Received: 27 January 2023 | Revised: 29 April 2023 | Accepted: 23 May 2023

DOI: 10.4218/etrij.2023-0035

ORIGINAL ARTICLE

ETRI Journal WILEY

Proportionally fair load balancing with statistical quality of service provisioning for aerial base stations

Shengqi Jiang[1] | **Ying Loong Lee**[1] | **Mau Luen Tham**[1] | **Donghong Qin**[2] | **Yoong Choon Chang**[1] | **Allyson Gek Hong Sim**[3]

[1]Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Selangor, Malaysia

[2]School of Artificial Intelligence, Guangxi University for Nationalities, Guangxi, China

[3]SEEMOO Lab, Technische Universität Darmstadt, Darmstadt, Germany

Correspondence
Ying Loong Lee, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Kajang, Selangor, Malaysia.
Email: leeyingl@utar.edu.my

Funding information
Deutsche Forschungsgemeinschaft; Initiation of International Collaboration within the 6G Multi-RATs Communication Systems (6G-RATs) project; LOEWE Initiatives; Universiti Tunku Abdul Rahman, Grant/Award Number: IPSR/RMC/UTARRF/2020-C2/L07

Abstract

Aerial base stations (ABSs) seem promising to enhance the coverage and capacity of fifth-generation and upcoming networks. With the flexible mobility of ABSs, they can be positioned in air to maximize the number of users served with a guaranteed quality of service (QoS). However, ABSs may be overloaded or underutilized given inefficient placement, and user association has not been well addressed. Hence, we propose a three-dimensional ABS placement scheme with a delay-QoS-driven user association to balance loading among ABSs. First, a load balancing utility function is designed based on proportional fairness. Then, an optimization problem for joint ABS placement and user association is formulated to maximize the utility function subject to statistical delay QoS requirements and ABS collision avoidance constraints. To solve this problem, we introduce an efficient modified gray wolf optimizer for ABS placement with a greedy user association strategy. Simulation results demonstrate that the proposed scheme outperforms baselines in terms of load balancing and delay QoS provisioning.

KEYWORDS
aerial base station, base station placement, load balancing, quality of service, user association

1 | INTRODUCTION

Unmanned aerial vehicles (UAVs) have recently emerged as a promising technology for beyond-fifth-generation (5G) networks, enabling fast communication and computing services. By mounting a base station on a UAV, and aerial base station (ABS) can be established and hover in three-dimensional (3D) space, and its position can be flexibly adjusted to optimize mobile network coverage [1, 2]. Thus, quick ABS deployment and coverage enhancement can be achieved, especially when ground base stations are nonfunctional in disasters or hotspot areas requiring a capacity boost. Therefore, extensive research and development on ABS mobile networks have been conducted regarding various aspects, such as network coverage enhancement, ABS trajectory optimization, energy efficiency maximization, and power control optimization. Recently, load balancing among ABSs has gained increasing attention in academia and industry. Because each ABS has a lower computational capacity than a ground base station, the number of users that each ABS can serve is limited [3]. Thus, inefficient ABS

placement and user association can easily lead to load imbalance among ABSs and poor quality-of-service (QoS) provisioning, thereby overloading some ABSs and underutilizing others. Although several QoS-aware load balancing schemes have been developed for traditional ground networks [4–6], that is, conventional cellular networks, these schemes are designed for fixed base stations and cannot handle flexible 3D placement of ABSs for load balancing. QoS-aware load balancing among ABSs is more challenging than that among fixed base stations because the 3D placement and user associations must be jointly considered. This inevitably increases the dimensionality of the load balancing problem, that is, the number of load balancing variables.

Several studies have attempted to address the load balancing problem of ABSs [7–10]. In [9], a joint user association and power control scheme is proposed to achieve load balancing and user fairness in UAV-enabled cellular networks based on matching game theory. In [7], an ABS deployment scheme based on machine learning is devised for load balancing in small-cell multi-ABS networks. In [8], load balancing among ABSs and data rate fairness among users in multi-ABS networks are jointly optimized using the virtual force field method and successive convex optimization. In [10], a software-defined networking-based load balancing technique is proposed, demonstrating improved load balancing among ABSs with QoS guarantees.

Existing load balancing techniques have not adequately considered the limited user capacity of ABSs. In fact, an ABS may become overloaded if the number of associated users exceeds its capacity. In addition, most studies have neglected QoS provisioning for load balancing [7–9]. This is vital for scenarios with real-time multimedia traffic that requires low latency. In addition, ABSs can be deployed in disaster-stricken areas to provide emergency connectivity and support multimedia services to improve the rescue efficiency [11]. ABSs are also expected to be deployed to support live audio/video streaming in hotspot areas where events such as concerts or sports are held. Therefore, delay QoS provisioning is necessary for load balancing among ABSs.

We modeled the delay QoS performance of a user by adopting a statistical delay QoS metric called the effective capacity (EC) [12]. The EC is channel dependent and characterized by a delay-related parameter that reflects the QoS requirements. This is interpreted as the maximum constant arrival rate that a given time-varying service process can support while satisfying the QoS requirements. The EC is particularly convenient for analyzing the statistical QoS performance of wireless transmissions. The EC of ABSs has been investigated in [11, 13]. In [11], a joint ABS placement and resource allocation scheme based on convex optimization is introduced to maximize the EC of ABS networks, whereas in [13], the aggregate EC is maximized under heterogeneous statistical delay-bounded QoS requirements for downlink and uplink transmission groups in ABS networks. However, these studies do not address load balancing among ABSs. On the other hand, we aim to address load balancing among ABSs while ensuring delay QoS provisioning for every user to fill gaps of available studies. Load balancing is challenging owing to the channel dynamics and several degrees of freedom of ABSs. To overcome these challenges, we introduce a joint ABS placement and user association scheme with statistical delay QoS provisioning for load balancing among ABSs. The proposed scheme is focused on downlink transmission.

The main contributions of this study are summarized as follows:

1. We formulate a joint ABS placement and user association problem to maximize a proportionally fair load balancing utility function subject to delay QoS requirements and collision avoidance constraints.
2. Given the mixed-integer programming nature of the problem, we propose a modified gray wolf optimizer (GWO) for ABS placement with greedy EC-based user association to efficiently solve load balancing.
3. We analyze and demonstrate that the computational complexity of the proposed GWO-based load balancing scheme is linear, and the network performance using the proposed scheme is evaluated in terms of probability of blocking, EC, and load balance.

The remainder of this paper is organized as follows. Problem modeling is presented in Section 2. Section 3 describes the proposed load balancing scheme. Simulation results are reported in Section 4. Finally, Section 5 presents our conclusions.

2 | PROBLEM MODELING

In this section, we first present a multi-ABS network scenario followed by descriptions of the channel model, EC modeling, and problem formulation for load balancing in the network.

