IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 7, 2021, accepted April 28, 2021, date of publication May 24, 2021, date of current version June 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3082862

Fairness-Aware Link Optimization for Space-Terrestrial Integrated Networks: A Reinforcement Learning Framework

ATEFEH HAJIJAMALI ARANI[1], PENG HU[1,2], (Senior Member, IEEE), AND YEYING ZHU[1], (Member, IEEE)

[1]Department of Statistics and Actuarial Science, University of Waterloo, ON N2L 3G1, Canada
[2]Digital Technologies Research Center, National Research Council Canada, Ottawa, ON K1A 0R6, Canada

Corresponding authors: Peng Hu (peng.hu@nrc-cnrc.gc.ca) and Atefeh Hajijamali Arani (ahajijam@uwaterloo.ca)

This work was supported by the High-Throughput and Secure Networks (HTSN) Challenge Program through the National Research Council of Canada under Grant CH-HTSN-418.

ABSTRACT The integration of space and air components considering satellites and unmanned aerial vehicles (UAVs) into terrestrial networks in a space-terrestrial integrated network (STIN) has been envisioned as a promising solution to enhancing the terrestrial networks in terms of fairness, performance, and network resilience. However, employing UAVs introduces some key challenges, among which backhaul connectivity, resource management, and efficient three-dimensional (3D) trajectory designs of UAVs are very crucial. In this paper, low-Earth orbit (LEO) satellites are employed to alleviate the backhaul connectivity issues with UAV networks, where we address the problem of jointly determining backhaul-aware 3D trajectories of UAVs, resource management, and associations between users, satellites and base stations (BSs) in an STIN while satisfying ground users' quality-of-experience requirements and provisioning fairness concerning users' data rates. The proposed approach maximizes a novel objective function with joint consideration for BS's load and fairness, which can be categorized as a non-deterministic polynomial time hard (NP-hard) problem. To tackle this issue, we leverage a reinforcement learning framework, in which our problem is modeled as a multi-armed bandit problem. Accordingly, BSs learn the environment and its dynamics and then make decisions under an upper confidence bound based method. Simulation results show that our proposed approach outperforms the benchmark methods in terms of fairness, throughput, and load.

INDEX TERMS Space-terrestrial integrated networks, space-air-ground integrated networks, unmanned aerial vehicles, fairness, reinforcement learning, community networks.

I. INTRODUCTION

With the recent advancements in the satellite, aerial, and terrestrial networks, future space-terrestrial integrated networks (STINs) are expected to ubiquitously employ intelligence and heterogeneity as a foundation for new Internet infrastructures. STINs have a great potential of improving the quality of experience (QoE) for all satellite-dependent Internet users and services in metropolitan, rural, and remote areas across the world [1].

Currently, satellite-dependent Internet users suffer from various Internet access interruptions, while autonomous solutions to effectively resolving the access issues are lacking in the literatures. The 3rd generation partnership project (3GPP) study group on the fifth generation of telecommunication networks (5G) [2] and International Telecommunications Union (ITU) focus group on machine learning (ML) for future networks including 5G (FG-ML5G) [3] have identified some general architectures for future networks, but no actual solutions have been proposed yet. One way to address this challenge is through fault compensation techniques using aerial components considering unmanned aerial vehicles (UAVs), where alternative links can be put in service to overcome link outage issues in an STIN. One or more UAVs can be purposed to temporarily provide alternative links to ensure continuous underlying connections in case of link outages between satellite and terrestrial components. Furthermore, UAVs can assist terrestrial networks in providing ubiquitous connectivity for under-served and under-connected areas (e.g., rural and disaster-affected areas) [4]. In these cases, to achieve high QoE provided by a UAV-assisted link, throughput and fairness as two important performance metrics need to be met at the same time. In other words, it is not desirable to serve certain users most of the time

The associate editor coordinating the review of this manuscript and approving it for publication was Shafiqul Islam.

while leaving others under-served on the network. With the expected global coverage using the advanced STIN, the Internet users, including those in rural and remote communities will have increasing reliance on the available connectivity options provided by an STIN. In this case, STINs that ensure the fault compensation and fairness are indispensable components as the bedrocks of serving the next-generation Internet infrastructure.

A. RELATED WORK

Space-terrestrial networks have been used in some applications which are referred as ''alternative networks'' in RFC 7962 [5]. Such alternative networks are considered to be self-managed and self-sustained. The RIFE project [6] has explored an architecture for a sustainable Internet access consisting of a satellite backhaul and information-centric networking (ICN) fronthaul. Recently, the integration of satellite and telecommunications networks have also been discussed under the umbrella of space information networks (SINs) or space-air-ground integrated networks (SAGINs), where the recent low-Earth orbit (LEO) satellite constellation and UAVs are new components for realizing backhaul links. In the rest of the paper, we will consider SAGIN as a variant of STIN, and use two terms interchangeably.

UAVs can assist cellular networks in providing reliable connectivity in under-connected areas such as rural and disaster-affected areas where terrestrial communication infrastructure is often damaged and/or there is no existing ground infrastructure. One of the major issues in integrating UAVs into terrestrial networks in the role of aerial base stations (BSs) is optimizing their three-dimensional (3D) locations and managing radio resource which can adapt to the dynamics of the networks. The existing works have studied a number of problems related to UAV-enabled systems including UAV placement, trajectory design and resource allocation problems such as in [7]–[10]. To maximize the data rates of users, a joint user association and UAV horizontal location problem was investigated in [7]. The problem was modeled as a mixed-integer non-convex optimization problem, and was decomposed into two sub-problems. Then, an iterative algorithm was employed. In [8], the problem of the 3D trajectory design and resource allocation for a single solar-powered UAV was investigated. The objective of the optimization problem was to maximize the throughput of the system during a finite period of time. An adaptive UAV deployment approach was proposed in [9], in which a single UAV adjusts its location in one-dimensional (1D) and two-dimensional (2D) planes to maximize the average throughput of users. In [10], the altitude of a single UAV was optimized to achieve reliable communication and maximum coverage by minimizing the outage probability. Accordingly, a height-dependent closed-form expression for the outage probability was derived.

