Digital Communications and Networks 10 (2024) 25–37

Contents lists available at ScienceDirect

Digital Communications and Networks

journal homepage: www.keaipublishing.com/dcan

Joint position optimization, user association, and resource allocation for load balancing in UAV-assisted wireless networks

Daosen Zhai [a,b], Huan Li [a], Xiao Tang [a], Ruonan Zhang [a], Haotong Cao [c,*]

[a] School of Electronics and Information, Northwestern Polytechnical University, Xi'an, 710072, China
[b] State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, 710071, China
[c] Department of Computing, Hong Kong Polytechnic University, Hong Kong SAR, China

ARTICLE INFO

Keywords:
Load balance
Unmanned aerial vehicle
User association
Resource management

ABSTRACT

Unbalanced traffic distribution in cellular networks results in congestion and degrades spectrum efficiency. To tackle this problem, we propose an Unmanned Aerial Vehicle (UAV)-assisted wireless network in which the UAV acts as an aerial relay to divert some traffic from the overloaded cell to its adjacent underloaded cell. To fully exploit its potential, we jointly optimize the UAV position, user association, spectrum allocation, and power allocation to maximize the sum-log-rate of all users in two adjacent cells. To tackle the complicated joint optimization problem, we first design a genetic-based algorithm to optimize the UAV position. Then, we simplify the problem by theoretical analysis and devise a low-complexity algorithm according to the branch-and-bound method, so as to obtain the optimal user association and spectrum allocation schemes. We further propose an iterative power allocation algorithm based on the sequential convex approximation theory. The simulation results indicate that the proposed UAV-assisted wireless network is superior to the terrestrial network in both utility and throughput, and the proposed algorithms can substantially improve the network performance in comparison with the other schemes.

1. Introduction

With the popularity of the Fifth-Generation Mobile Communication (5G) technologies, a variety of data-intensive services (e.g., ultra-high definition video, virtual reality, immersive game, etc.) emerge constantly, which lead to a sharp increase in wireless traffic volume [1]. The ever-growing wireless traffic imposes heavy pressure on cellular networks. As a tough problem, the unbalanced traffic distribution caused by the mobility of users results in congestion and degrades the spectrum efficiency. To tackle this problem, load balancing technologies have attracted extensive attention and inspired thorough research in both the industry and academia [2,3]. It usually solves the congestion of the hot-spot area by pushing some traffic to the adjacent underloaded cells. By this way, more users can be well served and the spectrum is made full use of, such that the overall network performance is enhanced. Therefore, it is important to investigate the load balancing technologies for 5G and the future wireless networks.

There have been a lot of works investigating the load balancing problems in the Heterogeneous Networks (HetNets), where [4–9] aims to improve the system throughput [9], aims to improve the coverage [10–13], focus on the Energy Efficiency (EE) and Spectrum Efficiency (SE) maximization, and some others deal with the multi-objective optimization problems [14,15]. To achieve load balance, the biased approach and the user association optimization [16] are usually adopted with some other control policies such as beamforming, power control, channel allocation, user scheduling, and so on. Besides, for 5G and the future wireless networks, some advanced technologies are incorporated with the load balancing technologies, e.g., Coordinated Multi-Point (CoMP) transmission [6], Millimeter Wave (mmWave) [8], Massive Multiple Input Multiple Output (mMIMO) [10], Non-Orthogonal Multiple Access (NOMA) [12], etc.

Although the existing works enhance the network performance, there are still some problems remained to be tackled. If the Base Station (BS) distance is large, it is unpractical to change the user association from one BS to another, as the large path loss counteracts the gain of load balancing. Under this circumstance, we can deploy an Unmanned Aerial Vehicle (UAV) as an air relay above the users to divert some traffic from the overloaded BSs to their neighbors quickly and efficiently. In

* Corresponding author.
E-mail addresses: zhaidaosen@nwpu.edu.cn (D. Zhai), lihuan1@mail.nwpu.edu.cn (H. Li), tangxiao@nwpu.edu.cn (X. Tang), rzhang@nwpu.edu.cn (R. Zhang), haotong.cao@polyu.edu.hk (H. Cao).

https://doi.org/10.1016/j.dcan.2022.03.011
Received 11 August 2021; Received in revised form 20 January 2022; Accepted 11 March 2022
Available online 25 March 2022

comparison with the terrestrial network, the UAV can usually establish Line-Of-Sight (LoS) links [17], and hence provides better coverage. Furthermore, due to the flexibility of UAVs, the spatial position or the moving trajectory of UAVs can be optimized to further improve the network performance [18,19], which is incomparable to the terrestrial networks. These advantages make UAVs have potential in dealing with the unbalanced load problem in cellular networks.

In recent years, the UAV-assisted wireless communications have been extensively studied. We can use UAV as aerial BS for wide-area coverage [20–22], air relay for long-distance transmission [23,24], and information fusion center for data collection [25,26]. Furthermore, UAV can also be used to assist the terrestrial networks to make up for their shortcomings [27–30]. However, to the best of our knowledge, none of the existing works investigates the load balancing problem in the air-and-ground cooperative networks. As discussed above, UAVs have the potential in assisting terrestrial networks to deal with the unbalanced load. Motivated by this, this paper studies how to utilize the UAV to adjust the traffic distribution among the terrestrial cells and how to optimize the system parameters to improve the network performance. The main contributions of this paper are summarized as follows.

- We propose an air-and-ground cooperative network to tackle the unbalanced problem in the terrestrial cellular networks. Specifically, when one cell is overloaded, some users in this cell are associated with its neighboring underloaded cell by the relaying of the UAV. To mitigate interference, we assume that the UAV relay utilizes the licensed band to receive data from one cell, while using the unlicensed band and the Decode-And-Forward (DF) protocol to transmit data to the designated users in another cell. By this way, the traffic is balanced among cells, and the spectrum is fully utilized.
- We formulate a joint UAV Position Optimization (PO), User Association (UA), Spectrum Allocation (SA), and Power Allocation (PA) problem with the objective to maximize the sum-log-rate of all users. The joint optimization problem is mixed-integer and non-convex, which is hard to tackle. To effectively solve the formulated problem, we first analyze its characters and adopt the genetic-based algorithm to optimize the UAV position. Then exploiting the structure of the problem, we design a branch-and-bound based algorithm to get the optimal user association and spectrum allocation schemes. Afterward, we devise an iterative power allocation algorithm according to the sequential convex approximation theory. Our proposed algorithms exhibit good performance with low complexity.
- We conduct extensive simulations to evaluate the performance of our proposed network and algorithms. The simulation results indicate that the air-and-ground cooperative network outperforms the traditional terrestrial network in terms of utility and throughput, which reflects that the traffic load can be balanced with the aid of the UAV relay. Furthermore, the gain derived by each control policy (i.e., PO, UA, SA, or PA) is analyzed by comparing the performance of different schemes. Moreover, it shows that the network performance can be further upgraded significantly by our proposed algorithms.

The rest of the paper is organized as follows. In Section 2, we introduce the load balancing and UAV related works. Section 3 presents the network model and the problem formulation. In Section 4, we describe the UAV position optimization algorithm, the joint user association and spectrum allocation algorithm, and the power allocation algorithm respectively. The simulation results are provided in Section 5. Finally, we conclude our paper in Section 6.

2. Related works

Abundant works have been conducted on the load balancing problems in traditional terrestrial networks with different objectives, such as system throughput [4–8], EE [10,11], SE [12,13], and multi-objective optimization [14,15]. The authors in Ref. [4] proposed a low-complexity user association scheme to improve the sum rate of the HetNets in a load balance way. In Ref. [5], the authors presented a Video Load Balancing Solution (ViLBaS), which improves the user Quality Of Experience (QoE) over multi-hop wireless mesh networks [6]. utilized the CoMP technique To upgrade the Signal to Interference-Plus-Noise Ratio (SINR) of all users and adopted a biased approach to achieve load balance. In Ref. [7], a Utility-based Mobility Load Balancing (UMLB) algorithm was proposed to minimize standard deviation with a higher average-user data rate [8]. introduced a load balancing user association scheme for the mmWave MIMO cellular networks, which enhances the data rates of cell-edge users under the max-min fairness. In Ref. [10], the authors regulated the downlink power consumption to achieve energy-efficient load balance in the cell-free massive MIMO networks. To achieve the energy-saving objective in the two-layer HetNets [11], derived a distributed load balancing algorithm based on the message-passing framework. In Ref. [12], a joint user association and Resource Block (RB) allocation scheme was presented for the NOMA HetNet, wherein the fairness index and spectrum efficiency are enhanced. In Ref. [13], a spectrum allocation method combining load balancing and Quality Of Service (QoS) guarantee was proposed for the HetNet [14]. developed both centralized and distributed user association algorithms, which achieve high EE and SE and maintain backhaul load balancing [15]. combined two biased Q-Learning Based Selection Strategies (QSS) to relief load imbalance and energy imbalance in the HetNets.

On the other hand, the UAV related technologies such as flight control, battery, material, etc., have got rapid development in recent years [31]. UAV equipped with communication devices can act as aerial relay or BS to provide temporary communication in a low-cost and high-mobility manner [18,19,32]. There have been extensive researches conducted on the UAV-assisted wireless communications, which can be mainly classified into three aspects [18]: UAV aerial BS [20–22], UAV-aided relaying [23,24] and UAV-aided information dissemination and data collection [25,26]. For the first category, an on-demand density-driven UAV 3D placement algorithm was proposed in Ref. [20]. In Ref. [21], the authors proposed a Flow-Level Model (FLM) to realistically characterize the UAV aerial BS performance, and adopted the Deep Reinforcement Learning (DRL) approach to optimize the UAV trajectory. In Ref. [22], UAVs were deployed to assist the terrestrial BSs in increasing the user coverage probability. Regarding the second category, UAV relay can achieve significant throughput gain because of its mobility compared with ground static relays [23]. studied the UAV relay to support ground node communication, and optimized the UAV's transmission power, trajectory, and flying speed to maximize the network energy efficiency [24]. investigated a hybrid satellite-terrestrial network where multiple UAV relays assist a satellite to communicate with ground UEs, and improved system performance by optimizing the 3-D deployment of the UAV relays. In terms of the third category [25], provided a scheme for using the NOMA technology to achieve UAV-assisted data collection on the Internet Of Things (IoT), while [26] tended to adopt Deep Q-Learning Based Resource Management (DQL-RM) scheme.

