Drone-Small-Cell-Assisted Resource Slicing for 5G Uplink Radio Access Networks

Hang Shen, *Member, IEEE*, Qiang Ye, *Member, IEEE*, Weihua Zhuang, *Fellow, IEEE*, Weisen Shi, *Graduate Student Member, IEEE*, Guangwei Bai, and Geng Yang, *Member, IEEE*

***Abstract*—Radio resource slicing is critical to customize service provisioning in fifth-generation (5G) uplink radio access networks (RANs). Using drone-small-cells (DSCs) as aerial support for terrestrial base stations can enhance the flexibility for resource provisioning in response to traffic distribution variations. In this paper, we study a multi-DSC-assisted radio resource slicing problem for 5G uplink RANs, with the objective of minimizing the total uplink resource consumption under differentiated quality-of-service (QoS) constraints for both human-type and machine-type communication services. We begin with an interference-aware graph model to formulate the joint DSC three-dimension (3D) placement and device-DSC association problem for uplink radio resource slicing and prove that the proposed problem is NP-hard. A complexity-adjustable problem approximation is presented via screening candidate DSC deployment positions, which incorporates flight height adaptation to balance the uplink communication coverage and resource utilization. A lightweight approximation using a fixed DSC flight altitude is also provided with reduced complexity. For mathematical traceability, the DSC placement and device-DSC associations in each approximation are transformed as a special weight clique problem. An upgraded clique algorithm is then developed to determine how to deploy DSCs for a given number of DSCs. Simulation results demonstrate the proposed scheme's effectiveness in terms of resource utilization, network coverage, and drone dispatching cost.**

***Index Terms*—Uplink, radio access networks (RANs), radio resource slicing, drone-small-cell (DSC) deployment, differentiated quality-of-service (QoS).**

Manuscript received February 1, 2021; revised May 3, 2021; accepted May 18, 2021. Date of publication May 25, 2021; date of current version July 20, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 61502230, 61501224, 61701222, 61802176, and 61802177, in part by the National Project Funding for Key R&D Programs under Grant 2018YFC0808500, in part by the Natural Science Foundation of Jiangsu Province under Grants BK20201357 and BK20180696, in part by the Six Talent Peaks Project in Jiangsu Province under Grant RJFW-020, and in part by the Jiangsu Key Laboratory Program of Big Data Security, and Intelligent Processing (Nanjing University of Posts, and Telecommunications) under Grant BDSIP1910. The review of this article was coordinated by Dr. Phone Lin. (*Corresponding author: Qiang Ye.*)

Hang Shen and Guangwei Bai are with the Department of Computer Science and Technology, Nanjing Tech University, Nanjing 211816, China, and with Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: hshen@njtech.edu.cn; bai@njtech.edu.cn).

Qiang Ye is with the Department of Electrical and Computer Engineering and Technology, Minnesota State University, Mankato MN 56001 USA (e-mail: qiang.ye@mnsu.edu).

Weihua Zhuang and Weisen Shi are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo ON N2L 3G1, Canada (e-mail: wzhuang@uwaterloo.ca; w46shi@uwaterloo.ca).

Geng Yang is with the Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: yangg@njupt.edu.cn).

Digital Object Identifier 10.1109/TVT.2021.3083255

I. Introduction

THE fifth-generation (5G) radio access networks (RANs) are envisioned to encompass heterogeneous end devices supporting human-type data communication services and machine-type sensing and intelligent control services for Internet-of-Things (IoT) [1], [2]. Unlike traditional communication networks where downlink data traffic is predominant, a distinct feature of 5G is that a significant portion of the data traffic is carried in the uplink [3]. 5G data services incur more intensive uplink data traffic (e.g., live video and cloud backup) at mobile user devices (MUDs). Besides, 5G IoT applications are realized by supporting an enormous amount of machine-type devices (MTDs) uploading monitored data with stringent QoS requirements [4]. Using small-cells underlaying macro base stations (MBSs) is a potential solution to accommodate increasing traffic by exploiting spatial multiplexing [5]. However, this conventional multi-tier terrestrial RAN architecture may lead to low cell resource utilization while incurring extra infrastructure deployment costs due to imbalanced device distribution and network load [6]. Hence, it becomes inevitable to explore flexible network architecture with an agile, scalable, and cost-effective resource slicing framework.

Drones, a.k.a. unmanned aerial vehicles (UAVs), equipped with specific wireless transceivers can form drone-small-cells (DSCs) to assist 5G RANs, which advance in three aspects: 1) DSCs have a high probability of establishing short-distance line-of-sight (LoS) communication links with a high signal-to-noise ratio (SNR) [7]; 2) DSCs behaving as air relays between end devices and terrestrial base stations (BSs) can enhance the coverage of hotspot areas at the MBS edge in response to spatial and temporal traffic load unevenness [4]; 3) Benefiting from low transmit power and flexible maneuverability, DSCs can facilitate spectrum reuse to alleviate resource pressure via effective deployment. Despite the benefits, taking advantage of DSCs faces some challenging issues: 1) Integrating the MBS with DSCs leads to various network coverages and device-BS association patterns, which complicate resource management; 2) Interference fluctuations during the DSC deployment may reduce resource utilization; 3) Due to the unique characteristic of ground-to-DSC (G2D) and DSC-to-MBS (D2M) channels, it is difficult to determine an appropriate DSC flight altitude to balance the effective cell coverage and resource utilization.

Differentiated services' coexistence in 5G RANs requires guaranteeing performance isolation, preventing QoS violations, especially when a massive number of MTDs access radio channels [8]. Based on network function virtualization (NFV) [9] and software-defined networking (SDN) [10], radio resource slicing is an essential technological innovation in network resource management, supporting differentiated service

deliveries and achieving QoS isolation among services over a common underlying physical infrastructure [6]. With NFV, radio access and processing functions are decoupled from proprietary hardware, realized as software instances, managed centrally by an SDN-enabled virtualization controller [11]. Depending on SDN programmability, the controller determines the amount of radio resources allocated at each BS and further slices the resources into multiple isolated slices of varying sizes, which are customized to different service types with various characteristics and QoS requirements.

Researchers have conducted many studies on radio resource slicing for terrestrial IoT systems. A downlink radio resource slicing framework is proposed to maximize network utility while providing differentiated QoS guarantees [12]. The downlink spectrum utilization for vehicular networks is improved in [13], where spectrum slicing and transmit power control are jointly considered. Papa *et al.* investigate a downlink RAN slicing scheme to reduce resource usage while satisfying each slice's isolation, average rate, and delay requirements [14]. However, compared with the downlink control information dissemination, the performance bottleneck of IoT systems is usually the uplink rather than the downlink since uplink data transmissions are more resource consuming. An energy-efficient uplink resource allocation scheme is proposed in [15] for an MUD/MTD co-existence scenario to maximize bits-per-joule capacity subject to QoS constraints, where MTD gateways are employed for network architectural enhancement.

Considering the high spatio-temporal dynamics of the communication demands, deploying DSCs to assist terrestrial RANs by providing enhanced communication coverages with increased resource utilization has started to gain attention from the research community, where the joint optimization of resource slicing and DSC deployment is a challenging research issue. Most existing works consider downlink RAN scenarios to support mobile data service deliveries. Shi *et al.* propose a joint multi-DSC trajectory planning and resource allocation scheme to maximize the accumulative network throughput in a high-mobility scenario [16]. Yan *et al.* study drone access selection and resource allocation problems in drone-assisted communications for IoT and present a hierarchical Stackelberg game framework to balance network performance and service costs [17]. To support more diversified IoT services in a dynamic network environment, Lyu *et al.* propose an online control framework to slice the spectrum resource of space-air-ground integrated RANs, which maximizes system revenue and guarantees service queueing stabilization [18].

DSC-assisted uplink resource provisioning problems are investigated in literature for accommodating more machine-to-machine communication traffic. A drone-assisted cellular networking solution is proposed to deal with an increasing uplink MTD traffic volume with special traffic characteristics [19]. Wang *et al.* explore a resource allocation problem for uplink transmissions in space-air-ground integrated networks, where drones as relays upload the data from end devices to low earth orbit satellites [20]. Ali *et al.* investigate drone-assisted resource allocation to minimize the average queuing delay of traffic offloading in a drone/WiFi co-existence scenario [21]. However, some issues remain unsolved. First, for a given number of DSCs, how to deploy DSCs is non-trivial. Excessive DSC deployment can lead to resource wastage and even network performance degradation, especially in a low network traffic load condition. If the number of DSCs is insufficient, the QoS performance may not be satisfied. Second, most existing works omit the resource consumption of D2M communications and ignore the impact of DSC flying height on the effective cell coverage with proper device-BS association patterns. Third, radio resource slicing and sharing for finer-grained QoS differentiation, considering the distinctive drone channel and diversified data traffic features, need further investigation.

In this paper, we study a multi-DSC-assisted uplink radio resource slicing problem. By considering where and how many DSCs are placed, we propose a joint DSC placement and device association scheme, focusing on minimizing the total amount of consumed resources while satisfying differentiated QoS requirements of MUDs and MTDs. The main contributions are three folded:

- An interference-aware graph model is established to characterize multi-DSC three-dimension (3D) placement. Based on the model, we mathematically formulate the joint optimization problem of DSC placement and device-BS association for uplink resource slicing with DSC number and differentiated QoS constraints. We prove that the proposed problem is NP-hard and analyze the impact of DSC deployment on resource consumption;
- By dynamically selecting DSC placement positions based on their performance gains, we present a complexity-adjustable problem approximation incorporating an adaptive control policy for DSC flight altitude to balance uplink coverage and resource utilization. A lightweight approximation with a constant flying height is also provided that maximizes the communication coverages.
- In each approximation, the joint problem is transformed into a special weight clique problem, and a performance-aware clique algorithm is developed for problem-solving. Simulation results are presented to compare the proposed scheme's performance with various baseline methods for low and high device density cases, demonstrating that the proposed solution is superior to existing approaches.

The rest of this paper is organized as follows. The system model under consideration is presented in Section II. The problem formulation and analysis are given in Section III, and problem approximation and algorithm design are given in Section IV. We conduct performance evaluation in Section V, followed by the concluding remarks in Section VI. The main notations and variables are listed in Table I. Proof of the theorems and corollaries are given in appendices.

II. System Model

Consider a two-tier uplink heterogeneous RAN where a single MBS is deployed at the center of the network underlaid by multiple DSCs, as illustrated in Fig. 1. The MBS with high transmit power supports wide-area communication coverage. The DSCs with low transmit power and small coverage range are placed in the MBS's 3D coverage space. Due to the limited DSC coverage, end devices are associated with the MBS in most cases. When covered by a DSC, end devices can select to connect to either the homing DSC or the MBS, depending on network conditions. All packets generated by end devices are uploaded to reach the MBS via wireless propagation channels. There are three types of uplinks: ground-to-MBS (G2M) links, G2D links, and D2M links. A G2M link is a connection between an end device and the MBS, while a G2D link connects an end device and a DSC. Data received by a DSC is relayed through a D2M link to the MBS.

TABLE I
MAIN NOTATIONS AND VARIABLES

Symbols	Definition	Symbols	Definition
$a_{i,j,k}$	Association indicator for device i to a DSC at $v_{j,k}$	$v_{j,k}$	Candidate DSC deployment position $(x_j, y_j, z_{j,k})$
$g_{i,m}$	Channel gain from device i to the MBS	$w_{j,k}$	weight of $v_{j,k}$
$g_{i,j,k}$	Channel gain from device i to the DSC at $v_{j,k}$	W_1/W_2	Resource slice for the MBS and each DSC
$g_{j,k,m}$	Channel gain from a DSC at $v_{j,k}$ to the MBS	$\mathcal{X}_1/\mathcal{X}_2$	set of MUDs/MTDs
$h_{i,j}$	Horizontal distance between device i and $v_{j,k}$	Y_1/Y_2	Minimum rate requirement of an MUD/MTD
$h_{i,m}$	Horizontal distance between device i and the MBS	$z^{(\max)}$	Flight height that maximizes effective coverage
$\mathcal{H}_j/H_j$	Set/Number of candidate DSC height indexes	α	Number of available DSCs
$\mathcal{I}_{j,k}/I_{j,k}$	Set/Number of devices covered by a DSC at $v_{j,k}$	$\beta_{j,k}$	Performance gain for a DSC at $v_{j,k}$
$\mathcal{I}_{j,k,1}/I_{j,k,1}$	Set/Number of MUDs covered by a DSC at $v_{j,k}$	χ	Free space path-loss threshold for G2D links
$\mathcal{I}_{j,k,2}/I_{j,k,2}$	Set/Number of MTDs covered by a DSC at $v_{j,k}$	δ	Amount of resources for creating W_2
L_a/L_d	MTD/MUD packet size	$\delta_{j,k}$	Resource consumption of a DSC at $v_{j,k}$ from W_2
$\mathcal{N}/N$	Set/Number of x-y coordinate position indexes	$\delta_{j,k,1}/\delta_{j,k,2}$	Resource consumption of MUDs/MTDs from $\delta_{j,k}$
$p_i/p_{j,k}$	Transmit power on device i/the DSC at $v_{j,k}$	λ_a/λ_d	Poission/Periodic packet arrival rate
$q_{j,k}$	DSC deployment indicator for $v_{j,k}$	ξ	LoS probability threshold for G2D links
$\mathcal{Q}/Q$	Clique (DSC deployment plan)/Cardinality of $\mathcal{Q}$	ε	Bound of statistical delay violation probability
$r_{i,m}$	Spectrum efficiency at the MBS from device i	μ	Total uplink resource consumption
$r_{i,j,k}$	Spectrum efficiency at $v_{j,k}$ from device i	σ^2	Average background noise power
$r_{j,k,m}$	Spectral efficiency at the MBS from a DSC at $v_{j,k}$	ψ	Interference distance threshold
$R_{j,k}$	Effective coverage radius of a DSC at altitude $z_{j,k}$	ω	Amount of resources for creating W_1
$R^{(\max)}$	Maximum effective coverage radius of a DSC	ω_1/ω_2	Resource consumption of MUDs/MTDs from ω

