ij}^{fea}|}(u_n-b_{in})x_{in} \tag{35}$$

$$\text{s.t.} \sum_{n=1}^{|D_{ij}^{fea}(t)|} b_{in}x_{in} \leq \delta_i, \tag{36}$$

$$x_{in}\in\{0,1\}, \forall n=1,2,\ldots,|D_{ij}^{fea}(t)|, \tag{37}$$

$$\begin{aligned} \text{KP2: } \max &\sum_{n\in n_{KP1}^*}\left[(u_n-b_{jn})-(u_n-b_{in})\right]x_{jn} \\ &+ \sum_{n\notin n_{KP1}^*}(u_n-b_{jn})x_{jn} \\ &= \sum_{n\in n_{KP1}^*}(b_{in}-b_{jn})x_{jn} + \sum_{n\notin n_{KP1}^*}(u_n-b_{jn})x_{jn} \end{aligned} \tag{38}$$

$$\text{s.t.} \sum_{n=1}^{|D_{ij}^{fea}(t)|} b_{jn}x_{jn} \leq \delta_j, \tag{39}$$

$$x_{jn}\in\{0,1\}, \forall n=1,2,\ldots,|D_{ij}^{fea}(t)|, \tag{40}$$

where sets n_{KP1}^* and n_{KP2}^* are the solutions of the KP1 and KP2 problems, respectively. This means that $x_{in}=1$ and $x_{jn}=1$ if $n\in n_{KP1}^*$ and $n\in n_{KP2}^*$, respectively. It is noteworthy that some elements in set n_{KP2}^* might belong to set n_{KP1}^*. Hence, set n_{KP1}^* must be updated by $n_{KP1}^* = n_{KP1}^*\backslash n_{KP2}^*$.

The proposed scheme is summarized in Algorithm 1. In this scheme, we initialize the value of $u_n^0 = 0$, $n = 1,2,\ldots,|D_{ij}^{fea}(t)|$ in the first step. In each step k, we solve the two Knapsack problems in (32) with the given u^k in sequence to find the maximum value of $f(x)$. The new value u^{k+1} is updated following the gradient value $\nabla\varphi(u^k) = 1+\sum_{m\in\{i,j\}} x_{mn}$ with respect to x^k at step k. According to the steepest descent method, the value of $\varphi(u)$ increases after each iteration and converges to the near optimal value. When the magnitude $||\nabla\varphi(u^k)||$ is less than a very small tolerance ξ, the iteration stops. Subsequently, we obtain the near optimal value x^*, which results in $W_{ij}(t), d_{ij}^*(t)$, and $d_{ji}^*(t)$. The complexity of the dynamic heuristic algorithm is $O(\xi^{-2}nW)$, where $n=|D_{ij}^{fea}|$ and $W=max(\delta_i,\delta_j)$. Because $|D_{ij}^{fea}|$ is large for a massive

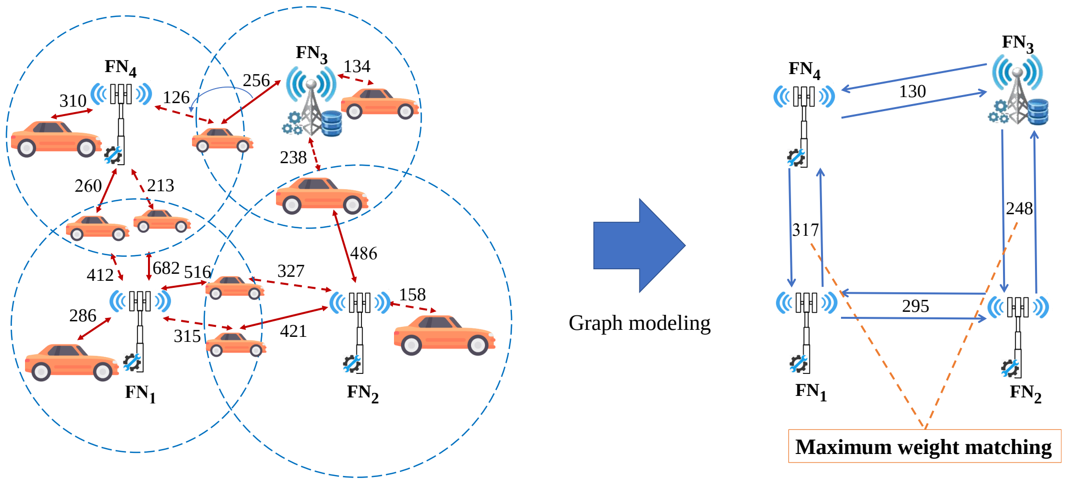

Fig. 2. Example of service migration using the dynamic resource orchestration algorithm.

Algorithm 1 Dynamic Heuristic Algorithm.