Although a large amount of researches investigate the UAV communications, few works focus on the air-and-ground cooperative networks [27–30]. In Ref. [27], the authors proposed a UAV-assisted D2D underlaid cellular network framework, in which each D2D pair either selects direct or UAV-assisted relay transmission, and all D2D transmission channels are shared with cellular users. To enhance the indoor coverage of urban high-rise buildings [28], proposed an architecture to introduce floating UAV relays into cellular systems, in which users can connect to both the relay and BS at the same time. In Ref. [29], the authors designed a relay protocol to provide better service for cell-edge users through the cooperation of BSs and UAVs. Besides, they proposed a joint height optimization and resource allocation algorithm to maximize the total data rate. Adopting NOMA technology, the authors in Ref. [30] used multiple UAVs to assist the ground BSs to serve the terminals simultaneously, and minimized the aggregate gap between the target rates and

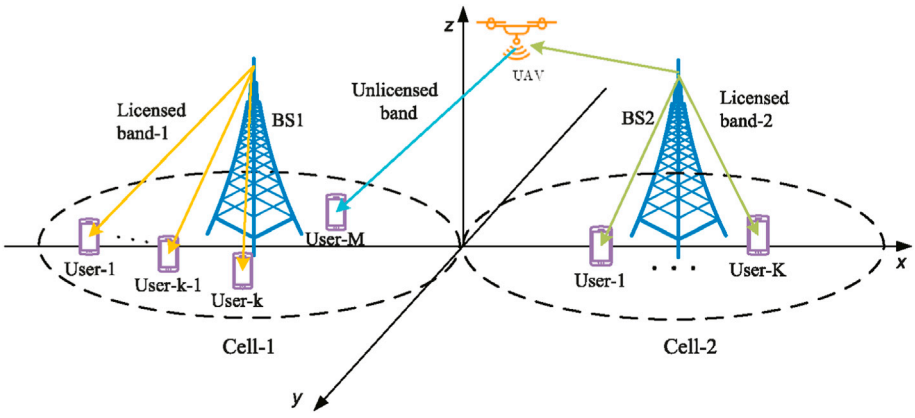


Fig. 1. Air-and-ground cooperative network for load balancing.

the throughput of terminals by optimizing the 3D positions of the UBSs and the resources (power, time, and bandwidth) of the network.

3. Network model and problem formulation

In this section, we begin with describing the proposed air-and-ground cooperative network, followed by defining the concerned control variables, and then formally give the problem formulation.

3.1. Network model

As shown in Fig. 1, we consider a downlink cellular network, which consists of two BSs named BS1 and BS2, a UAV relay, and multiple User Equipments (UEs). The two BSs use orthogonal frequencies to avoid inter-cell interference. Without loss of generality, we adopt the Cartesian coordinate system and define the midpoint of the two BSs as the origin, as depicted in Fig. 1. Let H_B denote the antenna height of the BSs, R denote the coverage radius of the BSs, and L denote the distance between BS1 and BS2, so the positions of the signals transmitted by BS1 and BS2 are $w_{B1} = (-\frac{L}{2}, 0, H_B)$ and $w_{B2} = (\frac{L}{2}, 0, H_B)$ respectively. In addition, we assume that there are M UEs in the coverage of BS1 (represented by $\mathbb{U}_1$) and K UEs in the coverage of BS2 (represented by $\mathbb{U}_2$). We use i as the UE index in cell-1, j as the UE index in cell-2, and k as the UE index in both cells. In our considered network, M is much larger than K, that is, we consider an unbalanced scenario. Due to the unbalanced traffic distribution, the communication resource of BS1 is insufficient while that of BS2 is surplus, such that the service capability of the whole network is not fully exploited.

Although the user association status can be changed by setting different bias for the two BSs, the cross-cell path loss is large and thus, the network performance is still poor. To tackle this problem, we propose an air-and-ground cooperative network, wherein an UAV relay is deployed to assist the ground network. The position of the UAV is denoted as $w_U = (x_U, y_U, H_U)$. To avoid the interference between the UAV relay and the BSs, the UAV relay uses the unlicensed band (e.g., WiFi [30]) to transmit data from BS2 to some of the UEs in cell-1 by the DF relay protocol.[1] Through the two-hop transmission, the problem of the large cross-cell path loss is tackled, and thereby more UEs in cell-1 can be associated with BS2 for load balancing. On the other hand, when the UAV deployment height is much higher than the ground, the BS-UAV links are usually LOS, which can enhance the coverage. Furthermore, the position of the UAV can be more easily adjusted according to the UEs' distribution to provide high quality of service. Therefore, the UAV relay can achieve better performance in comparison with the ground relay.

In this paper, we only consider the large-scale fading in the channel models, which can be seen as the time-averaged performance. Specifically, we use different path loss models to describe the BS-UE channel and the BS-UAV channel. More specifically, the BS-UAV channel is adopted as the experimental model given in Ref. [34], which is expressed as

$$PL = 38.5 + 22\lg(d[m]) + 20\lg(f_c[GHz]) + 1.005 \times 10^{-4}(H_U[m])^2 - 0.0286H_U[m] \quad (1)$$

where d is the distance between the transmitter and receiver, and f_c is the carrier frequency.

For the BS-UE and UAV-UE[2] links, we use the typical UMa NLOS model [35], which is specified as

$$PL = 32.4 + 30\lg(d[m]) + 20\lg(f_c[GHz]) \quad (2)$$

According to the above models, we define G_k^{B1}, G_k^U, and G_k^{B2} as the Channel Power Gains (CPGs) of the links from BS1, UAV, and BS2 to UE k, and define G_U^{B2} as the CPG of the link from BS2 to UAV.

The network performance can be enhanced by appropriately adjusting the system parameters. In this paper, we consider the user association, spectrum allocation, and power allocation. The user association is denoted as $\mathbf{X} = \{x_i^U, x_i^{B1}\}$, where x_i^U and x_i^{B1} are binary variables. If UE i is associated with the UAV relay, $x_i^U = 1$ and $x_i^{B1} = 0$, otherwise $x_i^U = 0$ and $x_i^{B1} = 1$. Since all the UEs in cell-2 are associated with BS2, we only consider the user associations of the UEs in cell-1. The spectrum allocation is denoted as $\mathbf{S} = \left\{s_i^{B1}, s_U^{B2}, s_i^U, s_j^{B2}\right\}$, where $s_i^{B1}, s_U^{B2}, s_i^U, s_j^{B2} \in [0, 1]$ represents the spectrum allocation ratio. More specifically, s_i^{B1} represents the spectrum ratio acquired by UE i in cell-1 allocated by BS1. The definitions of $s_U^{B2}, s_i^U, s_j^{B2}$ are similar to s_i^{B1}. It is worth noting that when the user association scheme is determined, part of s_i^{B1} or s_i^U are equal to 0. For example, if UE i is associated with BS1, then $s_i^U = 0$, and vice versa. The power allocation is denoted as $\mathbf{P} = \left\{p_i^{B1}, p_U^{B2}, p_i^U, p_j^{B2}\right\}$, where p_i^{B1} represents the transmission power from BS1 to UE i. It is similar to p_U^{B2}, p_i^U, and p_j^{B2}.

According to the above definitions, the achievable data rates of the UEs in cell-1 from BS1 can be written as

$$R_i^{B1} = x_i^{B1} s_i^{B1} \log_2\left(1 + \frac{p_i^{B1} G_i^{B1}}{\sigma^2}\right), i \in \mathbb{U}_1 \quad (3)$$

[1] We hope that the signal after the delivery of the UAV can still maintain a high SINR, but the AF scheme amplifies the useful signal while also amplifying the noise, reducing the SNR of the received signal. In addition, The simulation results in Ref. [33] show that the performance of the DF protocol is better than the AF protocol in terms of power loss, outage probability, and bit error rate (BER).

[2] In this paper, we consider low altitude UAV, the altitude of which is set as 50 m. As such, the channel model is almost the same as the ground network.

where σ^2 denotes the power of the general Additive White Gaussian Noise (AWGN).

Similarly, the achievable data rates of the UEs in cell-1 from the UAV relay can be expressed as

$$R_i^U = x_i^U s_i^U \log_2\left(1+\frac{p_i^U G_i^U}{\sigma^2}\right), i \in \mathbb{U}_1 \tag{4}$$

According to (3) and (4), we can get the data rate of the UEs in cell-1, i.e.,

$$\begin{aligned} R_i &= R_i^{B1} + R_i^U \\ &= x_i^{B1} s_i^{B1} \log_2\left(1+\frac{p_i^{B1} G_i^{B1}}{\sigma^2}\right) \\ &+ x_i^U s_i^U \log_2\left(1+\frac{p_i^U G_i^U}{\sigma^2}\right), i \in \mathbb{U}_1 \end{aligned} \tag{5}$$

Since the UEs in cell-2 can only be associated with BS2, their achievable data rates are given by

$$R_j = R_j^{B2} = s_j^{B2} \log_2\left(1+\frac{p_j^{B2} G_j^{B2}}{\sigma^2}\right), j \in \mathbb{U}_2 \tag{6}$$

In addition, the achievable date rate of the UAV relay from BS2 is given by

$$R_U^{B2} = s_U^{B2} \log_2\left(1+\frac{p_U^{B2} G_U^{B2}}{\sigma^2}\right) \tag{7}$$

3.2. Problem formulation

The position of the UAV w_U has a great effect on the CPGs G_U^{B2} and G_i^U which further affects the achievable data rate of the UEs. Besides, optimizing the user association **X**, spectrum allocation **S**, and power allocation **P** to achieve load balancing can also greatly enhance the network performance. On this basis, we investigate a utility maximization problem by jointly optimizing w_U, **X**, **S**, and **P**. The problem is mathematically formulated as

$$\begin{aligned} \max_{w_U,\mathbf{X},\mathbf{S},\mathbf{P}} & \sum_{i=1}^{M} \ln(R_i) + \sum_{j=1}^{K} \ln(R_j) \\ \text{s.t.C1}: & \ x_{\min} \le x_U \le x_{\max} \\ \text{C2}: & \ y_{\min} \le y_U \le y_{\max} \\ \text{C3}: & \ x_i^U + x_i^{B1} = 1, \forall i \\ \text{C4}: & \ x_i^U, x_i^{B1} \in \{0,1\}, \forall i \\ \text{C5}: & \ \sum_{i=1}^{M} s_i^{B1} \le 1 \\ \text{C6}: & \ \sum_{i=1}^{M} s_i^U \le 1 \\ \text{C7}: & \ \sum_{j=1}^{K} s_j^{B2} + s_U^{B2} \le 1 \\ \text{C8}: & \ 0 \le s_i^{B1}, s_i^U, s_j^{B2}, s_U^{B2} \le 1, \forall i, \forall j \\ \text{C9}: & \ \sum_{i=1}^{M} p_i^{B1} \le P_{B1}^{\max} \\ \text{C10}: & \ \sum_{i=1}^{M} p_i^U \le P_U^{\max} \\ \text{C11}: & \ \sum_{i=1}^{K} p_j^{B1} + p_U^{B2} \le P_{B2}^{\max} \\ \text{C12}: & \ p_i^{B1}, p_i^U, p_j^{B2}, p_U^{B2} \ge 0, \forall i, \forall j \\ \text{C13}: & \ \sum_{i=1}^{M} R_i^U \le R_U^{B2} \end{aligned} \tag{8}$$

The objective function is to maximize the log-sum-rate of all UEs, which takes fairness into consideration. The logarithmic function is an increasing concave function, which makes the contribution rate of larger independent variables to the function lower. Therefore, in the optimization process, all independent variables can be maintained at a similar level by using the form of sum-log-rate, which effectively ensures the fairness of user downlink data rate.

