

Latency Aware 3D Placement and User Association in Drone-assisted Heterogeneous Networks with FSO-Based Backhaul

Shuai Zhang, *Student Member, IEEE*, and Nirwan Ansari, *Fellow, IEEE*

Abstract—Drone Base Station (DBS) is expected to play an important role in next generation cellular networks due to its maneuverability in establishing air-to-ground Line of Sight (LoS) links. Working as a mobile relay node between the Macro Base Station (MBS) and the ground users, the DBS can significantly reduce the latency of ground users by offloading traffic loads from the MBS to the DBS since a better channel condition can be provided. In this paper, we focus on how to determine the 3D location of the DBS, the user association and the bandwidth allocation policy between the MBS and the DBS in order to minimize the total average latency ratio of all users with the constraint of each user's QoS requirement and total available bandwidth. The formulated problem is a mixed integer nonconvex optimization problem, a very challenging and difficult problem. We propose a cyclic iterative algorithm to efficiently solve it by decomposing the primal problem into two subproblems, i.e., the user association and bandwidth allocation problem and the 3D DBS placement problem. In each iteration, the two sub-problems are alternatively optimized by using the output of one as the input for the other. Numerical simulation results demonstrate the significant latency ratio reduction of our proposed algorithm as compared to other baseline schemes.

I. INTRODUCTION

Nowadays, there have been increasing interests [1]–[4] in using Drone Base Stations (DBSs) as aerial platforms to facilitate terrestrial communications owing to their high maneuverability and flexibility for on-demand deployment. DBS is envisioned as a promising solution to improve the performance of current cellular networks for its ease of establishing Line of Sight (LoS) communication links with ground users [5]–[7]. Therefore, DBS-assisted heterogeneous network has many potential use cases, such as ground MBS traffic offloading, temporary events and Internet of Things (IoT) communication.

Free Space Optics (FSO), with its potential high capacity, can be utilized to build the connection between the DBS and the MBS (i.e., the backhaul link) in a DBS-assisted heterogeneous network [8]–[11]. In the meantime, employing FSO in the backhaul link will not cause interference or decrease the available bandwidth (because FSO works on a frequency range different from the radio frequency signal) in the access link. FSO, as shown in Fig. 1, is deployed to serve as a dedicated backhaul from an MBS to a DBS.

In the downlink communication, the mobile data (carried in the optical beam) is sent from the FSO transmitter to the FSO receiver, which are mounted on the MBS and the DBS, respectively. The DBS decodes the mobile data, encodes them on an Radio Frequency (RF) signal and finally transmits them to the corresponding ground user.

Deploying a DBS in an existing cellular network can significantly boost network throughput and improve QoS of users, but it must overcome the following challenges.

1) How to determine the location of the DBS?

The deployment of the DBS not only affects the user QoS but also influences the association policy (i.e., whether a user should be associated with the MBS or the DBS). For users associated with the DBS, moving the DBS closer to one user may decrease its latency ratio (i.e., the amount of time a user must be sacrificed in waiting for a unit service time) but at the expense of the latency ratio of another user. Therefore, the location of the DBS should be optimized to minimize the overall latency ratio while satisfying the user QoS requirements.

2) What is the association policy of the users?

Associating with an access point which has a better channel condition (i.e., a lower pathloss) will increase the data rate. Intuitively, more users should be associated with the DBS since a LoS communication link is more likely to be established between the DBS and a user. However, too many users associating with the DBS may increase the traffic load and thus significantly increase the latency ratio of the users. Therefore, it is necessary to properly design the user association policy to minimize the overall latency ratio.

3) What is the optimal bandwidth allocation scheme between the MBS and DBS?

Allocating more bandwidth to the MBS or DBS will decrease the latency ratio of its associated users; however, the latency ratio of users which are not associated with the access point will increase (less bandwidth becomes available for them). Meanwhile, the placement of the DBS, the user association policy and the bandwidth allocation scheme are mutually dependent. Therefore, the three subproblems mentioned above should be jointly considered to minimize the overall latency ratio.

To tackle the above challenges, we propose to decompose the joint optimization problem into two subproblems, i.e., the user association and bandwidth allocation problem and the DBS placement problem, to minimize the overall latency ratio. The main contributions of this paper are summarized as

The authors are with the Advanced Networking Laboratory, Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: sz355@njit.edu; nirwan.ansari@njit.edu).

This work was supported in part by the U.S. National Science Foundation under Grant No. CNS-1814748.

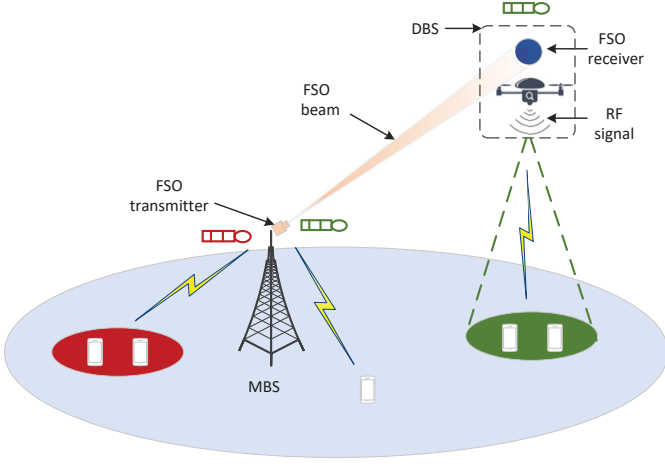


Fig. 1. DBS-assisted heterogeneous network with FSO-based backhaul.

follows.

1) We propose to deploy a DBS to serve the ground users by offloading their traffic from the MBS to the DBS. The FSO technique is employed to avoid interference with the access link and provide a high-capacity backhaul link. A latency ratio minimization problem is formulated to optimize the DBS's location, user association and bandwidth allocation subject to user QoS requirements and limited available bandwidth constraints.

2) We design a cyclic iterative algorithm to tackle the formulated problem where the primal problem is decoupled into two sub-problems, which are the bandwidth allocation and user association problem and 3D DBS placement problem, respectively. We prove the former problem to be a sum-of-ratios problem and solve it by utilizing the Lagrangian and modified Newton method. In the 3D DBS placement problem, we first determine the altitude that maximizes the coverage area. Then, we exhaustively search all the candidate locations to determine the optimal horizontal location which incurs the minimum latency ratio of all users. The performance of the proposed algorithm is validated via extensive simulations.

The rest of the paper is organized as follows. The related works are summarized in Section II. In Section III, we present the system model, and formulate the DBS placement, user association and spectrum resource allocation as an optimization problem. A cyclic iterative algorithm is designed to solve our problem in Section IV. Numerical results are presented to demonstrate the effectiveness of the proposed algorithm in Section V. Finally, we conclude the paper in Section VI.

II. RELATED WORKS

Many researches have been done on how to deploy DBSs in existing cellular networks. Duan *et al.* [12] studied the deployment of multi-Unmanned Aerial Vehicles (UAVs) to facilitate task offloading in a heterogeneous cloud aided system. A power consumption minimization problem is formulated by optimizing the task scheduling and resource allocation. Alzidaneen *et al.* [13] proposed to place multiple UAVs and tethered balloons (TBs) to serve users in a network where the MBS is malfunctioned. They formulated an optimization

problem aiming to maximize the end-to-end throughput by optimizing the UAVs' placement, transmit power and user associations. Huang *et al.* [14] tried to maximize the data rate of a device-to-device (D2D) pair by jointly optimizing the transmit power and 3D location of the UAV and the allocated bandwidth of ground terminals. However, they only considered one D2D pair in the network and assumed ground terminals are dominated by LoS connections. The assumption is not practical when ground terminals are in an urban area where high-rise buildings impair the LoS connection between the UAV and ground terminals. Xie *et al.* [15] utilized a UAV as the flying base station to serve ground users periodically in an infrastructure-less environment. The UAV employs the radio frequency wireless power transfer (WPT) to provide convenient and reliable energy supply to low-power IoT devices. They jointly optimized the UAV trajectory, the transmit power and mode to maximize the uplink minimum throughput subject to the UAV's maximum speed and users' energy neutrality constraints. Mozaffari *et al.* [16] proposed to deploy multiple UAVs working as mobile aerial base stations to dynamically serve IoT devices using uplink communication links. They designed a novel framework to minimize the total transmit power of the IoT devices by optimizing the 3D placement and mobility of the UAVs and the device-UAV association. Sun and Ansari [17] tried to deploy a DBS to help the MBS deliver traffic to the ground users. They formulated an problem to maximize the spectral efficiency of all users by optimizing the user association and the altitude of a DBS. However, the horizontal location of the DBS is assumed to be fixed, which is not appropriate if the user distribution does not follow a uniform process. Al-Hourani *et al.* [18] derived the optimal altitude of a DBS to maximize the coverage area of the DBS. Mozaffari *et al.* [26] studied the coverage rate and performance of a DBS-assisted heterogeneous network where the DBS coexists with underlaid D2D communication links. They analyzed two key cases, i.e., static DBS and mobile DBS, where the average downlink coverage probabilities for downlink users and D2D users are derived.