In more practical scenarios, the recent works have considered multiple UAVs and/or integration of UAVs into terrestrial networks such as in [11]–[16]. In [11], the authors derived approximate expressions for the coverage probability and average achievable rate for UAVs located at fixed altitudes. In [12], a set of UAVs were deployed as relays to provide reliable wireless connection for ground users. In order to improve the spectrum efficiency of the system, the 2D trajectories of the UAVs were optimized. In [13], the coverage area for UAVs in the presence of co-channel interference was maximized. In [14], a set of UAVs were deployed to assist a macro BS (MBS) in downlink for overload situations. To solve the problem in a distributed manner, reinforcement learning algorithms were proposed. In [15], a number of UAVs were distributed in a 3D space to support terrestrial networks during temporary mass events or post-disaster situations. The problem of 3D placement of UAVs was decomposed into two separate sub-problems: i) 2D placement to find the horizental locations of UAVs, and ii) 1D optimization to adjust the altitudes of UAVs. Firstly, it was assumed that the altitudes of UAVs are fixed at random values. Then, using a K-means algorithm the locations of UAVs in the horizontal plane were determined, in which a prior knowledge of the users' locations is required. However, this assumption on the availability of users' locations is impractical in real time situations, and providing it for UAVs can be very challenging. Finally, the altitudes of UAVs were optimized given the 2D locations of UAVs. In [16], UAVs coexist with small BSs (SBSs) to maximize the satisfaction of users with provided data rates through optimizing the 3D locations of UAVs. To this end, two heuristic algorithms based on the genetic algorithm and particle swarm optimization were developed. In [17], the problem of resource allocation for UAV networks was investigated. However, the trajectories of UAVs were predefined according to the pre-programmed flight plans.

Another main design challenge in UAV-assisted networking is the backhaul connectivity which is not addressed in the aforementioned work. The authors in [18] obtained the placement of a single UAV, in which ground BSs provide backhaul connectivity for the UAV. In [19], UAVs receive backhaul data from a MBS, in which a non-orthogonal multiple access (NOMA) scheme was employed for the backhaul transmissions. To maximize the sum rate of users, a mechanism for optimizing the location of UAVs and resource allocation in the MBS and UAVs were proposed. In [20], the 2D placement of UAVs and resource allocation problem were investigated, in which UAVs are connected to the core network through a cellular BS. The problem was decomposed into UAV placement and resource allocation sub-problems which are solved using iterative algorithms. In [21], a MBS provides the backhaul connectivity for UAVs in the system, in which they aim at maximizing their throughput through optimizing their 2D trajectories based on a learning algorithm. In [22], the problem of the backhaul operation dynamic link rerouting for UAV-based relay networks was investigated. In [23], a set of ground BSs provide backhaul connectivity for a single UAV. To maximize the total network

profit, the problems of the UAV's location and bandwidth allocation to users were addressed. To solve the problems, a heuristic algorithm with low computational complexity was developed. The authors in [24] addressed the problem of resource allocation for an integrated network of a single satellite, ground BSs and UAVs, in which the satellite and MBSs provide backhaul connectivity. The problem is formulated as a competitive market aiming at maximizing the total profit in the system, and an iterative heavy ball algorithm is applied to find the solution. However, the 3D locations of UAVs were not optimized jointly with the resource allocation problem. In [25], the authors proposed an interference management algorithm to maximize the overall sum rate gains in downlink by optimizing the user-UAV association and resource allocation. To consider the dependency between backhaul and access links, a binary model was employed, in which if the received signal to interference plus noise ratio (SINR) at backhaul links are below a certain threshold, there will be no transmission in access links. In [26], a set of LEO satellites provide backhaul connectivity for a set of ground and aerial BSs which serve users in downlink. To maximize the system throughput, the UAVs aim at optimizing their trajectories, and also all the BSs manage the radio resource.

On the other hand, fairness provisioning is one of the major objective in wireless networks. Adaptive system design without considering fairness may cause poor service to some users since resource may be distributed to users with relatively good channel conditions. In this regard, when adaptive system design strategies are employed, fairness issue is required to be taken into account. Specifically, there are different fairness criteria such as max–min fairness [27], proportional fairness [28], and Jain's index [29]. However, the aforementioned work has not taken into account the issue of the fairness in SAGINs, and there are a few works in UAV networks that have considered this issue such as in [30]–[33]. In [30], the optimal altitude of a UAV was optimized to maximize the fairness between users. In [31], the authors proposed a method for UAV control flying at a fixed altitude, in which the target region is divided into several cells and each cell needs to be covered by at least a UAV based on a fairness criteria. In [32], a UAV as a relay was employed to extend coverage for two disconnected far vehicles. The 2D trajectory design of the UAV was optimized based on the throughput fairness between vehicles and among different time slots. The fairness issue is considered in [33], where a proportional fairness metric is used to maintain fairness among users through optimizing the location of a single UAV. However, the previous studies on fairness [30]–[33] have not explicitly considered fairness in conjunction with load balancing, 3D trajectory of UAVs, BS-satellite/user-BS association and resource management in SAGINs.