Constraints C1 and C2 specify the flight range of the UAV. This article is more concerned with the use of UAVs to solve the load balancing problem, and the optimization of the UAV height only increases the overall throughput of the system, which has no intuitive improvement on the unbalanced load. In order to reduce the computational complexity of the problem, the optimization of the UAV height has not been considered. It is noted that although we do not consider the height optimization in this paper, our proposed algorithm can be easily extended to the 3-dimensional position optimization problem. Constraints C3 and C4 together ensure that each UE in cell-1 is associated with the UAV or the BS1. C5–C7 respectively give the total spectrum allocation constraint for BS1, UAV, and BS2 (Without loss of generality, the bandwidth is normalized to one). Constraints C9–C11 limit the total transmission power of BS1, UAV, and BS2. Since the total data rate from UAV to UEs cannot exceed the rate from BS2 to UAV, we impose the constraint C13.

4. Algorithm design

The joint optimization problem in (8) is a mixed-integer and no-convex programming problem, which is very hard to tackle. To overcome this difficulty, we in this section divide it into three tractable subproblems (i.e., UAV position optimization subproblem, joint user association and spectrum allocation subproblem, and power allocation subproblem), and then propose three efficient algorithms to solve them sequentially.

4.1. UAV position optimization

In this subsection, we assume that the control variables **X**, **S**, and **P** are given in advance, and focus on the UAV position optimization. To achieve load balancing, a uniform user association scheme is concerned, that is, we associate the $(M-K)/2$ UEs closest to the BS2 with the UAV relay. By this way, each BS serves $(M+K)/2$ UEs. Under the uniform user association, the transmission power of the BS1, the BS2, and the UAU relay are equally allocated among the UEs. When **X** and **P** are determined, we can analytically obtain the optimal spectrum allocation scheme. The conclusions are summarized in Theorem 1 and Theorem 2, which are also used to reduce computational complexity in subsection III-B.

Theorem 1. *In a cell, to maximize the sum-log-rate of the UEs associated with the BS, the spectrum should be equally allocated among the UEs.*

Proof. See Appendix A. □

Remark 1. According to *Theorem 1*, we set $s_i^{B1} = \frac{2}{M+K}$ if UE i is associated with the BS1. Since partial UEs in cell-1 are indirectly associated with the BS2, we set $s_j^{B2} = \frac{2}{M+K}$ and $s_U^{B2} = \frac{M-K}{M+K}$.

Theorem 2. To maximize the sum-log-rate of the UEs associated with the UAV relay, the spectrum should be equally allocated among the UEs if $\sum_{k=1}^{N} R_k^U \le R_U^{B2}$ where $R_k^U = \frac{x_k^U}{N}\log_2\left(1+\frac{p_k^U G_k^U}{\sigma^2}\right)$ and N is the number of the UEs associated with the UAV relay, otherwise the rate should be equally allocated among the UEs.

Proof. *See* Appendix B. □

Remark 2. *According to Theorem 2, we set $s_i^U = \frac{2}{M-K}$ if UE i is associated with the UAV relay and $\sum_{i=1}^{M} R_i^U \le R_U^{B2}$. On the contrary, if $\sum_{i=1}^{M} R_i^U > R_U^{B2}$, we set $R_i^U = \frac{2R_U^{B2}}{M-K}$ and the corresponding s_i^U can be derived.*

It is noted that the conclusions in Theorems 1 and 2 always hold for any w_U, **X**, and **P**. Given **X**, **S**, and **P**, the UAV position optimization problem is transformed into

$$\begin{aligned} \max_{x_U,y_U} \quad & \sum_{i=1}^{M} \ln(R_i) + \sum_{j=1}^{K} \ln(R_j) \\ \text{s.t.} \quad & \text{C1}: \ x_{\min} \le x_U \le x_{\max} \\ & \text{C2}: \ y_{\min} \le y_U \le y_{\max} \end{aligned} \tag{9}$$

To solve the above non-convex optimization problem, we design a genetic-based algorithm (referred to as GBA). GBA is a computational model of the biological evolution process that simulates the natural

selection and genetic mechanism of Darwin's biological evolution theory. It is a method of searching for the optimal solution by simulating the natural evolution process. The algorithm uses computer simulation to transform the search process into cross-mutation of chromosomal genes in biological evolution. Compared with some conventional optimization algorithms, better results can be obtained more quickly when solving complex combinatorial optimization problems. Genetic algorithm has been widely used in the fields of combinatorial optimization, machine learning, signal processing, adaptive control, and artificial life.

For the position optimization in (9), we define $W_k = \{w_1^k, w_2^k, \ldots, w_n^k\}$ as the k-th iterative solution and W_0 as the initial solution which is constituted of n random UAV positions in the appointed range (C1 and C2 in (9)). For each iteration, every solution is evaluated in terms of its fitness value (i.e., the objective function in (9)), which is defined as

$$f_l^k = \sum_{i=1}^{M} \ln\left(R_i\left(w_l^k\right)\right) + \sum_{j=1}^{K} \ln\left(R_j\left(w_l^k\right)\right) \tag{10}$$

We define $F_k = \left\{f_1^k, f_2^k, \ldots, f_n^k\right\}$ as the fitness of all solutions in the k-th iteration. To simulate natural selection, we define p_S as the selection rate, and $v = n \times p_S$ solutions are selected by roulette selection. These solutions can survive to the next iteration and we define them as $W_{k+1}^1 = \left\{w_1^{k+1}, w_2^{k+1}, \ldots, w_v^{k+1}\right\}$. In order to simulate the crossover and mutation of chromosomes, we first convert W_{k+1}^1 to 22-digit binary set C_{k+1}^1 which represents the chromosomes. Then, $(n-v)/2$ pairs of chromosomes are randomly selected with replacement as parents. For each pair of them, two integers t_1 and t_2 from 1 to 22 are randomly generated. The two chromosomes exchange the part after t_1 according to the crossover rate p_C, and reverse the original number (i.e., 0 or 1) of t_2 according to the mutation rate p_M [36]. The new generated chromosomes are represented by $C_{k+1}^2 = \left\{w_{v+1}^{k+1}, w_{v+2}^{k+1}, \ldots, w_n^{k+1}\right\}$. Finally, convert C_{k+1}^2 back to the decimal set W_{k+1}^2. After completing these operations, W_{k+1}^1 and W_{k+1}^2 constitute the new solution set W_{k+1}. After completing s iterations, the optimal solution is selected from the final population W_s according to F_s. The detailed procedures are summarized in Algorithm 1.

Algorithm 1 The GBA for UAV position optimization.

```
Initialize X, S, P, n, s, p_S, p_C and p_M.
Randomly create an initial solution set W_0 .
for k = 1; k ≤ s; k++ do
    Calculate F_k according to (10).
    Select W_{k+1}^1 according to F_k and p_S .
    Convert W_{k+1}^1 to a binary set C_{k+1}^1.
    Select (n − v)/2 pairs of chromosomes from C_{k+1}^1 .
    for i = 1; i ≤ (n − v)/2; i++ do
        Generate two integers t_1 and t_2 from 1 to 22.
        Exchange the parts after t_1 according to p_C.
        Reverse the number of t_2 according to p_M.
        C_{k+1} = {C_{k+1}^1, C_{k+1}^2}.
        Convert C_{k+1} to a decimal set W_{k+1}.
    end for
end for
Calculate F_s according to (10).
Find the optimal UAV position w_U^* through F_s.
```

4.2. User association and spectrum allocation

In this subsection, we aimed at solving the joint user association $\mathbf{X}$ and spectrum allocation $\mathbf{S}$ problem. However, even w_U and $\mathbf{P}$ are given, the problem in (8) is still mixed-integer and nonconvex. To tackle this hard problem, we first transform it into a simplified form. In particular, $\mathbf{X}$ and $\mathbf{S}$ have the following relationship.

$$\begin{cases} 0 \le s_i^U \le x_i^U \\ 0 \le s_i^{B1} \le x_i^{B1} \end{cases} \tag{11}$$

With (11), the rate equation in (5) can be simplified as

$$\begin{aligned} R_i = {} & s_i^{B1} \log_2\left(1 + \frac{p_i^{B1} G_i^{B1}}{\sigma^2}\right) \\ & + s_i^U \log_2\left(1 + \frac{p_i^U G_i^U}{\sigma^2}\right), i \in \mathbb{U}_1 \end{aligned} \tag{12}$$

In this way, we can eliminate the product term of the integer variables $\mathbf{X}$ and the continuous variables $\mathbf{S}$, and thereby the problem in (8) can be equivalently reformulated as

$$\begin{aligned} \max_{\mathbf{X},\mathbf{S}} \quad & \textstyle\sum_{i=1}^{M} \ln(R_i) + \sum_{j=1}^{K} \ln\left(R_j\right) \\ \text{s.t.} \quad & \text{C1}: x_i^U + x_i^{B1} = 1, \forall i \\ & \text{C2}: x_i^U, x_i^{B1} \in \{0,1\}, \forall i \\ & \text{C3}: \textstyle\sum_{i=1}^{M} s_i^{B1} \le 1 \\ & \text{C4}: \textstyle\sum_{i=1}^{M} s_i^U \le 1 \\ & \text{C5}: \textstyle\sum_{j=1}^{K} s_j^{B2} + s_U^{B2} \le 1 \\ & \text{C6}: 0 \le s_i^{B1}, s_i^U, s_j^{B2}, s_U^{B2} \le 1, \forall i, \forall j \\ & \text{C7}: 0 \le s_i^U \le x_i^U, \forall i \\ & \text{C8}: 0 \le s_i^{B1} \le x_i^{B1}, \forall i \\ & \text{C9}: \textstyle\sum_{i=1}^{M} R_i^U \le R_U^{B2} \end{aligned} \tag{13}$$

To optimally solve the above problem, we employ the branch-and-bound method (BBA) to design a low-complexity algorithm. The designed algorithm consists of three main procedures, i.e., branching, bounding, and pruning. In what follows, we will introduce these procedures in detail.