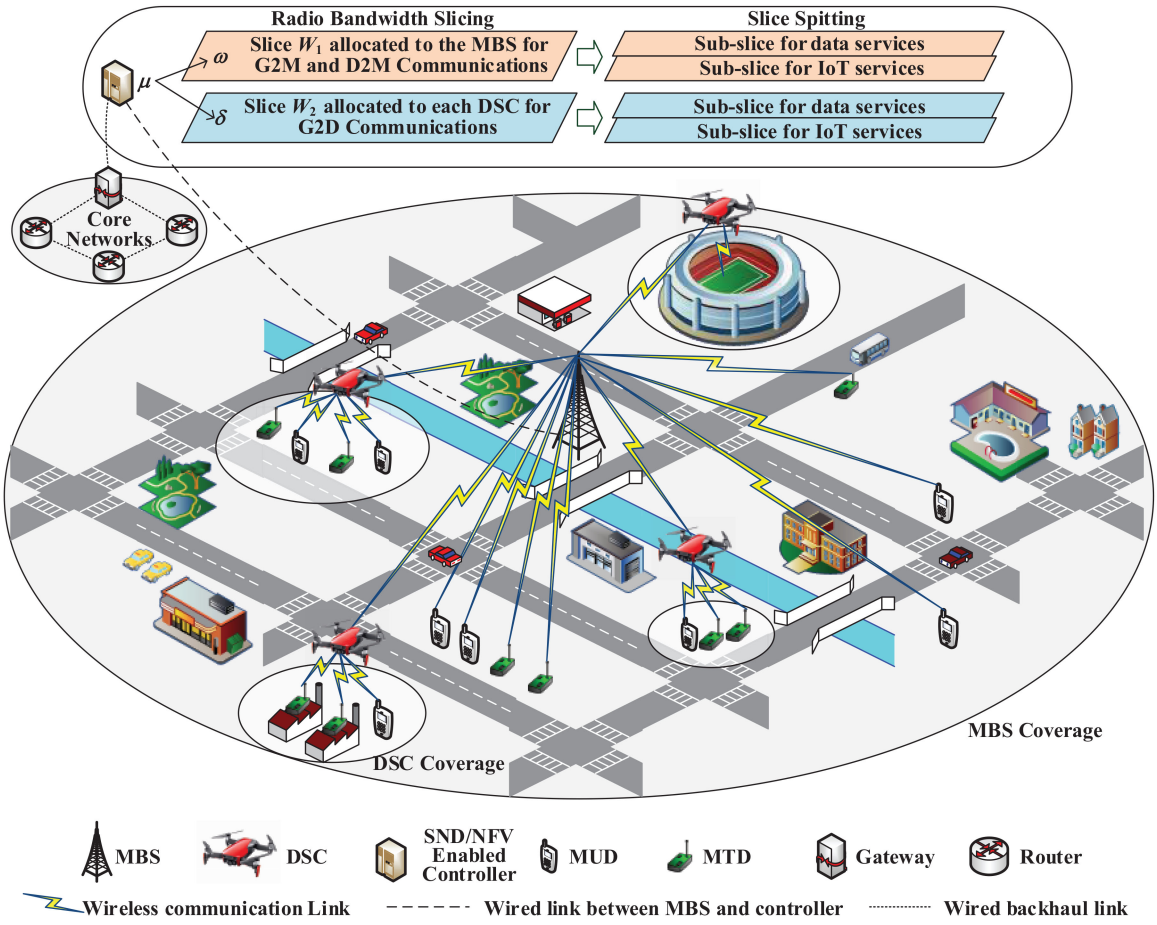

Fig. 1. Radio resource slicing and splitting in a multi-DSC-assisted RAN.

There are two types of end devices, MTDs and MUDs, with differentiated QoS requirements. The former has delay-sensitive machine-type traffic requiring high transmission reliability, and the latter generates data traffic requiring high throughput. Consider MUDs with low-to-moderate mobility and MTDs more or less stationary [22]. Denote MUD and MTD sets as $\mathcal{X}_1$ and $\mathcal{X}_2$. The position coordinates of end device $i \in \mathcal{X}_1 \cup \mathcal{X}_2$ and the MBS are denoted as $(x_i, y_i, 0)$ and (x_m, y_m, z_m). Let $v_{j,k} = (x_j, y_j, z_{j,k})$ represent a candidate DSC 3D deployment position, where $j \in \mathcal{N}$ is the index of the projection position on the x-y coordinate plane (i.e., $(x_j, y_j, 0)$), $k \in \mathcal{H}_j$ is the height index for j, and $\mathcal{N}$ and $\mathcal{H}_j$ denote the index sets with N and H_j being their cardinalities.

A. Communication Models

Consider a general path-loss model for G2M links, in which the signal power decays at a rate of $(\sqrt{h_{i,m}^2 + z_m^2})^\gamma$ over the propagation distance $\sqrt{h_{i,m}^2 + z_m^2}$, γ being the path-loss exponent, and $h_{i,m} = \sqrt{(x_i - x_m)^2 + (y_i - y_m)^2}$ being the horizontal distance between device i and the MBS. Denote p_i and σ^2 as the transmit power of end device i and average received background noise power, respectively. Based on Shannon's capacity formula, the spectral efficiency from end device i destined for the MBS is

$$r_{i,m} = \log_2\left(1 + \frac{p_i g_{i,m}}{\sigma^2}\right) \tag{1}$$

where $g_{i,m} = 10^{-\frac{(\sqrt{h_{i,m}^2 + z_m^2})^\gamma}{10}}$ is the channel gain from end device i to the MBS.

Different from G2M links, the channel characteristics of the G2D links rely on DSC flight altitude, elevation angle, and type of propagation environment with different LoS or non-line-of-sight (NLoS) occurrence probabilities [23]. Based on the aerial channel model proposed in [24], the LoS probability of the G2D link from device i to a DSC at $v_{j,k}$ is

$$F_{\mathrm{LoS}}(z_{j,k}, h_{i,j}) = \frac{1}{1 + o\exp\left(-b\arctan\left(\frac{z_{j,k}}{h_{i,j}}\right) - o\right)} \tag{2}$$

where $h_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ is the horizontal distance between device i and $v_{j,k}$, o and b are constant related to the urban environment. Based on [25], the additional path-loss (dB) of the G2D link from device i to $v_{j,k}$ for LoS reception, denoted by $\eta_{\mathrm{LoS}}(z_{j,k}, h_{i,j})$, and that for non-LoS reception, denoted by $\eta_{\mathrm{NLoS}}(z_{j,k}, h_{i,j})$, are given by

$$\begin{cases} \eta_{\mathrm{LoS}}(z_{j,k}, h_{i,j}) = \kappa_{\mathrm{LoS}} + \rho_{\mathrm{LoS}}\log_{10}\sqrt{h_{i,j}^2 + z_{j,k}^2} \\ \eta_{\mathrm{NLoS}}(z_{j,k}, h_{i,j}) = \kappa_{\mathrm{NLoS}} + \rho_{\mathrm{NLoS}}\log_{10}\sqrt{h_{i,j}^2 + z_{j,k}^2} \end{cases} \tag{3}$$

where κ_{LoS} (κ_{NLoS}) is path-loss at a reference distance under an LoS (NLoS) connection, and ρ_{LoS} (ρ_{NLoS}) is the path-loss exponent under an LoS (NLoS) connection. The free space path-loss from device i to $v_{j,k}$ is [25]

$$F_{\mathrm{PL}}(z_{j,k}, h_{i,j}) = 20\log_{10}\left(\frac{4\pi f\sqrt{z_{j,k}^2 + h_{i,j}^2}}{c}\right) \tag{4}$$

where f is the carrier frequency in Hz, c is the speed of light in m/s. Based on (2), (3), and (4), the average path-loss is given by [26]

$$\begin{aligned} \phi(z_{j,k}, h_{i,j}) = {} & F_{\mathrm{PL}}(z_{j,k}, h_{i,j}) \\ & + F_{\mathrm{LoS}}(z_{j,k}, h_{i,j})\eta_{\mathrm{LoS}}(z_{j,k}, h_{i,j}) \\ & + (1 - F_{\mathrm{LoS}}(z_{j,k}, h_{i,j}))\eta_{\mathrm{NLoS}}(z_{j,k}, h_{i,j}). \end{aligned} \tag{5}$$

Denote $\mathcal{Q}$ as the selected DSC deployment position set and $R_{j,k}$ as the effective coverage radius of a DSC hovering at $v_{j,k}$. The DSC experiences interference due to simultaneous uplink transmissions to other DSCs, whose deployment positions are contained in $\Theta_{j,'k'}(\psi) = \{v_{j,'k'} \in \mathcal{Q} \backslash \{v_{j,k}\} | h_{j,j'} - (R_{j,k} + R_{j,'k'}) < \psi\}$, where ψ represents an interference distance threshold [27] and $h_{j,j'} = \sqrt{(x_j - x_{j'})^2 + (y_j - y_{j'})^2}$ is the horizontal distance between $v_{j,k}$ and $v_{j,'k'}$. Thus, the spectral efficiency of the DSC at $v_{j,k}$ from end device i, similar to [28], is expressed as

$$r_{i,j,k} = \log_2\left(1 + \frac{p_i g_{i,j,k}}{\sum_{v_{j,'k'} \in \Theta_{j,'k'}(\psi)} p_i g_{i,j,'k'} + \sigma^2}\right) \tag{6}$$

where $g_{i,j,k} = 10^{-\frac{\phi(z_{j,k}, h_{i,j})}{10}}$ represents the channel gain from device i to the DSC at $v_{j,k}$.

Normally, the height of a DSC is higher than that of the MBS, and the link from the DSC to the MBS belongs to an LoS connection. Thus, the average path-loss of the D2M link from a DSC at $v_{j,k}$ to the MBS is expressed as

$$\begin{aligned} & \eta_{\mathrm{LoS}}(z_{j,k} - z_m, h_{j,m}) \\ & = \kappa_{\mathrm{LoS}} + \rho_{\mathrm{LoS}}\log_{10}\left(\sqrt{(h_{j,m})^2 + (z_{j,k} - z_m)^2}\right). \end{aligned} \tag{7}$$

If the traffic is delivered via a DSC at $v_{j,k}$, the spectral efficiency at the MBS is calculated as

$$r_{j,k,m} = \log_2\left(1 + \frac{p_{j,k} g_{j,k,m}}{\sigma^2}\right) \tag{8}$$

where $g_{j,k,m} = 10^{-\frac{\eta_{\mathrm{LoS}}(z_{j,k} - z_m, h_{j,m})}{10}}$ is the channel gain from the DSC to the MBS.

B. DSC Uplink Coverage Model

The effective uplink coverage radius of a DSC is constrained by both both LoS probability and free space path-loss, which satisfies [29], [30]

$$\begin{cases} F_{\mathrm{LoS}}(z_{j,k}, h_{i,j}) \geq \xi \\ F_{\mathrm{PL}}(z_{j,k}, h_{i,j}) \leq \chi \end{cases} \tag{9}$$

where ξ is the LoS probability threshold and χ is the free space path-loss threshold. The values of these two thresholds are determined by the minimum signal-to-noise ratio for signal decoding of DSCs. It is assumed that all DSCs have the same sensitivity. Substituting (2) and (4) into (9), we have

$$\begin{cases} h_{i,j} \leq \dfrac{z_{j,k}}{\tan\left(o - \frac{1}{b}\ln\frac{1-\xi}{o\xi}\right)} \triangleq J_{\mathrm{LoS}}(z_{j,k}) \\ h_{i,j} \leq \sqrt{\left(\dfrac{c10^{\frac{\chi}{20}}}{4\pi f}\right)^2 - z_{j,k}^2} \triangleq J_{\mathrm{PL}}(z_{j,k}) \end{cases} \tag{10}$$

and $z_{j,k} \in [0, \frac{c10^{\frac{\chi}{20}}}{4\pi f}]$. If the hovering height of a DSC is $z_{j,k}$, the effective coverage radius of the DSC is determined as

$$R_{j,k} = \min\{J_{\mathrm{LoS}}(z_{j,k}), J_{\mathrm{PL}}(z_{j,k})\}. \tag{11}$$

If $h_{i,j} \leq R_{j,k}$, end device i is under the effective coverage of the DSC at $v_{j,k}$.