2.1 | Network scenario

We consider an orthogonal frequency-division multiple-access network with multiple ABSs deployed in a service

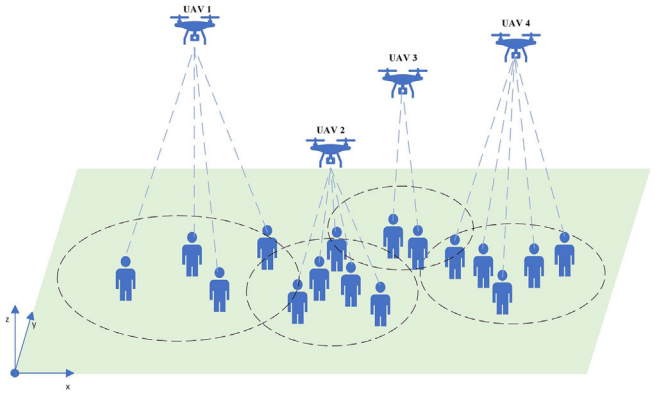


FIGURE 1 Multi-ABS network model.

area with several ground users, as shown in Figure 1. Let $A=\{1, ..., i, ..., m\}$ be the set of m ABSs and $G=\{1, ..., k, ..., n\}$ be the set of n ground users. We consider orthogonal spectrum allocation among ABSs; that is, the subchannels available to each ABS are orthogonal to those available to the other ABSs. We assume that each user associated with an ABS is assigned a subchannel. Let the number of subchannels available to ABS i be denoted by $N_i^{\max}$. Then, the maximum number of users that can be supported by ABS i is also $N_i^{\max}$. For ABS placement, we denote the 3D location coordinates of ABS i by (x_i, y_i, z_i), whereas those of user k are denoted by $(x_k, y_k, 0)$, noting that every user is assumed to remain at ground level, that is, $z_k=0$. Next, we define the following user association variables:

$$c_{i,k}=\begin{cases}1 & \text{if user } k \text{ associates with ABS } i,\\ 0 & \text{otherwise,}\end{cases} \quad \forall k\in G. \tag{1}$$

We assume that ABS placement and user association are jointly performed periodically over a long period, during which the channel conditions (e.g., fast fading) are averaged [5, 6]. This scenario can describe static or low-mobility users, which is the focus of our study. This scenario is relevant for disaster-struck and damaged areas, where the movement of ground users is limited in speed and distance. In these areas, the propagation path loss between users and ABSs is stable [14] over a long period. Similar assumptions were considered in [15] and [16] for joint ABS placement and user association.

2.2 | Channel model

For channel modeling, we calculate the distance between ABS i and user k as

$$d_{i,k}=\sqrt{(x_i-x_k)^2+(y_i-y_k)^2+z_i^2} \quad \forall i\in A,\ k\in G, \tag{2}$$

and the distance between any two ABSs i and j as

$$l_{i,j}=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2} \quad \forall i,j\in A,\ i\neq j. \tag{3}$$

Based on the air-to-ground channel model in [17], the line-of-sight (LoS) path loss, $\mathrm{PL}_{i,k}^{\mathrm{LoS}}$, and non-LoS (NLoS) path loss, $\mathrm{PL}_{i,k}^{\mathrm{NLoS}}$, between ABS i and user k can be obtained as

$$\mathrm{PL}_{i,k}^{\mathrm{LoS}}=20\log\left(\frac{4\pi f_c d_{i,k}}{v}\right)+\eta_{\mathrm{LoS}}, \tag{4}$$

$$\mathrm{PL}_{i,k}^{\mathrm{NLoS}}=20\log\left(\frac{4\pi f_c d_{i,k}}{v}\right)+\eta_{\mathrm{NLoS}}, \tag{5}$$

where η_{LoS} and η_{NLoS} are the additional mean loss values owing to the LoS and NLoS links, respectively, v is the speed of light in vacuum, and f_c is the carrier frequency [17]. Next, the probability of LoS between ABS i and user k is calculated as

$$P_{i,k}^{\mathrm{LoS}}=\frac{1}{1+ae^{-b(\phi_{i,k}-a)}}, \tag{6}$$

where $\phi_{i,k}=\frac{180}{\pi}\tan^{-1}\left(z_i/\left(\sqrt{(x_i-x_k)^2+(y_i-y_k)^2}\right)\right)$ represents the elevation angle from user k at the ground level to ABS i, and a and b are environmental parameters [17]. The probability of NLoS between ABS i and user k can be expressed as $P_{i,k}^{\mathrm{NLoS}}=1-P_{i,k}^{\mathrm{LoS}}$. Thus, the mean path loss in decibels of the signal from ABS i to user k can be estimated as

$$\mathrm{PL}_{i,k}^{\mathrm{Avg}}=P_{i,k}^{\mathrm{LoS}}\times\mathrm{PL}_{i,k}^{\mathrm{LoS}}+P_{i,k}^{\mathrm{NLoS}}\times\mathrm{PL}_{i,k}^{\mathrm{NLoS}}. \tag{7}$$

From (7), the average path loss between ABS i and user k can be rewritten as [17]

$$\mathrm{PL}_{i,k}^{\mathrm{Avg}}=\frac{M}{1+ae^{-b(\phi_{i,k}-a)}}+20\log(d_{i,k})+N, \tag{8}$$

where $M=\eta_{\mathrm{LoS}}-\eta_{\mathrm{NLoS}}$ and $N=20\log((4\pi f_c)/v)+\eta_{\mathrm{NLoS}}$. Using (8), the received signal-to-noise ratio for user k from ABS i can be obtained as

$$\gamma_{i,k}=\frac{pg_{i,k}|h_{i,k}|^2}{N_0B}, \tag{9}$$

where p is the downlink transmit power of each ABS, N_0 is the single-sided noise power spectral density, and $h_{i,k}$ and $g_{i,k} = 10^{(\mathrm{PL}_{i,k}^{\mathrm{Avg}}/10)}$ are the small-scale fading channel coefficient and average path loss between ABS i and user k, respectively. The signal-to-noise ratio is assumed to be averaged over the ABS placement and user association periods and to remain constant throughout the period regardless of the channel dynamics because we also consider the average fast fading over the period.

2.3 | EC

The instantaneous achievable data rate (in bits per frame) between ABS i and user k can be expressed as [18]

$$R_{i,k} = TB\log_2(1+\gamma_{i,k}), \tag{10}$$

where T is the timeframe length and B is the bandwidth of a subchannel. For statistical delay QoS provisioning, we employ the EC QoS metric [12] to model the latency requirement of each user. The EC is a link-layer channel model characterized by the probability of a nonempty buffer and QoS requirement of a connection. Alternatively, it can be interpreted as the maximum constant arrival rate supported by a given departure or service process. In the EC, delay QoS exponent θ determines the strictness of the latency requirement for a traffic flow.