B. CONTRIBUTIONS

The main contributions of this paper include to propose a novel framework for joint 3D trajectory design of UAVs, BS-satellite/user-BS association, and managing resource among UAVs and terrestrial BSs while ensuring the fairness among users and minimizing the load of BSs. Besides, providing backhaul connectivity is a major challenge especially for disaster-recovering and rural areas. Thus, we assume that LEO satellites provide backhaul connectivity to BSs. To optimize access links, we formulate our problem as a multi-armed bandit (MAB) problem, in which a reward function capturing the fairness and load is defined. Our main contributions include:

- We model a novel framework to jointly address BS-satellite associations, user-BS association, resource allocation among BSs, and 3D trajectories of UAVs within an STIN. Due to the high complexity of the problem, we decompose it into the backhaul link and the access link problem. In our system, a set of LEO satellites provide backhaul connectivity for SBSs and UAVs. UAVs, as aerial BSs, coexist with SBSs, and both UAVs and SBSs provide data to ground users in downlink. Moreover, we assume that users move according to a mobility model.
- We formulate our problem in access links, i.e., the joint resource allocation and 3D trajectory design of UAVs, as a MAB problem, in which SBSs and UAVs aim at maximizing an objective function capturing the fairness and the loads of BSs. Meanwhile, UAVs and SBSs receive the data in backhaul links through connecting the satellite offering the best signal strength. To take into account the loads of BSs, we use a rigorous definition of the load instead of using simple definitions such as the number of users associated with a BS that do not properly reflect the loads of BSs.
- To solve our MAB problem, we employ the upper confidence bound (UCB) policy. It can provide an effective balance between the exploration and exploitation to adapt the system dynamic without prior and full information of the system.
- The simulation results reveal that the proposed approach generally outperforms the benchmark algorithms in terms of fairness, throughput and load.

The rest of this paper is organized as follows. In Section II, we describe our system model and the user association policy. In Section III, we formulate our problem for BS-satellite association in backhaul links, and resource allocation and 3D trajectory design of UAVs in access links. Section IV describes our learning based approach to solve the problem in access links. In Section V, we evaluate the performance of our proposed approach. Finally, Section VI concludes the paper.

Notations: The regular and boldface symbols refer to scalars and matrices, respectively. For any finite set $\mathcal{A}$, the cardinality of set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. $\boldsymbol{X}^T$ denotes the transpose of matrix $\boldsymbol{X}$. The floor of a real number x is denoted by $\lfloor x \rfloor$ which maps x to the greatest integer less than or equal to x. The function $\mathbb{1}_{\phi}$ denotes the indicator function which equals 1 if event ϕ is true and 0, otherwise. $\mathbb{R}^M$ and $\log_2(.)$

TABLE 1. List of notations.

System Notations			
Notation	**Description**	**Notation**	**Description**
$\mathcal{L}$	Set of LEO satellites	$\mathcal{S}$	Set of SBSs
$\mathcal{U}$	Set of UAVs	$\mathcal{K}$	Set of users
$\mathcal{B}$	Set of total BSs	T	Total time
T_{s}	Duration of each time slot	$\boldsymbol{a}_l^{\mathrm{L}}(t)$	3D location of satellite l
$\boldsymbol{a}_b^{\mathrm{B}}(t)$	3D location of BS b	$\boldsymbol{a}_k^{\mathrm{K}}(t)$	3D location of user k
V_{L}	Orbital speed of satellites	H_{L}	Altitude of satellites
G	Gravitational constant	$\mathrm{M_E}$	mass of Earth
$\mathrm{R_E}$	Radius of Earth	T_{L}	Orbital period of the LEO satellites
ω_{BCK}	Total bandwidth for backhaul links	$L_{l,b}(t)$	Path loss model between satellite l and BS b at time t
$d_{l,b}(t)$	Distance between satellite l and BS b at time t	p_l	Transmit power of satellite l
$g_{l,b}(t)$	Channel gain between satellite l and BS b at time t	G_l^{T}	Transmit gain of satellite l's antenna
G_b^{R}	Receive gain of BS b	$r_b^{\max}$	Maximum distance of a satellite above the horizon of BS b
$r_{b,\mathrm{O}}$	Distance of BS b from Earth's center	$r_{b,\mathrm{L}}$	minimum distance from a satellite to BS b
$C_{l,b}(t)$	Data rate of BS b associated to satellite l at time t	σ_0^2	Noise power
$a_{l,b}^{\mathrm{BCK}}(t)$	Association relation element satellite l and BS b	$V_k^{\min}$	Minimum speed of user k
$V_k^{\max}$	Maximum speed of user k	$C_{b,k}(t)$	Achievable data rate to user k provided by BS b
$\gamma_{b,k}(t)$	SINR at the receiver of user k associated to BS b	p_b	Transmit power of BS b
$g_{b,k}(t)$	Channel gain between BS b and user k at time t	$a_{b,k}^{\mathrm{ACC}}(t)$	Association relation element BS b and user k
$\mathcal{K}_b(t)$	Set of users associated to BS b	$\rho_b(t)$	load of BS b at time t
$q_b(t)$	Transmit channel of BS b	f_c	Carrier frequency
$\bar{C}_k(t)$	Total data rate for user k until time t	ϑ_k	Packet arrival rate of user k
ζ_k	Mean packet size of user k	$r_{b,k}(t)$	Horizontal distance between BS b and user k at time t
$\mathrm{pr}_{b,k}^{\mathrm{LoS}}$	Probability of LoS between user k and BS b	$h_b(t)$	Altitude of BS b at time t
h_k	Height of user k	$L_{b,k}^z(t)$	Channel gain between BS b and user k at time t
$\mathcal{F}(t)$	Jain's fairness index	$\boldsymbol{q}(t)$	Channel vector
$\boldsymbol{A}^{\mathrm{U}}(t)$	Matrix of UAVs' locations at time t	$\boldsymbol{A}_t^{\mathrm{BCK}}$	Association matrix for backhaul links
$\boldsymbol{A}_t^{\mathrm{ACC}}$	Association matrix for access links	ϕ_b and ψ_b	Weight parameters in reward function
$h_{\min}$	Minimum altitude of UAVs	$h_{\max}$	Maximum altitude of UAVs
$\mathcal{A}_{\mathrm{S}}$	Action set of SBSs	$\mathcal{Z}$	Set of movement in different directions
$\mathcal{A}_{\mathrm{U}}$	Action set of UAVs	$a_{s,i}$	Action i^{th} of SBS s
$a_{u,i}$	Action i^{th} of UAV u	$\bar{R}_{b,i}(t)$	Average reward from action $a_{b,i}$ for player b at time t
$n_{b,i}(t)$	Number of times that action $a_{b,i}$ has been selected by BS b until time t	$a_b^{\mathrm{UCB}}(t)$	Selected action by BS b at time t

denote the space of M-dimensional real-valued vectors, and the logarithm with base 2, respectively. Additionally, Table 1 presents the list of notations.