Let z be a continuous variable with respect to DSC flight height. With an increase of z, $J_{\mathrm{LoS}}(z)$ monotonically increases and $J_{\mathrm{PL}}(z)$ monotonically decreases. Accordingly, the flight height that maximizes the effective coverage radius must be at the unique intersection of $J_{\mathrm{LoS}}(z)$ and $J_{\mathrm{PL}}(z)$, obtained when $J_{\mathrm{LoS}}(z) = J_{\mathrm{PL}}(z)$, given by

$$\begin{aligned} z^{(\max)} &= z^* = \arg(J_{\mathrm{LoS}}(z) = J_{\mathrm{PL}}(z)) \\ &= \frac{c10^{\frac{\chi}{20}} \tan\left(o - \frac{1}{b}\ln\frac{1-\xi}{o\xi}\right)}{4\pi f\sqrt{1+\tan\left(o - \frac{1}{b}\ln\frac{1-\xi}{o\xi}\right)^2}}. \end{aligned} \tag{12}$$

Due to the different monotonicity of $J_{\mathrm{LoS}}(\cdot)$ and $J_{\mathrm{PL}}(\cdot)$, the largest effective coverage radius when the DSC is at altitude $z^{(\max)}$ is expressed as

$$R^{(\max)} = J_{\mathrm{LoS}}(z^{(\max)}) = J_{\mathrm{PL}}(z^{(\max)}). \tag{13}$$

C. Traffic Model

An uplink transmission queue is considered at each end device. The link-layer packetized traffic [31] is used to model packet arrivals of human-type and machine-type traffic for QoS characterization. It is assumed that human-type packets arrive at the queue in an MUD periodically with an average rate of λ_d packet/s and the size of L_d bits. The MUD's uplink transmission throughput requirement can be met if the allocated transmission rate of each link reaches $Y_1 = \lambda_d L_d$. Since machine-type packets arrive at an MTD in an event-driven manner with a low packet arrival rate and a small packet size [22], the transmission delay can only be satisfied in a probabilistic way. As suggested in [32], machine-type packet arrivals at the queue in an MTD are modeled as a Poisson process with an average rate of λ_a packet/s and the size of L_a bits. Due to the randomness of packet arrival, the effective bandwidth theory is applied for each MTD to calculate the minimum transmission rate, Y_2, to provide a probabilistic guarantee of a packet transmission delay bound [12], [15], [31]. The effective bandwidth for a machine-type traffic source, with a QoS exponent, τ, is derived as [33]

$$B(\tau) = \frac{\lambda_a(\exp(\tau) - 1)}{\tau}. \tag{14}$$

Based on the large deviation theory [34], the probability of uplink transmission delay, D, for an MTD packet on the G2D/G2M/D2M link exceeding a delay bound, $D^{(\max)}$, is expressed as

$$\Pr(D \geq D^{(\max)}) \approx \exp(-\frac{Y_2 \tau D^{(\max)}}{L_a}) \tag{15}$$

where $\frac{Y_2}{L_a}$ is the achievable rate at an target BS. Let ε denote a delay bound violation probability threshold, which is set to $\exp(-\frac{Y_2 \tau D^{(\max)}}{L_a})$. Then, we have

$$Y_2 = -\frac{L_a \log \varepsilon}{D^{(\max)}\tau} \tag{16}$$

where τ is obtained when $Y_2 = B(\tau)$, calculated as

$$\tau = \log\left(1 - \frac{L_a \log \varepsilon}{\lambda_a D^{(\max)}}\right). \tag{17}$$

D. Resource Slicing Framework

Through SDN and NFV, the physical radio resources from heterogeneous BSs are abstracted as a centralized virtual radio resource pool. As shown in Fig. 1, an SDN/NFV-enabled controller, connected to the MBS, centrally manages the slicing of the virtualized uplink radio resources, the amount of which is denoted by μ, among all the heterogeneous BSs under consideration to further create resource slices reserved to different services for customized QoS provisioning. The amount of radio resources allocated to each slice needs to be adjusted in response to both network conditions (including end device and DSC locations, device-DSC association patterns, and interference levels) and service-level characteristics (i.e., traffic statistics and QoS constraints). Two levels of resource partitioning are considered.

At the network level, the determined bandwidth resources are physically partitioned into two bandwidth slices W_1 and W_2, the amounts of which are denoted as ω and δ (with $\mu = \omega + \delta$). They are mutually orthogonal to avoid inter-slice interference. The W_1 slice is reserved to the MBS to support G2M and D2M communications, where DSCs in D2M communications can also be treated as gateways/relays associated with the MBS from end devices. The W_2 slice is reused by each DSC for G2D communications by keeping interference distances among the DSCs. For a DSC placed at $v_{j,k}$, the actual amount of resources allocated from W_2, denoted by $\delta_{j,k}$, cannot be larger than δ.

At the service level, the bandwidth slice allocated to each BS is further partitioned into two sub-slices for human-type and machine-type communication services. The amounts of resources split from ω for the two types of services are denoted as ω_1 and ω_2 (with $\omega = \sum_{s\in\{1,2\}} \omega_s$). Similarly, the amounts of resources separated from $\delta_{j,k}$ are denoted as $\delta_{j,k,1}$ and $\delta_{j,k,2}$ (with $\delta_{j,k} = \sum_{s\in\{1,2\}} \delta_{j,k,s}$).

III. Optimal Uplink Resource Slicing

In this section, a graph model are presented to characterize the multi-DSC placement. Based on the model, we mathematically formulate the optimal resource slicing problem.

A. Graph Model for Multi-DSC Placement

The position variation (involving plane projection and flight altitude) of any DSC may lead to the interference fluctuation of slice W_2 and affect the placement of other DSCs. We construct an interference graph model to coordinate multi-DSC 3D placement for interference control.

Let $G(V, E)$ denote a graph, in which each vertex represents a candidate DSC deployment position. The vertex set is expressed as

$$V(G) = \{v_{j,k} | j \in \mathcal{N}, k \in \mathcal{H}_j\}. \tag{18}$$

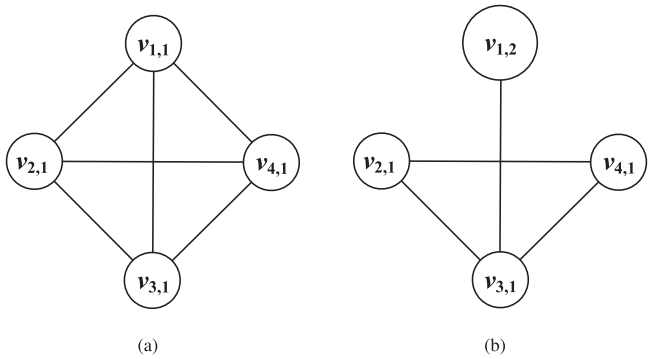

Fig. 2. Comparison of different interference graphs. (a) Case-1. (b) Case-2.

Denote $\mathcal{I}_{j,k,1} = \{i \in \mathcal{X}_1 | h_{i,j} \leq R_{j,k}\}$ and $\mathcal{I}_{j,k,2} = \{i \in \mathcal{X}_2 | h_{i,j} \leq R_{j,k}\}$ as the groups of MUDs and MTDs under the coverage of a DSC at $v_{j,k}$. Let $\mathcal{I}_{j,k}$ be $\mathcal{I}_{j,k,1} \cup \mathcal{I}_{j,k,2}$.

From (6), if the distance between any two DSCs at $v_{j,k}$ and $v_{j,'k'}$ satisfies

$$h_{j,j'} - (R_{j,k} + R_{j,'k'}) \geq \psi \tag{19}$$

the interference of each DSC from the uplink transmission to other DSCs can be controlled. Under (19), the distance between devices $i \in \mathcal{I}_{j,k}$ and $i' \in \mathcal{I}_{j,'k'}$ must be larger than ψ. For any two different vertices $(v_{j,k}, v_{j,'k'}) \subseteq V(G)$, there exists an edge $(v_{j,k}, v_{j,'k'})$ if (19) is met. The edge set of the graph is expressed as

$$\begin{aligned} E(G) = \{(v_{j,k}, v_{j,'k'}) | h_{j,j'} - (R_{j,k} + R_{j,'k'}) \geq \psi, \\ v_{j,k}, v_{j,'k'} \in V(G)\}. \end{aligned} \tag{20}$$

An example is used to illustrate the graph model. The graph shown in Fig. 2(a) consists of three vertices $v_{1,1}$, $v_{2,1}$, $v_{3,1}$, and $v_{4,1}$. Each circle represents the effective coverage area of a DSC deployed at the vertex. Taking the circle surrounding $v_{1,1}$ as an example, the radius of the circle is $R_{1,1}$. It is assumed that $h_{1,2} - (R_{1,1} + R_{2,1}) = h_{2,3} - (R_{2,1} + R_{3,1}) = h_{3,4} - (R_{3,1} + R_{4,1}) = h_{4,1} - (R_{4,1} + R_{1,1}) = \psi$ and $h_{1,3} - (R_{1,1} + R_{3,1}) = h_{2,4} - (R_{2,1} + R_{2,1}) \geq \psi$. A connecting edge exists between each pair of vertices because each pair satisfies (19). Assuming that $v_{1,1}$ is replaced by $v_{1,2}$ and $R_{1,1} < R_{2,1}$, we have $h_{1,2} - (R_{1,2} + R_{2,1}) = h_{1,4} - (R_{1,2} + R_{4,1}) < \psi$. Under (19), there is no edge existed between $v_{1,2}$ and $v_{2,1}$, and $v_{1,2}$ and $v_{4,1}$, as shown in Fig. 2(b).

Based on the graph model, we express the selected DSC deployment position set, $\mathcal{Q} = \{v_{j_1,k_1}, v_{j_2,k_2}, \ldots, v_{j_Q,k_Q}\}$, as a clique (complete subgraph) in $G(V, E)$, representing a vertex subset in which every two vertices are adjacent, with Q being the cardinality of $\mathcal{Q}$. For example, Fig. 2(a) has four cliques with three vertices, i.e., $\{v_{1,1}, v_{2,1}, v_{3,1}\}$, $\{v_{1,1}, v_{3,1}, v_{4,1}\}$, $\{v_{1,1}, v_{2,1}, v_{4,1}\}$, and $\{v_{2,1}, v_{3,1}, v_{4,1}\}$, while Fig. 2(b) contains a single clique with three vertices, i.e., $\{v_{2,1}, v_{3,1}, v_{4,1}\}$.

We use $q_{j,k}$ to indicate whether $v_{j,k}$ is included in $\mathcal{Q}$, given by

$$q_{j,k} = \begin{cases} 1, \text{if } v_{j,k} \in \mathcal{Q} \\ 0, \text{otherwise.} \end{cases} \tag{21}$$

We choose $v_{j,k}$ to deploy a DSC if and only if $q_{j,k} = 1$. Each clique corresponds to a candidate DSC placement plan.

B. Resource Partitioning

Resource partitioning depends on DSC placement and device association selection. Let indication variable $a_{i,j,k}$ represent the association pattern between device $i \in \mathcal{I}_{j,k}$ and the DSC placed at $v_{j,k}$. If device i establishes a connection with the DSC, $a_{i,j,k}$ is set to 1; otherwise, set to 0. If device i is associated with the MBS, the amount of resources required by the device is $\frac{Y_s}{r_{i,m}}$. If $a_{i,j,k} = 1$, the amount of resources required by the device to submit data through the G2D link and the D2M link are $\frac{Y_s}{r_{i,j,k}}$ and $\frac{Y_s}{r_{j,k,m}}$, respectively.

For the quantification of ω, we divide the end devices in $\mathcal{X}_1 \cup \mathcal{X}_2$ into two groups, located outside and inside DSC coverage areas. In the former case, since the corresponding end devices can only connect to the MBS, the amount of resources required from ω is calculated as

$$\sum_{s\in\{1,2\}} \sum_{i\in\mathcal{X}_s} \frac{Y_s}{r_{i,m}} - \sum_{v_{j,k}\in V(G)} q_{j,k} \sum_{s\in\{1,2\}} \sum_{i\in\mathcal{I}_{j,k,s}} \frac{Y_s}{r_{i,m}}. \tag{22}$$

In the latter case, the association pattern for device $i \in \mathcal{I}_{j,k}$ is determined by $\sum_{v_{j,k}\in V(G)} q_{j,k} a_{i,j,k}$, which equals to 0 if it is associated with the MBS and 1 if it connects with a DSC. The amount of resources required by the end devices associated with the MBS for G2M communications is calculated as

$$\sum_{v_{j,k}\in V(G)} q_{j,k} \sum_{s\in\{1,2\}} \sum_{i\in\mathcal{I}_{j,k,s}} (1 - a_{i,j,k}) \frac{Y_s}{r_{i,m}} \tag{23}$$

and that consumed by the end devices associated with DSCs for D2M communications is calculated as

$$\sum_{v_{j,k}\in V(G)} q_{j,k} \sum_{s\in\{1,2\}} \sum_{i\in\mathcal{I}_{j,k,s}} a_{i,j,k} \frac{Y_s}{r_{j,k,m}}. \tag{24}$$

Summing (22), (23), and (24), we have

$$\begin{aligned} \omega = \sum_{s\in\{1,2\}} \omega_s = \sum_{s\in\{1,2\}} \sum_{i\in\mathcal{X}_s} \frac{Y_s}{r_{i,m}} \\ + \sum_{v_{j,k}\in V(G)} q_{j,k} \sum_{s\in\{1,2\}} \sum_{i\in\mathcal{I}_{j,k,s}} a_{i,j,k} \left(\frac{Y_s}{r_{j,k,m}} - \frac{Y_s}{r_{i,m}} \right). \end{aligned} \tag{25}$$

For a DSC is deployed at $v_{j,k}$ ($q_{j,k} = 1$), we have

$$\delta_{j,k,s} = \sum_{i\in\mathcal{I}_{j,k,s}} a_{i,j,k} \frac{Y_s}{r_{i,j,k}}, s \in \{1, 2\}. \tag{26}$$

Since slice W_2 is reused among DSCs, the amount of resources allocated to this slice is determined by the DSC consuming the largest amount of resources. The position of such DSC is contained in $\mathcal{Q}$, whose indexes are expressed as

$$j^*, k^* = \arg \max_{v_{j,k}\in\mathcal{Q}} \sum_{s\in\{1,2\}} \sum_{i\in\mathcal{I}_{j,k,s}} a_{i,j,k} \frac{Y_s}{r_{i,j,k}}. \tag{27}$$

Combining (26) and (27), we have

$$\delta = \sum_{s\in\{1,2\}} \delta_{j^*,k^*,s} = \sum_{s\in\{1,2\}} \sum_{i\in\mathcal{I}_{j^*,k^*,s}} a_{i,j^*,k^*} \frac{Y_s}{r_{i,j^*,k^*}}. \tag{28}$$

C. Problem Formulation

In the proposed resource slicing framework, a challenging issue is to determine where and how many DSCs are placed, as well as device-BS association patterns to minimize the total uplink resource consumption, which is the the summation of (25) and (28), expressed as

$$\mu = \omega + \delta = \sum_{s\in\{1,2\}} \sum_{i\in\mathcal{X}_s} \frac{Y_s}{r_{i,m}} + \sum_{v_{j,k}\in V(G)} q_{j,k} w_{j,k} \tag{29}$$

where

$$w_{j,k} = \begin{cases} \sum\limits_{i\in\mathcal{I}_{j,k}} a_{i,j,k} \left(\dfrac{Y_s}{r_{i,j,k}} + \dfrac{Y_s}{r_{j,k,m}} - \dfrac{Y_s}{r_{i,m}} \right), \text{ if } v_{j,k} = v_{j^*,k^*} \\ \sum\limits_{i\in\mathcal{I}_{j,k}} a_{i,j,k} \left(\dfrac{Y_s}{r_{j,k,m}} - \dfrac{Y_s}{r_{i,m}} \right), \text{ if } v_{j,k} \in V(G)\backslash\{v_{j^*,k^*}\}. \end{cases} \tag{30}$$

$w_{j,k}$ is equivalent to the weight of vertex $v_{j,k}$.