The branching procedure is to create smaller feasible regions by constantly trying the user association scheme $\mathbf{X}$ for each x_i^U (or x_i^{B1}). In our concerned problem, each x_i^U (or x_i^{B1}) has two states, thereby the branching procedure can be described as a binary tree as shown in Fig. 2. The whole binary tree is denoted by Γ, wherein Γ_k represents the set of branches in the k-th layer. As depicted in Fig. 2, Γ_k contains at most 2^k branches. For each branch, we solve the following relaxed problem.

$$\begin{aligned} \max_{\mathbf{X},\mathbf{S}} \quad & \textstyle\sum_{i=1}^{M} \ln(R_i) + \sum_{j=1}^{K} \ln\left(R_j\right) \\ \text{s.t.} \quad & \text{C1}: x_i^U + x_i^{B1} = 1, \forall i \\ & \text{C2}: 0 \le x_i^U, x_i^{B1} \le 1, \forall i \\ & \text{C3}: \textstyle\sum_{i=1}^{M} s_i^{B1} \le 1 \\ & \text{C4}: \textstyle\sum_{i=1}^{M} s_i^U \le 1 \\ & \text{C5}: \textstyle\sum_{j=1}^{K} s_j^{B2} + s_U^{B2} \le 1 \\ & \text{C6}: 0 \le s_i^{B1}, s_i^U, s_j^{B2}, s_U^{B2} \le 1, \forall i, \forall j \\ & \text{C7}: 0 \le s_i^U \le x_i^U, \forall i \\ & \text{C8}: 0 \le s_i^{B1} \le x_i^{B1}, \forall i \\ & \text{C9}: \textstyle\sum_{i=1}^{M} R_i^U \le R_U^{B2} \end{aligned} \tag{14}$$

To further reduce the computational complexity, we simplify the problem in (14) by theoretical analysis. To maximize the objective function in (14), the condition $\sum_{j=1}^{K} s_j^{B2} + s_U^{B2} = 1$ must hold in C5. This is because if $\sum_{j=1}^{K} s_j^{B2} + s_U^{B2} < 1$, we can allocate the remainder spectrum to UE j, such that R_j can be improved. As a consequence, the objective function value is improved as well. On the other hand, Lemma 1 indicates that the optimal $\left\{s_j^{B2}\right\}$ are equal. Thus, we can get $s_j^{B2} = \frac{1 - s_U^{B2}}{K}, \forall j \in \mathbb{U}_2$ and thereby the problem in (14) can be reformulated as

$$
\begin{aligned}
\max_{\mathbf{X},\mathbf{S}} \quad & \sum_{i=1}^{M} \ln(R_i) + \sum_{j=1}^{K} \ln\left(\frac{1-s_U^{B2}}{K}\log_2\left(1+\frac{p_j^{B2}G_j^{B2}}{\sigma^2}\right)\right) \\
\text{s.t.} \quad & \text{C1}: x_i^U + x_i^{B1} = 1, \forall i \\
& \text{C2}: 0 \le x_i^U, x_i^{B1} \le 1, \forall i \\
& \text{C3}: \sum_{i=1}^{M} s_i^{B1} \le 1 \\
& \text{C4}: \sum_{i=1}^{M} s_i^U \le 1 \\
& \text{C5}: 0 \le s_i^{B1}, s_i^U, s_U^{B2} \le 1, \forall i, \forall j \\
& \text{C6}: 0 \le s_i^U \le x_i^U, \forall i \\
& \text{C7}: 0 \le s_i^{B1} \le x_i^{B1}, \forall i \\
& \text{C8}: \sum_{i=1}^{M} R_i^U \le R_U^{B2}
\end{aligned} \tag{15}
$$

Compared with (14), the variables $\left\{s_j^{B2}\right\}$ and the constraint C5 in (14) are crossed out in (15), and thus the problem becomes more easy to tackle.

Theorem 3. *The problem in (15) is a convex optimization problem.*

Proof. See Appendix C. □

According to Theorem 3, we can use some software (e.g., CVX) to solve the problem in (15). We define Q as the optimal solution of (15), and $h(Q)$ as its optimal value. In addition, in order to find the feasible solution of (13) as soon as possible, we choose the unselected UE which is with the largest uncertainty in $\mathbf{X}$. Specifically, the selection rule for the next branch UE is

$$
m = \arg\min_{i \notin I} |x_i^U - x_i^{B1}| \tag{16}
$$

where we define I as the UEs that has been branched [37].

In our designed algorithm, a lower bound is preserved and updated constantly, which is denoted by LB. In the initialization phase, we set $LB = -\infty$, such that LB can be updated by any feasible solutions. The main role of LB is utilized to prune the branches without the optimal solutions. According to the differences of $h(Q)$ and Q, we divide the results of the relaxed convex programming into three cases and give the corresponding operations respectively.

Case 1: $h(Q) > LB$ and $x_i^U = 0$ or 1, $\forall i$, which means that Q is an optimal solution. There is no need to test the user association schemes for the remainder UEs. Therefore, we terminate this branch and update Q^* as Q, and LB as $h(Q)$.

Case 2: $h(Q) > LB$ and $x_i^U \neq 0$ or 1, $\exists i$, which means that Q is not a feasible solution, but there is still potential for an optimal solution in this branch. Therefore, we need to preserve this branch and continue to branch according to (13). Besides, LB remains unchanged.

Case 3: $h(Q) \le LB$, which means that there is no optimal solution in this branch. So we have to prune it.

As shown in Γ_3 in Fig. 2, branch Q_4 meets Case 1, so it is terminated, and the optimal solution is updated. Branches Q_1 and Q_3 meet Case 2, and they are preserved for the next branching. Branch Q_2 is pruned, because it meets Case 3, which means that there is no optimal solution in this branch. In fact, during the execution of the algorithm, the existence of Case 1 and Case 3 can greatly reduce the search space. As shown in Fig. 2, Q_2 and Q_4 are pruned in Γ_3, such that the searching space is reduced by half. Therefore, our algorithm can optimally solve the joint user association and spectrum allocation problem in (13) with high efficiency. The detailed procedures are summarized in Algorithm 2.

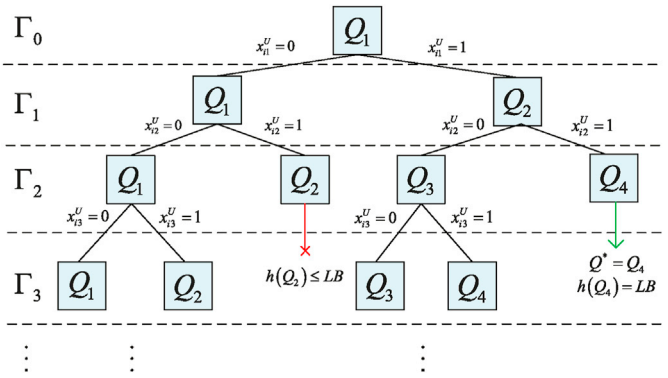


Fig. 2. Illustration for the branch-and-bound procedure.

Algorithm 2 The BBA for user association and spectrum allocation.

```
Initialize: Set LB = −∞ and A = ∅.
Add X_ini = {0, 0, ..., 0}_M into A.
while A ≠ ∅ do
    Take one branch X from A.
    Choose m according to (16).
    for l = 0; l ≤ 1; l++ do
        Set X_l = X, unless x_m = l.
        Solve the relaxed problem (15) as Q_l and h(Q_l).
        if h(Q_l) > LB then
            if x_l=0 or 1, ∀i then
                Q* = Q_l and LB= h(Q_l).
            else
                Add X_l into A.
            end if
        end if
    end for
end while
Output the final user association scheme X* and spectrum allocation scheme S* according to Q*.
```

4.3. Power allocation

After getting w_U^*, $\mathbf{X}^*$, and $\mathbf{S}^*$, we optimize the power allocation $\mathbf{P}$ for the BS1, the BS2, and the UAV relay in this subsection. The power allocation problem is reformulated as

$$
\begin{aligned}
\max_{\mathbf{P}} \quad & \sum_{i=1}^{M} \ln(R_i) + \sum_{j=1}^{K} \ln(R_j) \\
\text{s.t.} \quad & \text{C1}: \sum_{i=1}^{M} p_i^{B1} \le P_{B1}^{\max} \\
& \text{C2}: \sum_{i=1}^{M} p_i^U \le P_U^{\max} \\
& \text{C3}: \sum_{j=1}^{K} p_j^{B2} + p_U^{B2} \le P_{B2}^{\max} \\
& \text{C4}: p_i^{B1}, p_i^U, p_j^{B2}, p_U^{B2} \ge 0, \forall i, \forall j \\
& \text{C5}: \sum_{i=1}^{M} R_i^U - R_U^{B2} \le 0
\end{aligned} \tag{17}
$$

In constraint C5, $\sum_{i=1}^{M} R_i^U - R_U^{B2}$ is a standard d.c. (difference between two concave functions) function [38]. The trend of this function is uncertain and may be non-convex, so it cannot be solved by the traditional convex optimization method. Therefore, we need to approximate the constraint first to make it a convex constraint. To deal with it, we can

exploit the first order convex approximation. Specifically, we define $\mathbf{P}_U = \{p_i^U\}$ and $g(\mathbf{P}_U)$ as

$$g(\mathbf{P}_U) = \sum_{i=1}^{M} R_i^U = \sum_{i=1}^{M} x_i^U s_i^U \log_2\left(1 + \frac{p_i^U G_i^U}{\sigma^2}\right) \tag{18}$$