II. SYSTEM MODEL

We consider a network composed of the LEO satellites, the UAVs, and the terrestrial BSs to provide service for ground users in the downlink in a particular area $\mathcal{R} \in \mathbb{R}^2$, as illustrated in Fig. 1. In the space segment, the set of the LEO satellites $\mathcal{L}$ provide the capacity for the UAVs and the terrestrial BSs in backhaul links. The set of UAVs is denoted by $\mathcal{U}$ which act as the aerial BSs in the aerial segment, and fly with speed V_{U}. The ground segment consists of a set of users $\mathcal{K}$ and a set of SBSs $\mathcal{S}$. The UAVs and the SBSs provide service connectivity over the access links, and the satellites are expected to cause negligible interference on the access links. We assume that the total time T is discretized into N equally spaced time instants with duration T_{s} which is chosen to be sufficiently small. Therefore, the locations of the UAVs are assumed to be constant during each time instant t.

Let the set of users which are associated to BS $b \in \mathcal{B}$ at time instant $t \in N$ is denoted by $\mathcal{K}_b(t)$, where $\mathcal{B} = \mathcal{S} \cup \mathcal{U}$ is the set of total BSs in the system.[1] We denote the 3D coordinates of satellite $l \in \mathcal{L}$, BS $b \in \mathcal{B}$, and user $k \in \mathcal{K}$ at time instant t, respectively, by $\boldsymbol{a}_l^{\mathrm{L}}(t) = \big(x_l(t), y_l(t), h_l(t)\big)$, $\boldsymbol{a}_b^{\mathrm{B}}(t) = \big(x_b(t), y_b(t), h_b(t)\big)$, and $\boldsymbol{a}_k^{\mathrm{K}}(t) = \big(x_k(t), y_k(t), h_k\big)$.

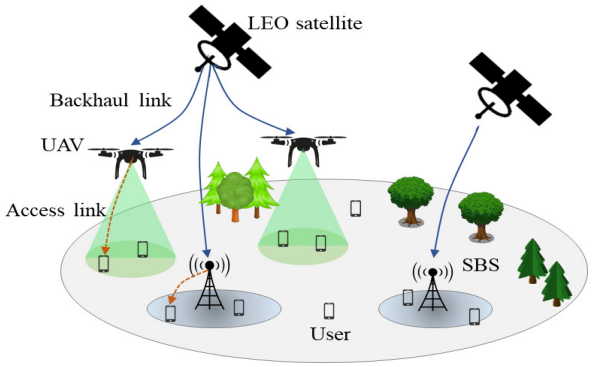

FIGURE 1. An illustration of the system model.

[1] In what follows, the term *BS* denotes both a terrestrial BS and a UAV.

A. BACKHAUL LINK

Here, we focus on backhaul communication over the millimeter-wave (mmWave) links. We consider a wireless backhaul network composed of the set of the LEO satellites $\mathcal{L} = \{1, \ldots, |\mathcal{L}|\}$ and the set of the BSs $\mathcal{B} = \{1, \ldots, |\mathcal{B}|\}$.

We assume that a circular orbit for the equally distributed LEO satellites moving in the direction of y-axis at the fixed altitude H_{L} above the surface of Earth [34]. Let G and $\mathrm{M_E}$ denote the gravitational constant and the mass of Earth in kilograms, respectively. Thus, the orbital speed of the satellites in the orbital plane can be determined as follows [35]:

$$V_{\mathrm{L}} = \sqrt{\frac{\mathrm{G} \cdot \mathrm{M_E}}{(H_{\mathrm{L}} + \mathrm{R_E})}} \ \text{[m/s]}, \tag{1}$$

where $\mathrm{R_E}$ is the radius of Earth in meters. Therefore, the orbital period of the LEO satellites can be calculated as follows [35]:

$$T_{\mathrm{L}} = \frac{2\pi\ (H_{\mathrm{L}} + \mathrm{R_E})}{V_{\mathrm{L}}} \ \text{[sec]}. \tag{2}$$

We assume that the total bandwidth ω_{BCK} is allocated for the backhaul communications which is divided into $|\mathcal{L}|$ orthogonal channels equally with bandwidth $\omega_{\mathrm{BCK}}/|\mathcal{L}|$. Let f_c be the carrier frequency in GHz. Therefore, the free space path loss model between each satellite $l \in \mathcal{L}$ and each BS $b \in \mathcal{B}$ can be expressed as [36]

$$L_{l,b}(t) = 32.45 + 20\log_{10}(f_c) + 20\log_{10}(d_{l,b}(t)) \ \text{[dB]}, \tag{3}$$

where $d_{l,b}(t)$ is the distance between satellite l and BS b at time t which is determined as follows:

$$d_{l,b}(t) = \sqrt{(x_l(t) - x_b(t))^2 + (y_l(t) - y_b(t))^2 + (h_l(t) - h_b(t))^2}. \tag{4}$$

Let p_l and $g_{l,b}(t)$ denote the transmit power of satellite l and the channel gain between satellite l and BS b at time t which is given by

$$g_{l,b}(t) = \begin{cases} 10^{\frac{-L_{l,b}}{10}} G_l^{\mathrm{T}} G_b^{\mathrm{R}}, & \text{if } d_{l,b} \le r_b^{\max} \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

where G_l^{T} and G_b^{R} are the transmit gain of satellite l's antenna and the receive gain of BS b, respectively. Here, $r_b^{\max}$ indicates the maximum distance of a satellite above the horizon of BS b, in which the satellite and the BS are able to communicate. Parameter $r_b^{\max}$ can be calculated as follows [37]:

$$r_b^{\max} = \sqrt{2.r_{b,\mathrm{O}}.r_{b,\mathrm{L}} + r_{b,\mathrm{L}}^2}, \tag{6}$$

where $r_{b,\mathrm{O}}$ and $r_{b,\mathrm{L}}$ denote the distance of BS b from Earth's center and the minimum distance from a satellite to BS b, respectively.