Input: $D_{ij}^{fea}(t)$
Output: $W_{ij}(t)$, $d_{ij}^*(t)$, $d_{ji}^*(t)$
• Initialization
u^k and x^k are u and x at *k-th* step, respectively, $u^0 = 0$
ξ is the tolerance, γ is the step size
$\nabla\varphi(u) = 1 + \sum_{m\in\{i,j\}} x_{mn}$ is the gradient of $\varphi(u)$
repeat
  Given u^k, solve the *Knapsack* problem in (35) and (38) to derive x^k and $max_{f^k(x)}$
  $u^{k+1} = u^k + \gamma\nabla\varphi(u^k)$
  $k = k + 1$
until $||\nabla\varphi(u^k)|| \le \xi$
$u^* = u^k$ and $x^* = x^k$
From x^*, obtain $D_i^*(t), D_j^*(t)$ and $W_{ij}(t)$ by (20), and $d_{ij}^*(t), d_{ji}^*(t)$ by (21)

number of connected vehicles, the dynamic heuristic algorithm is relatively effective and allows a fast derivation of the near optimal value as compared to the BAB method, the computational complexity of which is $O(m2^m)$, where m is the number of connected vehicles served in the network.

C. Dynamic Resource Orchestration

By obtaining the optimal service migration between two FNs, we can determine whether the services should be migrated between these two FNs to minimize the amount of their occupied RBs. With the aim of optimizing the resource utilization efficiency, as well as to minimize the amount of occupied RBs for the entire network, we need to find the pairs of FNs that allow optimal service migration. The problem is presented as maximum weight matching in (7) and considered an NP-hard problem. An exact algorithm was presented by Edmonds [36] to address the maximum weight matching problem having a complexity of $O(em^2)$, where e and m are the number of edges and vertices of the graph, respectively. Although the Edmonds algorithm provides a method to achieve the exact solution in polynomial time, its complexity is high. With the objective of decreasing the complexity, we propose a greedy algorithm for finding the maximum weight matching having a complexity of $O(e\log|m|)$. A DRO scheme is proposed for

Algorithm 2 Dynamic Resource Orchestration.

Input: Set of the FNs Ω, $G(V, E)$
Output: Set of the matching FNs conducting service migration Φ
Find $W_{ij}, i, j = 1, 2, \ldots, N$ by Algorithm 1
$\Phi = \varnothing$
repeat
  $e_{ij} = \{e \in E | W_{ij}$ is the largest$\}$
  $\Phi = \Phi \cup \ e_{ij}$
  $E = E\backslash\{e_{ij}$ and all edges incident to $e_{ij}\}$
until $E == \varnothing$

migrating services among FNs such as resource harmonization and efficient resource utilization are achieved, resulting in a maximized service capability. The DRO scheme is described in Algorithm 2.

In the first step, the DRO scheme finds all matching weights (i.e., the weight of the edge) W_{ij} between two FNs according to Algorithm 1 presented in Section III.B. In the next step, the DRO scheme addresses the problem in (7) by finding the edges (i.e., the matching FNs) in a descending sequence of the matching weights. The edge that has the largest matching weight is chosen and updated to the matching FN set Φ. Because matching involves a set of edges without common vertices, the chosen edge and all edges incident to it are removed from the edge set E to guarantee the next finding iteration. The iteration is completed if the edge set E is empty. The DRO scheme obtains the set of matching FNs. In other words, the services of the connected vehicles matching FNs are migrated to achieve RB occupation minimization, which results in an optimized resource utilization efficiency.

Fig. 2 illustrates an example of the DRO scheme for minimizing the number of occupied RBs among FNs. The red solid lines between connected vehicles and FNs indicate that the connected vehicles are currently served by the FNs. Meanwhile, the red dashed lines show that the services of the connected vehicles are possibly migrated to the FNs. The numbers on the red lines represent the number of RBs that the connected vehicles require from the FNs to satisfy their desired data rates. Moreover, the numbers on the blue lines determine the maximum resource reduction when optimal service migrations are conducted in each pair of FNs. The optimal service migration for maximizing the occupied RB reduction is derived by the dynamic heuristic algorithm. In Fig. 2, two connected vehicles in the overlapped area of FN_1 and FN_4 are feasible for service migration. In particular, the total numbers of RBs occupied by these vehicles before and after conducting service migration are 942 (i.e., 260 + 682) and 625 (i.e., 412 + 213), respectively. Therefore, the weight matching (i.e., maximum resource reduction) if migrating services between FN_1 and FN_4 is calculated by 932 - 625 = 317. Similarly, we can obtain the weight matching for the remaining pairs such as $\{FN_1, FN_2\}$, $\{FN_2, FN_3\}$, and $\{FN_3, FN_4\}$. According to Algorithm 2, the selection of FN pairs for service migration is based on a descending order of the matching weights. Therefore, the pair $\{FN_1, FN_4\}$ is selected first owing to its largest matching weight. Consequently, the next selected pair is $\{FN_2, FN_3\}$. As a result, the maximum weight matching of the network is 565 (i.e., 317 + 248), and the matching set involves $\{FN_1 \leftrightarrow FN_4, FN_2 \leftrightarrow FN_3\}$. This means that the services of the connected vehicles should be migrated between FN_1 and FN_4 and between FN_2 and FN_3 to minimize the number of RBs occupied by the connected vehicles.