According to the large deviation principle, the buffer length overflow probability, that is, the probability that the delay exceeds the maximum delay bound, can be expressed as

$$P\{D(t) \geq D_{\max}\} \approx \beta(t)e^{-\theta\Lambda(\theta)D_{\max}}, \tag{11}$$

where $D(t)$ is the delay experienced by a packet arriving at time t, $D_{\max}$ is the delay bound required by the connection, $\beta(t) = P\{D(t) \geq 0\}$ is the probability of an event in which the queue is nonempty, and $\Lambda(\theta)$ is the achievable EC of the traffic flow. In (11), a large θ leads to a high decay rate for the probability, implying that the QoS requirement is stringent. On the other hand, a small θ indicates a low decay rate for the probability in (11), indicating a loose QoS requirement. We define the EC between ABS i and user k as [12]

$$\Lambda_{i,k} = -\frac{1}{\theta_k}\ln\mathbb{E}\left(e^{-\theta_k R_{i,k}}\right), \tag{12}$$

where $\mathbb{E}(\cdot)$ represents the expectation operator with respect to the fading state and θ_k is the delay QoS exponent of user k. As we consider small-scale Rayleigh fading, $|h_{i,k}|^2$ is treated as a random variable of Q with the following exponential probability distribution function:

$$f_Q(q) = \frac{1}{\mu}e^{-\frac{q}{\mu}}, \tag{13}$$

where $Q = |h_{i,k}|^2$ and μ is the mean of the exponential probability distribution function. For simplicity, we assume that $\mu = 1$ to obtain the following proposition.

Proposition 1. *The EC between the ABS i and user k can be expressed in the following closed form:*

$$\Lambda_{i,k} = -\frac{1}{\theta_k}\ln\left[\frac{e^{\frac{1}{K_{i,k}}}}{K_{i,k}}E_{J_k}\left(\frac{1}{K_{i,k}}\right)\right], \tag{14}$$

where $K_{i,k} = pg_{i,k}/N_0B$ and $E_{J_k}(\cdot)$ is an exponential integral function with $J_k = \theta_k TB/\ln 2$.

Proof. Note that $\mathbb{E}\left(e^{-\theta_k R_{i,k}}\right)$ can be obtained as

$$\begin{aligned}\mathbb{E}\left(e^{-\theta_k R_{i,k}}\right) &= \int_0^{+\infty} e^{-\theta_k B\log_2(1+\gamma_{i,k}(q))}f_Q(q)dq \\ &= \int_0^{+\infty} e^{-\theta_k B\log_2\left(1+\frac{pg_{i,k}q}{N_0B}\right)}e^{-q}dq.\end{aligned} \tag{15}$$

Let $t_{i,k} = 1 + K_{i,k}q$. Hence, (15) can be expressed as [19]

$$\mathbb{E}\left(e^{-\theta_{i,k}R_{i,k}}\right) = \int_1^{+\infty} t^{-J_k}e^{-\frac{t-1}{K_{i,k}}}dt \approx \frac{e^{\frac{1}{K_{i,k}}}}{K_{i,k}}E_{J_k}\left(\frac{1}{K_{i,k}}\right). \tag{16}$$

Substituting (16) into (12), (14) is obtained. □

2.4 | Problem formulation

We define the load of ABS i as the ratio of the number of users served by the ABS to its maximum user capacity:

$$\Delta_i = \frac{\sum_{k\in G} c_{i,k}}{N_i^{\max}}, \forall i \in A, \tag{17}$$

where Δ_i is the load of ABS i and the numerator represents the current number of users served by ABS i. To balance the load among ABSs, an appropriate utility

function must be designed and optimized. We aim to achieve proportionally fair load balancing among ABSs by adopting the following logarithmic utility function of the ABS load: $\varphi_i(\Delta_i) = \log(\Delta_i)$. The load balancing utility function can be designed as follows:

$$U(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{c}) = \sum\nolimits_{i \in A} \varphi_i(\Delta_i), \tag{18}$$

where $\mathbf{x} = \{x_1, ..., x_i, ..., x_m\}$, $\mathbf{y} = \{y_1, ..., y_i, ..., y_m\}$, $\mathbf{z} = \{z_1, ..., z_i, ..., z_m\}$, and $\mathbf{c} = \{c_{1,1}, ..., c_{i,k}, ..., c_{m,n}\}$. Because the logarithmic function has a diminishing return property with respect to Δ_i, the maximization of (18) promotes proportional fairness in load balancing [20]. Next, we formulate the load balancing problem for multi-ABS networks as the following joint 3D ABS placement and user association problem that maximizes (18):

$$\mathbf{P}: \max_{\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{c}} U(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{c}) \tag{19}$$

subject to

$$\sum_{i \in A} c_{i,k} \le 1 \ \forall k \in G, \tag{20}$$

$$\sum_{i \in A} c_{i,k} \Lambda_{i,k} \ge R_k \ \forall k \in G, \tag{21}$$

$$l_{i,j} > d_{\text{sec}} \ \forall i, j \in A, \ i \ne j, \tag{22}$$

$$x_{\min} \le x_i \le x_{\max} \ \forall i \in A, \tag{23}$$

$$y_{\min} \le y_i \le y_{\max} \ \forall i \in A, \tag{24}$$

$$z_{\min} \le z_i \le z_{\max} \ \forall i \in A, \tag{25}$$

$$c_{i,k} \in \{0, 1\} \ \forall i \in A, \ k \in G, \tag{26}$$

$$0 \le \Delta_i \le 1 \ \forall i \in A, \tag{27}$$

where R_k is the required EC of user k and d_{sec} is the security distance between any two ABSs to avoid collisions. Constraint (20) ensures that each user is associated with only one ABS. Constraint (21) ensures that each user satisfies the EC requirements for statistical delay QoS guarantees. To prevent collisions between ABSs, constraint (22) ensures that the 3D distance between any two ABSs exceeds d_{sec}. As the locations of the ABSs are bounded by the service area and altitude, constraints (23), (24), and (25) are enforced to ensure that the positions of ABSs are within the service areas and allowable altitude. Constraint (26) guarantees binary-valued user association variables. Owing to the limited user capacity of each ABS, $\sum_{k \in G} c_{i,k} \le N_i^{\max}$. Thus, Δ_i cannot exceed 1, as indicated by constraint (27).

3 | PROPOSED SCHEME

Problem **P** is a mixed-integer programming problem, in which $\mathbf{x},\mathbf{y},\mathbf{z}$ are continuously valued and $\mathbf{c}$ is binary valued. This problem is difficult to solve because of its nonconvex nature, and especially because it consists of several continuous and binary variables. Exhaustive brute-force methods are computationally impractical given that 3D ABS placement and user association must be efficiently performed within a limited fixed timeframe. Swarm intelligence algorithms are promising for efficient optimization of nonconvex problems with several variables. Notably, the GWO, recently proposed in [21] as a new swarm intelligence algorithm, can successfully solve various optimization problems.