According to Shannon's capacity formula, the achievable data rate BS b associated to satellite l at time instant t is given by

$$C_{l,b}(t) = \frac{\omega_{\mathrm{BCK}}}{|\mathcal{L}|} \log_2\left(1 + \frac{a_{l,b}^{\mathrm{BCK}}(t)\, p_l\, g_{l,b}(t)}{\sigma_0^2}\right) \ \text{[bps]}, \tag{7}$$

where σ_0^2 represents the noise power. The binary element $a_{l,b}^{\mathrm{BCK}}(t) \in \{0, 1\}$ denotes the association relation between satellite l and BS b such that $a_{l,b}^{\mathrm{BCK}}(t) = 1$ indicates that BS b is associated to satellite l at time instant t, otherwise $a_{l,b}^{\mathrm{BCK}}(t) = 0$. Furthermore, for $d_{l,b} > r_{b,\mathrm{O}}$, the association element $a_{l,b}^{\mathrm{BCK}}(t)$ is equal to zero. Thus, we define the association matrix for the backhaul links as $\boldsymbol{A}_t^{\mathrm{BCK}} = [a_{l,b}^{\mathrm{BCK}}(t)]_{|\mathcal{L}|\times|\mathcal{B}|}$ which is updated based on the association elements $a_{l,b}^{\mathrm{BCK}}(t)$ at each time instant t.

B. ACCESS LINK

In the access links, the set of BSs $\mathcal{B}$ serve the ground users in the downlink direction. We assume that the users move according to the random walk mobility model at each time instant [38]. Thus, each user $k \in \mathcal{K}$ selects a speed uniformly distributed from the ranges $[V_k^{\min}, V_k^{\max}]$ and a movement angle uniformly from the ranges $[0, 2\pi]$, and performs a movement based on the selected speed and direction. Here, $V_k^{\min}$ and $V_k^{\max}$ denote the minimum and maximum speed of user k, respectively.

1) DATA RATE AND LOAD

For the access links, we consider the total bandwidth ω_{ACC} which is equally divided into $|\mathcal{Q}|$ orthogonal channels with bandwidth $\omega_{\mathrm{ACC}}/|\mathcal{Q}|$, where $\mathcal{Q}$ is the set of available channels for communications in the access links. Therefore, the maximum achievable data rate to user k provided by BS b which is associated to satellite l is given by [21]

$$C_{b,k}(t) = \min\left(\frac{\omega_{\mathrm{ACC}}}{|\mathcal{Q}|} \log_2(1 + \gamma_{b,k}(t)), C_{l,b}(t)\right) \ \text{[bps]}, \tag{8}$$

where $\gamma_{b,k}(t)$ denotes the SINR at the receiver of user k associated to BS b, which can be defined as follows [39]:

$$\gamma_{b,k}(t) = \frac{a_{b,k}^{\mathrm{ACC}}(t).p_b.g_{b,k}(t)}{\sum_{b' \in \mathcal{B}\setminus b} p_{b'}.g_{b',k}(t)\rho_{b'}(t)\mathbb{1}_{(q_b(t)=q_{b'}(t))} + \sigma_0^2}, \tag{9}$$

where p_b and $g_{b,k}(t)$ denote the transmit power of BS b and the channel gain between BS b and user k at time t, respectively. Here, $q_b(t)$ is the channel which BS b transmits over it at time instant t. The binary element $a_{b,k}^{\mathrm{ACC}}(t) \in \{0, 1\}$ indicates the association between BS b and user k at time t, such that

$$a_{b,k}^{\mathrm{ACC}}(t) = \begin{cases} 1, & \text{if user } k \text{ is associated to BS } b, \\ 0, & \text{o.w.} \end{cases} \tag{10}$$

According to the association elements $a_{b,k}^{\mathrm{ACC}}(t)$, we define the association matrix for the access links as $\boldsymbol{A}_t^{\mathrm{ACC}} = [a_{b,k}^{\mathrm{ACC}}(t)]_{|\mathcal{B}|\times|\mathcal{K}|}$. Furthermore, the set of users associated to BS b can be defined as follows:

$$\mathcal{K}_b(t) = \{k | k \in \mathcal{K}, a_{b,k}^{\mathrm{ACC}}(t) = 1\}. \tag{11}$$

Parameter $\rho_b(t)$ denotes the load of BS b at time t, which can be obtained as follows [40]:

$$\rho_b(t) = \sum_{k \in \mathcal{K}_b(t)} \frac{\vartheta_k}{\zeta_k\, C_{b,k}(t)} \triangleq f_b(\boldsymbol{\rho}(t)), \tag{12}$$

where ϑ_k and ζ_k denote the packet arrival rate and the mean packet size of user k, respectively. Here, ϑ_k/ζ_k represents the user rate requirement. Therefore, we can assume that the users are heterogeneous in nature, in which each user can have a different QoE requirement based on its packet arrival rate and mean packet size. Moreover, $\boldsymbol{\rho}(t) = \big(\rho_1(t), \ldots, \rho_{|\mathcal{B}|}(t)\big)$ denotes the BS load vector consisting of the load values of all the BSs, and function $f_b(.)$ defined in (12) represents the load of BS b as a function of the BS load vector. We can rewrite (12) in the format of vector as follows [41]:

$$\boldsymbol{\rho}(t) = \boldsymbol{f}(\boldsymbol{\rho}(t)), \tag{13}$$

where $\boldsymbol{f}(\boldsymbol{\rho}(t)) = \Big(f_1(\boldsymbol{\rho}(t)), \ldots, f_{|\mathcal{B}|}\boldsymbol{\rho}(t)\Big)^T$ denotes the vector of load functions. Since the BS load vector appears in both sides of (13), we are not able to solve it as a fixed-point solution in a closed form. Therefore, we use the fixed point iteration algorithm, based on the fact that $\boldsymbol{f}(\boldsymbol{\rho}(t))$ is a standard interference function (SIF). Due to the availability of the limited amount of the resource in the network, the load of each BS b is bounded by the full load, i.e. $\rho_b(t) \leq 1, b \in \mathcal{B}$ [42]. When $\rho_b > 1$, BS b sorts its associated users based on the load density $\frac{\vartheta_k}{\zeta_k C_{b,k}(t)}, \forall k \in \mathcal{K}_b(t)$ in descending order. Then, it drops users from the top of the list (interpreted as outage users) until $\rho_b \leq 1$. Then, it associates the remaining resource to the outage users.

Definition 1: A function $I(\boldsymbol{n})$ is a SIF if for all $n \geq 0$ the following properties are satisfied [43]:

1) *Positivity:* $I(\boldsymbol{n}) > 0$,
2) *Monotonicity:* $\boldsymbol{n} \geq \boldsymbol{n}' \Rightarrow I(\boldsymbol{n}) \geq I(\boldsymbol{n}')$,
3) *Scalability:* $\alpha I(\boldsymbol{n}) > I(\alpha \boldsymbol{n})$ *for* $\alpha > 1$,

To numerically obtain $\boldsymbol{\rho}(t)$ in (13), we use an iterative algorithm for the load of the BSs as follows. Specifically, we start from an arbitrary initial BS load vector $\boldsymbol{\rho}^0 > 0$, and calculate the output of the mth iteration of the algorithm as follows [44]:

$$\boldsymbol{\rho}^m = \min\Big(\boldsymbol{f}(\boldsymbol{\rho}^{m-1}), 1\Big), \tag{14}$$

where $\boldsymbol{\rho}^m$ is the BS load vector at iteration $m \in \{1, \ldots, M\}$. Parameter M denotes the total number of iterations. Lemma 1 indicates that the BS load vector $\boldsymbol{\rho}^M$ converges to the fixed point solution of (13).

Lemma 1: If the fixed point of (13) exists, then it is unique, and can be iteratively obtained by (14) as M goes to infinity.

Proof: In [45], it is proved that $f_b(\boldsymbol{\rho}(t))$ is a SIF. Furthermore, Theorem 7 in [43] prove that $\min(f_b(\boldsymbol{\rho}), 1)$ is a SIF. Then, by using Theorem 2 in [43], the convergence is proved. ■

2) PROPAGATION CHANNEL

We consider a line-of-sight (LoS)/non-LoS propagation channel for each access link. Let the horizontal distance between each BS $b \in \mathcal{B}$ and each user $k \in \mathcal{K}$ at time t be denoted by $r_{b,k}(t) = \sqrt{(x_b(t) - x_k(t))^2 + (y_b(t) - y_k(t))^2}$. According to the ITU model, the probability of LoS between user $k \in \mathcal{K}$ and BS $b \in \mathcal{B}$ can be written as follows [46]:

$$\mathrm{pr}_{b,k}^{\mathrm{LoS}} = \prod_{j=0}^{J}\left[1 - \exp\left(-\frac{\left[h_b(t) - \frac{(j+\frac{1}{2})(h_b(t)-h_k)}{J+1}\right]^2}{2\gamma^2}\right)\right], \tag{15}$$

where $J = \lfloor \frac{r_{b,k}(t)\sqrt{\alpha\beta}}{1000} - 1 \rfloor$ The three statistical parameters α, β and γ characterize different types of environments [47, Table 1]. Parameter α is the ratio of land area covered by buildings to total land area, β is the mean number of buildings per unit area, and γ denotes the distribution of building height. Note that the blockage model defined in (15) can be used for any transmitter/receiver heights for UAV-to-ground and ground-to-ground transmissions and for a wide spectrum range [48]. Furthermore, the non-LoS probability can be determined as $\mathrm{pr}_{b,k}^{\mathrm{NLoS}} = 1 - \mathrm{pr}_{b,k}^{\mathrm{LoS}}$.

Let $d_{b,k}(t) = \sqrt{r_{b,k}^2(t) + (h_b(t) - h_k)^2}$ be the 3D distance between BS b and user k at time t. The channel gain between BS b and user k is given by

$$L_{b,k}^z(t) = \delta_b^z + \eta_b^z \log_{10} d_{b,k}(t) + \chi_b^z \ \ [\mathrm{dB}], \tag{16}$$

where superscript $z \in \{\mathrm{LoS}, \mathrm{NLoS}\}$ denotes LoS and non-LoS components. Here, δ_b^z and η_b^z indicate the reference path loss and the path loss exponent, respectively. Parameter χ_b^z is a zero-mean Gaussian random variable with a standard deviation $\sigma_{b,\mathrm{SF}}^z$ in dB.

3) USER-BS ASSOCIATION

A main problem in wireless networks is to associate the users to the BSs. Since we assume that the users move in the system, they require to periodically assess their actual performances. Therefore, if they are not satisfied with their current associations, they may change their serving BSs, and perform new associations. In this regard, at each time instant t, the set of outage users $\mathcal{O} \subset \mathcal{K}$ perform new association processes in order to assign to new BSs. Assuming the fixed UAVs' locations and BSs' channels, each user $k \in \mathcal{O}$ is associated to BS $b_k^*(t)$ according to the following user association policy:

$$b_k^*(t) = \underset{b \in \mathcal{B}}{\mathrm{argmax}}\{p_b g_{b,k}(t)\}. \tag{17}$$

Therefore, the association element in matrix $\boldsymbol{A}_t^{\mathrm{ACC}}$ is updated as $a_{b,k}^{\mathrm{ACC}}(t) = b_k^*(t)$.