In our earlier paper [27], we considered the scenario where users are located in a hotspot area and tried to maximize the sum rate of all users in a DBS-assisted heterogeneous network with FSO-based backhaul by optimizing the DBS placement and resource allocation.

Different from the above works and our earlier work, in this paper, we try to jointly determine the user association, the spectrum allocation and the 3D location of the DBS to minimize the user latency ratio by considering the constraints of user QoS requirements and limited available bandwidth.

III. SYSTEM MODEL AND PROBLEM FORMULATION

As shown in Fig. 1, the latency of users can be reduced by offloading traffic from the MBS to the DBS since a better channel condition can be provisioned. The FSO link is working as the backhaul due to its high potential capacity. We assume that the Frequency Division Multiple Access (FDMA) mode is adopted in the access link to avoid interference among users.

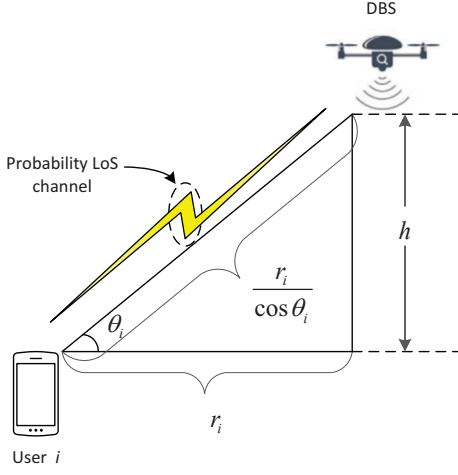


Fig. 2. Probabilistic pathloss model.

A. Pathloss model between a DBS and a user

Denote $\mathcal{I}$ as the set of users and i as the index of users. The wireless propagation channel between the DBS and user i is modeled by a probabilistic LoS channel, and the probability of having an LoS connection between the DBS and user i is [18]

$$P_i^{LoS} = \frac{1}{1 + a \exp(-b(\frac{180}{\pi} \arctan \varphi_i - a))}, \quad (1)$$

where a and b are constant values determined by the surrounding environment, and φ_i is the elevation angle between user i and the DBS. As shown in Fig. 2, we assume the DBS is flying at the altitude of h and the horizontal distance between the DBS and user i is r_i . The probability of establishing an NLoS link can be obtained as $P_i^{NLoS} = 1 - P_i^{LoS}$. Therefore, the average pathloss between user i and the DBS is [18]

$$\begin{aligned} \eta_i^D &= \mathcal{G}(\varphi_i, r_i) \\ &= 20 \log \left(\frac{4\pi f \frac{r_i}{\cos(\varphi_i)}}{c} \right) + P_i^{LoS} \eta_{LoS} + P_i^{NLoS} \eta_{NLoS}, \end{aligned} \quad (2)$$

where η_{LoS} and η_{NLoS} are the average additional pathloss for LoS and NLoS connections, respectively, f denotes the carrier frequency, and c stands for the speed of light.

Definition 1. We define the maximum allowed horizontal distance between the DBS and user i (i.e., r_i^{max}) as the maximum value of r_i where the pathloss requirement of user i is still satisfied (i.e., $\mathcal{G}(\varphi_i, r_i) \leq \phi_i$, where ϕ_i is the pathloss requirement).

The maximum value of r_i^{max} can be obtained when $\mathcal{G}(\varphi_i, r_i) = \phi_i$. By setting the first derivative of $\mathcal{G}(\varphi_i, r_i)$ w.r.t. φ_i equal to zero, i.e., $\frac{\partial \mathcal{G}(\varphi_i, r_i)}{\partial \varphi_i} = 0$, the optimal elevation angle (i.e., φ_i^*) between user i and DBS can be derived by solving [19]:

$$\frac{\pi}{9 \ln(10)} \tan \varphi_i^* + \frac{ab(\eta_{LoS} - \eta_{NLoS}) \exp(-b(\frac{180}{\pi} \varphi_i^* - a))}{(a \exp(-b(\frac{180}{\pi} \varphi_i^* - a)) + 1)^2} = 0. \quad (3)$$

By substituting $\varphi_i = \varphi_i^*$ into $\mathcal{G}(\varphi_i, r_i) = \phi_i$, r_i^{max} can be expressed as

$$r_i^{max} = \frac{c \cos \varphi_i^*}{4\pi f} 10^{\frac{\phi_i - P_i^{LoS} \eta_{LoS} - P_i^{NLoS} \eta_{NLoS}}{20}}. \quad (4)$$

Definition 2. Given user i ' pathloss requirement ϕ_i , the optimal flying height of the DBS w.r.t. user i is defined as the flying height which maximizes r_i , i.e., $h_i^* = r_i^{max} \varphi_i^*$.

Lemma 1. Assuming the flying height of DBS is h_i^* , and the horizontal distance between the DBS and user i is r_i^{max} , we have $\mathcal{G}(\varphi_i, r_i^{max}) = \phi_i$. Then, user i 's pathloss requirement can always be satisfied (i.e., $\mathcal{G}(\varphi_{i'}, r_{i'}) \leq \phi_{i'}$) if user i 's pathloss requirement $\phi_{i'} \geq \phi_i$ and its horizontal distance to the DBS $r_{i'} \leq r_i^{max}$.

Proof: Since $r_{i'} \leq r_i^{max}$ and the flying height of DBS is h_i^* , we can derive $\varphi_{i'} \geq \varphi_i^*$ (note that $\varphi_i = \arctan(\frac{h}{r_i})$, and further $P_{i'}^{LoS} \geq P_i^{LoS}$ (note that P^{LoS} is an increasing function of the elevation angle φ). Using the fact that $\eta_{LoS} \leq \eta_{NLoS}$, we can easily have

$$\mathcal{G}(\varphi_{i'}, r_{i'}) \leq \mathcal{G}(\varphi_i^*, r_i^*). \quad (5)$$

Given that ϕ_i , i.e., $\mathcal{G}(\varphi_i^*, r_i^*) = \phi_i$, it is not difficult to obtain the following equation

$$\mathcal{G}(\varphi_{i'}, r_{i'}) \leq \phi_i \leq \phi_{i'}. \quad (6)$$

That is, the pathloss requirement of user i can always be satisfied. ■

B. Traffic load model of the MBS

We assume that the traffic of user i is generated according to a Poisson process with the arrival rate λ_i . The traffic sizes of all requests follow a general distribution with the average value of v_i . Therefore, the average traffic load of user i can be obtained as $\lambda_i v_i$.