The GWO is inspired by the behavior of gray wolves, who prefer to live in groups under a strict socially dominant hierarchy. In addition, the wolf pack exhibits the social behavior of pack hunting, which can describe an optimization mechanism. Compared with other swarm intelligence algorithms, the GWO involves relatively fewer parameters and is simple, easy to use, flexible, and stable. It has been applied to ABS deployment and path planning [22, 23]. Motivated by the strengths of GWO, we developed a GWO-based ABS placement and user association algorithm to solve problem **P**. In the following section, we describe ABS placement and user association under the GWO algorithmic framework.

3.1 | ABS placement

In the GWO, the position of each gray wolf represents a candidate solution. The gray wolf with the best solution is denoted as alpha (α), whereas those with the second- and third-best solutions are denoted as beta (β) and delta (δ), respectively, while the remaining gray wolves are denoted as omega (ω). Similar to the wolf pack hunting behavior, α, β, and δ guide the hunting behavior of the wolf pack. To mathematically model the behavior of gray wolves surrounding the prey during hunting, the following equations were derived by [21]:

$$\mathbf{D}(t) = |\mathbf{C} \cdot \mathbf{X}_{\text{p}}(t) - \mathbf{X}_w(t)|, \tag{28}$$

$$\mathbf{X}_w(t+1) = \mathbf{X}_{\text{p}}(t) - \mathbf{A} \cdot \mathbf{D}(t), \tag{29}$$

where t is the iteration index representing the timestep, $\mathbf{X}_w$ is the position vector of the gray wolf, and $\mathbf{X}_{\text{p}}$ is the

position vector of the prey. Coefficient vectors **A** and **C** are obtained as

$$\mathbf{A} = 2\mathbf{a} \cdot \mathbf{r}_1 - \mathbf{a}, \tag{30}$$

$$\mathbf{C} = 2\mathbf{r}_2, \tag{31}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are vectors with each element randomly set within [0, 1] and **a** is a vector whose elements linearly decrease from 2 to 0 over the iterations to obtain the solution. By decreasing the value of the elements in **a**, vector **A** can drive the gray wolf to move closer to the prey to attack it. Regarding optimization, this behavior encourages local exploitation, where the search agent attempts to find a better solution near the current best solution. However, with an increased value of the elements in **a**, the gray wolf tends to move away from the prey to search for a better solution. This behavior is equivalent to the global exploration of solutions during optimization. Vector **C** introduces randomness in the movement of the gray wolf to promote global exploration and solution diversity.

To implement the GWO, N gray wolves are created to search for the solution. We denote $\mathbf{X} = \{\mathbf{X}_1, ..., \mathbf{X}_w, ..., \mathbf{X}_N\}$ as the set of position vectors of the gray wolves, where $\mathbf{X}_w$ represents the position vector (i.e., candidate solution) of the w-th gray wolf. To determine the solution quality for a gray wolf, we evaluate its *fitness* position vector. Let $F(\mathbf{X}_w)$ be the fitness function of the w-th gray wolf. In each iteration, The fitness value of each gray wolf is evaluated. The GWO identifies gray wolves with the three highest fitness values and stores the consecutive position vectors as alpha (X_α), beta (X_β), and delta (X_δ). Then, the position of each gray wolf is updated as follows:

$$\mathbf{D}_\alpha(t) = |\mathbf{C}_1 \cdot \mathbf{X}_\alpha(t) - \mathbf{X}_w(t)|, \tag{32}$$

$$\mathbf{X}_{\alpha 1}(t) = \mathbf{X}_\alpha(t) - \mathbf{A}_1 \cdot \mathbf{D}_\alpha(t), \tag{33}$$

$$\mathbf{D}_\beta(t) = |\mathbf{C}_2 \cdot \mathbf{X}_\beta(t) - \mathbf{X}_w(t)|, \tag{34}$$

$$\mathbf{X}_{\beta 2}(t) = \mathbf{X}_\beta(t) - \mathbf{A}_2 \cdot \mathbf{D}_\beta(t), \tag{35}$$

$$\mathbf{D}_\delta(t) = |\mathbf{C}_3 \cdot \mathbf{X}_\delta(t) - \mathbf{X}_w(t)|, \tag{36}$$

$$\mathbf{X}_{\delta 3}(t) = \mathbf{X}_\delta(t) - \mathbf{A}_3 \cdot \mathbf{D}_\delta(t), \tag{37}$$

$$\mathbf{X}_w(t+1) = \frac{\mathbf{X}_{\alpha 1}(t) + \mathbf{X}_{\beta 2}(t) + \mathbf{X}_{\delta 3}(t)}{3}, \tag{38}$$

where $\mathbf{C}_1$, $\mathbf{C}_2$, and $\mathbf{C}_3$ are calculated based on (31) and $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{A}_3$ are calculated based on (30). The entire process is repeated until the maximum number of iterations is reached.

Because **P** is a constrained optimization problem with continuous and binary variables, the conventional GWO cannot be applied directly. Therefore, we propose a modified GWO-based ABS placement and user association scheme for problem **P**, which is designed as follows. The position of the w-th gray wolf is given by

$$\mathbf{X}_w = \left\{x_1^{(w)}, y_1^{(w)}, z_1^{(w)}, ..., x_i^{(w)}, y_i^{(w)}, z_i^{(w)}, ..., x_m^{(w)}, y_m^{(w)}, z_m^{(w)}\right\},$$

where $\{x_i^{(w)}, y_i^{(w)}, z_i^{(w)}\}$ are the 3D coordinates of ABS i stored in the position vector of the w-th gray wolf. To deal with constraints (23)–(25), we implement the following strategy:

$$X_w(j) = r_2(X_{\max}(j) - X_{\min}(j)) + X_{\min}(j), \tag{39}$$

where $X_\omega(j)$ is the j-th element of $\mathbf{X}_\omega$, r_2 is a random value ranging in [0, 1], and

$$X_{\max}(j) = \begin{cases} x_{\max} & \text{if } X_\omega(j) \in \{x_i | i \in \mathcal{A}\} \\ y_{\max} & \text{if } X_\omega(j) \in \{y_i | i \in \mathcal{A}\} \\ z_{\max} & \text{if } X_\omega(j) \in \{z_i | i \in \mathcal{A}\}, \end{cases}$$

$$X_{\min}(j) = \begin{cases} x_{\min} & \text{if } X_\omega(j) \in \{x_i | i \in \mathcal{A}\} \\ y_{\min} & \text{if } X_\omega(j) \in \{y_i | i \in \mathcal{A}\} \\ z_{\min} & \text{if } X_\omega(j) \in \{z_i | i \in \mathcal{A}\}. \end{cases}$$

Through (39), ABSs that are placed at locations beyond the boundaries of the service area are randomly repositioned within the boundaries. Algorithm 1 summarizes the steps enforcing constraints (23)–(25). In principle, Algorithm 1 checks whether each variable in the position vector of each gray wolf, which is updated using (38), exceeds the boundary values defined in constraints (23)–(25), as described by step 4. If the boundary value of a variable is exceeded, the algorithm invokes step 5 to update the variable using (39).