Because the part of $\sum_{s\in\{1,2\}} \sum_{i\in\mathcal{X}_s} \frac{Y_s}{r_{i,m}}$ in (29) is constant given $\mathcal{X}_1$ and $\mathcal{X}_2$, the objective is expressed as minimizing $\sum_{v_{j,k}\in V(G)} q_{j,k} w_{j,k}$. Then, the optimal uplink radio resource slicing is formulated as an integer linear program in $\mathcal{P}1$.

$$\mathcal{P}1: \underset{\mathcal{Q},\mathcal{A}}{\text{Minimize}}: \sum_{v_{j,k}\in V(G)} q_{j,k} w_{j,k}$$

$$\text{s.t.} \begin{cases} \sum\limits_{v_{j,k}\in V(G)} q_{j,k} \le \alpha & \text{(31a)} \\ q_{j,k} + q_{j,'k'} \le 1, \forall (v_{j,k}, v_{j,'k'}) \in \overline{E(G)} & \text{(31b)} \\ q_{j,k} \in \{0,1\}, \forall v_{j,k} \in V(G) & \text{(31c)} \\ a_{i,j,k} \in \{0,1\}, \forall i \in \mathcal{I}_{j,k}, v_{j,k} \in V(G). & \text{(31d)} \end{cases}$$

In the problem formulation, $\mathcal{A} = \{\mathcal{A}_{j,k} | i \in \mathcal{I}_{j,k}, v_{j,k} \in \mathcal{Q}\}$ denotes the set of device-DSC association patterns. In (31a), α represents the number of available DSCs, which determines the upper bound of Q. Actually, due to the unevenness of device distribution, the optimal number of DSCs deployed may be less than α. Taking Fig. 2 as an example, in some scenarios, when $\alpha = 3$, the optimal DSC deployment plan may be $\{v_{1,2}\}$ instead of $\{v_{1,1}, v_{2,1}, v_{3,1}, v_{4,1}\}$ or $\{v_{2,1}, v_{3,1}, v_{4,1}\}$. Constraint (31b) ensures that $v_{i,j}$ and $v_{j,'k'}$ cannot be included in the same clique if $(v_{j,k}, v_{j,'k'}) \notin E(G)$.

Problem $\mathcal{P}1$ is a joint clique partition and device association problem. In the clique partition, the clique size (i.e., the number of DSCs) and vertex weight need to be determined, distinguishing it from the traditional clique problem with preset constant clique size and vertex weights.

Given DSC deployment plan $\mathcal{Q}$, the objective of Problem $\mathcal{P}1$ can be transformed as

$$\varphi(\mathcal{Q}) = \sum_{v_{j,k}\in\mathcal{Q}} w_{j,k}. \tag{32}$$

The lower bound of (32) is expressed as

$$\varphi^{(\min)}(\mathcal{Q}) = \sum_{v_{j,k}\in\mathcal{Q}} w_{j,k}^{(\min)}. \tag{33}$$

where $w_{j,k}^{(\min)} = \min_{s\in\{1,2\}, i\in\mathcal{I}_{j,k,s}, a_{i,j,k}\in\{0,1\}} w_{j,k}$ represents the lower bound of $w_{j,k}$. Then, we have Proposition 1.

Proposition 1: Under (34), shown at the bottom of this page, we have $\varphi(\mathcal{Q}) = \varphi^{(\min)}(\mathcal{Q})$.

The Proof of Proposition 1 is given in Appendix VI-A.

Proposition 1 implies that (34) is determined by $\mathcal{Q}$. Further, we have Proposition 2.

Proposition 2: If $\mathcal{Q}' \subseteq \mathcal{Q}$, then $\varphi^{(\min)}(\mathcal{Q}) \le \varphi^{(\min)}(\mathcal{Q}')$.

The Proof of Proposition 2 is given in Appendix VI-B.

Proposition 2 indicates that under (34), supplementing new DSCs (if possible) will not increase resource consumption. In the case of $\mathcal{Q} = \emptyset$, the resource consumption, calculated as $\varphi(\mathcal{Q}) = \sum_{s\in\{1,2\}} \sum_{i\in\mathcal{X}_s} \frac{Y_s}{r_{i,m}}$, is at its maximum.

Theorem 1: $\mathcal{P}1$ is NP-hard.

The Proof of Theorem 1 is given in Appendix VI-C.

From Theorem 1, the computational complexity of finding the optimal solution of $\mathcal{P}1$ exponentially increases with $|V(G)| = \sum_{j\in\mathcal{N}} H_j$. Although the optimal solution can be obtained by solving the given integer program, the computational complexity becomes high when $|V(G)|$ is large. Hence, it is necessary to design an approximation mechanism for a suboptimal solution.

IV. Clique Based Solution

We first explore vertex filtering for complexity-adjustable multi-DSC deployment. Then, we establish problem approximations for DSC flight height adaptive and fixed cases. Finally, an upgraded clique algorithm is proposed for problem-sloving.

A. Flight Altitude Selection

The association indicator in (34) is approximated as in (35) to reduce complexity.

$$\tilde{a}_{i,j,k} = \begin{cases} 1, \text{if} \dfrac{1}{r_{i,j,k}} + \dfrac{1}{r_{j,k,m}} \le \dfrac{1}{r_{i,m}} \\ 0, \text{otherwise}. \end{cases} \tag{35}$$

If a DSC is deployed at $v_{j,k}$, the performance gain obtained by the DSC is defined as

$$\begin{aligned} \beta_{j,k} &\triangleq \sum_{i\in\mathcal{I}_{j,k}} \tilde{a}_{i,j,k} \left(\frac{1}{r_{i,m}} - \left(\frac{1}{r_{i,j,k}} + \frac{1}{r_{j,k,m}} \right) \right) \\ &= \sum_{i\in\mathcal{I}_{j,k}} \left| \frac{1}{r_{i,m}} - \left(\frac{1}{r_{i,j,k}} + \frac{1}{r_{j,k,m}} \right) \right| \end{aligned} \tag{36}$$

which reflects the amount of resources required for devices associated with the DSC at $v_{j,k}$ to achieve the same transmission rate. The larger the value of $\beta_{j,k}$, the higher the performance gain. Given plane index j, the height index paired with it is

$$a_{i,j,k} = \begin{cases} 1, \left(\dfrac{1}{r_{i,j,k}} + \dfrac{1}{r_{j,k,m}} \le \dfrac{1}{r_{i,m}}, \text{if } v_{j,k} = v_{j^*,k^*} \right) \text{or} \left(\dfrac{1}{r_{j,k,m}} \le \dfrac{1}{r_{i,m}}, \text{if } v_{j,k} \in \mathcal{Q}\backslash\{v_{j^*,k^*}\} \right) \\ 0, \text{otherwise} \end{cases} \tag{34}$$

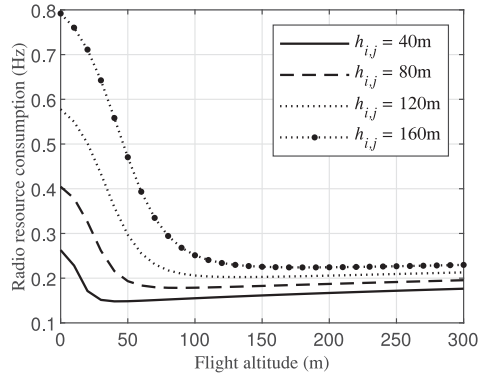

Fig. 3. Radio resource consumption.

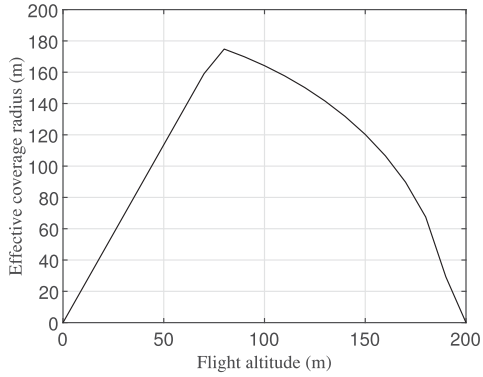

Fig. 4. Effective DSC coverage radius.

selected to maximize the DSC's performance gain, given by

$$k_j^* = \arg\max_{k\in\mathcal{H}_j} \beta_{j,k}. \tag{37}$$

Now, we discuss how to efficiently find k_j^* from $\mathcal{H}_j$ through simulation, in which parameters are configured with ξ and χ of 89 dB and 0.5, and o and b of 9.61 and 0.16. If device $i \in \mathcal{I}_{j,k,s}$ connects to a DSC at $v_{j,k}$, the amount of resource consumption of the device is $Y_s(\frac{1}{r_{i,j,k}} + \frac{1}{r_{j,k,m}})$. Fig. 3 shows the value of $\frac{1}{r_{i,j,k}}+\frac{1}{r_{j,k,m}}$ as a function of $z_{j,k}$, where $h_{i,j}$ is set to 40 m, 80 m, 120 m, and 160 m, respectively. Given plane position index j, there exists a unique flight altitude that consumes the least resources. The altitude's index is expressed as

$$k_{i,j}^* = \arg\min_{k\in\mathcal{H}_j} \left(\frac{1}{r_{i,j,k}}+\frac{1}{r_{j,k,m}}\right), i \in \mathcal{I}_{j,k_{i,j}^*}. \tag{38}$$

Fig. 4 shows that the effective coverage radius of the DSC at $v_{j,k}$, $R_{j,k}$, increases first and decreases with the increase of flight altitude, demonstrating a unique elevation, $z^{(\max)}$, that produces the maximum coverage radius, $R^{(\max)}$. Based on (12) and (13), we have $z^{(\max)}$=77.4 m and $R^{(\max)}$=175.9 m, respectively. Given the environmental parameters, the DSC's effective coverage is determined by flying height.

Regarding the DSC flight height, we can summarize from Fig. 3 and Fig. 4 that 1) when a DSC associates with an end device, the DSC's flight height that minimizes the end device's resource consumption is unique, and 2) the DSC's flight height that maximizes effective coverage is also unique. Inspired by the properties, we eliminate suboptimal flight height indexes by

Algorithm 1: Filter_altitude ($\mathcal{H}_j$).

1 **for** $l \leftarrow 2$ **to** H_j **do**
2 **if** $R_{j,k_{l-1}} < R_{j,k_l}$ &&$z_{j,k_l} \le \min_{i\in\mathcal{I}_{j,k_{l-1}}} z_{j,k_{i,j}^*}$ **then**
3 $\mathcal{H}_j \leftarrow \mathcal{H}_j \backslash \{k_{l-1}\}$;
4 **else if** $R_{j,k_{l-1}} > R_{j,k_l}$ &&$z_{j,k_{l-1}} \ge \max_{i\in\mathcal{I}_{j,k_l}} z_{j,k_{i,j}^*}$ **then**
5 $\mathcal{H}_j \leftarrow \mathcal{H}_j \backslash \{k_l, k_{l+1}, ..., k_{H_j}\}$;
6 Break;

Theorems 2 and 3. It is assumed that all indexes are sorted in an ascending order of heights in the theorems.

Theorem 2: If $R_{j,k_{l-1}} < R_{j,k_l}$ and $z_{j,k_l} \le \min_{i\in\mathcal{I}_{j,k_{l-1}}} z_j$, $k_{i,j}^*$, then $k_j^* \neq k_{l-1}$.

The Proof of Theorem 2 is given in Appendix VI-D.

Theorem 3: If $R_{j,k_{l-1}} > R_{j,k_l}$ and $z_{j,k_{l-1}} \ge \max_{i\in\mathcal{I}_{j,k_l}} z_j$, $k_{i,j}^*$, then $k_j^* \notin \{k_l, k_{l+1}, \ldots, k_{H_j}\}$.

Proof of Theorem 3 is given in Appendix VI-E.

A fast vertex screening algorithm, summarized in Algorithm 1, is developed to preclude suboptimal height indexes from $\mathcal{H}_j$. Initially, $l \in [1, H_j]$ is set to be 2. For a given l, k_{l-1} is removed (line 3) if k_l satisfies the condition of Theorem 2 (line 2); k_l and its subsequent are excluded (line 5) if k_l meets the condition of Theorem 3 (line 4).

B. Dynamic Vertex Filtering

Due to the uneven end device distribution, some candidate deployment positions' performance gains are meager. A natural step is to exclude the vertices of low performance gain from $V(G)$ based on (36), which reduces the decision space for the DSC deployment while hardly sacrificing performance. If β_{j,k_j^*} is large, v_{j,k_j^*} is highly likely to be included in the output and should be preserved in the graph. If β_{j,k_j^*} is small, the negative effect of excluding v_{j,k_j^*} on performance is negligible. In the worst case where no end device connects to the DSC at v_{j,k_j^*} (with $\beta_{j,k_j^*} = 0$), v_{j,k_j^*} should be excluded. Next, we explain how to implement elastic vertex filtering.

Let ζ denote a threshold of performance gain. If $\beta_{j,k_j^*} \le \zeta$, v_{j,k_j^*} is removed from $V(G)$. The principle of setting the threshold is to ensure that the probability of $\beta_{j,k_j^*} > \zeta$ is not greater than θ, which is given by

$$\Pr(\beta_{j,k_j^*} > \zeta) \le \theta. \tag{39}$$

By applying one-sided Chebyshevs inequality, we have

$$\Pr(\beta_{j,k_j^*} > \zeta) \le \frac{(\Delta\beta_{j,k_j^*})^2}{(\Delta\beta_{j,k_j^*})^2 + (\zeta - \overline{\beta_{j,k_j^*}})^2} \tag{40}$$

where $\overline{\beta_{j,k_j^*}}$ is the mean of β_{j,k_j^*} and $\Delta\beta_{j,k_j^*}$ is the variance of β_{j,k_j^*}. The probabilistic guarantee in (39) leads to

$$\frac{(\Delta\beta_{j,k_j^*})^2}{(\Delta\beta_{j,k_j^*})^2 + (\zeta - \overline{\beta_{j,k_j^*}})^2} = \theta \tag{41}$$

from which we have

$$\zeta = \overline{\beta_{j,k_j^*}} + \Delta\beta_{j,k_j^*}\sqrt{\frac{1-\theta}{\theta}}. \tag{42}$$

This policy can guarantee high-quality vertices while filtering vertices by adjusting θ.