The data rate of user i which is associated with the MBS, denoted as r_i^M , can be expressed as

$$r_i^M = \beta^M \log_2 \left(1 + \frac{P^M 10^{-\frac{\eta_i^M}{10}}}{\sigma^2} \right) \quad (7)$$

where P^M is the transmit power of the MBS, β^M is the amount of bandwidth allocated to the MBS, σ^2 is the environment noise power, and η_i^M is the pathloss between user i and the MBS, which can be modeled as

$$\eta_i^M = \alpha + \gamma \log_{10}(d_i^M). \quad (8)$$

Here, α is the path loss at the reference distance and γ is the path loss exponent, and d_i^M is the distance between the MBS and user i , i.e.,

$$d_i^M = \sqrt{(x_i - x^M)^2 + (y_i - y^M)^2}, \quad (9)$$

where (x_i, y_i) and (x^M, y^M) are the locations of user i and the MBS, respectively. The average utilization of the MBS that indicates the fraction of time during which the MBS is busy serving user i can be calculated as

$$\rho_i^M = \frac{\lambda_i v_i}{r_i^M}. \quad (10)$$

Thus, by summing up the traffic of all users associated with the MBS, the average utilization of the MBS can be calculated as

$$\rho^M = \sum_{i=1}^{|I|} \rho_i^M (1 - \theta_i) = \frac{\lambda_i v_i (1 - \theta_i)}{r_i^M}, \quad (11)$$

where θ_i is the binary variable to indicate whether user i is associated with the DBS (i.e., $\theta_i = 1$) or not (i.e., $\theta_i = 0$).

Assume that the traffic arrival of users associated with the MBS are independent. The arrival stream, which is formed by merging all the traffic from users associated with the MBS, is also a Poisson process (recall that the traffic arrival of each user is a Poisson process). Thus, user i 's (associated with the MBS) service time can be obtained as $s_i^M = \frac{v_i}{r_i^M}$. In addition, the traffic size v_i follows a general distribution such that the service time of user i also satisfies a general distribution. Therefore, based on queuing theory, the MBS's downlink transmission process realizes an M/G/1 processor sharing queue. For user i which is associated with the MBS, the corresponding average traffic delivery time, including the waiting time and service time, is [20]

$$T_i^M = \frac{s_i^M}{1 - \rho^M}. \quad (12)$$

Based on the average traffic delivery time, the average traffic latency ratio τ_i^M can be calculated as

$$\tau_i^M = \frac{T_i^M - s_i^M}{s_i^M} = \frac{\rho^M}{1 - \rho^M}, \quad (13)$$

where the value of τ_i^M implies the waiting time of user i to receive a unit service time. From (13), we can observe that τ_i^M is independent of user index i , which indicates that the average latency ratio of users associated with the MBS share the same value, i.e.,

$$\tau^M = \frac{\rho^M}{1 - \rho^M}. \quad (14)$$

C. Traffic load model of the DBS

Note that a two-hop communication is incurred when users are associated with the DBS. Therefore, two queues are generated in the access link (the link between a user and the DBS) and the backhaul link (the link between the DBS and the MBS), respectively. Following the similar derivation process, the average latency ratio of users associated with the DBS in the access link can be obtained as

$$\tau^{D,a} = \frac{\rho^{D,a}}{1 - \rho^{D,a}}, \quad (15)$$

where $\rho^{D,a}$ is the average utilization of users (associated with the DBS) in the access link, i.e.,

$$\rho^{D,a} = \sum_{i=1}^{|I|} \rho_i^{D,a} \theta_i = \frac{\lambda_i v_i \theta_i}{r_i^{D,a}}, \quad (16)$$

where $r_i^{D,a}$ is user i 's data rate of the access link. Thus, we have

$$r_i^{D,a} = \beta^M \log_2 \left(1 + \frac{P^D 10^{-\frac{\eta_i^D}{10}}}{\sigma^2} \right), \quad (17)$$

where P^D is the transmit power of the DBS, β^D is the amount of bandwidth allocated to the DBS and η_i^D is the pathloss between the DBS and user i .

Similarly, the average latency ratio in the backhaul link can be obtained as

$$\tau^{D,b} = \frac{\rho^{D,b}}{1 - \rho^{D,b}}, \quad (18)$$

where $\rho^{D,b}$ is the average utilization of the backhaul link, i.e.,

$$\rho^{D,b} = \sum_{i=1}^{|I|} \rho_i^{D,b} \theta_i = \frac{\lambda_i v_i \theta_i}{r_i^{D,b}}, \quad (19)$$

where $r_i^{D,b}$ is user i 's data rate of the FSO-based backhaul link, which can be obtained as [21], [22]

$$r_i^{D,b} = \frac{P_t \eta_t \eta_r 10^{-\frac{L_{atm}}{10}} 10^{-\frac{L_{geo}}{10}}}{E_p N_b}, \quad (20)$$

where P_t is the transmission power of the laser, and η_t and η_r denote the transmitting efficiency and receiving efficiency, respectively. $E_p = h_p c / \lambda_c$ is the photon energy. λ_c is the carrier wavelength, h_p denotes Planck's constant. N_b implicates the receiver sensitivity (photons/bit). $L_{geo} = 10 \log \left(\frac{\pi r^2}{\pi (\psi_t l / 2)^2} \right)$ is the geometrical loss in dB, l is the distance between the laser transmitter and receiver in Km, r is the radius of the receiver's aperture in m, and ψ_t denotes the transmitting divergence angle. $L_{atm} = \frac{17}{V} \left(\frac{\lambda}{550nm} \right)^{-q}$ stands for the atmospheric attenuation caused by bad weather conditions, where L_{atm} is in dB/Km, q is the size distribution of the scattering particles under different weather conditions, V is the visibility in Km. The value of q is determined by the value of the visibility distance [23]:

$$q = \begin{cases} 1.6, & V > 50 \text{ km} \\ 1.3, & 6 \text{ km} < V \leq 50 \text{ km} \\ 0.16V + 0.34, & 1 \text{ km} < V \leq 6 \text{ km} \\ V - 0.5, & 0.5 \text{ km} < V \leq 1 \text{ km} \\ 0, & V \leq 0.5 \text{ km} \end{cases} \quad (21)$$

D. Problem formulation

As mentioned previously, the deployment of the DBS reduces users' latency by enabling the traffic offloaded from the MBS to the DBS since a better channel condition is provided. Our goal is to minimize the overall latency ratio of users while considering the QoS requirements of users and the limited available resource. Specifically,

$$\begin{aligned}
\mathbf{P0}: \quad & \min_{x,y,h,\theta_i,\beta^M,\beta^D} \sum_{i=1}^{|\mathcal{I}|} \tau^{D,a} + \tau^{D,b} + \tau^M \quad (22) \\
\text{s.t.} \quad & r_{ij}^{D,a} \theta_i + r_i^M (1 - \theta_i) \geq T_i, \forall i \in \mathcal{I}, \quad (23) \\
& \tau^{D,a} + \tau^{D,b} \leq \epsilon, \forall i \in \mathcal{I}, \quad (24) \\
& \tau^M \leq \epsilon, \forall i \in \mathcal{I}, \quad (25) \\
& \beta^M + \beta^D = B, \quad (26) \\
& \beta^M, \beta^D \geq 0, \quad (27) \\
& \theta_i \in \{0, 1\}, \forall i \in \mathcal{I}, \quad (28) \\
& 0 \leq \rho^{D,a}, \rho^{D,b}, \rho^M < 1, \quad (29)
\end{aligned}$$

where T_i indicates user i 's data rate requirement, and x, y , and h are the 3D location of the DBS. ϵ denotes the latency ratio requirement. B is the total available bandwidth. Constraints (23), (24) and (25) enforce the QoS and latency requirement of each user. Constraint (26) stands for the resource limitations. Constraint (27) imposes the bandwidth allocated to the MBS and DBS to be non-negative. Constraint (28) imposes θ_i to be a binary variable. Constraint (29) ensures the stability of the queuing system.