D. Computational Complexity Analysis

As analyzed in Sections III.B and III.C, the proposed DRO scheme finds the optimal service migration among FNs by using a dynamic heuristic algorithm and then utilizes the maximum weight matching algorithm to determine the pairs of FNs for realizing these service migrations. The complexities of these algorithms are $O(\xi^{-2}nW)$ and $O(e \log |m|)$, respectively, where ξ is the tolerance, $n = |D_{ij}^{fea}|$, $W = max(\delta_i, \delta_j)$, and m is the number of served connected vehicles in the network. Hence, the complexity of the DRO scheme is $O(\xi^{-2}mnW)$. Meanwhile, the complexity of the exact algorithm, i.e., branch and bound (BAB) [38], is $O(m2^m)$. Because $n \ll m$, the DRO scheme introduces a much lower complexity than the BAB scheme while maintaining an approximate performance (see the following sections for detailed comparisons).

V. Performance Evaluation

The simulation and evaluation of the performance of our proposed DRO scheme are described in this section. We operated a network model including 10~90 FNs deployed in a region measuring $(500\sim1500)^2$ square meters. During each timeslot unit, an arbitrary number of service connections from the connected vehicles arrived at and departed from the network. The duration of each simulation was 300 timeslots. One hundred Monte-Carlo simulations were conducted and the average results were obtained. Table II summarizes the details of the simulation parameters. To evaluate the effects of the DRO scheme on individual device associations, we conducted simulations for three association schemes: the SA, capacity-aware CaA, and CoA schemes. We also performed the simulation using additional migration schemes, such as the ARB [27] and BAB [38] (i.e., the exact solution), to demonstrate the performance of our method. The effectiveness of the proposed DRO scheme was determined in terms of the

TABLE II
Simulation Parameters

Parameter	Value
Network area	$(500\sim1500)^2$ m^2
Number of FNs	10~90
FN coverage radius (R)	100~500 m
FN bandwidth	{10, 15, 20} MHz
Cumulative number of service connections	6000
Mean arrival rate (λ)	20 service connections/s
Mean departure rate (μ)	10 service connections/s
Desired data rate (r_j)	0.5~2 Mbps
Timeslot duration	1 s

service capability, serviceability, availability, and throughput. The serviceability and availability of the network are defined as the ability of the network to serve connected vehicles within the desired and minimum requirements (e.g., the throughput and delay), respectively [27]. In other words, the serviceability and availability are defined by the percentage of connected vehicles that are served per cumulative number of connected vehicles arriving at the network during an interval of time within the desired and minimum requirements, respectively. In this study, for determining network availability the minimum data rate of the connected vehicles was considered the minimum requirement.

Fig. 3 shows the effects of the DRO scheme as compared to those of the corresponding device association schemes in terms of network service capability. During the first 100 timeslots, the service capability rapidly decreases, because the radio resource is occupied to serve numerous connected vehicles arriving at the network. In the CaA and CoA approaches, the resource is not efficiently managed, because their device association prefers FNs that have a high resource availability and interesting contents, respectively. This means that the FNs may allocate a larger amount of RBs for the connected vehicles to guarantee the data rate requirements because of the lower SINR. This leads FNs to reach sooner the state where they are unable to serve connected vehicles, because the resource is exhausted. Meanwhile, the SA approach overcomes the shortcomings of the two aforementioned approaches and achieves better service capability, because the spectral resource efficiency is considered when the connected vehicles are associated with the FNs having a high SINR. However, the problem arises that these FNs may attract an enormous number of users, resulting in an overcapacity issue, and then these FNs cannot provide services to new incoming connected vehicles. This means that the later incoming connected vehicles may be served by FNs that have a lower SINR, and thus, the network service capability decreases. By migrating the services of the connected vehicles among FNs, the resource in the DRO scheme is orchestrated with the aim of maximizing the spectral resource utilization, which results in better service capability. The simulation results show that the DRO scheme significantly improves the network service capability by up to 37.92%, 44.87%, and 8.04% as compared to the standalone CaA, CoA, and SA approaches, respectively.

Fig. 4 depicts the aggregate network throughput achieved by each device association scheme. A comparison of the six

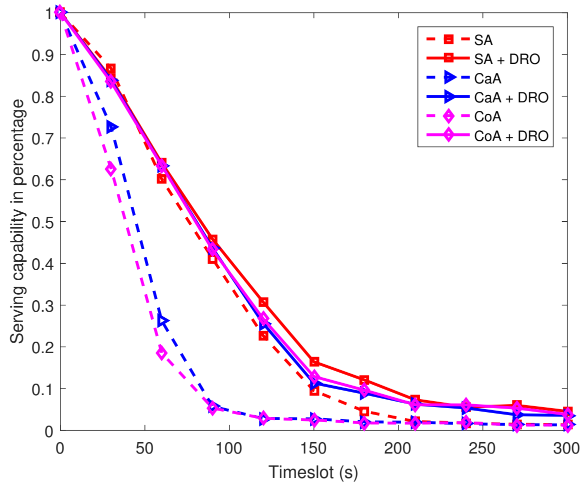

Fig. 3. Network service capability of six device association schemes.