C. Problem Approximation

Incorporating the proposed flight altitude control and vertex filtering, we present a general problem approximation. In the approximation, the vertex set is reduced to

$$V'(G) = \{v_{j,k_j^*} | j \in \mathcal{N}, \beta_{j,k_j^*} \leq \zeta\} \tag{43}$$

where ζ is set as in (42). The flight height index of each vertex in (43) is determined by the plane position, which reduces the size of the decision space. The corresponding edge set is

$$E'(G) = \{(v_{j,k_j^*}, v_{j,'k_{j'}^*}) | h_{j,j'} \geq R_{j,k_j^*} + R_{j,'k_{j'}^*} + \psi, \\ v_{j,k_j^*}, v_{j,'k_{j'}^*} \in V'(G)\}. \tag{44}$$

Based on (43) and (44), $\mathcal{P}1$ is approximated as clique problem $\mathcal{P}2$ related to DSC deployment.

$$\mathcal{P}2: \underset{\mathcal{Q}}{\text{Minimize}}: \sum_{v_{j,k_j^*} \in V'(G)} q_{j,k_j^*} \tilde{w}_{j,k_j^*}$$

$$\text{s.t.} \begin{cases} \sum\limits_{v_{j,k_j^*} \in V'(G)} q_{j,k_j^*} \leq \alpha & \text{(45a)} \\ q_{j,k_j^*} + q_{j,'k_{j'}^*} \leq 1, \forall (v_{j,k_j^*}, v_{j,'k_{j'}^*}) \in \overline{E'(G)} & \text{(45b)} \\ q_{j,k_j^*} \in \{0,1\}, \forall v_{j,k_j^*} \in V'(G) & \text{(45c)} \end{cases}$$

In the minimization term, $\tilde{w}_{j,k_j^*}$ represents the value of w_{j,k_j^*} where a_{i,j,k_j^*} is replaced by $\tilde{a}_{i,j,k_j^*}$. The essence of problem $\mathcal{P}1$ is to find plane index set $\{j_1, j_2, \ldots, j_Q\}$ to minimize $\varphi(\mathcal{Q})$. Let $|\cdot|$ denote the cardinality of a set. Because $|V'(G)| \leq |V(G)|$, the complexity of solving $\mathcal{P}2$ is lower than that of solving $\mathcal{P}1$.

Further, we consider a lightweight approximation as an alternative, where the flight altitude of all DSCs is fixed to $z^{(\max)}$. In this case, the performance gain of deploying a DSC at $v_j = (x_j, y_j, z^{(\max)})$ is simplified from (36) to $\sum_{i \in \mathcal{I}_j} \tilde{a}_{i,j}$, where $\mathcal{I}_j$ and $\tilde{a}_{i,j}$ represent $\mathcal{I}_{j,k}$ and $\tilde{a}_{i,j,k}$ when $z_{j,k} = z^{(\max)}$. In this approximation, the vertex index set is compressed to

$$V''(G) = \{v_j | j \in \mathcal{N}, \sum\nolimits_{i \in \mathcal{I}_j} \tilde{a}_{i,j} \leq \zeta'\} \tag{46}$$

where the role of ζ' is identical to ζ in (43). The corresponding edge index set is

$$E''(G) = \{(v_j, v_{j'}) | h_{j,j'} \geq 2R^{(\max)} + \psi, v_j, v_{j'} \in V''(G)\}. \tag{47}$$

Let q_j and $\tilde{w}_j$ denote $q_{j,k}$ and $\tilde{w}_{j,k}$ when $z_{j,k}$ is set to $z^{(\max)}$. $\mathcal{P}2$ is further approximated as $\mathcal{P}3$.

$$\mathcal{P}3: \underset{\mathcal{Q}}{\text{Minimize}}: \sum_{v_j \in V''(G)} q_j \tilde{w}_j$$

$$\text{s.t.} \begin{cases} \sum\limits_{v_j \in V''(G)} q_j \leq \alpha & \text{(48a)} \\ q_j + q_{j'} \leq 1, \forall (v_j, v_{j'}) \in \overline{E''(G)} & \text{(48b)} \\ q_j \in \{0,1\}, \forall v_j \in V''(G) & \text{(48c)} \end{cases}$$

When a DSC hovers to height $z^{(\max)}$, the probability of adjacent DSCs staying within the DSC's interference range becomes the highest. Even with $|V''(G)| = |V'(G)|$, we have $|E''(G)| \leq |E'(G)|$, indicating that the number of cliques in $G(V,'' E'')$ is less than that in $G(V,' E')$, with reduced DSC deployment complexity, along with suboptimal performance.

Problems $\mathcal{P}1$, $\mathcal{P}2$, and $\mathcal{P}3$ are different from the existing maximum weight clique problem [35] or maximum weight α-clique problem (the output is a clique of α vertices with the largest sum of all vertex weights) [36]. Since the number of DSCs to be placed is uncertain ($Q \leq \alpha$) and vertex weight $w_{j,k}$ is determined by $\mathcal{Q}$ instead of each vertex, it is inappropriate to find $\mathcal{Q}$ directly with existing clique algorithms. Hence, we need to design an upgraded clique algorithm to determine the DSC deployment strategy.

D. Algorithm Design

A performance-aware algorithm is designed to solve the size-constrained clique problem. Given a graph and the number of available DSCs, the algorithm outputs a clique, determining the amount of DSCs to be deployed and the corresponding deployment plan. We use $G(V,' E')$ as a general input case to show the procedure of solving $\mathcal{P}2$, summarized in Algorithm 2. The input graph to solve $\mathcal{P}3$ is $G(V,'' E'')$.

Let Φ_m denote the set for w-clique in the graph. Enumerating all the edges, we can find all the 2-cliques at the initial stage (line 1). Each 2-clique in the graph is put into Φ_2 for subsequent 3-clique searches. The elements contained in Φ_w can be used to find $(w+1)$-cliques by comparing all the pairs of w-cliques. If there are $(w-1)$ overlapped vertices between two different w-cliques and Φ_2 contains the missing edge (line 7), we can combine each pair of w-cliques into a new $(w+1)$-clique and put it in Φ_{w+1} (line 8). Repeat the execution until w equals to α or there is no $(w+1)$-clique (line 11). If Φ_2 is empty, the output is a vertex (line 15). Device-BS association patterns and resource allocation strategy can be obtained based on the algorithm's output.

The complexity of the proposed algorithm is different from determining whether $G(V,' E')$ contains an α-clique (belonging to an NP-complete problem). The naive algorithm of α-clique has a polynomial computational complexity of $\binom{|V'(G)|}{\alpha} = O(|V'(G)|^\alpha)$ to examine all α-tuples of vertices [36]. As finding the optimal clique $\mathcal{Q}$ in the given graph based on Algorithm 2 needs to examining all cliques whose number of vertices is less than α, the complexity of the entire traversal process is $\sum_{j=1}^{\alpha} \binom{|V'(G)|}{\alpha} = O(\sum_{j=1}^{\alpha} |V'(G)|^\alpha)$.

V. Performance Evaluation

Simulation results are presented to verify the effectiveness of the proposed solution. All the simulations are carried out through Python and MATLAB. In the multi-DSC-assisted RAN, the height and coverage radius of the MBS are 10 m and 800 m. The coverage area of the MBS is divided into a set of grid areas. Each grid area has a length and width of 200 m. Each intersection point of grid lines represents the x-y coordinate plane projection of a candidate DSC deployment position. The total number of the intersection points equals to the number of plane position indexes on the x-y axis, i.e., N. The DSC flight height range is set to [z_m, 200m] with a fixed step size of 10 m for adjacent heights (with $H_j = \frac{200\text{m}}{10\text{m}} = 20$) for $j \in \mathcal{N}$. Each DSC has the same

TABLE II
PARAMETER SETTINGS

Parameters	Values
Altitude of the MBS (z_m)	10m [25]
End device transmit power (p_i)	3.2mW [37]
DSC transmit power ($p_{j,k}$)	250mW [25]
MUD/MTD packet arrival rate (λ_d/λ_a)	20pkt/s [12]/5pkt/s [38]
MUD/MTD packet length (L_d/L_a)	9000bit [12]/2000bit [38]
Vertex filtering threshold (θ)	0.98
MTD packet delay bound ($D^{(\max)}$)	100ms [38]
Delay bound violation probability (ε)	10^{-3} [38]
Number of MUDs/MTUs (U/M)	50/50, 100/100
Urban environment parameter (o/b)	9.61/0.16 [26]
Carrier frequency (f)	3.5GHz [39]
Interference distance (ψ)	200m [27]
LoS probability threshold (ξ)	0.5
Free space path-loss threshold (χ)	89dB [29]
LoS/NLoS path-loss ($\kappa_{\mathrm{LoS}}/\kappa_{\mathrm{NLoS}}$)	103.4dB/131.4dB [25]
$\kappa_{\mathrm{LoS}}/\kappa_{\mathrm{NLoS}}$ exponent ($\rho_{\mathrm{LoS}}/\rho_{\mathrm{NLoS}}$)	24.2dB/42.8dB [25]

Algorithm 2: Search_clique ($G(V', E')$, α).

$w \leftarrow 2$; $\Phi_w \leftarrow E'(G)$;
for $n \leftarrow 3$ **to** α **do**
  $\Phi_n \leftarrow \emptyset$;
if $\Phi_2 \neq \emptyset$ **then**
  while $w < \alpha$ **do**
    foreach $(\mathcal{Q}, \mathcal{Q}') \in \Phi_w$ **do**
      if $!((Q - Q') - (w-1))\&\&$ $(\mathcal{Q} \cup \mathcal{Q}')\backslash(\mathcal{Q} \cap \mathcal{Q}') \in \Phi_2$ **then**
        $\Phi_{w+1} \leftarrow \mathcal{Q} \cup \mathcal{Q}'$;
        $\Phi_w \leftarrow \Phi_w \backslash (\mathcal{Q} \cup \mathcal{Q}')$;
    if $!|\Omega_{w+1}|$ **then**
      Break;
    $w \leftarrow w + 1$;
  $\mathcal{Q} \leftarrow \arg \min_{u \in \{2,\ldots,w\}, \mathcal{Q} \in \Phi_u} \sum_{v_{j,k_j^*} \in V'(G)} q_{j,k_j^*} \tilde{w}_{j,k_j^*}$;
else if $!|\Phi_2|$ **then**
  $\mathcal{Q} \leftarrow \{v_{j^*,k^*}\} \leftarrow \arg \min_{v_{j,k_j^*} \in V'(G)} \tilde{w}_{j,k_j^*}$;
return $\mathcal{Q}$;

uplink transmit power of 250mW [25], while end devices have a low transmit power set to 3.2mW [37]. Two types of end devices are randomly distributed in the coverage region of the MBS. The periodic data packet arrival rate λ_d is 20 packet/s, while the average rate λ_a of machine-type packet arrivals is 5 packet/s [38]. The lengths of the machine-type packet and data packet (L_a and L_d) are 2000 b [38] and 9000 b. The machine-type packet deadline bound $D^{(\max)}$ and deadline bound violation probability ε are 100 s and 10^{-3} [38]. The detailed parameter setting is listed in Table II.

The proposed joint DSC deployment and resource allocation scheme (D2RA) is categorized into D2RA-1 and D2RA-2, corresponding to the solutions of problems $\mathcal{P}2$ and $\mathcal{P}3$. For further comparison and verification, we consider four baseline approaches in the following:

1) Maximum DSC Number First (MDNF), which incorporates the proposed flight altitude selection policy to deploy as many DSCs as possible (without the constraint on the number of available DSCs);
2) Most Association First (MAF) [25], which maximizes the number of device-DSC associations;
3) Maximum Device-Coverage First (MDCF) [29], which maximizes the number of end devices in the coverage region of DSCs;
4) Maximum Rate-Coverage First (MRCF), which maximizes the summation of transmission rates from end devices under the coverage of DSCs.

Each baseline approach is incorporated into our graph model to support the multi-DSC deployment. Of these methods, only the MDNF has no constraint on the number of available DSCs. We summarize these baseline strategies in a unified model as shown in Table III, where I_j, $I_{j,s}$, and $\tilde{a}_{i,j}$ represent $I_{j,k}$, $I_{j,k,s}$, and $\tilde{a}_{i,j,k}$ when $z_{j,k}$ is set to $z^{(\max)} = 77.4$ m to maximize the effective coverage. Using the proposed interference graph framework, the DSC deployment can be incorporated into these strategies using one formulation with different inputs. For a fair comparison, the number of vertices in the D2RA-1 ($|V'(G)|$) is used to unify the number of vertices in each method. Taking the MAF as an example, we guarantee $|V'(G)| = |V_2(G)|$ by adjusting ζ_2.

A. Case Study I

The number of end devices within the MBS coverage is set to 100. Fig. 5 shows the network performance with an increasing number of DSCs.

In Fig. 5(a), the total resource consumption using the D2RA and MAF is almost the same when α is small. In particular, the D2RA-1 is more resource-efficient and can accommodate more DSCs than the D2RA-2 and MAF. Since the adaptive altitude selection in the D2RA-1 improves and spectrum reuse among DSCs, the D2RA-1 can save up to 16% of resources when $\alpha > 4$. Although the D2RA-2 and MAF cover more devices and increase the probability of device-DSC associations, their flight height selection is unreasonable. Due to the interference distance constraint, there is an upper bound on the number of DSCs that can be accommodated in the MBS's 3D coverage space. The flight height adaptation in the D2RA-1 and MDNF helps capture potential spectrum reuse opportunities while accommodating more DSCs. In Fig. 5(a), the approaches using fixed flight height can accommodate up to 5 DSCs, and their results remain unchanged if $\alpha > 5$. In the case of $\alpha = 6$, the D2RA-1 with 5 DSCs consumes lower resources than other methods with 6 DSCs, indicating that the D2RA-1 can reduce DSC dispatching cost while saving resources.

Figs. 5(b) and (c) show results on the resource consumption from MUDs and MTDs. Because an MUD requires a higher rate than an MTD, the former consumes more resources than the latter. The trend in Fig. 5(d) is similar to that of Fig. 5(a), since most devices connect to the MBS. In Fig. 5(e), the increase of α has a slight impact on the resource consumption of G2D links through spectrum reusing. In Fig. 5(f), more resources are consumed for D2M communications without spectrum reuse. In Fig. 5(g), the amount of resources reused by DSCs for G2D links is calculated as $\sum_{v_{j,k} \in \mathcal{Q} \backslash \{v_{j^*,k^*}\}} \delta_{j,k}$, which is proportional to α.