Since the latency ratio is a monotonically increasing function of utilization if $0 \leq \rho < 1$ (remember $\tau = \frac{\rho}{1-\rho}$), we can minimize the latency ratio by minimizing the utilization. Meanwhile, we omit the latency ratio incurred in the backhaul link since the capacity of FSO is sufficiently high. According to [21], [22], the data rate of a FSO link can reach 10 Gbps under clear weather condition within the range of 1 Km. Therefore, **P0** can be transformed into the following problem

$$\begin{aligned}
\mathbf{P0-a}: \quad & \min_{x,y,h,\theta_i,\beta^M,\beta^D} \rho^{D,a} + \rho^M \\
\text{s.t.} \quad & \tau^{D,a} \leq \epsilon, \forall i \in \mathcal{I}, \quad (30) \\
& (23), (25) - (29).
\end{aligned}$$

Unfortunately, **P0-a** is a mixed integer non-linear non-convex problem and thus challenging to solve considering Eq. (2) and Constraint (23). Next, we propose an efficient framework to solve this optimization problem. In essence, we partition the entire decision variables into two blocks. In each iteration, the bandwidth allocation and user association policy, and the location of the DBS are alternately optimized, i.e., one block is optimized at each iteration while keeping the other block fixed. We iteratively optimize these two blocks until the user association, DBS's location and bandwidth allocation are determined.

IV. DBS PLACEMENT, BANDWIDTH ALLOCATION AND USER ASSOCIATION

To make **P0-a** more tractable, we decompose this joint optimization problem into two subproblems, i.e., the bandwidth allocation and user association problem and the DBS placement problem.

A. Bandwidth allocation and user association

In this subproblem, for any fixed location of the DBS, we determine the amount of bandwidth allocated to the MBS and DBS and whether each user should be associated with the MBS or DBS to minimize the overall latency ratio. The bandwidth allocation and user association problem can be formulated as

$$\begin{aligned}
\mathbf{P1}: \quad & \min_{\theta_i, \beta^M, \beta^D} \sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i \theta_i}{r_{ij}^{D,a}} + \sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i (1 - \theta_i)}{r_i^M} \quad (31) \\
\text{s.t.} \quad & (23), (25) - (30).
\end{aligned}$$

Since $\tau = \frac{\rho}{1-\rho} \leq \epsilon$ is equivalent to $\rho \leq \frac{\epsilon}{1+\epsilon}$, we can derive that Constraint (29) is always satisfied if Constraints (30) and (25) are satisfied. Problem **P1** can be simplified by removing Constraint (29). The problem is still challenging due to the non-continuity of θ_i . We relax θ_i as a continuous variable (i.e., $0 \leq \theta_i \leq 1, \forall i \in \mathcal{I}$). Finally, Problem **P1** can be rewritten as

$$\begin{aligned}
\mathbf{P1-a}: \quad & \min_{\theta_i, \beta^M, \beta^D} \sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i \theta_i}{\beta^D \iota_{ij}} + \sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i (1 - \theta_i)}{\beta^M \kappa_i} \quad (32) \\
\text{s.t.} \quad & \beta^M \kappa_i \theta_i - \beta^D \iota_{ij} \theta_i - \beta^M \kappa_i + T_i \leq 0, \quad (33) \\
& \sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i}{\iota_{ij}} \theta_i - \frac{\epsilon}{\epsilon + 1} \beta^D \leq 0, \quad (34) \\
& \sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i}{\kappa_i} (1 - \theta_i) - \frac{\epsilon}{\epsilon + 1} \beta^M \leq 0, \quad (35) \\
& 0 \leq \theta_i \leq 1, \forall i \in \mathcal{I}, \quad (36) \\
& (26), (27),
\end{aligned}$$

where $\iota_{ij} = \log_2(1 + \frac{P^D 10^{-\frac{\eta_{ij}^D}{10}}}{\sigma^2})$ and $\kappa_i = \log_2(1 + \frac{P^M 10^{-\frac{\eta_i^M}{10}}}{\sigma^2})$. We can see that all the constraints are linear except Constraint (33). If we can prove that Constraint (33) is a convex set, **P1-a** is a sum-of-ratios problem [24] since the numerators are all convex and the denominators are all concave (note that a linear function is both convex and concave).

Next, we prove that Constraint (33) is a convex function. Since $-\beta^M \kappa_i + T_i$ is linear, it will not affect the convexity of Constraint (33). Thus, proving Constraint (33) is convex is equivalent to proving $\beta^M \kappa_i \theta_i - \beta^D \iota_{ij} \theta_i$ is convex. Its Hessian matrix w.r.t. β^D, β^M and θ_i can be derived as

$$H = \begin{bmatrix} 0 & 0 & -\iota_{ij} \\ 0 & 0 & \kappa_i \\ -\iota_{ij} & \kappa_i & 0 \end{bmatrix}. \quad (37)$$

Since ι_{ij}, κ_i are all positive, it can be easily proved that H is positive semi-definite; so Constraint (33) is a convex set, and **P1-a** is a sums-of-ratios problem which can be resolved by applying algorithm proposed in [25]. It is easy to see that **P1-a** is equivalent to the following problem

$$\mathbf{P1-b:} \quad \min_{\theta_i, \omega_i, \beta^M, \beta^D} \sum_{i=1}^{2|\mathcal{I}|} \omega_i \quad (38)$$

$$s.t. \quad \frac{\lambda_i v_i \theta_i}{\beta^D \iota_{ij}} \leq \omega_i, \quad 1 \leq i \leq |\mathcal{I}| \quad (39)$$

$$\frac{\lambda_i v_i (1 - \theta_i)}{\beta^M \kappa_i} \leq \omega_{i+|\mathcal{I}|}, \quad 1 \leq i \leq |\mathcal{I}| \quad (40)$$

(26), (27), (33) – (36),

where ω_i is the introduced relaxation variable. Then, the Lagrangian function of **P1-b** is

$$\begin{aligned} \mathcal{L}(\theta, \beta, \omega, \mu, \varrho) = & \sum_{i=1}^{2|\mathcal{I}|} \omega_i + \sum_{i=1}^{|\mathcal{I}|} \mu_i (\lambda_i v_i \theta_i - \omega_i \beta^D \iota_{ij}) \\ & + \sum_{i=1}^{|\mathcal{I}|} \mu_{i+|\mathcal{I}|} [\lambda_i v_i (1 - \theta_i) - \omega_{i+|\mathcal{I}|} \beta^M \kappa_i] \\ & + \sum_{i=1}^{|\mathcal{I}|} \varrho_{i+2|\mathcal{I}|} (\beta^M \kappa_i \theta_i - \beta^D \iota_{ij} \theta_i - \beta^M \kappa_i + T_i) \\ & + \varrho_{3|\mathcal{I}|+1} \left(\sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i}{\iota_{ij}} \theta_i - \frac{\epsilon}{\epsilon+1} \beta^D \right) \\ & + \varrho_{3|\mathcal{I}|+2} \left[\sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i}{\kappa_i} (1 - \theta_i) - \frac{\epsilon}{\epsilon+1} \beta^M \right] \\ & + \varrho_{3|\mathcal{I}|+3} (\beta^M + \beta^D - B) \\ & + \sum_{i=1}^{|\mathcal{I}|} \varrho_i (-\theta_i) + \sum_{i=1}^{|\mathcal{I}|} \varrho_{i+|\mathcal{I}|} (\theta_i - 1) \end{aligned} \quad (41)$$

where μ and ϱ are the Lagrangian multipliers of the constraints.