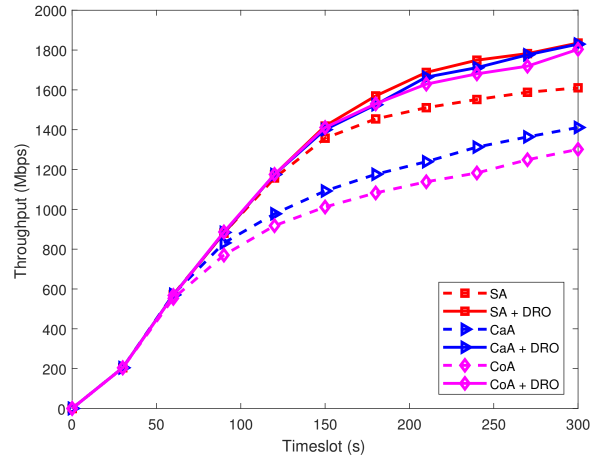

Fig. 4. Aggregate network throughput achieved by each device association scheme.

schemes in more than the first 50 timeslots reveals that the network throughput steadily increases and the difference in throughput is small because the capacity of almost all the FNs is sufficient to serve the incoming connected vehicles with their required data rate. Thereafter, the cumulative number of connected vehicles arriving at the network increases rapidly. This leads to some interesting FNs that have a high SINR, favorite contents, and/or high resource availability suffering from an overload. This means that the new incoming connected vehicles may associate with the poorer FNs or drop from the network, which results in a network throughput decrease. The SA approach is aimed at maximizing the data sum-rate, and thus obtains a better throughput performance than the CaA and CoA approaches. By migrating the services among FNs, the DRO scheme harmonizes resource occupation, alleviates the burden on the preferred FNs, and guarantees resource availability to provide service to the greatest possible number of connected vehicles, which results in better network throughput. The simulation results demonstrate the effectiveness of the DRO scheme in that it improves the network throughput by up to 43.09%, 34.28%, and 13.96% as compared to the CoA, CaA, and SA approaches, respectively.

The proposed DRO scheme also significantly improves the network performance in terms of serviceability, as shown on the left hand of Fig. 5. The network serviceability is almost maintained at the maximum value in the first 40 timeslots. This means that the FNs can ensure services to all incoming connected vehicles within the expected data rate requirement in the 0.5–2 Mbps range, because the resource is available. Subsequently, when the connected vehicles arrive the network increases. Consequently, the preferred FNs reach overload and cannot provide services to the new incoming connected vehicles. This leads to a decrease in the serviceability of the network. The DRO scheme maximizes the resource availability, and thus derives the best serviceability. The simulation results reveal that the proposed DRO scheme improves performance by 14.95%, 11.39%, and 4.81% as compared to the CoA, CaA, and SA approaches, respectively. Similarly, the right part of

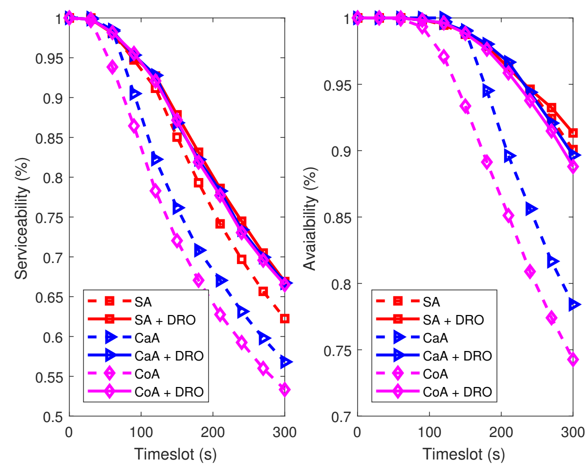

Fig. 5. Comparison of network serviceability and availability of six device association schemes.

Fig. 5 presents the network availability satisfying the minimum connected vehicle data rate of 512 Kbps. Because FNs allocate only the resource amount that ensures the minimum rate, the capacity is sufficient to serve all incoming connected vehicles until timeslot 100. As a consequence, the network availability decreases when the cumulative number of connected vehicles arriving at the network increases. It is noteworthy that the DRO scheme may achieve the best improvement when applied to the SA approach when the FNs rapidly reach overcapacity and a large number of connected vehicles are associated with the poorer FNs. In this case, the FNs slowly reach overload in the SA approach. Hence, the DRO scheme slightly improves the network availability. The DRO scheme's effectiveness is demonstrated in terms of improving network availability by 14.53%, 11.25%, and 1.25% as compared to the standalone CoA, CaA, and SA approaches, respectively.