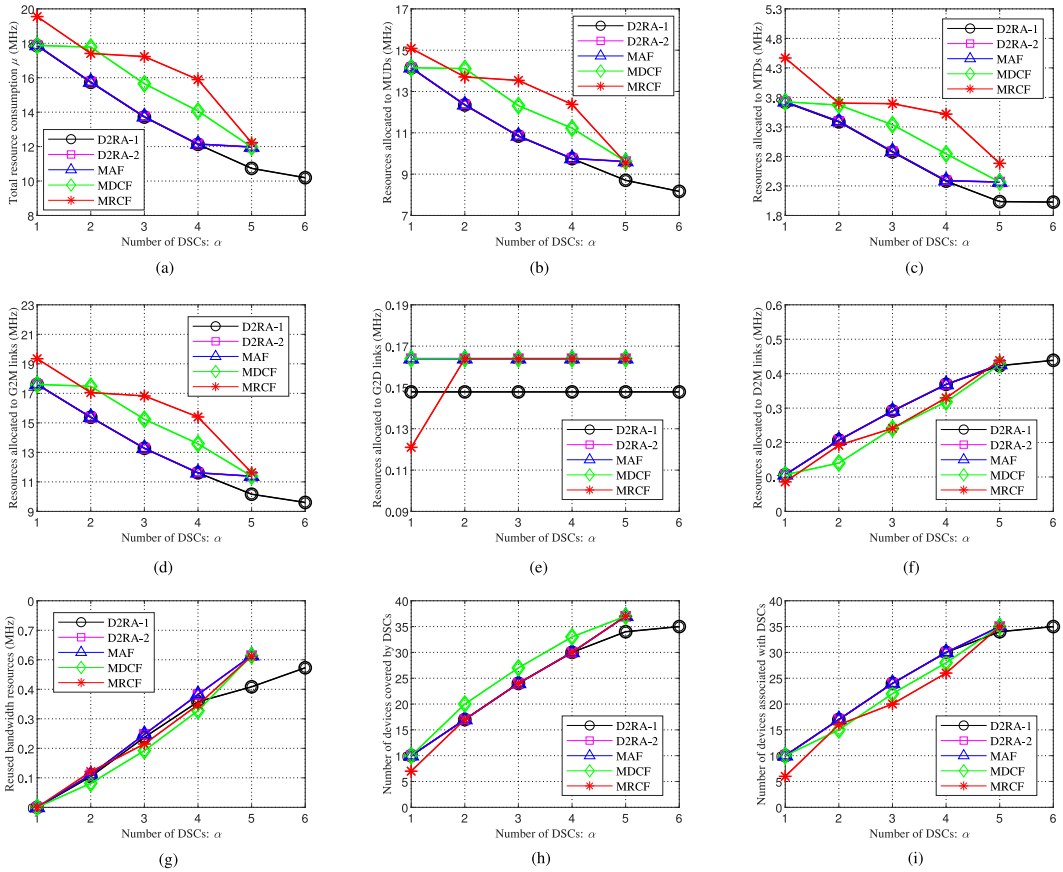

Fig. 5. Impact of the number of available DSCs (Case Study I). (a) Total resource consumption μ. (b) Resources allocated to MUDs. (c) Resources allocated to MTDs. (d) Resources allocated to G2M links. (e) Resources allocated to G2D links. (f) Resources allocated to D2M links. (g) Reused resources. (h) Number of devices covered by DSCs. (i) Number of devices-DSC associations.

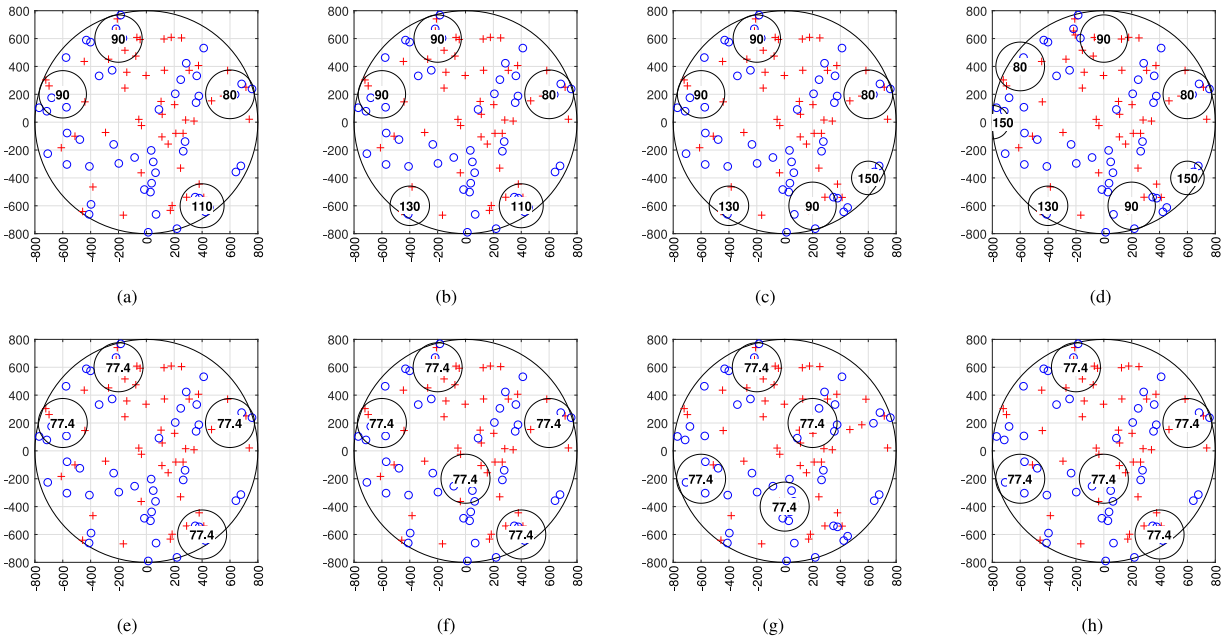

Fig. 6. DSC placement scenarios (Case Study I). (a) DARA-1 (α= 4). (b) DARA-1 (α= 5). (c) DARA-1 (α>6). (d) MDNF. (e) DARA-2 (α = 4). (f) DARA-2 ($\alpha \geq$5). (g) MRCF (α = 4). (h) MRCF ($\alpha \geq$5).

TABLE III
BASELINE IMPLEMENTATION LOGIC

Baseline	Set of candidate DSC deployment positions	Optimization objective for DSC deployment	Available DSC num. (α)
MDNF	$V_1(G) = \{v_{j,k_j^*} \mid j \in \mathcal{N}, \beta_{j,k_j^*} \leq \zeta_1\}$	$\underset{\mathcal{Q}}{\text{Maximize}} : \sum_{v_{j,k_j^*} \in V_1(G)} q_{j,k_j^*} \tilde{w}_{j,k_j^*}$	N/A
MAF	$V_2(G) = \{v_j \mid j \in \mathcal{N}, \sum_{i \in I_j} \tilde{a}_{i,j} \leq \zeta_2\}$	$\underset{\mathcal{Q}}{\text{Maximize}} : \sum_{v_j \in V_2(G)} q_j \sum_{i \in I_j} \tilde{a}_{i,j}$	$1, 2, 3, \ldots, 6$
MDCF	$V_3(G) = \{v_j \mid j \in \mathcal{N}, I_j \leq \zeta_3\}$	$\underset{\mathcal{Q}}{\text{Maximize}} : \sum_{v_j \in V_3(G)} q_j I_j$	$1, 2, 3, \ldots, 6$
MRCF	$V_4(G) = \{v_j \mid j \in \mathcal{N}, \sum_{s \in \{1,2\}} Y_s I_{j,s} \leq \zeta_4\}$	$\underset{\mathcal{Q}}{\text{Maximize}} : \sum_{s \in \{1,2\}} Y_s \sum_{v_j \in V_4(G)} q_j I_{j,s}$	$1, 2, 3, \ldots, 6$

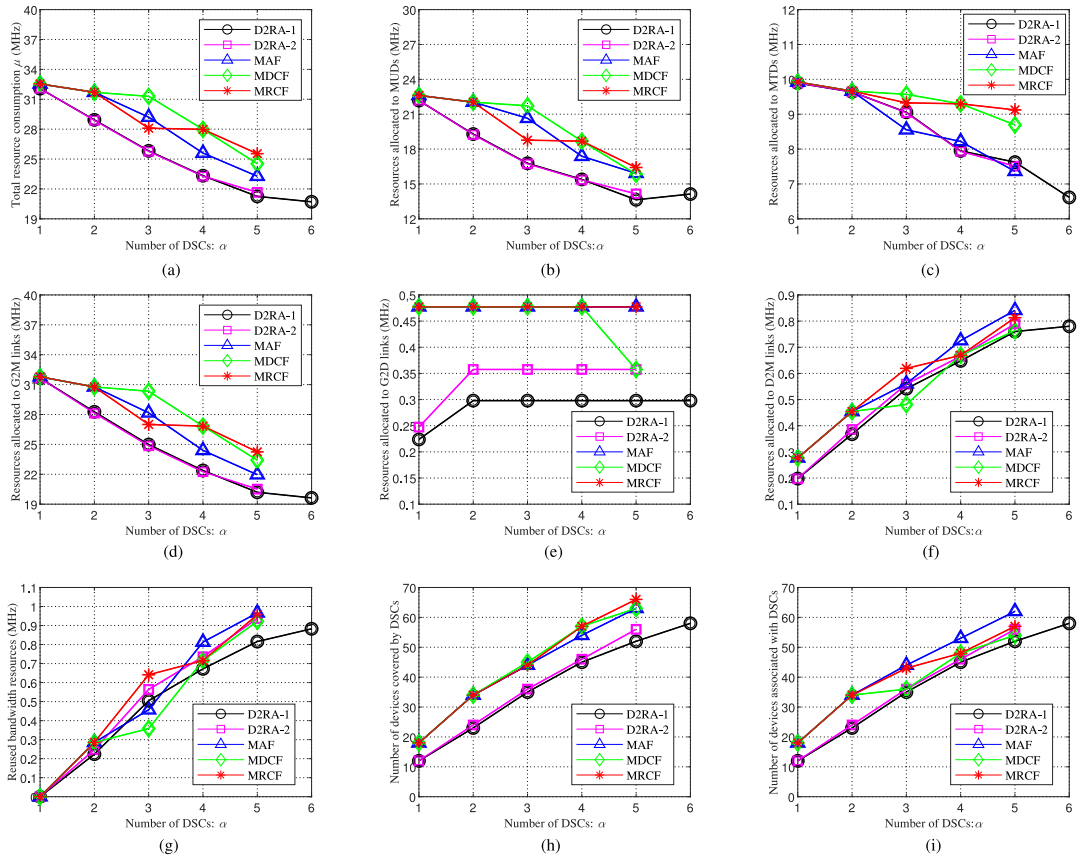

Fig. 7. Impact of the number of available DSCs (Case Study II). (a) Total resource consumption μ. (b) Resources allocated to MUDs. (c) Resources allocated to MTDs. (d) Resources allocated to G2M links. (e) Resources allocated to G2D links. (f) Resources allocated to D2M links. (g) Reused resources. (h) Number of devices covered by DSCs. (i) Number of device-DSC associations.

In Figs. 5(h) and 5(i), the MDCF has the most significant number of devices but a small number of device-DSC associations. Compared with the MDCF and MRCF, the D2RA and MAF maintain a higher number of device-DSC associations. Once a DSC covers a device, the device connects to the DSC with a high probability.

Fig. 6 shows various DSC deployment scenarios. The big (small) black circle represents the coverages of the MBS (DSC). The number surrounded by the small circle represents the height of the corresponding DSC. Those randomly distributed blue circles (red crosses) under the BS coverage areas represent MUDs (MTDs).

In Figs. 6(a)-6(c), the DSC deployment determined by the D2RA-1 changes with the increase of α. When a new DSC joins, the placement plan needs to be adjusted to mitigate inter-DSC interference. Under the distance constraint, the increase of α means that each DSC has to reduce coverage radius to mitigate inter-DSC interference, which may decrease the number of device-DSC associations. In some cases, the performance may be deteriorated if excessive DSCs are dispatched. The D2RA strategies comprehensively consider the number of DSCs launched and the deployment strategy, while other methods make deployment strategies for a given number of DSCs. In the simulation, the D2RA-1 with 6 DSCs performs better than the that with 7 DSCs, and therefore, the deployment with 6 DSCs is chosen by the D2RA-1. Figs. 6(e) and 6(f) show that the D2RA-2 compensates for low resource utilization by increasing the number of device-DSC associations. The

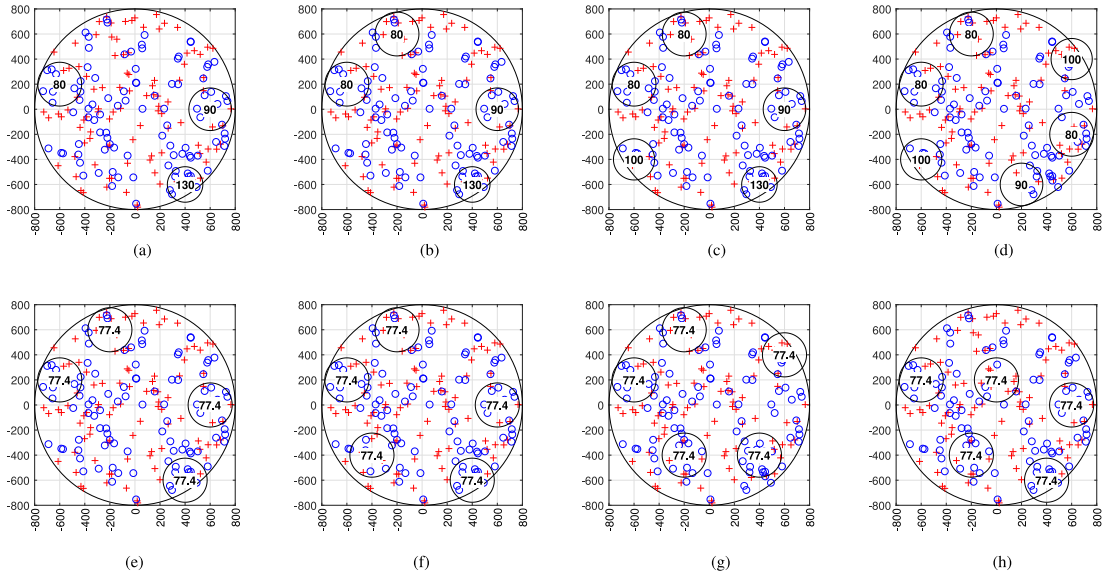

Fig. 8. DSC placement scenarios (Case Study II). (a) DARA-1 ($\alpha = 3$). (b) DARA-1 ($\alpha = 4$). (c) DARA-1 ($\alpha = 5$). (d) DARA-2 ($\alpha \geq 6$). (e) DARA-2 ($\alpha = 4$). (f) DARA ($\alpha \geq 5$). (g) MAF ($\alpha \geq 5$). (h) MDCF ($\alpha \geq 5$).