The Karush-Kuhn-Tucker (KKT) conditions are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \theta_i} = & \mu_i \lambda_i v_i - \mu_{i+|\mathcal{I}|} \lambda_i v_i + \varrho_{i+2|\mathcal{I}|} (\beta^M \kappa_i - \beta^D \iota_{ij}) \\ & + \varrho_{3|\mathcal{I}|+1} \frac{\lambda_i v_i}{\iota_{ij}} - \varrho_{3|\mathcal{I}|+2} \frac{\lambda_i v_i}{\kappa_i} - \varrho_i + \varrho_{i+|\mathcal{I}|} = 0, \end{aligned} \quad (42)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \beta^D} = & \sum_{i=1}^{|\mathcal{I}|} -\mu_i \omega_i \iota_{ij} - \sum_{i=1}^{|\mathcal{I}|} \varrho_{i+2|\mathcal{I}|} \iota_{ij} \theta_i \\ & - \varrho_{3|\mathcal{I}|+1} \frac{\epsilon}{\epsilon+1} + \varrho_{3|\mathcal{I}|+3} = 0, \end{aligned} \quad (43)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \beta^M} = & \sum_{i=1}^{|\mathcal{I}|} -\mu_{i+|\mathcal{I}|} \omega_{i+|\mathcal{I}|} \kappa_i - \sum_{i=1}^{|\mathcal{I}|} \varrho_{i+2|\mathcal{I}|} (\kappa_i \theta_i - \kappa_i) \\ & - \varrho_{3|\mathcal{I}|+2} \frac{\epsilon}{\epsilon+1} + \varrho_{3|\mathcal{I}|+3} = 0, \end{aligned} \quad (44)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \omega_i} = & \begin{cases} 1 - \mu_i \beta^D \iota_{ij}, & 1 \leq i \leq |\mathcal{I}|, \\ 1 - \mu_i \beta^M \kappa_{i-|\mathcal{I}|}, & |\mathcal{I}| + 1 \leq i \leq 2|\mathcal{I}|, \end{cases} \\ & = 0, \end{aligned} \quad (45)$$

$$\begin{aligned} \mu_{i+|\mathcal{I}|} [\lambda_i v_i (1 - \theta_i) - \omega_{i+|\mathcal{I}|} \beta^M \kappa_i] &= 0, \\ \mu_i (\lambda_i v_i \theta_i - \omega_i \beta^D \iota_{ij}) &= 0, \quad 1 \leq i \leq |\mathcal{I}|, \end{aligned} \quad (46)$$

$$\varrho_i (-\theta_i) = 0, \quad \varrho_{i+|\mathcal{I}|} (\theta_i - 1) = 0, \quad (47)$$

$$\varrho_{i+2|\mathcal{I}|} (\beta^M \kappa_i \theta_i - \beta^D \iota_{ij} \theta_i - \beta^M \kappa_i + T_i) = 0, \quad (48)$$

$$\varrho_{3|\mathcal{I}|+1} \left(\sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i}{\iota_{ij}} \theta_i - \frac{\epsilon}{\epsilon+1} \beta^D \right) = 0, \quad (49)$$

$$\varrho_{3|\mathcal{I}|+2} \left[\sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i}{\kappa_i} (1 - \theta_i) - \frac{\epsilon}{\epsilon+1} \beta^M \right] = 0, \quad (50)$$

$$\varrho_{3|\mathcal{I}|+3} (\beta^M + \beta^D - B) = 0, \quad (51)$$

(26), (27), (33) – (36).

Eq. (45) is equivalent to

$$\begin{cases} \mu_i = \frac{1}{\beta^D \iota_{ij}}, & 1 \leq i \leq |\mathcal{I}|, \\ \mu_{i+|\mathcal{I}|} = \frac{1}{\beta^M \kappa_i}, & 1 \leq i \leq |\mathcal{I}|, \end{cases} \quad (52)$$

which is substituted into (46) to yield

$$\begin{cases} \lambda_i v_i (1 - \theta_i) - \omega_{i+|\mathcal{I}|} \beta^M \kappa_i = 0, & 1 \leq i \leq |\mathcal{I}|, \\ \lambda_i v_i \theta_i - \omega_i \beta^D \iota_{ij} = 0, & 1 \leq i \leq |\mathcal{I}|. \end{cases} \quad (53)$$

Note that Eqs.(42)-(44), Eqs. (47)-(51) and Eqs. (33)-(36) are KKT conditions of problem **P2** if ω_i and μ_i are fixed.

$$\begin{aligned} \mathbf{P2:} \quad \min_{\theta_i, \beta^M, \beta^D} \quad & \sum_{i=1}^{|\mathcal{I}|} \mu_i (\lambda_i v_i \theta_i - \omega_i \beta^D \iota_{ij}) \\ & + \sum_{i=1}^{|\mathcal{I}|} \mu_{i+|\mathcal{I}|} [\lambda_i v_i (1 - \theta_i) - \omega_{i+|\mathcal{I}|} \beta^M \kappa_i] \\ s.t. \quad & (26), (27), (33) - (36). \end{aligned} \quad (54)$$

For fixed μ and ϱ , the objective function of **P2** is convex and thus **P2** is a convex optimization problem. Therefore, we conclude that the solution of problem **P2** can be obtained by finding those satisfying Eqs. (52) and (53) among the solutions of the convex problem **P2**.

Denote $\mathbf{x} = [\omega; \mu]$ and $\theta_i(x)$, $\beta^D(x)$ and $\beta^M(x)$ as the solution of problem **P2**. To satisfy Eqs. (52) and (53), we have ($1 \leq i \leq |\mathcal{I}|$)

$$\begin{cases} \mu_i \beta^D \iota_{ij} - 1 = 0, \\ \mu_{i+|\mathcal{I}|} \beta^M \kappa_i - 1 = 0, \\ \omega_i \beta^D \iota_{ij} - \lambda_i v_i \theta_i = 0, \\ \omega_{i+|\mathcal{I}|} \beta^M \kappa_i - \lambda_i v_i (1 - \theta_i) = 0. \end{cases} \quad (55)$$

Let $\phi_i(x) = \mu_i \beta^D \iota_{ij} - 1$, $\phi_{i+|\mathcal{I}|}(x) = \mu_{i+|\mathcal{I}|} \beta^M \kappa_i - 1$, $\phi_{i+2|\mathcal{I}|}(x) = \omega_i \beta^D \iota_{ij} - \lambda_i v_i \theta_i$ and $\phi_{i+3|\mathcal{I}|}(x) = \omega_{i+|\mathcal{I}|} \beta^M \kappa_i - \lambda_i v_i (1 - \theta_i)$. Thus, we have

$$\begin{aligned}\phi(\mathbf{x}) &= [\phi_1(\mathbf{x}), \dots, \phi_{|\mathcal{I}|}(\mathbf{x}), \dots, \phi_{2|\mathcal{I}|}(\mathbf{x}), \\ &\quad \dots, \phi_{3|\mathcal{I}|}(\mathbf{x}), \dots, \phi_{4|\mathcal{I}|}(\mathbf{x})] \\ &= 0.\end{aligned}\quad (56)$$

Thus, the solution of **P1-b** can be obtained by finding those satisfying Eq. (56) among the solutions of **P2**.

Lemma 2. $\phi(\mathbf{x})$ is strongly monotone.

Proof: Let $f_i(\mathbf{x}) = \beta^D \iota_{ij} > 0$ and $f_{i+|\mathcal{I}|}(\mathbf{x}) = \beta^M \kappa_i > 0$ for simplicity. The Jacobian matrix of $\phi(\mathbf{x})$ can be derived based on Eqs. (55) and (56) as

$$\phi'(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & f_{2|\mathcal{I}|}(\mathbf{x}) & 0 & \cdots & 0 \\ 0 & \cdots & 0 & f_1(\mathbf{x}) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & f_{2|\mathcal{I}|}(\mathbf{x}) \end{bmatrix}$$

We can observe that $\phi'(\mathbf{x})$ is a positive definite matrix because $f_i(\mathbf{x}) > 0$, ($1 \leq i \leq 2|\mathcal{I}|$). Therefore, $\phi(\mathbf{x})$ is strongly monotone.

The modified Newton method can be used to obtain the solution of Eq. (56). In each iteration, we update $\mathbf{x}$ by calculating

$$\mathbf{x}_{m+1} = \mathbf{x}_m + \delta_m \zeta_m, \quad (57)$$

where m denotes the iteration number, and ζ_m denotes the search direction defined as

$$\begin{aligned}\zeta_m &= -[\phi'(\mathbf{x})]^{-1} \phi(\mathbf{x}) \\ &= - \begin{bmatrix} \frac{1}{f_1(\mathbf{x})} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{f_{2|\mathcal{I}|}(\mathbf{x})} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \frac{1}{f_1(\mathbf{x})} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \frac{1}{f_{2|\mathcal{I}|}(\mathbf{x})} \end{bmatrix} \begin{bmatrix} \phi_1(\mathbf{x}) \\ \vdots \\ \vdots \\ \phi_{4|\mathcal{I}|}(\mathbf{x}) \end{bmatrix} \\ &= - \begin{bmatrix} \phi_1(\mathbf{x}) & \cdots & \phi_{2|\mathcal{I}|}(\mathbf{x}) & \phi_{2|\mathcal{I}|+1}(\mathbf{x}) & \cdots & \phi_{4|\mathcal{I}|}(\mathbf{x}) \\ \frac{1}{f_1(\mathbf{x})} & \cdots & \frac{1}{f_{2|\mathcal{I}|}(\mathbf{x})} & \frac{1}{f_1(\mathbf{x})} & \cdots & \frac{1}{f_{2|\mathcal{I}|}(\mathbf{x})} \end{bmatrix}^T\end{aligned}$$

where $[\cdot]^T$ denotes the transpose operation and $\delta_m \in (0, 1)$ is the step length.