Fig. 6 shows a comparison of the six schemes in terms of the resource utilization efficiency (eff_{RB}) of the network. The height of the box represents the distribution of the eff_{RB}

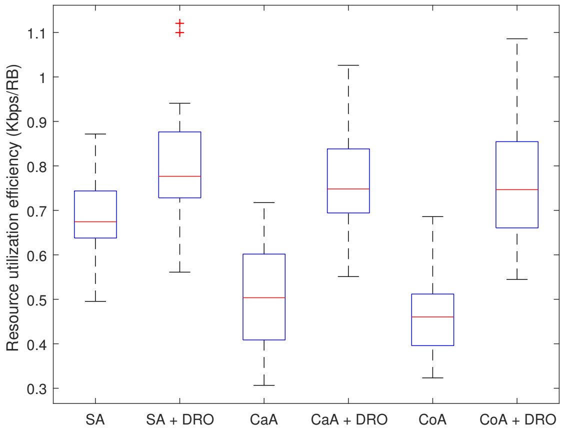

Fig. 6. Resource utilization efficiency of the network.

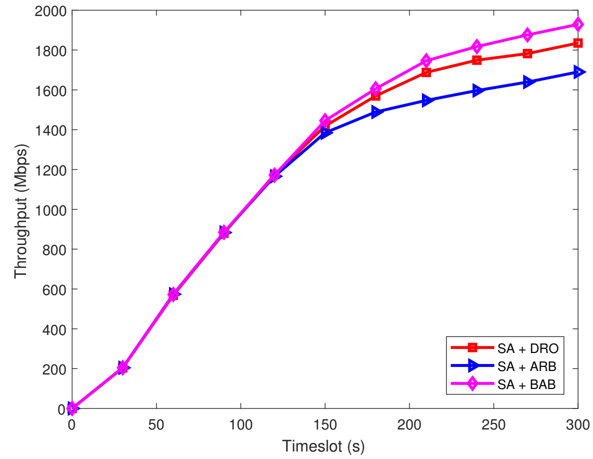

Fig. 7. Network throughput of three migration algorithms.

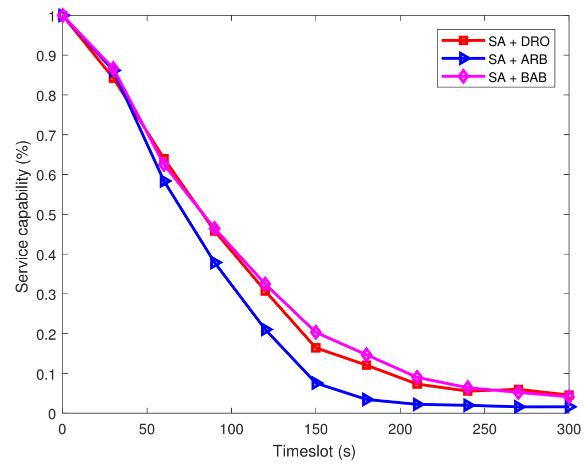

Fig. 8. Service capability of three migration algorithms.

of FNs, and the red median line inside the box represents the eff_{RB} of the entire network. It is readily observed that the DRO scheme provides a better performance in terms of resource utilization. The DRO scheme can minimize the amount of occupied RBs of the FNs, thus achieving a better spectral efficiency. Among the device association schemes, the SA approach obtains the best eff_{RB}, because it considers the signal quality. The simulation results shows that the application of the DRO scheme improves resource utilization efficiency by up to 30.33%, 25.70%, and 11.71% as compared to the CoA, CaA, and SA schemes, respectively. In addition, a significant communication latency (l_{RB}) reduction is obtained as a result. Assume that each connected vehicle remains the number of assigned RBs from the FNs within all association schemes. Therefore, a given desired data rate of a connected vehicle $r_j = \#\text{RB} \times \text{eff}_{RB} \times l_{RB}$. Since r_j and #RB are constant, the l_{RB} is inversely proportional to the eff_{RB}. In other word, communication latency reduction is equal to the eff_{RB} as aforementioned.

Fig. 7 and Fig. 8 present the network throughput and service capability in comparison with those of other migration schemes, respectively. During the early timeslots, the FNs possess a sufficient resource amount to satisfy the services required by the connected vehicles. Thus, there is little difference in the network throughput and service capability. Thereafter, the numerous connected vehicles arriving at the network cause the FNs to reach overload rapidly. It is observed that the proposed DRO scheme achieves a better performance than the ARB scheme and approximately the same performance as the BAB scheme in terms of throughput and service capability. This is because the ARB scheme simply migrates the services of the connected vehicles from the FNs that have a high resource availability to the FNs that have a low resource availability to obtain resource balancing among them. This results in inefficient resource utilization. Meanwhile, our proposed scheme migrates the optimal services to achieve resource optimization. The BAB scheme provides the best performance of all schemes. However, its computational complexity is high. The effectiveness of the DRO scheme when it has approximate results is demonstrated. Furthermore, its complexity is lower than that of to the BAB scheme. Similarly, Fig. 9 depicts the network serviceability achieved by each migration scheme. It can be seen that the DRO scheme obtains better serviceability than the ARB scheme. In particular, it almost achieves the optimal results of the BAB scheme. The simulation results show that the DRO scheme significantly improves the network throughput, service capability, and serviceability by up to 9.57%, 9.67%, and 3.3% and approximately 0.56%, 0.14%, and 0.08% as compared to the ARB and BAB schemes, respectively.