Then, the gradient of $g(\mathbf{P}_U)$ at any $\mathbf{P}_U$ can be expressed as

$$\nabla g(\mathbf{P}_U) = \left(\frac{\partial g(\mathbf{P}_U)}{\partial p_1^U}, \cdots \frac{\partial g(\mathbf{P}_U)}{\partial p_i^U}, \cdots \frac{\partial g(\mathbf{P}_U)}{\partial p_M^U}\right)^T \tag{19}$$

where $\frac{\partial g(\mathbf{P}_U)}{\partial p_i^U}$ is given by

$$\frac{\partial g(\mathbf{P}_U)}{\partial p_i^U} = \frac{s_i^U G_i^U}{\ln(2)\left(p_i^U G_i^U + \sigma^2\right)} \tag{20}$$

We approximate $g(\mathbf{P}_U)$ by its first order Taylor expansion at $\mathbf{P}_U^k$, i.e.,

$$g(\mathbf{P}_U) \approx g\left(\mathbf{P}_U^k\right) + \nabla g\left(\mathbf{P}_U^k\right)^T \left(\mathbf{P}_U - \mathbf{P}_U^k\right) \tag{21}$$

By the above approximation, the problem in (17) is transformed into

$$\begin{aligned} \max_{\mathbf{P}} \quad & \sum_{i=1}^{M} \ln(R_i) + \sum_{j=1}^{K} \ln\left(R_j\right) \\ \text{s.t.} \quad & \text{C1}: \sum_{i=1}^{M} p_i^{B1} \le P_{B1}^{\max} \\ & \text{C2}: \sum_{i=1}^{M} p_i^U \le P_U^{\max} \\ & \text{C3}: \sum_{j=1}^{K} p_j^{B2} + p_U^{B2} \le P_{B2}^{\max} \\ & \text{C4}: p_i^{B1}, p_i^U, p_j^{B2}, p_U^{B2} \ge 0, \forall i, \forall j \\ & \text{C5}: g\left(\mathbf{P}_U^k\right) + \nabla g\left(\mathbf{P}_U^k\right)^T \left(\mathbf{P}_U - \mathbf{P}_U^k\right) - R_U^{B2} \le 0 \end{aligned} \tag{22}$$

Theorem 4. *The problem in (22) is a convex optimization problem.*

Proof. See Appendix D. □

Based on the Sequential Convex Approximation (SCA) theory [39], we propose an iterative power allocation algorithm, which is summarized in Algorithm 3. For the proposed algorithm, we have the following lemma.

Theorem 5. *Algorithm 3 can always converge to a feasible solution of (17).*

Proof. See Appendix E. □

Algorithm 3 The SCA for power allocation.

1: Initialize $\mathbf{P}_U^0$ as the values given in subsection III-B.
2: Set $k = 0$, $\delta = 0.1$, and $f_k = 0$.
3: **repeat**
4: Construct the problem in (22) with $\mathbf{P}_U^k$.
5: Update $k = k + 1$.
6: Solve the problem in (22) and get the optimal solution $\mathbf{P}^k$ and the optimal value f_k.
7: Update $\mathbf{P}_U^k$ according to $\mathbf{P}^k$.
8: **until** $f_k - f_{k-1} \le \delta$
9: Output the final power allocation scheme $\mathbf{P}^* = \mathbf{P}^k$.

5. Simulation results

In this section, we provide abundant simulation results to evaluate the performance of our proposed network and algorithms. The positions of the UEs are randomly generated in the coverage area of the two BSs. Each UE has sufficient data to transmit (i.e., saturated traffic). The common simulation parameters are listed in Table I, and the others are specified in the simulation. For notational simplicity, we use the acronyms PO, UA, SA, and PA to represent the position optimization, user association, spectrum allocation, and power allocation respectively. To fully demonstrate the advantages of our scheme (referred to as the Joint Optimization), we compare it with the other four schemes (namely the Terrestrial Network, the UA + SA + PA, the PO + PA, and the PO + UA + SA) on the performance of utility (i.e., the objective function value in (8)) and throughput (i.e., the total data rate of all UEs). The detailed introduction of the four schemes is given in the following.

Table 1
Simulation parameters.

Height of the BS, H_B	30 m
Height of the UAV, H_U	50 m
Flight zone, $[x_{\min}, x_{\max}, y_{\min}, y_{\max}]$	[-3R,3R,-R,R]
Max power of the BS, $P_B^{\max}$	20 Watt
Noise power spectrum density	−174 dBm/Hz
Carrier frequency, f_c	3.5 GHz
Channel bandwidth	2 MHz

- Terrestrial Network: This scheme adopts the traditional terrestrial network without the aid of the UAV relay. All UEs are directly associated with their adjacent BS. According to Theorem 1, the spectrum should be equally allocated among the UEs, as it can maximize the utility function. The power allocation problem is convex and can be directly solved by the CVX. This scheme is similar with that given in Ref. [40]. We use this scheme as a benchmark to show the potential of the air-and-ground cooperative network in load balancing.
- UA + SA + PA: This scheme employs the UAV relay and optimizes the user association, spectrum allocation, and power allocation. However, the UAV position is randomly generated in the feasible region. This scheme is used to show the importance of the UAV position optimization.
- PO + PA: This scheme employs the UAV relay and optimizes the UAV position and power allocation. However, the user association and spectrum allocation are not optimized and set as the initial values given in subsection III-A. By comparing with this scheme, we aim to show the performance gain by jointly optimizing the user association and spectrum allocation.
- PO + PA: This scheme uses the UAV relay and optimizes the UAV position, user association, and spectrum allocation. However, the power allocation is not optimized and set as the initial values given in subsection III-A. We compare it with our scheme to illustrate the importance of the power allocation.

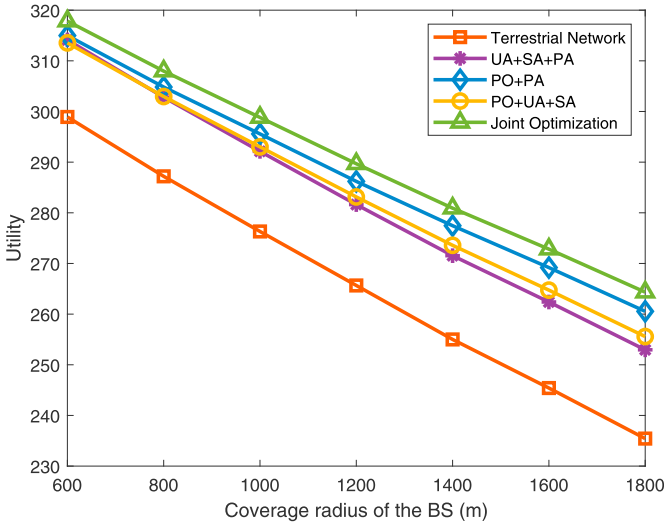
Fig. 3 plots the utility and throughput of different schemes versus the coverage radius of the BS R. Firstly, we can find that the utility and throughput of all schemes decrease with R. This is because with the increment of R, the transmitter-and-receiver distance increases, and the channel power gain decreases accordingly. As a consequence, the achievable data rates of all UEs (i.e., throughput) degrade, which further leads to the decrement of utility. In addition, this figure shows that the air-and-ground cooperative network is superior to the traditional terrestrial network even if the system parameters are not optimized. This indicates that the UAV relay can help the terrestrial BSs achieve better load balancing so as to improve the system performance. Moreover, the Joint Optimization outperforms the other schemes in terms of utility and throughput. This result verifies the efficiency of our proposed Algorithms 1-3 and demonstrates the importance of PO, UA, SA, and PA. By appropriately setting the system parameters, the performance of the air-and-ground cooperative network can be further enhanced.

Fig. 4 shows the effect of the BS distance on the utility and throughput. From Fig. 4(a), we observe that the utility of the air-and-ground cooperative network decreases with the BS distance. The reason is that when the BS distance is large, it becomes more difficult for the UEs in cell-1 to associate with the BS2 by the relay. This phenomenon is more obvious for the throughput as shown in Fig. 4(b). However, with respect to the UA + SA + PA, the performance of the Joint Optimization, the PO + PA, and the PO + PA degrades more slowly. This is because through optimizing the UAV position, it can weaken the effect of the BS distance. This result shows

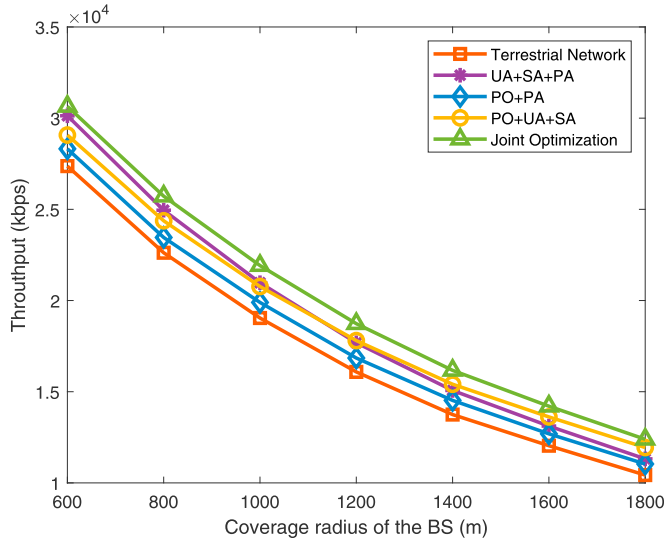
the importance of the UAV position optimization. It also demonstrates that the UAV relay is more flexible and can achieve better performance than the ground relay. Furthermore, we can find that our scheme can still achieve larger utility and throughput in comparison with the terrestrial network even when the BS distance is large. The performance gain mainly owes to the good air-to-ground channels.

Fig. 5 illustrates how the number of UEs in cell-1 M affects the utility and throughput. From this figure, we can see that the utility increases with M, while the throughput decreases. This is because the total spectrum is limited, the achievable data rate of each UE drops off when M is large, which results in the decline of the throughput. Since the utility function is the sum-log-rate form, the total value is still improved. Besides, it can be observed from Fig. 5(a) that the utility gap between the air-and-ground cooperative network and the terrestrial network becomes wider as M increases. The reason is when M is large, the unbalanced condition of the network becomes more serious. Through our proposed algorithms, the unbalanced problem can be alleviated, wherein the control policy of user association and spectrum allocation plays a significant role. As shown in Fig. 5(b), the throughput of the PO + PA decreases sharply with M, and even lower the Terrestrial Network. This tells us the fact that the uniform user association scheme is not always optimal.