B. 3D DBS placement

For fixed bandwidth allocation and user association, we optimize the 3D location of the DBS aiming to minimize the overall latency ratio, i.e.,

$$\mathbf{P3:} \min_{x,y,h} \frac{1}{\beta^D} \sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i}{\iota_{ij}} \quad (58)$$

$$\text{s.t.} \quad \beta^D \iota_{ij} \geq T_i, \forall i \in \mathcal{I}_1 \quad (59)$$

$$\sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i}{\iota_{ij}} \leq \frac{\epsilon}{\epsilon + 1} \beta^D, \quad (60)$$

where $\mathcal{I}_1$ is the set of users associated with the DBS. Note that the second part of the objective function is omitted (which is incurred by users associated with the MBS) since it will stay fixed once the bandwidth allocation and user association are given. As β^D and T_i are known in **P3**, Constraint (59) can be rewritten as

$$\eta_{ij}^D \leq -10 \lg[(2^{\frac{T_i}{\beta^D}} - 1) \frac{\sigma^2}{P^D}]. \quad (61)$$

Now the user data rate requirement is dictated by pathloss which is given by the above equation. Therefore, violating the pathloss requirement is equivalent to violating Constraint (59). We can further see that the minimal value of the objective function in **P3** should satisfy Constraint (60), otherwise, there would be no feasible solution for problem **P3**. Based on the above observations, we design an efficient heuristic algorithm to solve problem **P3**.

To make problem **P3** more tractable, we try to partition the decision variables of the 3D DBS location into two blocks (i.e., the horizontal location and vertical location) and solve them separately. In the vertical dimension, the flying height of the DBS is determined based on Definition 2. In the horizontal dimension, the longitude and latitude of the DBS is obtained by exhaustively searching for all the candidate locations that minimize the latency ratio of users that are associated with the DBS. Specifically,

1) In the horizontal dimension, we divide the coverage area into a number of locations with the same size. Denote $\mathcal{K}$ as the set of these locations and k as the index of these locations. These locations would be the candidate locations that the DBS can be placed in the horizontal dimension.

2) In the vertical dimension, find user $\bar{i}$ by calculating $\bar{i} = \arg \min_i \left\{ -10 \lg[(2^{\frac{T_i}{\beta^D}} - 1) \frac{\sigma^2}{P^D}] \mid i \in \mathcal{I}_1 \right\}$. Then, we set the flying height of the DBS be $h_{\bar{i}}^*$. Based on Lemma 1, we can conclude that the pathloss requirements (i.e., data rate requirements) of users that located within the coverage area of the DBS can always be satisfied.

3) We exhaustively search all candidate locations in the horizontal dimension with the flying height of $h_{\bar{i}}^*$. The optimal location index k^* of the DBS will be the one which incurs the minimum value of $\sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i}{\iota_{ij}}$, i.e.,

$$k^* = \arg \max_k \left\{ \sum_{i=1}^{|\mathcal{I}|} \frac{\lambda_i v_i}{\iota_{ij}} \mid k \in \mathcal{K} \right\}. \quad (62)$$

The joint DBS placement, bandwidth allocation and user association algorithm to solve the primal problem P0-a is embodied in Algorithm 1. In Line 1, we initialize the 3D location of the DBS. In Line 3, given the location, we obtain the bandwidth allocation and user association by solving problem P1-a. Given the bandwidth allocation and user association, Lines 4-5 determine the altitude of the DBS, while Lines 6-10 calculate the horizontal location of the DBS. We repeat Lines 3-10 until the algorithm converges.

The complexity of Step 3 is $O(|\mathcal{I}|^3)$; that of Steps 4-5 is $O(|\mathcal{I}|)$; that of Steps 6-8 is $O(|\mathcal{I}||\mathcal{K}|)$; that of Step 9 is also

$O(|\mathcal{I}||\mathcal{K}|)$ and they can repeat for at most $O(|\mathcal{K}|)$ times. Thus, the complexity of Algorithm 1 is $O(|\mathcal{I}|^3 + |\mathcal{K}|^2|\mathcal{I}|)$.

Algorithm 1

- 1: Initialize the location of the DBS x, y, h .
- 2: **while** the value of Eq. (31) decreases **do**
- 3: Compute θ_i, β^M and β^D by solving problem **P1-a**;
- 4: Identify user $\bar{i}$,
 where $\bar{i} = \arg \min_i \left\{ -10 \lg \left[\left(2^{\frac{T_i}{\beta^D}} - 1 \right) \frac{\sigma^2}{P^D} \right] \mid i \in \mathcal{I}_1 \right\}$
- 5: Let the flying height of the DBS $h = h_i^*$;
- 6: **for** $k = 1$ to $|\mathcal{K}|$ **do**
- 7: Calculate the value of Eq. (31) in each iteration k ;
- 8: **end for**
- 9: Obtain horizontal location index k^* based on Eq. (62);
- 10: Calculate x, y accordingly and let $h = h_i^*$.
- 11: **end while**

V. NUMERICAL RESULTS

We next present numerical results to demonstrate the effectiveness of our proposed algorithm. We consider an area with the size of 500 m by 500 m. The locations of the ground users follow the spatial Poisson point process with density of $\lambda_1 = 400 \text{ users/km}^2$. The size of each location k is 10 m by 10 m. Users' data rate requirements are generated based on the Poisson distribution with the expectation of 50 Kbps. Other parameters are given in Table II. MATLAB R2019a is used for the simulations, which are run on a macbook laptop with Quad-Core intel Core i5-8257U and 8 GB RAM. We repeat each simulation five times to obtain the average value. Table I shows the runtime of each scheme where the number of users are set to 100 and 200, respectively. The other parameters are the same as shown in Table II. We can see the proposed algorithm takes longer time to converge since a dynamic user association and bandwidth allocation policy is utilized to achieve better performance. The runtime of the 'Stationary DBS' scheme is longer than the 'MBS only' scheme because the bandwidth allocation and user association is the same as our proposed algorithm. While in the 'MBS only' scheme, all the bandwidth are allocated to the MBS and all the users are associated with the MBS.

TABLE I
RUNTIME EXPERIMENT RESULTS

Algorithms	100 users	200 users
	time (sec)	time (sec)
Proposed Algorithm	2.93	4.62
Stationary DBS	1.13	1.85
MBS only	0.91	1.47

Next, we compare the performance of our algorithm with the following two schemes: 1) Stationary DBS, where the DBS is placed at the geometrical center with the flying height $h = 30\text{m}$. Meanwhile, the bandwidth allocation and user association policy are optimized based on our proposed approach. 2) MBS only, where no DBS is deployed to assist the MBS in the existing cellular network. In this scheme,

TABLE II
SIMULATION PARAMETERS

Parameters	Definitions	Values
a	Environment parameter	9.61
b	Environment parameter	0.16
η_{LoS}	Average additional pathloss of LoS	1 dB
η_{NLoS}	Average additional pathloss of NLoS	20 dB
f	Carrier frequency	2 GHz
σ^2	Noise power	-140 dBm
B	Total available bandwidth	20 MHz
P^D	Transmit power of DBS	0.5 W
P^M	Transmit power of MBS	2 W
λ_i	Arrival rate of user i	0.1 request/s
v_i	Average traffic size of user i	100 Kb
ϵ	Latency ratio requirement	2

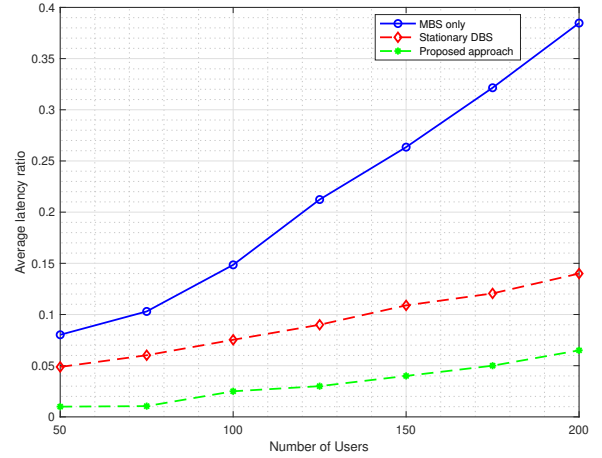


Fig. 3. The average latency ratio and number of users.

all the bandwidth is allocated to the MBS and all users are associated with the MBS.