Fig. 10 illustrates the main metrics, serviceability and network throughput, in various network configurations, where the network area is increased to 1500×1500 square meters. In this topology, the number of FNs and the FN coverage radius are adjusted in ranges of $(10 - 90)$ and $(100 - 500)$ m, respectively. The demonstrated results were captured after 150 timeslots in two cases, with and without the proposed DRO scheme on the SA algorithm. Both subfigures show a proportional increase of serviceability and throughput depending on the FN coverage

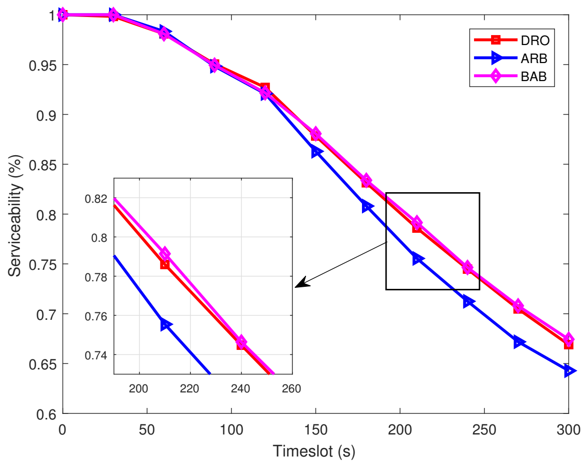

Fig. 9. Network serviceability of three migration algorithms.

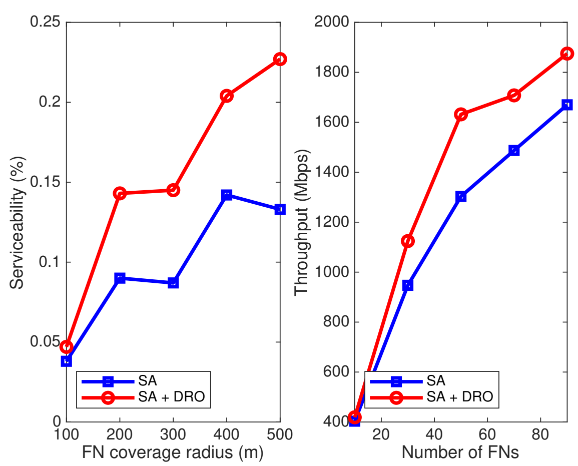

Fig. 10. Serviceability and network throughput in various network configurations.

radius and #FNs increases. In addition, the proposed DRO scheme provides up to approximate 70% and 25% increase for the SA algorithm in terms of serviceability and throughput, respectively.

VI. Concluding Remarks

In this paper, a dynamic resource orchestration scheme to schedule resource allocation in FCVNs by migrating connected vehicle services among FNs to maximize the service capability and resource utilization efficiency was proposed. On the basis of graph theory, the DRO scheme considers FNs as vertices, and the weight of an edge between two vertices is given by the maximum resource reduction when optimal service migration among the FNs is conducted. The optimal service migration between each pair of FNs for minimizing resource utilization is obtained by using the steepest descent method. Then, the maximum weight matching solution is used to determine the optimal pairs of FNs for migrating services, with the aim of maximizing network resource utilization efficiency. The results of simulation analyses demonstrate that the proposed DRO scheme achieves significant improvements in terms of service capability, resource utilization efficiency, and throughput as compared to existing algorithms. In future work, deep learning and power consumption will be considered as an extension of our current studies. In addition, a latency-aware task execution scheduling scheme in FNs will be a target.

References

[1] C. Huang, R. Lu, and K.-K. R. Choo, "Vehicular fog computing: architecture, use case, and security and forensic challenges," *IEEE Communications Magazine*, vol. 55, no. 11, pp. 105–111, 2017.

[2] X. Hou, Y. Li, M. Chen, D. Wu, D. Jin, and S. Chen, "Vehicular fog computing: A viewpoint of vehicles as the infrastructures," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3860–3873, 2016.

[3] N.-N. Dao, D.-N. Vu, Y. Lee, S. Cho, C. Cho, and H. Kim, "Pattern-identified online task scheduling in multitier edge computing for industrial IoT services," *Mobile Information Systems*, vol. 2018, p. 2101206, 2018.

[4] C. C. Byers, "Architectural imperatives for fog computing: Use cases, requirements, and architectural techniques for fog-enabled IoT networks," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 14–20, 2017.

[5] R. Moreno-Vozmediano, R. S. Montero, E. Huedo, and I. M. Llorente, "Cross-site virtual network in cloud and fog computing," *IEEE Cloud Computing*, vol. 4, no. 2, pp. 46–53, 2017.