TABLE IV
DETAILS FOR FIGS. 6(C) AND 6(D)

Approach	α	Q	Resource consumption	Device-DSC association num.
D2RA-1	7	6	10.193 MHz	35
MDNF	7	7	11.208 MHz	30

MRCF is sensitive to the increase in α regardless of where DSCs are deployed. In Figs. 6(g) and 6(h), the DSCs are not placed very close to the MBS edge with reduced performance gain.

Table IV details Figs. 6(c) and 6(d), where the D2RA-1 with 6 DSCs consumes fewer resources than the MDNF with 7 DSCs, more intuitively demonstrating that the D2RA-1 can reduce DSC dispatching cost without sacrificing performance.

B. Case Study II

We demonstrate the network performance in a scenario where the number of end devices covered by the MBS is 200.

Fig. 7(a) shows that the resource consumption of the D2RA is lower than that of the MAF, which is different from Fig. 5(a) in which both schemes are close in performance. In the case of high device density, reducing total resource consumption depends mainly on DSC deployment. Compared to a low-density case, each DSC has a better opportunity to boost performance gain. With the same number of device-DSC associations, DSCs at the edge of the MBS coverage are more effective in resource conservation. The D2RA-1 (D2RA-2) saves up to 11% (7%) of resources compared with the MAF. The difference between Fig. 7(b) (Fig. 7(c)) and Fig. 7(a) is mainly due to the randomness of device distribution.

In Figs. 7(d) and 7(e), the D2RA-1 saves radio resource of G2M and G2D links, by taking advantage of adaptive flight altitude. From Fig. 7(g), the resource reuse shows a growing trend as α increases. Since the D2RA-1 consumes the least resources, the reused resources are lower than that of the other methods. Figs. 7(h) and 7(i) show that maximizing the number of devices covered by DSCs or the number of device-DSC associations fails to improve overall performance.

In Figs. 8(a)-8(f), the DSCs' plane projection positions are approximately the same in the case of $\alpha = 4$ and $\alpha = 5$. The advantage of the D2RA-1 lies in its flight altitude adaptive policy, which can reuse spectrum resources more effectively than the D2RA-2 with fixed flight height. Comparing Fig. 8(g) and Fig. 8(h), we can see that the DSC deployment using the MAF is better than that using the MDCF. The MDCF can enable DSCs to cover the largest number of end devices. Still, it does not guarantee that most devices are associated with DSCs, especially when DSCs are located close to the MBS center. Because the MAF can effectively cover edge hotspots with an increased number of device-DSC associations, the entire network's resource utilization is improved.

VI. CONCLUSION

In this paper, we have proposed a multi-DSC assisted uplink radio resource slicing scheme, in which the resource consumption is minimized with differentiated QoS guarantees for MUDs and MTDs. An interference graph model is present to characterize the multi-DSC deployment while mitigating inter-DSC interference. On this basis, a joint optimization problem on the DSC placement and device-BS associations is formulated and proved to be NP-hard. By excluding candidate DSC deployment position with low-performance gain, we establish a complexity-adjustable problem approximation, which incorporates a flight height adaptive policy. A lightweight approximation with a fixed flight altitude is also presented to reduce complexity further. The joint optimization problem is transformed into a special weight clique problem, with a performance-aware clique algorithm designed to determine the DSC placement and device-DSC

association patterns. Simulation results demonstrate that the proposed DSC deployment and resource slicing scheme achieves significant performance improvement over the benchmarks.

APPENDIX A
PROOF OF PROPOSITION 1

In the case of $a_{i,j,k} \in \{0,1\}$ for $i \in \mathcal{I}_{j,k}$, the lower bound of $w_{j,k}$ is calculated as

$$w_{j,k}^{(\min)} = \begin{cases} -\sum\limits_{s\in\{1,2\}}\sum\limits_{i\in\mathcal{I}_{j,k,s}}\left|\frac{Y_s}{r_{i,j,k}}+\frac{Y_s}{r_{j,k,m}}-\frac{Y_s}{r_{i,m}}\right|, \text{if } v_{j,k}=v_{j^*,k^*} \\ -\sum\limits_{s\in\{1,2\}}\sum\limits_{i\in\mathcal{I}_{j,k,s}}\left|\frac{Y_s}{r_{j,k,m}}-\frac{Y_s}{r_{i,m}}\right|, \text{if } v_{j,k}\in\mathcal{Q}\backslash\{v_{j^*,k^*}\} \end{cases} \tag{49}$$

Under (34), we have $w_{j,k}=w_{j,k}^{(\min)}$. Since $\varphi^{(\min)}(\mathcal{Q}) = \sum_{v_{j,k}\in\mathcal{Q}} w_{j,k}^{(\min)}$ as in (33), the conclusion of Proposition 1 holds.

B. Proof of Proposition 2

Let $v_{j_{\mathcal{Q}}^*,k_{\mathcal{Q}}^*}$ and $v_{j_{\mathcal{Q}'}^*,k_{\mathcal{Q}'}^*}$ denote v_{j^*,k^*} under $\mathcal{Q}$ and $\mathcal{Q}'$. There are two cases (i.e., $v_{j_{\mathcal{Q}}^*,k_{\mathcal{Q}}^*}\in\mathcal{Q}'$ and $v_{j_{\mathcal{Q}}^*,k_{\mathcal{Q}}^*}\in\mathcal{Q}\backslash\mathcal{Q}'$) to be considered. For the first case where $v_{j_{\mathcal{Q}}^*,k_{\mathcal{Q}}^*}=v_{j_{\mathcal{Q}'}^*,k_{\mathcal{Q}'}^*}$, we have (50). For the second case where $v_{j_{\mathcal{Q}}^*,k_{\mathcal{Q}}^*}\neq v_{j_{\mathcal{Q}'}^*,k_{\mathcal{Q}'}^*}$, we have (51).

Similar to Proposition 1, under (34), the values of (50) and (51) are greater than 0.

$$\begin{aligned}\varphi^{(\min)}(\mathcal{Q})-\varphi^{(\min)}(\mathcal{Q}') &= \sum_{v_{j,k}\in(\mathcal{Q}\backslash\mathcal{Q}')\backslash\{v_{j_{\mathcal{Q}}^*,k_{\mathcal{Q}}^*}\}}\sum_{s\in\{1,2\}}\sum_{i\in\mathcal{I}_{j,k,s}} a_{i,j,k}\left(\frac{Y_s}{r_{j,k,m}}-\frac{Y_s}{r_{i,m}}\right) \\ &= \sum_{v_{j,k}\in(\mathcal{Q}\backslash\mathcal{Q}')\backslash\{v_{j_{\mathcal{Q}}^*,k_{\mathcal{Q}}^*}\}} w_{j,k} \end{aligned} \tag{50}$$

$$\begin{aligned}\varphi^{(\min)}(\mathcal{Q})-\varphi^{(\min)}(\mathcal{Q}') &= \sum_{v_{j,k}\in\mathcal{Q}'\backslash\{v_{j_{\mathcal{Q}}^*,k_{\mathcal{Q}}^*}\}}\sum_{s\in\{1,2\}}\sum_{i\in\mathcal{I}_{j,k,s}} a_{i,j,k}\left(\frac{Y_s}{r_{j,k,m}}-\frac{Y_s}{r_{i,m}}\right)-\sum_{s\in\{1,2\}} Y_s\sum_{i\in\mathcal{I}_{j_{\mathcal{Q}'}^*,k_{\mathcal{Q}'}^*,s}}\frac{a_{i,j_{\mathcal{Q}'}^*,k_{\mathcal{Q}'}^*}}{r_{i,j_{\mathcal{Q}'}^*,k_{\mathcal{Q}'}^*}} \\ &\quad+\sum_{s\in\{1,2\}}\sum_{i\in\mathcal{I}_{j_{\mathcal{Q}'}^*,k_{\mathcal{Q}'}^*,s}} a_{i,j_{\mathcal{Q}'}^*,k_{\mathcal{Q}'}^*}\left(\frac{Y_s}{r_{i,j_{\mathcal{Q}'}^*,k_{\mathcal{Q}'}^*}}+\frac{Y_s}{r_{j_{\mathcal{Q}'}^*,k_{\mathcal{Q}'}^*,m}}-\frac{Y_s}{r_{i,m}}\right) \\ &= w_{j^*,k^*}+\sum_{v_{j,k}\in\mathcal{Q}'\backslash\{v_{j_{\mathcal{Q}}^*,k_{\mathcal{Q}}^*}\}} w_{j,k}-\sum_{s\in\{1,2\}} Y_s\sum_{i\in\mathcal{I}_{j_{\mathcal{Q}'}^*,k,s}}\frac{a_{i,j_{\mathcal{Q}'}^*,k_{\mathcal{Q}'}^*}}{r_{i,j_{\mathcal{Q}'}^*,k_{\mathcal{Q}'}^*}} \\ &\le w_{j^*,k^*}+\sum_{v_{j,k}}\in\mathcal{Q}'\backslash\{v_{j_{\mathcal{Q}}^*,k_{\mathcal{Q}}^*}\} w_{j,k}. \end{aligned} \tag{51}$$

C. Proof of Theorem 1

Consider a special case of $\mathcal{P}1$, which satisfies: 1) $\delta_{j,k}=\delta_{j,'k'}=\delta, \forall v_{j,k}, v_{j,'k'}\in V(G)$, 2) $a_{i,j,k}=1, \forall i\in\mathcal{I}_{j,k}, \forall v_{j,k}\in V(G)$, and 3) α is sufficiently large. Since the value of δ is constant for any $\mathcal{Q}$, the objective is equivalent to minimizing $\sum_{v_{j,k}\in V(G)} q_{j,k}w_{j,k}$, where

$$w_{j,k}=\sum_{s\in\{1,2\}}\sum_{i\in\mathcal{I}_{j,k,s}}\left(\frac{Y_s}{r_{j,k,m}}-\frac{Y_s}{r_{i,m}}\right). \tag{52}$$

Because $w_{j,k}$ in (52) is determined by $v_{j,k}$, $\mathcal{P}1$ becomes the maximum weight clique problem [35], whose NP-complete nature is well known. So, $\mathcal{P}1$, as a general case, is NP-hard.

D. Proof of Theorem 2

If $R_{j,k_{l-1}}<R_{j,k_l}$, we have $\mathcal{I}_{j,k_{l-1}}\subseteq\mathcal{I}_{j,k_l}$ and

$$\begin{aligned}\beta_{j,k_l} &= \sum_{i\in\mathcal{I}_{j,k_{l-1}}}\tilde{a}_{i,j,k_l}\left(\frac{1}{r_{i,m}}-\left(\frac{1}{r_{i,j,k_l}}+\frac{1}{r_{j,k_l,m}}\right)\right) \\ &\quad+\sum_{i\in\mathcal{I}_{j,k_l}\backslash\mathcal{I}_{j,k_{l-1}}}\tilde{a}_{i,j,k_l}\left(\frac{1}{r_{i,m}}-\left(\frac{1}{r_{i,j,k_l}}+\frac{1}{r_{j,k_l,m}}\right)\right).\end{aligned} \tag{53}$$

Because each term in (53) is larger than 0, we have

$$\beta_{j,k_l}\ge\sum_{i\in\mathcal{I}_{j,k_{l-1}}}\tilde{a}_{i,j,k_l}\left(\frac{1}{r_{i,m}}-\left(\frac{1}{r_{i,j,k_l}}+\frac{1}{r_{j,k_l,m}}\right)\right). \tag{54}$$

If $z_{j,k_l}\le\min_{i\in\mathcal{I}_{j,k_{l-1}}} z_{j,k_{i,j}^*}$, then for each $i\in\mathcal{I}_{j,k_{l-1}}$, the value of $\frac{1}{r_{i,j,k_{l-1}}}+\frac{1}{r_{j,k_{l-1},m}}$ decreases in $(0, z_{j,k_l}]$ as l increases. Then, we have

$$\begin{aligned}\beta_{j,k_{l-1}} &= \sum_{i\in\mathcal{I}_{j,k_{l-1}}}\tilde{a}_{i,j,k_{l-1}}\left(\frac{1}{r_{i,m}}-\left(\frac{1}{r_{i,j,k_{l-1}}}+\frac{1}{r_{j,k_{l-1},m}}\right)\right) \\ &\le\sum_{i\in\mathcal{I}_{j,k_{l-1}}}\tilde{a}_{i,j,k_l}\left(\frac{1}{r_{i,m}}-\left(\frac{1}{r_{i,j,k_l}}+\frac{1}{r_{j,k_l,m}}\right)\right).\end{aligned} \tag{55}$$

Comparing (54) and (55), we have $\beta_{j,k_l}\ge\beta_{j,k_{l-1}}$, which completes the proof.