Fig. 3 illustrates how the total number of users affects the average latency ratio. It is easy to see that with the increase of the number of users, the average latency ratio increases in all schemes. The proposed approach achieves the best performance as compared to the other two algorithms. The following two factors explain the reason: 1) The average latency ratio can be reduced by associating users to the DBS (which has better channel condition as compared to associating with the MBS). 2) It can also be reduced by flexibly adjusting the DBS's location to those which incur lower average latency ratio (as compared to the 'Stationary DBS' scheme).

Fig. 4 illustrates the transmission power of the DBS versus the average latency ratio. From Fig. 4, we can see that the average latency ratio of the 'Stationary DBS' scheme and that of our designed approach decrease as the transmission power of DBS increases because the increase of transmission power will improve the achievable rate of the access link and thus reduce the average latency ratio of users associated with the DBS. The average latency ratio of the 'MBS only' scheme stays constant since no DBS is deployed in this case.

Fig. 5 and Fig. 6 show how the transmission power of the MBS and the total available bandwidth affect the average

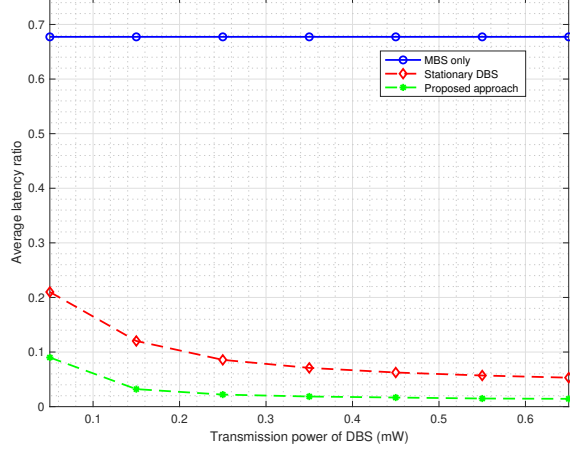


Fig. 4. The average latency ratio versus transmission power of DBS.

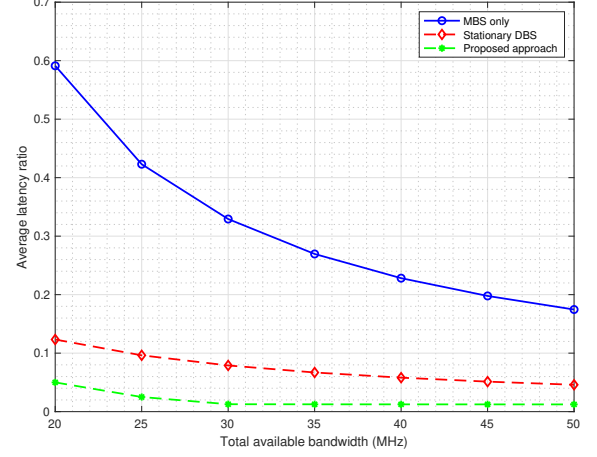


Fig. 6. The average latency ratio versus total available bandwidth.

latency ratio, respectively. In both figures, we can observe that the average latency ratio decreases as the amount of resource increases. In Fig. 5, the average latency ratio of the ‘MBS only’ scheme reduces significantly as the transmission power of MBS increases because all the users are associated with the MBS in this case and the increase of transmission power will affect all users. However, in the other two schemes where a DBS is deployed, the decrease is much less remarkable because the majority of users are associated with the DBS (since it provides a better channel condition) and the increase of transmission power of MBS does not influence such users.

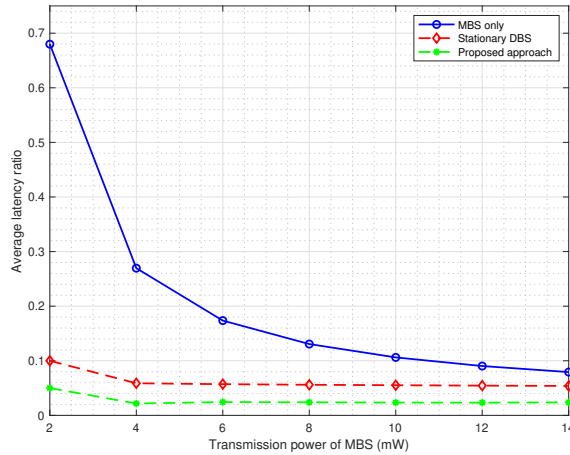


Fig. 5. The average latency ratio versus transmission power of MBS.

From Fig. 7, we can observe that with the increase of the average traffic size, the average latency ratio increases in all schemes because the service time of each use is increased, thus resulting in a larger latency ratio. Both schemes with DBSs outperform the ‘MBS only’ scheme since a better channel connection is provided. Furthermore, our proposed scheme achieves better performance as compared with the ‘Stationary DBS’ scheme because the latency can be reduced by adjusting

the location of the DBS and the resource allocation between the MBS and DBS.

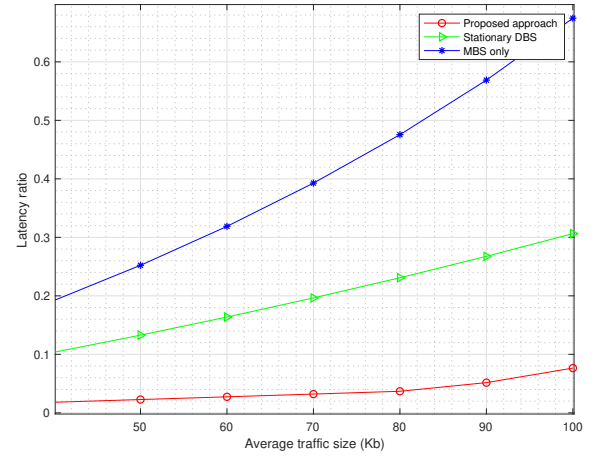


Fig. 7. The average latency ratio versus the average traffic size.

VI. CONCLUSION

In this paper, the problem of user association and bandwidth allocation and DBS placement in a DBS-assisted cellular network with FSO-based backhaul has been formulated and investigated, in which we aim to minimize the overall average latency ratio while meeting the user QoS requirements. To tackle the formulated problem, we have designed a cyclic iterative algorithm to decompose the primal problem into two subproblems, i.e., the bandwidth allocation and user association problem and 3D DBS placement problem. We have proven the former problem to be a sum-of-ratios problem and solve it by utilizing the Lagrangian and modified Newton method. For the 3D DBS placement problem, we first determine the altitude that maximize the coverage area. Then, we exhaustively search all the candidate locations to determine the optimal horizontal location which incurs the minimum latency ratio of all users.

Finally, we conducted extensive simulations to validate the effectiveness of the proposed algorithm. The results show that our proposed algorithm achieves better performance as compared to two baseline algorithms.