[6] M. Shojafar, N. Cordeschi, and E. Baccarelli, "Energy-efficient adaptive resource management for real-time vehicular cloud services," *IEEE Transactions on Cloud computing*, 2016.

[7] X. Zhang and Q. Zhu, "Game-theory based power and spectrum virtualization for optimizing spectrum efficiency in mobile cloud-computing wireless networks," *IEEE Transactions on Cloud Computing*, 2017.

[8] S. Parsaeefard, V. Jumba, A. Dalili, M. Derakhshani, and T. Le-Ngoc, "User association in cloud RANs with massive MIMO," *IEEE Transactions on Cloud Computing*, 2018.

[9] N.-N. Dao, D.-N. Vu, W. Na, J. Kim, and S. Cho, "SGCO: Stabilized green crosshaul orchestration for dense IoT offloading services," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 11, pp. 2538–2548, 2018.

[10] M. Tao, M. Dong, K. Ota, and Z. He, "Multiobjective network opportunistic access for group mobility in mobile internet," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–10, 2016.

[11] N.-N. Dao, M. Park, J. Kim, J. Paek, and S. Cho, "Resource-aware relay selection for inter-cell interference avoidance in 5G heterogeneous network for Internet of things systems," *Future Generation Computer Systems*, vol. 93, pp. 877–887, 2019.

[12] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K.-K. Wong, R. Schober, and L. Hanzo, "User association in 5G networks: A survey and an outlook," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1018–1044, 2016.

[13] D.-N. Vu, N.-N. Dao, Y. Jang, W. Na, Y.-B. Kwon, H. Kang, J. J. Jung, and S. Cho, "Joint energy and latency optimization for upstream IoT offloading services in fog radio access networks," *Transactions on Emerging Telecommunications Technologies*, vol. 28, p. e3497, 2018.

[14] C. Liu, M. Li, S. V. Hanly, and P. Whiting, "Joint downlink user association and interference management in two-tier HetNets with dynamic resource partitioning," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1365–1378, 2017.

[15] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Match to cache: Joint user association and backhaul allocation in cache-aware small cell networks," in *IEEE International Conference on Communications (ICC)*, 2015, pp. 3082–3087.

[16] Q. Ye, O. Y. Bursalioglu, H. C. Papadopoulos, C. Caramanis, and J. G. Andrews, "User association and interference management in massive MIMO HetNets," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 2049–2065, 2016.

[17] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.

[18] F. Boccardi, J. Andrews, H. Elshaer, M. Dohler, S. Parkvall, P. Popovski, and S. Singh, "Why to decouple the uplink and downlink in cellular networks and how to do it," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 110–117, 2016.

[19] Q. Zheng, K. Zheng, H. Zhang, and V. C. Leung, "Delay-optimal virtualized radio resource scheduling in software-defined vehicular networks via stochastic learning," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 7857–7867, 2016.

[20] M. M. Hasan, S. Kwon, and J.-H. Na, "Adaptive mobility load balancing algorithm for LTE small-cell networks," *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–1, 2018.

[21] T. Zhou, Y. Huang, L. Fan, and L. Yang, "Load-aware user association with quality of service support in heterogeneous cellular networks," *IET Communications*, vol. 9, no. 4, pp. 494–500, 2015.

[22] H. He, H. Shan, A. Huang, and L. Sun, "Resource allocation for video streaming in heterogeneous cognitive vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 7917–7930, 2016.

[23] J. Zheng, J. Li, N. Wang, and X. Yang, "Joint load balancing of downlink and uplink for eICIC in heterogeneous network," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6388–6398, 2017.

[24] D.-N. Vu, N.-N. Dao, and S. Cho, "Downlink sum-rate optimization leveraging Hungarian method in fog radio access networks," in *IEEE International Conference on Information Networking (ICOIN)*, Chiang Mai, Thailand, January 10–12, 2018, pp. 56–60.

[25] R. Q. Hu and Y. Qian, "An energy efficient and spectrum efficient wireless heterogeneous network framework for 5g systems," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 94–101, 2014.

[26] D. Liu, Y. Chen, K. K. Chai, T. Zhang, and M. Elkashlan, "Opportunistic user association for multi-service hetnets using nash bargaining solution," *IEEE Communications Letters*, vol. 18, no. 3, pp. 463–466, 2014.

[27] N.-N. Dao, J. Lee, D.-N. Vu, J. Paek, J. Kim, S. Cho, K.-S. Chung, and C. Keum, "Adaptive resource balancing for serviceability maximization in fog radio access networks," *IEEE Access*, vol. 5, pp. 14 548–14 559, 2017.

[28] X. Ge, X. Li, H. Jin, J. Cheng, and V. C. Leung, "Joint user association and user scheduling for load balancing in heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3211–3225, 2018.