Fig. 6 plots how the number of UEs in cell-2 K affects the utility and throughput. The over trend of Fig. 6 is identical to that of Fig. 6. The difference is that the utility gap between the air-and-ground cooperative network and the terrestrial network narrows with the increment of K. Since the traffic load of the two BSs is balanced when K is close to M, less UEs are associated with the UAV relay, and therefore the performance gain derived by the user association becomes smaller. Fig. 6(b) shows that the throughput of the PO + PA is smaller than the UA + SA + PA and the PO + UA + SA when K is small, and the condition is reverse when K is large. This result explains that the user association is more important when the system is unbalanced, and the power allocation is more important when the system is balanced with much UEs. By joint optimization, our scheme can always achieve the best performance in both utility and throughput under any conditions. Fig. 7 shows the utility and throughput versus the transmission power of the UAV. In our concerned system, the UAV utilizes the unlicensed bands and the BSs utilize the

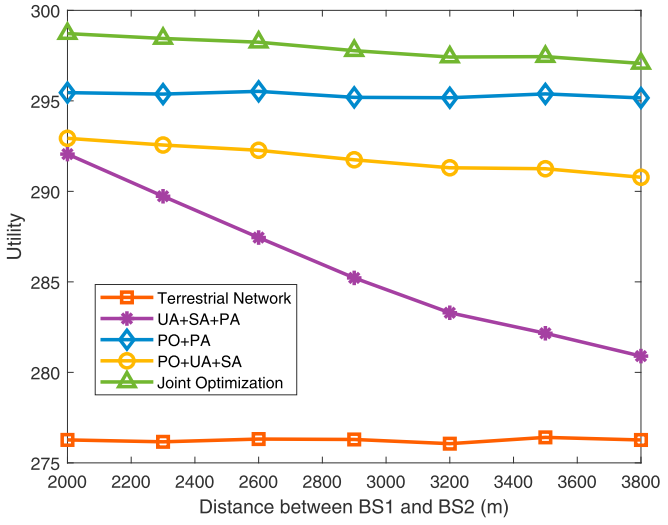


(a) Utility versus the BS coverage radius.

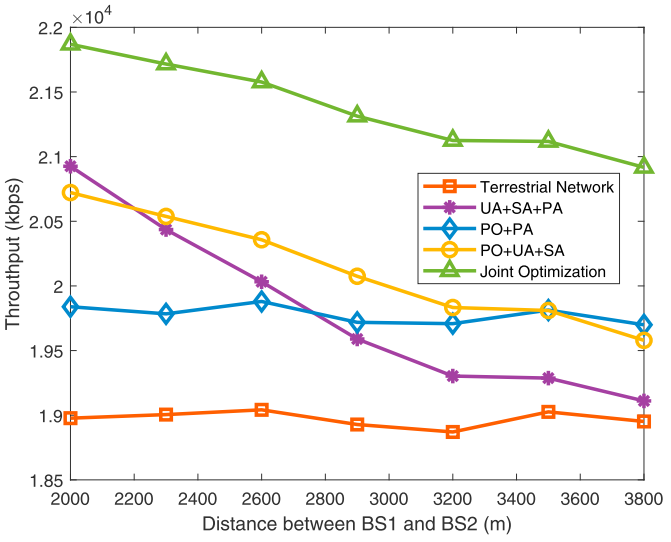


(b) Throughput versus the BS coverage radius.

Fig. 3. The effect of the BS coverage radius R on the utility and throughput ($L = 2R$, $M = 40$, $K = 10$, $P_U^{\max} = 2$ Watt).

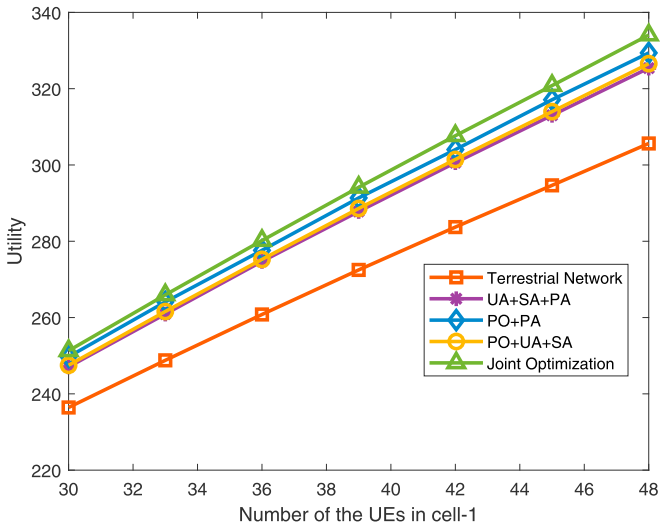


(a) Utility versus the station distance.

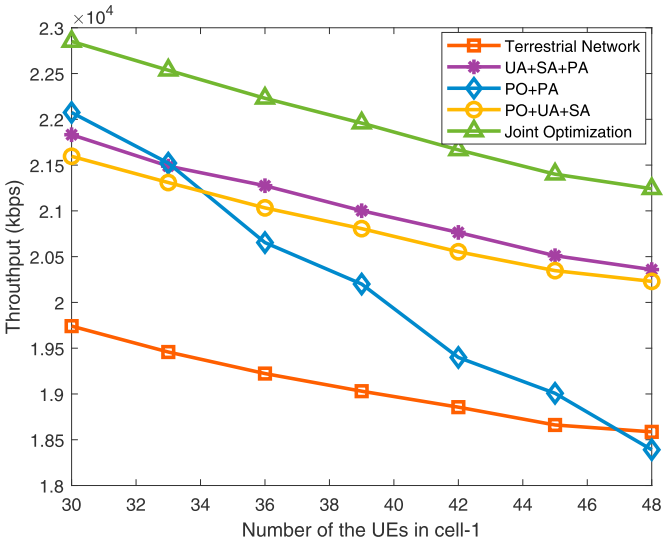


(b) Throughput versus the station distance.

Fig. 4. The effect of the station distance L on the utility and throughput ($R = 1$ km, $M = 40$, $K = 10$, $P_U^{\max} = 2$ Watt).

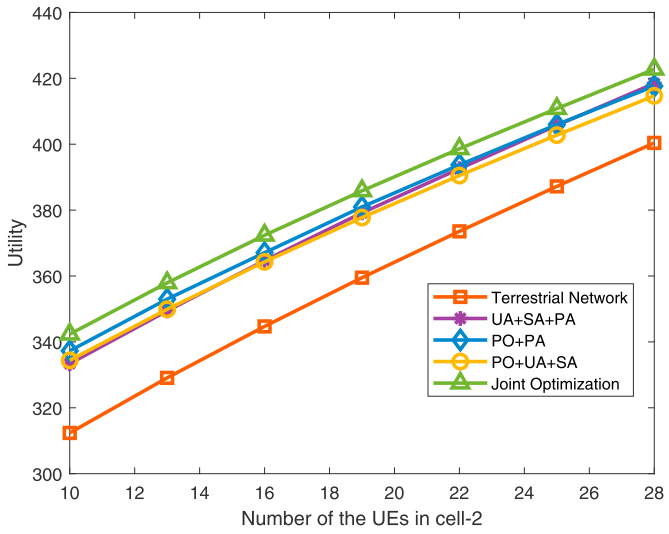


(a) Utility versus the number of UEs in cell-1.

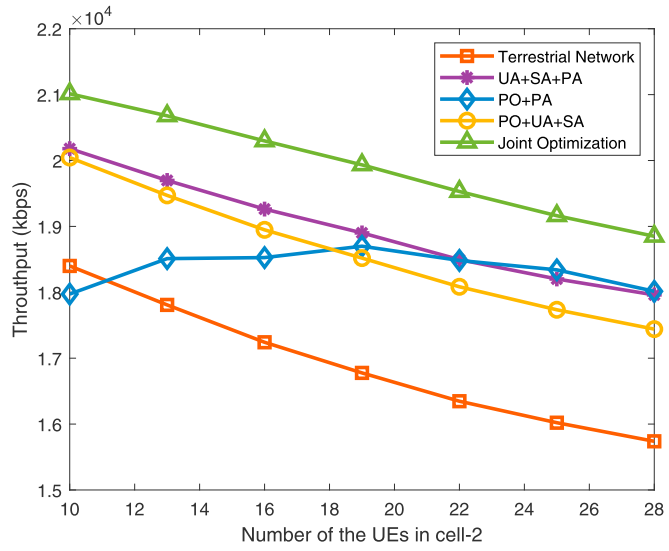


(b) Throughput versus the number of UEs in cell-1.

Fig. 5. The effect of the UE number in cell-1 M on the utility and throughput ($R = 1$ km, $L = 2$ km, $K = 10$, $P_U^{\max} = 2$ Watt).



(a) Utility versus the number of UEs in cell-2.



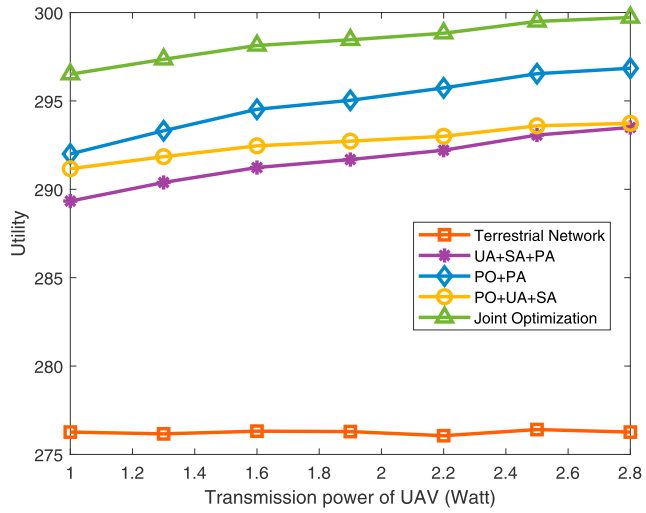
(b) Throughput versus the number of UEs in cell-2.