REFERENCES

- [1] L. Zhang and N. Ansari, "Optimizing the Deployment and Throughput of DBSs for Uplink Communications," (Invited Paper) *IEEE Open J. Veh. Tech.*, vol. 1, pp. 18-28, 2020.
- [2] S. Zhang, X. Sun and N. Ansari, "Placing Multiple Drone Base Stations in Hotspots," *2018 IEEE 39th Sarnoff Symposium*, 2018, pp. 1-6.
- [3] M. Alzenad, A. El-Keyi and H. Yanikomeroglu, "3-D Placement of an Unmanned Aerial Vehicle Base Station for Maximum Coverage of Users With Different QoS Requirements," *IEEE Wireless Comm. Lett.*, vol. 7, no. 1, pp. 38-41, Feb. 2018.
- [4] X. Liu and N. Ansari, "Resource Allocation in UAV-assisted M2M Communications for disaster Rescue," *IEEE Wireless Comm. Lett.*, vol. 8, no. 2, pp. 580-583, Apr. 2019.
- [5] M. M. Azari, F. Rosas, K. Chen and S. Pollin, "Ultra Reliable UAV Communication Using Altitude and Cooperation Diversity," *IEEE Trans. Comm.*, vol. 66, no. 1, pp. 330-344, Jan. 2018.
- [6] W. Mei, Q. Wu and R. Zhang, "Cellular-Connected UAV: Uplink Association, Power Control and Interference Coordination," *IEEE Trans. Wireless Comm.*, vol. 18, no. 11, pp. 5380-5393, Nov. 2019.
- [7] X. Sun and N. Ansari, "Latency Aware Drone Base Station Placement in Heterogeneous Networks," *IEEE Global Comm. Conf.*, 2017, pp. 1-6.
- [8] H. E. Nistazakis, T. A. Tsiftsis and G. S. Tombras, "Performance analysis of free-space optical communication systems over atmospheric turbulence channels," *IET Comm.*, vol. 3, no. 8, pp. 1402-1409, Aug. 2009.
- [9] N. Ansari, Q. Fan, X. Sun and L. Zhang, "SoarNet," *IEEE Wireless Comm.*, vol. 26, no. 6, pp. 37-43, December 2019.
- [10] D. Wu, X. Sun, and N. Ansari, "An FSO-based Drone Charging System for Emergency Communications," *IEEE Trans. Veh. Tech.*, vol. 69, no. 12, pp. 16155-16162, Dec. 2020.
- [11] D. Wu, X. Sun, and N. Ansari, "An FSO-based Drone Assisted Mobile Access Network for Emergency Communications," *IEEE Trans. Net. Sci. & Engrng.*, vol. 7, no. 3, pp. 1597-1606, Jul-Sep. 2020.
- [12] R. Duan, J. Wang, J. Du, C. Jiang, T. Bai and Y. Ren, "Power-Delay Trade-off for Heterogenous Cloud Enabled Multi-UAV Systems," *IEEE Int. Conf. Commun.*, Shanghai, China, 2019, pp. 1-6.
- [13] A. Alzidaneen, A. Alsharoa and M. Alouini, "Resource and Placement Optimization for Multiple UAVs using Backhaul Tethered Balloons," *IEEE Wireless Comm. Lett.*, vol. 9, no. 4, pp. 543-547, April 2020.
- [14] W. Huang *et al.*, "Joint Power, Altitude, Location and Bandwidth Optimization for UAV With Underlaid D2D Communications," *IEEE Wireless Comm. Lett.*, vol. 8, no. 2, pp. 524-527, April 2019.
- [15] L. Xie, J. Xu and R. Zhang, "Throughput Maximization for UAV-Enabled Wireless Powered Communication Networks," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1690-1703, April 2019.
- [16] M. Mozaffari, W. Saad, M. Bennis and M. Debbah, "Mobile Unmanned Aerial Vehicles (UAVs) for Energy-Efficient Internet of Things Communications," *IEEE Trans. Wireless Comm.*, vol. 16, no. 11, pp. 7574-7589, Nov. 2017.
- [17] X. Sun and N. Ansari, "Jointly Optimizing Drone-Mounted Base Station Placement and User Association in Heterogeneous Networks," *IEEE Int. Conf. Comm.*, Kansas City, MO, May 2018.
- [18] A. Al-Hourani, S. Kandeepan and S. Lardner, "Optimal LAP Altitude for Maximum Coverage," *IEEE Wireless Comm. Lett.*, vol. 3, no. 6, pp. 569-572, Dec. 2014.
- [19] M. Alzenad, A. El-Keyi, F. Lagum and H. Yanikomeroglu, "3-D Placement of an Unmanned Aerial Vehicle Base Station (UAV-BS) for Energy-Efficient Maximal Coverage," *IEEE Wireless Comm. Lett.*, vol. 6, no. 4, pp. 434-437, Aug. 2017.
- [20] L. Kleinrock, *Queueing Systems Vol.2: Computer Applications*. New York: Wiley, 1976.
- [21] A. K. Majumdar, "Free-Space Laser Communication Performance in the Atmospheric Channel," *Springer J. Optical and Fiber Commun. Reports*, vol. 2, no. 4, pp. 345-96, Oct. 2005.
- [22] M. Alzenad *et al.*, "FSO-Based Vertical Backhaul/Fronthaul Framework for 5G+ Wireless Networks," *IEEE Comm. Mag.*, vol. 56, no. 1, pp. 218-224, Jan. 2018.
- [23] I. I. Kim *et al.*, "Comparison of laser beam propagation at 785 nm and 1550 nm in fog and haze for optical wireless communications," *SPIE Proc. Opt. Wireless Comm. III*, vol. 4214, no. 2, pp. 26-37, 2001.
- [24] H. P. Benson, "Global optimization algorithm for the nonlinear sum of ratios problem," *J. Optimization Theory & App.*, vol. 112, no. 1, pp. 1-29, Jan. 2002.
- [25] Y. Jong, "Practical global optimization algorithm for the sum-of-ratios problem," *arXiv preprint arXiv:1207.1153 [math. OC]*, 2012.
- [26] M. Mozaffari, W. Saad, M. Bennis and M. Debbah, "Unmanned Aerial Vehicle With Underlaid Device-to-Device Communications: Performance and Tradeoffs," *IEEE Trans on Wireless Comm.*, vol. 15, no. 6, pp. 3949-3963, June 2016.
- [27] S. Zhang and N. Ansari, "3D Drone Base Station Placement and Resource Allocation With FSO-Based Backhaul in Hotspots," *IEEE Trans. Veh. Tech.*, vol. 69, no. 3, pp. 3322-3329, March 2020.



Shuai Zhang received the B.S. and M.S. degree from the Department of Electronic Engineering, University of Electronic Science and Technology of China, in 2014 and 2017, respectively. He is currently working towards the Ph.D. degree in Electrical Engineering at New Jersey Institute of Technology (NJIT), Newark, New Jersey. His research interests include drone-mounted base-station communications and wireless communications, network optimization and optical networks.



Nirwan Ansari (Fellow, IEEE), Distinguished Professor of Electrical and Computer Engineering at the New Jersey Institute of Technology (NJIT), received his Ph.D. from Purdue University, MSEE from the University of Michigan, and BSEE (summa cum laude with a perfect GPA) from NJIT. He is also a Fellow of National Academy of Inventors. He authored *Green Mobile Networks: A Networking Perspective* (Wiley-IEEE, 2017) with T. Han, and co-authored two other books. He has also (co-)authored more than 600 technical publications, over half published in widely cited journals/magazines. He has guest-edited a number of special issues covering various emerging topics in communications and networking. He has served on the editorial/advisory board of over ten journals including as Associate Editor-in-Chief of *IEEE Wireless Communications Magazine*. His current research focuses on green communications and networking, cloud computing, drone-assisted networking, and various aspects of broadband networks. He was elected to serve in the IEEE Communications Society (ComSoc) Board of Governors as a member-at-large, has chaired some ComSoc technical and steering committees, is current Director of ComSoc Educational Services Board, has been serving in many committees such as the IEEE Fellow Committee, and has been actively organizing numerous IEEE International Conferences/Symposia/Workshops. He has also been granted more than 40 U.S. patents.