[29] N. Cordeschi, D. Amendola, and E. Baccarelli, "Reliable adaptive resource management for cognitive cloud vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 6, pp. 2528–2537, 2015.

[30] H. H. M. Tam, H. D. Tuan, D. T. Ngo, T. Q. Duong, and H. V. Poor, "Joint load balancing and interference management for small-cell heterogeneous networks with limited backhaul capacity," *IEEE Transactions on Wireless Communications*, vol. 16, no. 2, pp. 872–884, 2017.

[31] T. K. Vu, M. Bennis, S. Samarakoon, M. Debbah, and M. Latva-aho, "Joint load balancing and interference mitigation in 5g heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 6032–6046, 2017.

[32] T. Zhou, Z. Liu, J. Zhao, C. Li, and L. Yang, "Joint user association and power control for load balancing in downlink heterogeneous cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2582–2593, 2018.

[33] Z. Bai, "Evolved universal terrestrial radio access (E-UTRA); physical layer procedures," *3GPP, Sophia Antipolis, Technical Specification 36.213 v. 11.4. 0*, 2013.

[34] ETSI, "Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation," *ETSI TS 136.211 V10*, 2011.

[35] J. Li, J. B. Song, and Z. Han, "Network connectivity optimization for device-to-device wireless system with femtocells," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3098–3109, 2013.

[36] R. Duan and S. Pettie, "Linear-time approximation for maximum weight matching," *Journal of the ACM (JACM)*, vol. 61, no. 1, p. 1, 2014.

[37] U. Pferschy and R. Scatamacchia, "Improved dynamic programming and approximation results for the knapsack problem with setups," *International Transactions in Operational Research*, vol. 25, no. 2, pp. 667–682, 2018.

[38] D. R. Morrison, S. H. Jacobson, J. J. Sauppe, and E. C. Sewell, "Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning," *Discrete Optimization*, vol. 19, pp. 79–102, 2016.

[39] J. Lenells, "The nonlinear steepest descent method: asymptotics for initial-boundary value problems," *SIAM Journal on Mathematical Analysis*, vol. 48, no. 3, pp. 2076–2118, 2016.

Duc-Nghia Vu is a senior engineer at Innowireless Inc., Seongnam, South Korea. He received his B.S. degree in Electronics and telecommunications from Hanoi University of Science and Technology, Viet Nam, in 2015. He also received the M.S. degree in computer science from Chung-Ang University, South Korea, in 2018. His research interests include wireless network and fog computing.

Nhu-Ngoc Dao (M'20) is an Assistant Professor at the Department of Computer Science and Engineering, Sejong University, Seoul, South Korea. He received his M.S. and Ph.D. degrees in computer science with the School of Computer Science and Engineering, Chung-Ang University, Seoul, South Korea, in 2016 and 2019, respectively. He received the B.S. degree in electronics and telecommunications from the Posts and Telecommunications Institute of Technology, Hanoi, Viet Nam, in 2009. Prior to being with the Sejong University, he was a postdoc researcher at the Institute of Computer Science, University of Bern, Switzerland from 2019 to 2020. His research interests include network softwarization, mobile cloudization, and Internet of things.

Woongsoo Na received the B.S., M.S., and Ph.D. degrees in computer science and engineering from Chung-Ang University, Seoul, South Korea, in 2010, 2012, and 2017, respectively. He is currently an Assistant Professor with the Division of Computer Science and Engineering, Kongju National University, Cheonan, South Korea. Prior to joining Kongju National University, he was an Adjunct Professor with the School of Information Technology, Sungshin Womens University, Seoul, South Korea, from 2017 to 2018, and a Senior Researcher with Electronics and Telecommunications Research Institute, Daejeon, South Korea, from 2018 to 2019. His current research interests include mobile edge computing, ying ad hoc networks, wireless mobile networks, and beyond 5G.

Sungrae Cho is a professor with the school of computer sciences and engineering, Chung-Ang University (CAU), Seoul. Prior to joining CAU, he was an assistant professor with the department of computer sciences, Georgia Southern University, Statesboro, GA, USA, from 2003 to 2006, and a senior member of technical staff with the Samsung Advanced Institute of Technology (SAIT), Kiheung, South Korea, in 2003. From 1994 to 1996, he was a research staff member with electronics and telecommunications research institute (ETRI), Daejeon, South Korea. From 2012 to 2013, he held a visiting professorship with the national institute of standards and technology (NIST), Gaithersburg, MD, USA. He received the B.S. and M.S. degrees in electronics engineering from Korea University, Seoul, South Korea, in 1992 and 1994, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2002.

His current research interests include wireless networking, ubiquitous computing, and ICT convergence. He has been a subject editor of IET Electronics Letter since 2018, and was an area editor of Ad Hoc Networks Journal (Elsevier) from 2012 to 2017. He has served numerous international conferences as an organizing committee chair, such as IEEE SECON, ICOIN, ICTC, ICUFN, TridentCom, and the IEEE MASS, and as a program committee member, such as IEEE ICC, MobiApps, SENSORNETS, and WINSYS.