III. PROBLEM FORMULATION

Given the model described in Section II, our objective is to improve the fairness among the users while minimizing the loads of all the BSs during their downlink transmissions. To do that, we decompose our problem into two sub-problems: i) the BS-satellite association problem in the backhaul links, and ii) the resource management and UAV trajectory design in the access links.

A. BACKHAUL LINK

For the backhaul links, we require to solve the following optimization problem, which involves finding the elements of matrix $\boldsymbol{A}_t^{\mathrm{BCK}}$ as follows:

$$\max_{\boldsymbol{A}_t^{\mathrm{BCK}}} \sum_{t\in N}\sum_{l\in\mathcal{L}}\sum_{b\in\mathcal{B}} C_{l,b}(t) \tag{18a}$$

$$\text{s.t. } a_{l,b}^{\mathrm{BCK}(t)} \in \{0,1\}, \quad \forall b\in\mathcal{B}, \forall l\in\mathcal{L} \tag{18b}$$

$$\sum_{l\in\mathcal{L}} a_{l,b}^{\mathrm{BCK}(t)} \le 1, \quad \forall b\in\mathcal{B}. \tag{18c}$$

To solve the above problem, each BS $b\in\mathcal{B}$ is associated to satellite $l_b^*(t)\in\mathcal{L}$ offering the highest signal strength, i.e.:

$$l_b^*(t) = \underset{l\in\mathcal{L}}{\mathrm{argmax}}\{p_l g_{l,b}(t)\}. \tag{19}$$

According to (19), the association element in matrix $\boldsymbol{A}_t^{\mathrm{BCK}}$ is updated as follows $a_{l,b}^{\mathrm{BCK}}(t) = l_b^*(t)$.

B. ACCESS LINK

Let the total data rate for user k until time instant t be expressed as follows:

$$\bar{C}_k(t) = \sum_{\tau\le t}\sum_{b\in\mathcal{B}} C_{b,k}(\tau)\mathbb{1}_{a_{b,k}^{\mathrm{ACC}}(\tau)}. \tag{20}$$

Therefore, the fairness among the users at time instant t can be expressed through the most widely-used fairness metric named as Jain's fairness index as follows [29], [49]:

$$\mathcal{F}(t) = \frac{\left(\sum_{k\in\mathcal{K}}\bar{C}_k(t)\right)^2}{|\mathcal{K}|\left(\sum_{k\in\mathcal{K}}\bar{C}_k(t)^2\right)}. \tag{21}$$

Obviously, the fairness index $\mathcal{F}(t)$ is in the range $[0,1]$. The higher value of the fairness index is the result of the smaller differences among the total data rates of the users $\{\bar{C}_k(t)\}_{k\in\mathcal{K}}$. Furthermore, it is a continuous function, and takes continuous values, in which a change in a user throughput can cause a change in the fairness index. It is independent of the number of users and can be applied to scenarios with a different number of users. Note that both the fairness and the loads of the BSs are unitless, and they are the functions of the trajectories of the UAVs and the resource allocation. Furthermore, at each time t, the configuration of the system can be determined by the channel vector $\boldsymbol{q}(t) = (q_1(t),\ldots,q_{|\mathcal{B}|}(t))$, the matrix of UAVs' locations $\boldsymbol{A}^{\mathrm{U}}(t) = (\boldsymbol{a}_1^{\mathrm{B}}(t),\ldots,\boldsymbol{a}_{|\mathcal{U}|}^{\mathrm{B}}(t))$, and the association matrices $\boldsymbol{A}_t^{\mathrm{BCK}}$ and $\boldsymbol{A}_t^{\mathrm{ACC}}$. In order to balance between the fairness and load, our objective is to maximize a utility function capturing both those performance metrics. Therefore, our problem can be formulated as follows:

$$\max_{\boldsymbol{q}(t),\boldsymbol{A}^{\mathrm{U}}(t)} \sum_{t\in N}\sum_{l\in\mathcal{L}}\sum_{b\in\mathcal{B}}\sum_{k\in\mathcal{K}_b(t)} \Big(\phi_b\mathcal{F}(t) + \psi_b(1-\rho_b(t))\Big) \tag{22a}$$

$$\text{s.t. } (x_u(t), y_u(t)) \in \mathcal{R}, \quad \forall u\in\mathcal{U}, \tag{22b}$$

$$h_u(t) \in [h_{\min}, h_{\max}], \quad \forall u\in\mathcal{U}, \tag{22c}$$

$$q_b(t) \in \mathcal{Q}, \quad \forall b\in\mathcal{B}, \tag{22d}$$

$$\rho_b(t) = f_b(\boldsymbol{\rho}), \quad \forall b\in\mathcal{B}, \tag{22e}$$

$$0 \le \rho_b(t) \le 1, \quad \forall b\in\mathcal{B}, \tag{22f}$$

$$a_{l,b}^{\mathrm{BCK}(t)} \in \{0,1\}, \quad \forall b\in\mathcal{B}, \forall l\in\mathcal{L}, \tag{22g}$$

$$\sum_{l\in\mathcal{L}} a_{l,b}^{\mathrm{BCK}(t)} \le 1, \quad \forall b\in\mathcal{B}, \tag{22h}$$

$$a_{b,k}^{\mathrm{ACC}(t)} \in \{0,1\}, \quad \forall b\in\mathcal{B}, \ \forall k\in\mathcal{K}, \tag{22i}$$

$$\sum_{b\in\mathcal{B}} a_{b,k}^{\mathrm{ACC}(t)} \le 1, \quad \forall k\in\mathcal{K}, \tag{22j}$$

where ϕ_b and ψ_b are the weight parameters that indicate the impact of the fairness index and the load of BS b on the objective function. $h_{\min}$ and $h_{\max}$ denote the minimum and the maximum altitude of the UAVs, respectively. The constraints in (22b)-(22c) determine the feasible region for the locations of the UAVs in the 3D space. The constraint on the set of available channels for the BSs in the access links is determined by (22d). The constraints in (22e)-(22f) correspond to the definition of the BS load vector. The constraints in (22g)-(22h) express the satellite-BS association for the backhaul links, and ensure that each BS b is associated to at most one satellite at each time instant t. Furthermore, the constraints in (22i)-(22j) are related to the user-BS association for the access links, and guarantee each user k is associated to at most one BS at each time instant t.