Algorithm 1 Boundary limitation algorithm

```
1: Inputs: positions of wolves
2: for w = 1 to N do
3:     for j = 1 to 3m do
4:         if X_w(j) ≤ X_max(j) or X_w(j) ≥ X_min(j) then
5:             Update X_w(j) using (39)
6:         end if
7:     end for
8: end for
```

To ensure constraint (22), we introduce the collision-avoidance algorithm described in Algorithm 2. In this algorithm, the 3D distance between ABS i and each of the other ABSs (that is, ABSs $j \in A \setminus \{i\}$) is calculated using (3) (step 5). If the distance is shorter than security distance d_{sec}, the 3D position of ABS i is randomly reset using (39), as described in steps 7–10.

Algorithm 2 Collision avoidance algorithm

1: Inputs: positions of wolves
2: **for** $w = 1$ to N **do**
3: **for** $i \in A$ **do**
4: **for** $j \in A$ and $j \neq i$ **do**
5: Calculate $l_{i,j}$ using (3)
6: **if** $l_{i,j} < d_{\text{sec}}$ **then**
7: Set $u = 3 \times i$;
8: Update $X_w(u)$ using (39)
9: Update $X_w(u-1)$ using (39)
10: Update $X_w(u-2)$ using (39)
11: **end if**
12: **end for**
13: **end for**
14: **end for**

3.2 | EC-based user association

The position vector of each gray wolf only represents the 3D coordinates of the ABSs, and user association variables are also required to evaluate the fitness function. As each user must satisfy constraint (20) and delay QoS requirement (21), we apply a greedy user association strategy, in which each user selects the ABS that provides the largest EC given the 3D coordinates of all the ABSs. Thus, the association for user k can be formulated as follows:

$$c_{i,k} = \begin{cases} 1 & i = \arg\max_{j \in A} \Lambda_{j,k} \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

The final user association result must also satisfy constraint (27) because the number of users served by each ABS i is limited to $N_i^{\max}$. Thus, the proposed greedy user association algorithm sequentially associates users with the ABS that provides the largest EC while ensuring that the selected ABS is not overloaded (see Algorithm 3). For each user, the algorithm first estimates the achievable EC of the user per ABS using (12) in steps 3–5. Next, user association described by steps 6–15 is executed. The ABS that provides the highest achievable EC is first identified. Then, if constraint (27) is satisfied when the ABS is associated to the user, user association is completed. Otherwise, user association is restarted, and the previously identified ABS is excluded.

Algorithm 3 Greedy EC-based user association algorithm

1: Initialize $\Omega = A$ and $c_{i,k} = 0$ for all $i \in A$ and $k \in G$
2: **for** $k \in G$ **do**
3: **for** $i \in A$ **do**
4: Calculate $\Lambda_{i,k}$ using (12)
5: **end for**
6: Reset $\Omega = A$
7: **while** $\Omega \neq \{\}$ **do**
8: Obtain $i = \arg\max_{j \in \Omega} \Lambda_{j,k}$
9: Set $c_{i,k} = 1$
10: **if** $\Delta_i \leq 1$ **then**
11: **break**
12: **else**
13: Set $c_{i,k} = 0$ and $\Omega \leftarrow \Omega \setminus \{i\}$
14: **end if**
15: **end while**
16: **end for**

3.3 | Fitness function

To evaluate the fitness of each gray wolf, the objective function in (19) is the natural choice of fitness function. However, each user must be placed within the effective coverage area of its associated ABS, such that constraint (21) in problem **P** is satisfied. The user association described by Algorithm 3 does not guarantee that the largest achievable EC provided to users by their associated ABS is sufficient to satisfy constraint (21). Thus, because the achievable EC depends on the distance between each user and its associated ABS, the ABS must be positioned such that every user can meet its target EC requirement. To this end, the degree of violation of constraint (21) must be reflected in the fitness value of each

gray wolf. Therefore, we convert this constraint into a penalty function and incorporate it into the fitness function as follows:

$$\mathcal{F}_w(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{c}) = \sum_{i \in A} \varphi(\Delta_i) - \psi \sum_{k \in G} \left(R_k - \sum_{i \in A} c_{i,k} \Lambda_{i,k} \right), \quad (41)$$

where $\mathcal{F}_w(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{c})$ is the fitness function of the w-th gray wolf and ψ is the weight of constraint (21), which determines the priority of the penalty function in the fitness evaluation. Note that variables $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are extracted from the position vector of the gray wolf. If constraint (21) is violated, the fitness value decreases, while the value increases otherwise. This fitness function encourages all gray wolves to search for better candidate solutions without incurring a high penalty on the fitness values.

3.4 | GWO-based load balancing

Algorithm 4 summarizes the proposed GWO-based ABS placement and user association scheme for load balancing. In step 1, a population of gray wolves is initialized with random position vectors. Then, user association is performed using Algorithm 3 followed by fitness evaluations of the gray wolves (steps 2–6). Next, in steps 9–11, the position vectors of the gray wolves are updated using (38). The position vectors are additionally bounded using Algorithm 1, and collision avoidance is performed using Algorithm 2 in steps 12–15. Subsequently, user association is performed again, followed by fitness evaluations of the gray wolves (steps 17–23). Given the fitness values of the gray wolves, the position vectors of wolves α, β, and δ are updated in step 24. Steps 8–26 are repeated until the maximum number of iterations, $T_{\max}$, is reached.

The time complexity of Algorithm 4 is analyzed. For initialization, Algorithm 4 has a complexity of $O(Nm)$. The calculation of the control parameters in Algorithm 4 has a complexity of $O(Nm)$. The update of the gray wolf positions has a complexity of $O(Nm)$. The calculation of the EC and user association has a complexity of $O(Nmn)$. The evaluation of the fitness value per gray wolf has a complexity of $O(Nm)$. Hence, the total time complexity for an iteration of Algorithm 4 is $O(Nm)$. Over the maximum number of iterations, the time complexity of Algorithm 4 is $O(NmT_{\max})$.