Fig. 6. The effect of the UE number in cell-2 K on the utility and throughput ($R = 1$ km, $L = 2$ km, $M = 50$, $P_U^{\max} = 2$ Watt).

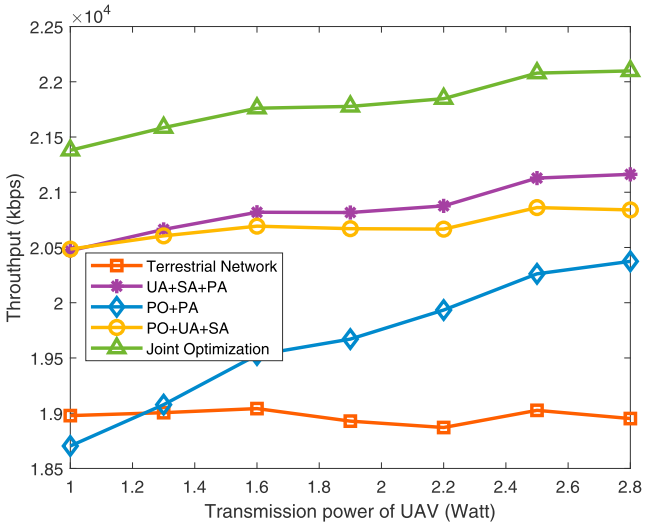
licensed bands, thus there is no inter-cell interference between the BS and the UAV relay. Therefore, the performance of the air-and-ground cooperative network is improved as the transmission power of the UAV increases. In the PO + UA + SA, the power allocation is not optimized and equally allocated among the UEs. As a consequence, the growth rate of its utility and throughput is the lowest. As depicted in Fig. 7(a), the utility of the PO + UA + SA is almost the same with the UA + SA + PA when $P_U^{\max} = 2.8$ watt, and it will be the lowest if $P_U^{\max}$ increases continuously. The similar trend can be seen in Fig. 7(b). This fully demonstrates the importance of the power allocation and the efficiency of our proposed Algorithm 3. Furthermore, we can find that the throughput of the PO + PA is smaller than the terrestrial network when $P_U^{\max}$ is small. Therefore, it is vital to optimize the user association and the spectrum allocation when the UAV power is limited.

6. Conclusion

In this paper, we have investigated how to incorporate the UAV with the cellular network and how to optimize the system parameters to achieve load balancing. Firstly, we have employed the UAV as an aerial relay to assist the terrestrial network to adjust the traffic distribution between two adjacent cells. Then, we have proposed three effective algorithms to jointly optimize the UAV position, user association, spectrum allocation, and power allocation with the objective to maximize the sum-log-rate of all users. Finally, the simulation results have been provided to indicate that the air-and-ground cooperative network is superior to the terrestrial network even without optimizing the system parameters, and the proposed algorithms can further improve the network performance significantly and outperform the other schemes in terms of throughput and utility.



(a) Utility versus the transmission power of the UAV.



(b) Throughput versus the transmission power of the UAV.

Fig. 7. The effect of the UAV transmission power $P_U^{\max}$ on the utility and throughput ($R = 1$ km, $L = 2$ km, $M = 40$, $K = 10$).

Funding

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1807003, in part by the National Natural Science Foundation of China under Grants 61901381, 62171385, and 61901378, in part by the Aeronautical Science Foundation of China under Grant 2020Z073053004, in part by the Foundation of the State Key Laboratory of Integrated Services Networks of Xidian University under Grant ISN21-06, in part by the Key Research Program and Industrial Innovation Chain Project of Shaanxi Province under Grant 2019ZDLGY07-10, and in part by the Natural Science Fundamental Research Program of Shaanxi Province under Grant 2021JM-069.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Proof of Theorem 1

We use B to represent BS1 or BS2 for simplicity and define

$$r_k^B = \log_2\left(1+\frac{p_k^B G_k^B}{\sigma^2}\right) \tag{A.1}$$

Then the achievable data rate of UE k can be expressed as

$$R_k^B = s_k^B r_k^B \tag{A.2}$$

Assume that D UEs are associated with the BS. The sum-log-rate of the UEs can be written as

$$\begin{aligned}\sum_{k=1}^{D}\ln\left(R_k^B\right) &= \sum_{k=1}^{D}\ln\left(s_k^B r_k^B\right)\\ &= \sum_{k=1}^{D}\ln\left(s_k^B\right)+\sum_{k=1}^{D}\ln\left(r_k^B\right)\end{aligned} \tag{A.3}$$

Note that r_k^B is a constant, we can get

$$\begin{aligned}&\arg\max_{\mathbf{S}_B}\ \sum_{k=1}^{D}\ln\left(R_k^B\right)\\ \Leftrightarrow&\arg\max_{\mathbf{S}_B}\ \sum_{k=1}^{D}\ln\left(s_k^B\right)\\ \Leftrightarrow&\operatorname{argmax}_{\mathbf{S}_B}\ \prod_{k=1}^{D}s_k^B\end{aligned} \tag{A.4}$$

where $\mathbf{S}_B = \left\{s_k^B\right\}$.

The optimization problem about $\mathbf{S}_B$ in (8) is equivalent to

$$\begin{aligned}\max_{\mathbf{S}_B}\quad & \prod_{k=1}^{D}s_k^B\\ \text{s.t.}\quad & \text{C1}:\ \sum_{k=1}^{D}s_k^B\le 1\\ & \text{C2}:\ 0\le s_k^B\le 1,\ \forall k\end{aligned} \tag{A5}$$

According to the Arithmetic Mean-Geometric Mean (AM-GM) inequality theorem, we can know that

$$\left(\prod_{k=1}^{D}s_k\right)^{\frac{1}{D}}\le\frac{\sum_{k=1}^{D}s_k}{D}\le\frac{1}{D} \tag{A.6}$$

$$\begin{aligned}&\max_{\mathbf{S}_B}\ \prod_{k=1}^{D}s_k=\left(\frac{\sum_{k=1}^{D}s_k}{D}\right)^{D}=\frac{1}{D^D}\\ &\Leftrightarrow s_k=\frac{1}{D},\forall k\end{aligned} \tag{A.7}$$

Therefore, only when the spectrum is equally allocated among the UEs, the sum-log-rate is maximized.

Appendix B. Proof of Theorem 2

Assume there are N UEs associated with the UAV relay and denote $\mathbf{S}_U = \left\{s_k^U\right\}$. The optimization problem about $\mathbf{S}_U$ in (8) can equivalently formulated as

$$\begin{aligned}\max_{\mathbf{S}_U}\quad & \prod_{k=1}^{N}R_k^U\\ \text{s.t.}\quad & \text{C1}:\ \sum_{k=1}^{N}s_k^U\le 1\\ & \text{C2}:\ 0\le s_k^U\le 1,\ \forall k\\ & \text{C3}:\ \sum_{k=1}^{N}R_k^U\le R_U^{B2}\end{aligned} \tag{B.1}$$

If C3 is a loose constraint (can be omitted), we can know that the optimal spectrum allocation scheme is $s_k^U=\frac{1}{N},\ \forall k$ according to Theorem 1. In this case, $R_k^U=\frac{1}{N}\log_2\left(1+\frac{p_k^U G_k^U}{\sigma^2}\right)=R_k^U\ \forall k$. That is when $\sum_{k=1}^{N}R_k^U\le R_U^{B2}$, the spectrum should be equally allocated among the UEs to maximize the sum-log-rate.

On the other hand, when $\sum_{k=1}^{N}R_k^U>R_U^{B2}$, C3 is a tight constraint (i.e., $\sum_{k=1}^{N}R_k^U=R_U^{B2}$ with the optimal solution). Now, we cannot set $s_k^U=\frac{1}{N}$, as it will make C3 unsatisfied.

Then, according to the Arithmetic Mean-Geometric Mean (AM-GM) inequality theorem, we know that

$$\left(\prod_{k=1}^{N}R_k^U\right)^{\frac{1}{N}}\le\frac{\sum_{k=1}^{N}R_k^U}{N}\le\frac{R_U^{B2}}{N} \tag{B.2}$$

$$\begin{aligned}&\max_{\mathbf{S}_U}\ \prod k=1^N R_k^U=\left(\frac{\sum_{k=1}^{N}R_k^U}{N}\right)^{N}=\left(\frac{R_U^{B2}}{N}\right)^{N}\\ &\Leftrightarrow R_k^U=\frac{R_U^{B2}}{N},\forall k\end{aligned} \tag{B.3}$$

Therefore, when $\sum_{k=1}^{N}R_k^U>R_U^{B2}$, the rate should be equally allocated among the UEs to maximize the log-sum-rate.

To this end, we have proved Theorem 2.

Appendix C. Proof of Theorem 3

$R_i = s_i^{B1}\log_2\left(1+\frac{p_i^{B1}G_i^{B1}}{\sigma^2}\right)+s_i^U\log_2\left(1+\frac{p_i^UG_i^U}{\sigma^2}\right)$ is a linear function in s_i^{B1} and s_i^U. $\ln(y)$ is concave and nondecreasing in y. According to the composition principle [41], we can get that $\ln(R_i)$ is concave in s_i^{B1} and s_i^U. Since $\frac{1-s_U^{B2}}{K}\log_2\left(1+\frac{p_j^{B2}G_j^{B2}}{\sigma^2}\right)$ is also a linear function in s_U^{B2}, $\ln\left(\frac{1-s_U^{B2}}{K}\log_2\left(1+\frac{p_j^{B2}G_j^{B2}}{\sigma^2}\right)\right)$ is a concave function. The objective function in (15) is the sum of $M+K$ concave functions, and thus is also concave. Besides, all constraints in (15) are linear. Thus, the problem in (15) is a convex optimization problem.

Appendix D. Proof of Theorem 4

As shown in (5), R_i is a concave function in p_i^{B1} or p_i^U. $\ln(y)$ is concave and nondecreasing in y. According to the composition principle [41], we can get that $\ln(R_i)$ is a concave function. Similarly, $\ln(R_j)$ is also a concave function. The sum of concave functions preserves concavity, thus the objective function in (22) is also concave. In (22), C1–C4 are all linear constraints. The left side of C5 is a linear function $g(\mathbf{P}_U^k)+\nabla g(\mathbf{P}_U^k)^T(\mathbf{P}_U-\mathbf{P}_U^k)$ plus a convex function $-R_U^{B2}$, therefore it is a convex function.