E. Proof of Theorem 3

If $R_{j,k_{l-1}}>R_{j,k_l}$, we have $\mathcal{I}_{j,k_l}\subseteq\mathcal{I}_{j,k_{l-1}}$ and

$$\begin{aligned}\beta_{j,k_{l-1}} &= \sum_{i\in\mathcal{I}_{j,k_l}}\tilde{a}_{i,j,k_{l-1}}\left(\frac{1}{r_{i,m}}-\left(\frac{1}{r_{i,j,k_{l-1}}}+\frac{1}{r_{j,k_{l-1},m}}\right)\right) \\ &\quad+\sum_{i\in\mathcal{I}_{j,k_{l-1}}\backslash\mathcal{I}_{j,k_l}}\tilde{a}_{i,j,k_{l-1}}\left(\frac{1}{r_{i,m}}-\left(\frac{1}{r_{i,j,k_{l-1}}}+\frac{1}{r_{j,k_{l-1},m}}\right)\right).\end{aligned} \tag{56}$$

Because each term in (56) is larger than 0, we have

$$\beta_{j,k_l}\ge\sum_{i\in\mathcal{I}_{j,k_l}}\tilde{a}_{i,j,k_{l-1}}\left(\frac{1}{r_{i,m}}-\left(\frac{1}{r_{i,j,k_{l-1}}}+\frac{1}{r_{j,k_{l-1},m}}\right)\right) \tag{57}$$

If $z_{j,k_{l-1}} \geq \max_{i\in\mathcal{I}_{j,k_l}} z_{j,k^*_{i,j}}$, then for each $i \in \mathcal{I}_{j,k_{l-1}}$, the value of $\frac{1}{r_{i,j,k_{l-1}}}+\frac{1}{r_{j,k_{l-1},m}}$ increases in $[z_{j,k_{l-1}}, \frac{c10^{\frac{\chi}{20}}}{4\pi f}]$ as l increases. Accordingly, we have

$$\beta_{j,k_l} \leq \sum_{i\in\mathcal{I}_{j,k_l}} \tilde{a}_{i,j,k_{l-1}} \left(\frac{1}{r_{i,m}} - \left(\frac{1}{r_{i,j,k_{l-1}}}+\frac{1}{r_{j,k_{l-1},m}}\right)\right). \tag{58}$$

Based on (57) and (58), we have $\beta_{j,k_l} \leq \beta_{j,k_{l-1}}$. Because $z_{j,k_{l+n}} > z_{j,k_{l+n+1}}$, we get $\beta_{j,k_{l+n}} \leq \beta_{j,k_{l+n-1}}$ given $n \in \{1, 2, \ldots, H_j - n\}$, which completes the proof.

REFERENCES

[1] W. Zhuang, Q. Ye, F. Lyu, N. Cheng, and J. Ren, "SDN/NFV-empowered future IoV with enhanced communication, computing, and caching," in *Proc. IEEE*, vol. 108, no. 2, pp. 274–291, 2020.

[2] Q. Ye, J. Li, K. Qu, W. Zhuang, X. S. Shen, and X. Li, "End-to-end quality of service in 5G networks: Examining the effectiveness of a network slicing framework," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 65–74, Jun. 2018.

[3] B. Li, J. Liu, and B. Ji, "Low-overhead wireless uplink scheduling for large-scale internet-of-things," *IEEE Trans. Mobile Comput.*, vol. 20, no. 2, pp. 577–587, Feb. 2021.

[4] X. Zhou, S. Durrani, J. Guo, and H. Yanikomeroglu, "Underlay drone cell for temporary events: Impact of drone height and aerial channel environments," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1704–1718, Apr. 2019.

[5] X. Zhang, W. Gu, H. Zhang, and M. Wang, "Hybrid communication path orchestration for 5G heterogeneous ultra-dense networks," *IEEE Netw.*, vol. 33, no. 4, pp. 112–118, Jul.–Aug. 2019.

[6] X. Shen *et al.*, "AI-assisted network-slicing based next-generation wireless networks," *IEEE Open J. Veh. Technol.*, vol. 1, pp. 45–66, Jan. 2020.

[7] W. Shi, J. Li, N. Cheng, F. Lyu, and X. Shen, "Multi-drone 3D trajectory planning and scheduling in drone assisted radio access networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8145–8158, 2019.

[8] C.-Y. Oh, D. Hwang, and T.-J. Lee, "Joint access control and resource allocation for concurrent and massive access of M2M devices," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4182–4192, Aug. 2015.

[9] B. Yi *et al.*, "A comprehensive survey of network function virtualization," *Comput. Netw.*, vol. 133, pp. 212–262, 2018.

[10] I. T. Haque and N. Abu-Ghazaleh, "Wireless software defined networking: A survey and taxonomy," *IEEE Commun. Surv. Tuts.*, vol. 18, no. 4, pp. 2713–2737, Oct.–Dec. 2016.

[11] Q. Duan, N. Ansari, and M. Toy, "Software-defined network virtualization: An architectural framework for integrating SDN and NFV for service provisioning in future networks," *IEEE Netw.*, vol. 30, no. 5, pp. 10–16, Sep.–Oct. 2016.

[12] Q. Ye, W. Zhuang, S. Zhang, A.-L. Jin, X. Shen, and X. Li, "Dynamic radio resource slicing for a two-tier heterogeneous wireless network," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9896–9910, Oct. 2018.

[13] H. Peng, Q. Ye, and X. Shen, "Spectrum management for multi-access edge computing in autonomous vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 7, pp. 3001–3012, Jul. 2020.

[14] A. Papa, M. Klugel, L. Goratti, T. Rasheed, and W. Kellerer, "Optimizing dynamic RAN slicing in programmable 5G networks," in *Proc. IEEE ICC*, 2019, pp. 1–7.

[15] A. Aijaz, M. Tshangini, M. R. Nakhai, X. Chu, and A.-H. Aghvami, "Energy-efficient uplink resource allocation in LTE networks with M2M/H2H co-existence under statistical QoS guarantees," *IEEE Trans. Wireless Commun.*, vol. 62, no. 7, pp. 2353–2365, Jul. 2014.

[16] W. Shi, J. Li, H. Wu, C. Zhou, N. Cheng, and X. Shen, "Drone-cell trajectory planning and resource allocation for highly mobile networks: A hierarchical DRL approach," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9800–9813, Jun. 2021.

[17] S. Yan, M. Peng, and X. Cao, "A game theory approach for joint access selection and resource allocation in UAV assisted IoT communication networks," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1663–1674, Apr. 2019.

[18] F. Lyu *et al.*, "Service-oriented dynamic resource slicing and optimization for space-air-ground integrated vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, to be published, doi: 10.1109/TITS.2021.3070542.

[19] Y. Li and L. Cai, "UAV-assisted dynamic coverage in a heterogeneous cellular system," *IEEE Netw.*, vol. 31, no. 4, pp. 56–61, Jul.–Aug. 2017.

[20] Y. Wang *et al.*, "Joint resource allocation and UAV trajectory optimization for space-air-ground internet of remote things networks," *IEEE Syst. J.*, to be published, doi: 10.1109/JSYST.2020.3019463.

[21] M. A. Ali, Y. Zeng, and A. Jamalipour, "Software-defined coexisting UAV and WiFi: Delay-oriented traffic offloading and UAV placement," *IEEE J. Sel. Area. Comm.*, vol. 38, no. 6, pp. 988–998, Jun. 2020.

[22] T. P. de Andrade, C. A. Astudillo, and N. L. da Fonseca, "Allocation of control resources for machine-to-machine and human-to-human communications over LTE/LTE-A networks," *IEEE Internet Things J.*, vol. 3, no. 3, pp. 366–377, Jun. 2016.

[23] A. A. Khuwaja, Y. Chen, N. Zhao, M.-S. Alouini, and P. Dobbins, "A survey of channel modeling for UAV communications," *IEEE Commun. Surv. Tuts.*, vol. 20, no. 4, pp. 2804–2821, Oct.–Dec. 2018.

[24] S. Sekander, H. Tabassum, and E. Hossain, "Multi-tier drone architecture for 5 G/B5G cellular networks: Challenges, trends, and prospects," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 96–103, Mar. 2018.

[25] X. Sun and N. Ansari, "Jointly optimizing drone-mounted base station placement and user association in heterogeneous networks," in *Proc. IEEE ICC*, 2018, pp. 1–6.

[26] R. I. Bor-Yaliniz, A. El-Keyi, and H. Yanikomeroglu, "Efficient 3-D placement of an aerial base station in next generation cellular networks," in *Proc. IEEE ICC*, 2016, pp. 1–5.

[27] A. Fotouhi, M. Ding, and M. Hassan, "Dynamic base station repositioning to improve spectral efficiency of drone small cells," in *Proc. IEEE WoWMoM*, 2017, pp. 1–9.

[28] H. Peng and X. Shen, "Multi-agent reinforcement learning based resource management in MEC- and UAV-assisted vehicular networks," *IEEE J. Sel. Area. Comm.*, vol. 39, no. 1, pp. 131–141, Jan. 2021.

[29] W. Shi *et al.*, "Multiple drone-cell deployment analyses and optimization in drone assisted radio access networks," *IEEE Access*, vol. 6, pp. 12 518–12 529, 2018.

[30] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.

[31] D. Wu and R. Negi, "Effective capacity-based quality of service measures for wireless networks," *Mobile Netw. Appl.*, vol. 11, no. 1, pp. 91–99, 2006.

[32] E. Soltanmohammadi, K. Ghavami, and M. Naraghi-Pour, "A survey of traffic issues in machine-to-machine communications over LTE," *IEEE Internet Things J.*, vol. 3, pp. 865–884, Dec. 2016.

[33] P. M. Rabinovitch, "Statistical estimation of effective bandwidth." Master's Thesis, Carleton Univ., 2000.

[34] A. Abdrabou and W. Zhuang, "Stochastic delay guarantees and statistical call admission control for IEEE 802.11 single-hop ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3972–3981, Oct. 2008.

[35] Y. Wang, S. Cai, and M. Yin, "Two efficient local search algorithms for maximum weight clique problem," in *Proc. AAAI*, 2016, pp. 805–811.

[36] V. Vassilevska, "Efficient algorithms for clique problems," *Inform. Process. Lett.*, vol. 109, no. 4, pp. 254–257, 2009.

[37] S. M. Hussain *et al.*, "Seven pillars to achieve energy efficiency in high-performance computing data centers," in *Proc. Recent Trends Adv. Wireless IoT-enabled Netw.*, 2019, pp. 93–105.

[38] 3GPP, "Feasibility study on new services and markets technology enablers for critical communications; stage 1 (Release 14)," *Gener. Partnership Project*, Sophia Antipolis Cedex, France, Tech. Rep., 2016.

[39] I.-P. Belikaidis, A. Georgakopoulos, E. Kosmatos, V. Frascolla, and P. Demestichas, "Management of 3.5-GHz spectrum in 5G dense networks: A hierarchical radio resource management scheme," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 57–64, 2018.

Hang Shen (Member, IEEE) received the Ph.D. degree (with honors) in computer science from the Nanjing University of Science and Technology, Nanjing, China. He is currently an Associate Professor with the Department of Computer Science and Technology, Nanjing Tech University, Nanjing, China. From 2018 to 2019, he was a Full-Time Postdoctoral Fellow with the Broadband Communications Research Lab, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include radio access network slicing, space-air-ground integrated networks, network security and privacy. He is an Associate Editor for the IEEE ACCESS, an Academic Editor of the *Mathematical Problems in Engineering*, and was a Guest Editor of the *Peer-to-Peer Networking and Applications*.

Qiang Ye (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2016. From December 2016 to November 2018, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, where he was a Research Associate from December 2018 to September 2019. Since September 2019, he has been an Assistant Professor with the Department of Electrical and Computer Engineering and Technology, Minnesota State University, Mankato, MN, USA. His current research interests include 5G networks, software-defined networking and network function virtualization, network slicing, artificial intelligence and machine learning for future networking, protocol design, and end-to-end performance analysis for the Internet of Things.

Weihua Zhuang (Fellow, IEEE) has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, since 1993, where she is currently a Professor and a Tier I Canada Research Chair of wireless communication networks. She is a Fellow of the Royal Society of Canada, the Canadian Academy of Engineering, and the Engineering Institute of Canada. She is also an Elected Member of the Board of Governors and VP Publications of the IEEE Vehicular Technology Society. She was the recipient of the 2017 Technical Recognition Award from the IEEE Communications Society Ad Hoc & Sensor Networks Technical Committee and a co-recipient of several best paper awards from IEEE conferences. She was the Editor-in-Chief of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2007 to 2013, the Technical Program Chair/Co-Chair of the IEEE Vehicular Technology Conference (VTC) Fall 2017 and Fall 2016, and the Technical Program Symposia Chair of the IEEE Global Communications Conference (Globecom) 2011.

Weisen Shi (Graduate Student Member, IEEE) received the B.S. degree from Tianjin University, Tianjin, China, in 2013, the M.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2016, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 2020. His interests include space-air-ground integrated networks, UAV communication and networking, and radio access network slicing.

Guangwei Bai received the B.Eng. and M.Eng. degrees in computer engineering from Xi'an Jiaotong University, Xi'an, China, in 1983 and 1986, respectively, and the Ph.D. degree in computer science from the University of Hamburg, Hamburg, Germany, in 1999. From 1999 to 2001, he was a Research Scientist with the German National Research Center for Information Technology, Germany. In 2001, he joined the University of Calgary, Calgary, AB, Canada, as a Research Associate. Since 2005, he has been a Professor of computer science with Nanjing Tech University, Nanjing, China. From October to December 2010, he was a Visiting Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include architecture and protocols design for communication networks, multimedia networking, network security, and location-based services. He is a Member of the ACM and a Distinguished Member of CCF.

Geng Yang (Member, IEEE) received the Ph.D. degree in computer science from Laval University, Quebec, QC, Canada, in 1994. He is currently a Professor with the Nanjing University of Posts and Telecommunications, Nanjing, China. From 1994 to 1996, he was a Postdoctoral Research Fellow with the Center for Research on Computation and its Applications, University of Montreal, Montreal, QC, Canada. He has authored or coauthored more than 100 papers and authored five books. His research interests include cloud computing and security, distributed computing, and mobile computing. He is a Member of the IEEE Computer Society and a Standing Member of the Chinese Computer Education Society. He is the Editor of the *Journal of China Universities of Posts and Telecommunications* and the *Journal of Nanjing University of Posts and Telecommunications (Natural Science)*.