Algorithm 4 Proposed load balancing scheme

1: Initialize gray wolf population $\mathbf{X}$, $\mathbf{a}$, $\mathbf{A}$, and $\mathbf{C}$
2: **for** $w = 1$ to N **do**
3: Apply Algorithm 3 to $\mathbf{c}$
4: Extract variables $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ from $\mathbf{X}_w$
5: Calculate $\mathcal{F}_w(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{c})$ using (41)
6: **end for**
7: Set $t = 0$
8: **while** $t < T_{\max}$ **do**
9: **for** $w = 1$ to N **do**
10: Update $\mathbf{X}_w$ with (38)
11: **end for**
12: **for** $w = 1$ to N **do**
13: Execute Algorithm 1
14: Execute Algorithm 2
15: **end for**
16: Update $\mathbf{A}$ and $\mathbf{C}$ using (30) and (31), respectively
17: **for** $w = 1$ to N **do**
18: **for** all $k \in G$ **do**
19: Apply Algorithm 3 to $\mathbf{c}$
20: Extract variables $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ from $\mathbf{X}_w$
21: Calculate $\mathcal{F}_w(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{c})$ using (41)
22: **end for**
23: **end for**
24: Update $\mathbf{X}_\alpha$, $\mathbf{X}_\beta$, and $\mathbf{X}_\delta$ using (28) and (29)
25: Update $t \leftarrow t + 1$
26: **end while**
27: Return $\mathbf{X}_\alpha$

4 | RESULTS AND DISCUSSION

4.1 | Simulation settings

For the performance evaluation, a geographical area of 500×500 m^2 with four ABSs was considered. We set $x_{\min} = 0$ m, $x_{\max} = 500$ m, $y_{\min} = 0$ m, $y_{\max} = 500$ m, $z_{\min} = 150$ m, $z_{\max} = 300$ m, $f_c = 2$ GHz, $N_i^{\max} = 50$ with a 10-MHz bandwidth [24] for all $i \in A$, and $N_0 = -174$ dBm/Hz [17]. In addition, we set $\theta_k = 0.001$ for all $k \in G$ and assumed an urban environment. The parameters related to path loss were set as follows: $a = 0.6$, $b = 0.11$, $\eta_{\text{LoS}} = 1$, and $\eta_{\text{NLoS}} = 20$ [25, 26]. For the proposed scheme, the weight of the fitness function was set to $\psi = 10^{-6}$. The population size of the GWO was $N = 20$, and the maximum number of iterations was set to $T_{\max} = 100$. Results were generated and averaged over 30 realizations.

For comparison, we also evaluated the following baseline schemes: (i) Random ABS placement and user association [27] denoted as RnD, (ii) geographical

partitioning-based ABS deployment and user association [28] denoted as PD, and (iii) QoS-oriented 3D multiple ABS deployment [29] denoted as QoS-Prior. The proposed scheme was denoted as GWO. We evaluated the proposed and baseline schemes in terms of the following performance metrics:

- Probability of blocking: Ratio of the number of blocked users, who are dropped from their associated ABS owing to unavailability of subchannels or violation of constraint (21), to the total number of users in the network, which can be calculated by $(n - \sum_{i \in A} \sum_{k \in G} c_{i,k})/n$.
- Total EC: Sum of EC over all the users in the network.
- Jain's fairness index [30]: Load balancing performance indicator for multi-ABS network. We define it as $\left(\sum_{i \in A} \Delta_i\right)^2 / m \sum_{i \in A} \Delta_i{}^2$. A higher fairness index indicates more balanced loading.

4.2 | Results

We first evaluated the convergence of the proposed scheme in two scenarios. First, the performance of the proposed and baseline schemes was evaluated for different numbers of users, with $R_k = 6$ Kbits/frame for all $k \in G$. Second, the performance was evaluated by varying the EC requirements per user for $n = 200$ users in the network.

4.2.1 | Convergence performance

Figure 2 shows the convergence of the proposed scheme over 100 iterations in a scenario with 200 users and $R_k = 6$ Kbits/frame for every user $k \in G$. The convergence values are averaged over 30 realizations. The fitness value increases with the number of iterations. Hence, the proposed scheme can maximize the fitness function. In addition, it converges after 80 iterations, which is reasonably fast. In addition, the 95% confidence intervals from the 80th iteration onward indicate that the proposed scheme is considerably stable at convergence.

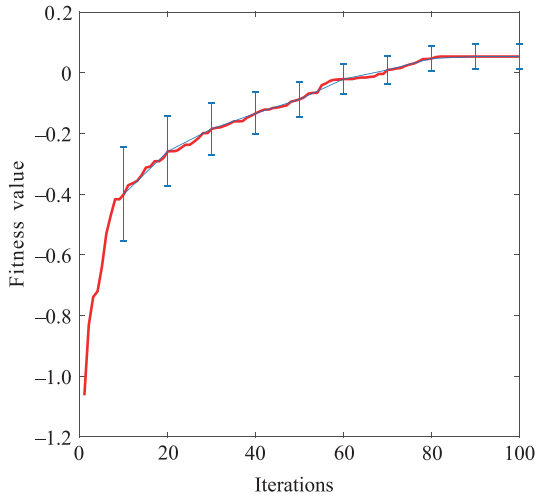


FIGURE 2 Convergence of proposed scheme.

4.2.2 | Scenario with varying numbers of users

Figure 3A shows the probability of blocking achieved by the proposed and baseline schemes. When the number of users in the network exceeds 200, the probability drastically increases. This is because the current network can serve up to 200 users because the maximum user capacity of each of the four ABSs is 50. Overall, the proposed scheme outperforms the baseline schemes because constraint (21) is a penalty function in the fitness function for the GWO. Hence, the proposed scheme increases the number of users associated with QoS guarantees. In contrast, the worst performance is provided by the QoS-Prior scheme because it only considers the users' signal-to-noise ratio requirements and tends to shrink the effective coverage area of each ABS. Consequently, it fails to guarantee adequate ECs to meet the minimum required EC imposed by constraint (22).

Figure 3B shows the EC of the proposed and baseline schemes. The proposed scheme performs better than the baseline schemes. When the number of users reaches 250, the performance of the proposed scheme stagnates because it focuses on balancing the load among ABSs and associates the users if their minimum EC requirement is satisfied. Consequently, a small cost of the total EC is incurred, particularly when the number of users is large. The PD and RnD schemes provide higher ECs when the number of users exceeds 200 but at the expense of a higher probability of blocking, as shown in Figure 3A. Meanwhile, the QoS-Prior scheme has the worst performance because it disregards the EC.

Figure 3C shows Jain's fairness index obtained using the proposed and baseline schemes. The proposed scheme is superior to the baseline schemes for different numbers of users. This is attributed to the optimization of the proportionally fair utility function in the proposed scheme, which contributes to load balancing in the network. Such function is not considered in the baseline schemes, which thus provide an inferior fairness performance.

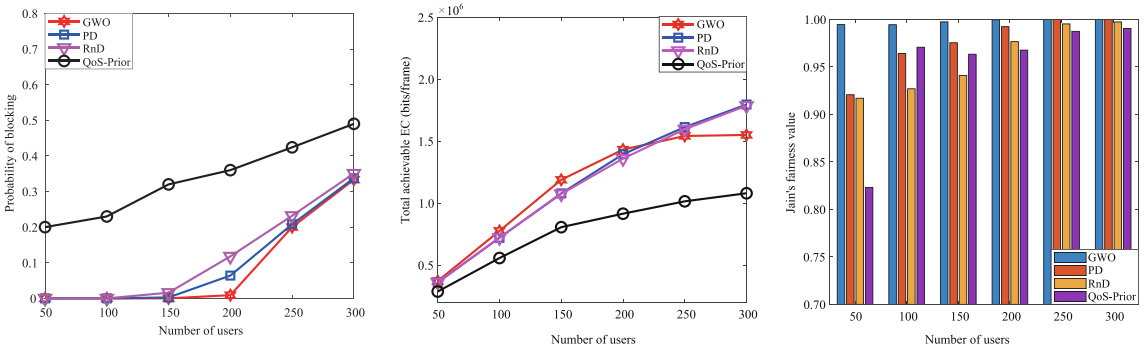


FIGURE 3 Performance of multi-ABS network with varying numbers of users in terms of (A) probability of blocking, (B) total achievable EC, and (C) Jain's fairness index.