According to the above, the problem in (22) satisfies the standard form of the convex optimization.

Appendix E. Proof of Theorem 5

$g(\mathbf{P}_U)$ is a concave function. Thus, its first order Taylor expansion is no smaller than itself, i.e.,

$$g(\mathbf{P}_U^k)+\nabla g(\mathbf{P}_U^k)^T(\mathbf{P}_U-\mathbf{P}_U^k)\geq g(\mathbf{P}_U) \tag{E.1}$$

If C5 in (22) is satisfied, we can get

$$\begin{aligned} &g(\mathbf{P}_U)-R_U^{B2}\\ &\leq g(\mathbf{P}_U^k)+\nabla g(\mathbf{P}_U^k)^T(\mathbf{P}_U-\mathbf{P}_U^k)-R_U^{B2}\\ &\leq 0 \end{aligned} \tag{E.2}$$

Therefore, the optimal solution of (22) must be a feasible solution of (17).

On the other hand, as shown in Algorithm 3, $\mathbf{P}_U^0$ is initialized as the values given in subsection III-B. Thus, the problem in (22) is always feasible, as the power allocation scheme given in subsection III-B satisfies all the constraints in (22). Furthermore, $\mathbf{P}^k$ is used as the Taylor expansion point to construct the approximate problem (22) in the k+1-th iteration. As such, $\mathbf{P}^k$ must be a feasible solution for the k+1-th approximate problem. After each iteration, the objective function value must be no smaller than the previous value, i.e.,

$$f_k-f_{k-1}\geq 0 \tag{E.3}$$

Since the optimal value of (17) is upper limited, Algorithm 3 can always converge to the optimal value or a fixed value after finite iterations.

To this end, we have proved Lemma 5.

References

[1] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J.J. Ramos-Munoz, J.M. Lopez-Soler, A survey on 5G usage scenarios and traffic models, IEEE Commun. Surveys Tuts. 22 (2) (2020) 905–929.

[2] J.G. Andrews, S. Singh, Q. Ye, X. Lin, H.S. Dhillon, An overview of load balancing in HetNets: old myths and open problems, IEEE Wireless Commun. 21 (2) (2014) 18–25.

[3] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K.-K. Wong, R. Schober, L. Hanzo, User association in 5G networks: a survey and an outlook, IEEE Commun. Surveys Tuts. 18 (2) (2016) 1018–1044.

[4] Z.Z.P. Han, Z. Wang, User association for load balance in heterogeneous networks with limited CSI feedback, IEEE Commun. Lett. 24 (5) (2020) 1095–1099.

[5] A. Hava, Y. Ghamri-Doudane, J. Murphy, G. Muntean, A load balancing solution for improving video quality in loaded wireless network conditions, IEEE Trans. Broadcast. 65 (4) (2019) 742–754.

[6] M.O.A.R.Q. Shaddad, N.A. Alsarori, T.M. Shami, Biased user association in 5G heterogeneous networks, in: 2021 International Conference of Technology, Science and Administration (ICTSA), 2021, pp. 1–4.

[7] K.M. Addali, S.Y.B. Melhem, Y. Khamayseh, Z. Zhang, M. Kadoch, Dynamic mobility load balancing for 5G small-cell networks based on utility functions, IEEE Access 7 (2019) 126998–127011.

[8] A. Alizadeh, M. Vu, Load balancing user association in millimeter wave MIMO networks, IEEE Trans. Wireless Commun. 18 (6) (2019) 2932–2945.

[9] D. Zhai, Q. Shi, R. Zhang, X. Tang, H. Cao, Coverage maximization for heterogeneous aerial networks, IEEE Wireless Commun. Lett.doi:10.1109/LWC.2021.3121076.

[10] E.B.T. Van Chien, E.G. Larsson, Joint power allocation and load balancing optimization for energy-efficient cell-free massive MIMO networks, IEEE Trans. Wireless Commun. 19 (10) (2020) 6798–6812.

[11] H.S.S.H. Lee, M. Kim, I. Lee, Belief propagation for energy efficiency maximization in wireless heterogeneous networks, IEEE Trans. Wireless Commun. 20 (1) (2021) 56–68.

[12] H.K.S. Sobhi-Givi, M.G. Shayesteh, N. Rajatheva, Resource allocation and user association for load balancing in NOMA-Based cellular heterogeneous networks, in: 2020 Iran Workshop on Communication and Information Theory (IWCIT), 2020, pp. 1–6.

[13] Y. C, et al., A spectrum allocation method based on load balancing for heterogeneous networks of smart grid, in: 2020 International Wireless Communications and Mobile Computing (IWCMC), 2020, pp. 111–115.

[14] M. Javad-Kalbasi, S. Valaee, Energy and spectrum efficient user association for backhaul load balancing in small cell networks, in: GLOBECOM 2020 - 2020 IEEE Global Communications Conference, 2020, pp. 1–6.

[15] D.X.C. Jialing, Y. Mingxi, J. Bingli, Q -learning based selection strategies for load balance and energy balance in heterogeneous networks, in: 2020 5th International Conference on Computer and Communication Systems (ICCCS), 2020, pp. 728–732.

[16] W. Teng, M. Sheng, K. Guo, Z. Qiu, Content placement and user association for delay minimization in small cell networks, IEEE Trans. Veh. Technol. 68 (10) (2019) 10201–10215.

[17] W. Khawaja, I. Guvenc, D.W. Matolak, U.-C. Fiebig, N. Schneckenburger, A survey of air-to-ground propagation channel modeling for unmanned aerial vehicles, IEEE Commun. Surveys Tuts. 21 (3) (2019) 2361–2391.

[18] Y. Zeng, R. Zhang, T.J. Lim, Wireless communications with unmanned aerial vehicles: opportunities and challenges, IEEE Commun. Mag. 54 (5) (2016) 36–42.

[19] Y. Li, L. Cai, UAV-assisted dynamic coverage in a heterogeneous cellular system, IEEE Netw. 31 (4) (2017) 56–61.

[20] C. Lai, C. Chen, L. Wang, On-demand density-aware UAV base station 3D placement for arbitrarily distributed users with guaranteed data rates, IEEE Wireless Commun. Lett. 8 (3) (2019) 913–916.

[21] V. Saxena, J. Jaldén, H. Klessig, Optimal UAV base station trajectories using flow-level models for reinforcement learning, IEEE Trans. Cogn. Commun. Netw. 5 (4) (2019) 1101–1112.

[22] Z. Chen, H. Zhang, UAV-assisted networks through a tunable dependent model, IEEE Commun. Lett. 24 (5) (2020) 1110–1114.
[23] S. Ahmed, M.Z. Chowdhury, Y.M. Jang, Energy-efficient UAV relaying communications to serve ground nodes, IEEE Commun. Lett. 24 (4) (2020) 849–852.
[24] P.K. Sharma, D. Deepthi, D.I. Kim, Outage probability of 3-D mobile UAV relaying for hybrid satellite-terrestrial networks, IEEE Commun. Lett. 24 (2) (2020) 418–422.
[25] W. Chen, S. Zhao, R. Zhang, Y. Chen, L. Yang, UAV-assisted data collection with nonorthogonal multiple access, IEEE Internet Things J 8 (1) (2021) 501–511.
[26] M. Yi, X. Wang, J. Liu, Y. Zhang, B. Bai, Deep reinforcement learning for fresh data collection in UAV-assisted IoT networks, in: IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2020, pp. 716–721.
[27] M. Monemi, H. Tabassum, Performance of UAV-assisted D2D networks in the finite block-length regime, IEEE Trans. Commun. 68 (11) (2020) 7270–7285.
[28] Y. Li, G. Feng, M. Ghasemiahmadi, L. Cai, Power allocation and 3-D placement for floating relay supporting indoor communications, IEEE Trans. Mobile Comput. 18 (3) (2019) 618–631.
[29] D. Zhai, H. Li, X. Tang, R. Zhang, Z. Ding, F.R. Yu, Height optimization and resource allocation for NOMA enhanced UAV-aided relay networks, IEEE Trans. Commun. 69 (2) (2021) 962–975.
[30] Y. Song, S.H. Lim, S.W. Jeon, S. Baek, On cooperative achievable rates of UAV assisted cellular networks, IEEE Trans. Veh. Technol. 69 (9) (2020) 9882–9895.
[31] A. Fotouhi, H. Qiang, M. Ding, M. Hassan, L.G. Giordano, A. Garcia-Rodriguez, J. Yuan, Survey on UAV cellular communications: practical aspects, standardization advancements, regulation, and security challenges, IEEE Commun. Surveys Tuts. 21 (4) (2019) 3417–3442.
[32] P. Yang, X. Xi, K. Guo, T.Q.S. Quek, J. Chen, X. Cao, Proactive uav network slicing for urllc and mobile broadband service multiplexing, IEEE J. Sel. Area. Commun. 39 (10) (2021) 3225–3244.
[33] Y. Chen, W. Feng, G. Zheng, Optimum placement of uav as relays, IEEE Commun. Lett. 22 (2) (2018) 248–251.
[34] K. Wang, R. Zhang, L. Wu, Z. Zhong, L. He, J. Liu, X. Pang, Path loss measurement and modeling for low-altitude UAV access channels, in: 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall), 2017, pp. 1–5.
[35] Document TR 38.901 V14.0.0: Study on Channel Model for Frequencies from 0.5 to 100 GHz (Release 14), Tech. rep., 3GPP Technical Specification Group Radio Access Network, July 2017.
[36] P. Guo, X. Wang, Y. Han, The enhanced genetic algorithms for the optimization design, in: Proc. Biomedical Engineering and Informatics Conference, vol. 7, 2010, pp. 2990–2994.
[37] C. Jiang, Y. Chen, Y. Gao, K.J. Ray Liu, Indian buffet game with negative network externality and non-bayesian social learning, IEEE Trans. Syst., Man, Cybern. Syst. 45 (4) (2015) 609–623.
[38] H. Tuy, Convex Analysis and Global Optimization, Kluwer Academic, 1998.
[39] Y. Li, M. Sheng, C. Yang, X. Wang, Energy efficiency and spectral efficiency tradeoff in interference-limited wireless networks, IEEE Commun. Lett. 17 (10) (2013) 1924–1927.
[40] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, J.G. Andrews, User association for load balancing in heterogeneous cellular networks, IEEE Trans. Wireless Commun. 12 (6) (2013) 2706–2716.
[41] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.