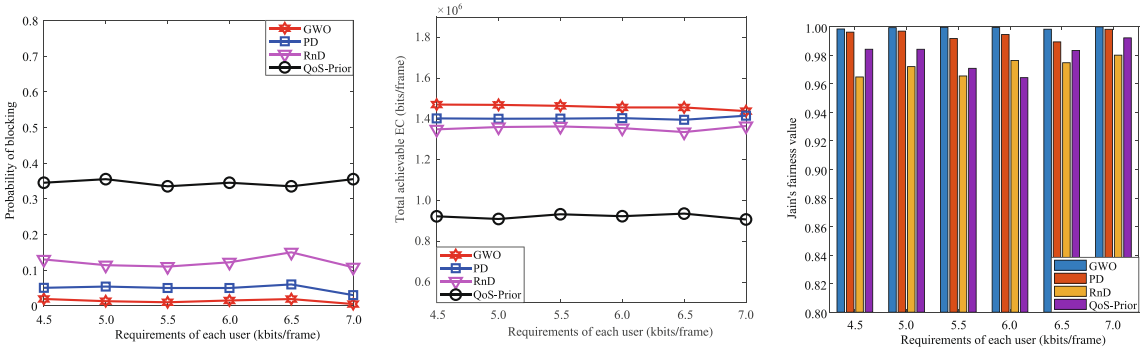


FIGURE 4 Performance of multi-ABS network with varying EC requirements in terms of (A) probability of blocking, (B) total achievable EC, and (C) Jain's fairness index.

4.2.3 | Scenario with varying EC requirements

Figure 4A shows that the proposed scheme outperforms the baseline schemes with the lowest probabilities of blocking across different EC requirements. Even with increasing EC demand, the proposed scheme ensures that more users are served in the network owing to the GWO.

Figure 4B shows that the proposed scheme achieves a higher total EC for the network than the baseline schemes. This is again attributed to the GWO that finds the ABS positions to meet the EC requirements of each user. In addition, the greedy user association algorithm ensures that every user associates with the ABS that maximizes the EC. Hence, the total EC of the network is high.

Figure 4C shows the load balancing performance of the proposed and baseline schemes in terms of Jain's fairness index. The proposed scheme achieves a substantial performance improvement over the baseline schemes. Hence, the proposed scheme can achieve better load balancing among ABSs compared with the baseline schemes, even with increasing EC requirements. Remarkably, the proposed scheme consistently achieves a Jain's fairness index above 99% across the different EC requirements.

5 | CONCLUSION

Load balancing in multi-ABS networks is vital for efficient network utilization and delay QoS provisioning. Thus, it requires effective and efficient joint ABS placement and user association. We propose a GWO-based ABS placement scheme with a greedy EC-based user association algorithm for proportionally fair load balancing among ABSs. The proposed scheme achieves a more balanced load among ABSs, lower probability of blocking, and higher EC compared with baseline schemes under scenarios with different numbers of users and EC

requirements. In future work, we will investigate the joint ABS placement and user association with handover management for load balancing in multi-ABS networks with high user mobility.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.

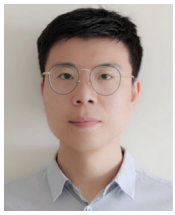
ORCID

Ying Loong Lee https://orcid.org/0000-0002-0841-449X

REFERENCES

1. Y. Zeng, Q. Wu, and R. Zhang, *Accessing from the sky: A tutorial on uav communications for 5G and beyond*, Proc. IEEE **107** (2019), no. 12, 2327–2375.
2. C. Zhang, L. Zhang, L. Zhu, T. Zhang, Z. Xiao, and X.-G. Xia, *3D deployment of multiple UAV-mounted base stations for uav communications*, IEEE Trans. Commun. **69** (2021), no. 4, 2473–2488.
3. M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, *A tutorial on UAVs for wireless networks: Applications, challenges, and open problems*, IEEE Commun. Surv. Tutorials **21** (2019), no. 3, 2334–2360.
4. P. Han, Z. Zhou, and Z. Wang, *User association for load balance in heterogeneous networks with limited CSI feedback*, IEEE Commun. Lett. **24** (2020), no. 5, 1095–1099.
5. Y. L. Lee, T. C. Chuah, A. A. El-Saleh, and J. Loo, *User association for backhaul load balancing with quality of service provisioning for heterogeneous networks*, IEEE Commun. Lett. **22** (2018), no. 11, 2338–2341.
6. Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, *User association for load balancing in heterogeneous cellular networks*, IEEE Trans. Wirel. Commun. **12** (2013), no. 6, 2706–2716.
7. J. Hu, H. Zhang, Y. Liu, X. Li, and H. Ji, *An intelligent UAV deployment scheme for load balance in small cell networks using machine learning*, (IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco), 2019, pp. 1–6.
8. Z. Luan, H. Jia, P. Wang, R. Jia, and B. Chen, *Joint UAVs' load balancing and UEs' data rate fairness optimization by diffusion UAV deployment algorithm in multi-UAV networks*, Entropy **23** (2021), no. 11, 1470.
9. M. Sami and J. N. Daigle, *User association and power control for UAV-enabled cellular networks*, IEEE Wirel. Commun. Lett. **9** (2019), no. 3, 267–270.
10. C. Singhal and K. Rahul, *LB-UAVnet: Load balancing algorithm for UAV based network using SDN*, (22nd International Symposium on Wireless Personal Multimedia Communications (WPMC), Lisbon, Portugal), 2019, pp. 1–5.
11. H. Niu, X. Zhao, and J. Li, *3D location and resource allocation optimization for UAV-enabled emergency networks under statistical QoS constraint*, IEEE Access **9** (2021), 41566–41576.
12. D. Wu and R. Negi, *Effective capacity: a wireless link model for support of quality of service*, IEEE Trans. Wirel. Commun. **2** (2003), no. 4, 630–643.
13. X. Zhang, W. Cheng, and H. Zhang, *Heterogeneous statistical qos provisioning over airborne mobile wireless networks*, IEEE J. Sel. Areas Commun. **36** (2018), no. 9, 2139–2152.
14. H. El Hammouti, M. Benjillali, B. Shihada, and M.-S. Alouini, *Learn-as-you-fly: A distributed algorithm for joint 3D placement and user association in multi-UAVs networks*, IEEE Trans. Wirel. Commun. **18** (2019), no. 12, 5831–5844.
15. Y. Sun, T. Wang, and S. Wang, *Location optimization and user association for unmanned aerial vehicles assisted mobile networks*, IEEE Trans. Veh. Technol. **68** (2019), no. 10, 10056–10065.
16. X. Xi, X. Cao, P. Yang, J. Chen, T. Quek, and D. Wu, *Joint user association and UAV location optimization for UAV-aided communications*, IEEE Wirel. Commun. Lett. **8** (2019), no. 6, 1688–1691.
17. S. Shakoor, Z. Kaleem, D.-T. Do, O. A. Dobre, and A. Jamalipour, *Joint optimization of UAV 3-D placement and path-loss factor for energy-efficient maximal coverage*, IEEE Internet Things J. **8** (2020), no. 12, 9776–9786.
18. T. Abrão, S. Yang, L. D. H. Sampaio, P. J. E. Jeszensky, and L. Hanzo, *Achieving maximum effective capacity in OFDMA networks operating under statistical delay guarantee*, IEEE Access **5** (2017), 14333–14346.
19. D. Zwillinger and A. Jeffrey, *Table of integrals, series, and products*, Elsevier, 2007.
20. F. Kelly, *Charging and rate control for elastic traffic*, Eur. Trans. Telecommun. **8** (1997), no. 1, 33–37.
21. S. Mirjalili, S. M. Mirjalili, and A. Lewis, *Grey wolf optimizer*, Adv. Eng. Softw. **69** (2014), 46–61.
22. C. Hu, L. Ding, B. Liu, S. Ding, J. Huang, H. Wang, Y. Liu, and M. Tan, *Multi unmanned aerial vehicle area coverage control based on enhanced alpha-guided grey wolf optimizer*, (International Conference on Electronic Information Technology and Smart Agriculture (ICEITSA), Huaihua, China), 2021, pp. 410–415.
23. X. Wang, H. Zhao, T. Han, H. Zhou, and C. Li, *A grey wolf optimizer using gaussian estimation of distribution and its application in the multi-UAV multi-target urban tracking problem*, Appl. Soft Comput. **78** (2019), 240–260.
24. M. M. Azari, G. Geraci, A. Garcia-Rodriguez, and S. Pollin, *UAV-to-UAV communications in cellular networks*, IEEE Trans. Wirel. Commun. **19** (2020), no. 9, 6130–6144.
25. M. Simunek, P. Pechac, and F. P. Fontán, *Excess loss model for low elevation links in urban areas for UAVs*, Radioengineering **20** (2011), no. 3, 561–568.
26. Z. Wei, Y. Cai, Z. Sun, D. W. K. Ng, J. Yuan, M. Zhou, and L. Sun, *Sum-rate maximization for IRS-assisted UAV OFDMA communication systems*, IEEE Trans. Wirel. Commun. **20** (2020), no. 4, 2530–2550.
27. S. Lim, H. Yu, and H. Lee, *Optimal tethered-UAV deployment in A2G communication networks: Multi-agent Q-learning approach*, IEEE Internet Things J. **9** (2022), no. 19, 18539–18549.
28. H. Huang and A. V. Savkin, *Deployment of heterogeneous UAV base stations for optimal quality of coverage*, IEEE Internet Things J. **9** (2022), no. 17, 16429–16437.
29. X. Luo, J. Xie, L. Xiong, Z. Wang, and C. Tian, *3-D deployment of multiple UAV-mounted mobile base stations for full coverage of IoT ground users with different QoS requirements*, IEEE Commun. Lett. **26** (2022), no. 12, 3009–3013.
30. R. Jain, *The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling*, John Wiley & Sons, 1990.

AUTHOR BIOGRAPHIES



Shengqi Jiang is currently pursuing a PhD degree at the Universiti Tunku Abdul Rahman. His current research interests include load balancing in UAV-assisted cellular networks, artificial intelligence-supported wireless network management, and optimization of 5G and future networks.

Ying Loong Lee received his BEng (Honors) degree in electronics majoring in telecommunications and PhD degree from Multimedia University in 2012 and 2017, respectively. He was a visiting researcher at the National Chiao Tung University, Taiwan, in 2015 under the NCTU Elite Intership Program, at the Guangxi University for Nationalities, China, in 2019, and at the Technical University of Darmstadt, Germany, in 2022. He is currently an assistant professor in the Department of Electrical and Electronics Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman. His current research interests include wireless communications for 5G and 6G networks, fixed–mobile convergence, and the applications of artificial intelligence to telecommunications.

Mau Luen Tham received his BEng and PhD degrees in telecommunication engineering from the University of Malaya. He is currently an assistant professor with Universiti Tunku Abdul Rahman. He has been a principal investigator and co-investigator for 20 research and development projects. This includes five international grants, two of which were simultaneously led as principal investigator and co-investigator with the support of ASEAN IVO and the British Council. He has published two IEEE Transactions papers as the lead author. His research interests include IoT, machine learning/deep learning/deep reinforcement learning, and beyond-5G communications.

Donghong Qin received his PhD degree from the Department of Computer Science and Technology, Tsinghua University, China in 2013. He has published more than 50 papers in international conferences and journals. He is a professor at the School of Artificial Intelligence, GuangXi Minzu University, Nanning, China. His current research interests include Internet architecture, Internet protocols, and algorithm design.

Yoong Choon Chang received his BEng (First Class Honors) in electrical & electronics engineering from the University of Northumbria at Newcastle (Northumbria University), UK, and MEngSc and PhD (engineering) degrees from Multimedia University, Malaysia. He is currently an associate professor at the Department of Electrical and Electronics Engineering, Universiti Tunku Abdul Rahman, and is a professional engineer registered with the Board of Engineers, Malaysia. In addition to academic teaching at the university, he regularly conducts consultations and training with professionals and engineers in the industry. His training clients include the Malaysia Rubber Council, PDRM, Motorola, Intel, Robert Bosch, Shrad Computing, Media Prima, Telekom Malaysia, and the Basic Human Needs Association of Japan. His research interests include IoT, digital transformation, and multimedia communications. He has more than 50 scientific publications in various international journals and conference proceedings. Over the past 15 years, 12 postgraduate students (10 master's and 2 PhD students) graduated under his supervision.

Allyson Gek Hong Sim joined SEEMOO Lab at the Technical University of Darmstadt since April 2016 as a Post Doctorate Researcher working with Prof. Dr. Eng. Matthias Hollick. She was a researcher at the Institute of IMDEA Networks from July 2011 to March 2016, where she completed her PhD studies. Dr. Sim received her PhD in telematics engineering from the University Carlos III of Madrid (UC3M) in March 2016, with the highest distinction. She also obtained a master's degree in telematics engineering from UC3M in October 2012 and another master's degree in telecommunications from Multimedia University, Malaysia in June 2011.

How to cite this article: S. Jiang, Y. L. Lee, M. L. Tham, D. Qin, Y. C. Chang, and A. G. H. Sim, *Proportionally fair load balancing with statistical quality of service provisioning for aerial base stations*, ETRI Journal **45** (2023), 887–898. DOI 10.4218/etrij.2023-0035