Note that the problem formulated in (22) is a non-convex optimization problem, and obtaining an optimal solution in an online manner is computationally intractable. Therefore, the important issue is to balance between performance and complexity. In this regard, we employ the tools of reinforcement learning algorithms, and model our problem as a MAB problem. Using MAB learning algorithm has several advantages in the following aspects. Firstly, it does not require the full knowledge of the environment. Secondly, the computational complexity of the MAB algorithm is linear with respect to the number of the BSs, while the exponential computational complexity of centralized mechanism makes it infeasible for a dynamic and complex environment.

IV. TOWARDS PROVISIONING FAIRNESS: REINFORCEMENT LEARNING FRAMEWORK

Here, we propose a reinforcement learning based mechanism to address the access link optimization problem stated in (22). In the following, we first give a brief description of reinforcement learning and the MAB problem. Then, we present our MAB problem and its solution.

A. AN OVERVIEW OF REINFORCEMENT LEARNING AND MAB

Reinforcement learning algorithms deal with the problem of designing policies for learners, named as *players*, who interact with their environment [50]. The players select actions over a sequence of discrete time-steps. Then, they observe their own results, named as *rewards*, which quantify the level of players' satisfactions. Reinforcement learning algorithms do not require the full knowledge of environments. Therefore,

they need to *explore* the possible actions aiming to enhance their future decisions. On the other hand, they may *exploit* actions that have been chosen in the past and found effective and turn out to be a good fit to learn what actions to be taken through balancing between the *exploration* and *exploitation*.

MAB is considered as one of the fundamental framework in reinforcement learning [51]. In a MAB problem, there are a set of bandits (i.e. players) with multiple arms (i.e. actions). At each time instant, a gambler pulls an arm from the set of arms, and obtains a reward. Then, according to the pulled arm and obtained reward, the gambler updates the average reward of the arm. Note that, it has no prior information about the rewards of the arms, and its goal is to maximize its total rewards. Thus, the gambler needs to do exploratory rounds to estimate the rewards of the arms at the expense of risking low reward. However, this information is essential for maximizing long-term reward.

B. LINK OPTIMIZATION AS A MAB PROBLEM

Here, we aim at maximizing the objective function defined in (22) through optimizing the trajectories of the UAVs and allocating the resource among the BSs. We use the framework provided by MAB problem, in which each BS is defined to be an intelligent agent of the algorithm. The three main components of our algorithm are the players, their actions, and their reward function which are defined as follows:

- **Players:** The players in the algorithm are the BSs $\mathcal{B}$ in the system.
- **Actions:** For each SBS $s \in \mathcal{S}$, we define its action $a_{s,i} = q_i$ as its transmit channel. Thus, the action set of SBS s, $\mathcal{A}_{\mathrm{S}}$, is the set of available channels which can be defined as

$$\mathcal{A}_{\mathrm{S}} = \{q_i \in \mathcal{Q} \mid i \in \{1, \ldots, |\mathcal{Q}|\}\}. \tag{23}$$

  Let $\mathcal{Z}$ denote the set of movement in different directions for the UAVs as follows:

$$\mathcal{Z} = \{\text{up,down,left,right,forward,backward, no movement}\}. \tag{24}$$

  For each UAV $u \in \mathcal{U}$, each action $a_{u,i}$ is composed of its movement direction z_u and channel q_u as follows:

$$a_{u,i} = (q_u, z_u),\ q_u \in \mathcal{Q}, z_u \in \mathcal{Z}, \tag{25}$$

  where $\mathcal{A}_{\mathrm{U}} = \mathcal{Q} \times \mathcal{Z}$ denote the action set of the UAVs, and $i \in |\mathcal{A}_{\mathrm{U}}|$.
- **Reward:** We define a reward function for each BS $b \in \mathcal{B}$ which captures the fairness and the load of the BS as follows:

$$R_b(t) = \phi_b \mathcal{F}(t) + \psi_b(1 - \rho_b(t)) \tag{26}$$

Now, we use the UCB policy to solve our MAB based problem.

Algorithm 1 Proposed Learning Based Approach

Initialization: $\bar{R}_{b,i}(t) = 0, n_{b,i}(t) = 0$ for $t = 0, \forall b \in \mathcal{B}$, $\forall a_{b,i} \in \mathcal{A}_b$ and $i \in \{1, \ldots, |\mathcal{A}_b|\}$
while $t < N$ **do**
  $t \leftarrow t + 1$
  for $\forall b \in \mathcal{B}$ **do**
    Find the serving satellite (update $a_{l,b}^{\mathrm{BCK}(t)}, \forall l \in \mathcal{L}$)
  end for
  for $\forall k \in \mathcal{K}$ **do**
    Find the serving BS (update $a_{b,k}^{\mathrm{ACC}(t)}, \forall b \in \mathcal{B}$)
  end for
  for $\forall b \in \mathcal{B}$ **do**
    if $\exists a_{b,i} \in \mathcal{A}_b$ s.t. $n_{b,i}(t) = 0$ **then**
      Select arm $a_{\mathrm{UCB}}^{b}(t) = a_{b,i}$
    else
      Select arm $a_b^{\mathrm{UCB}}(t)$ according to (27)
    end if
    Calculate $R_b(t)$ according to (26)
    for $\forall a_{b,i} \in \mathcal{A}_b$ **do**
      Update $a_{b,i}(t)$ as:
      $n_{b,i}(t) = n_{b,i}(t-1) + \mathbb{1}_{\{a_{b,i} = a_b^{\mathrm{UCB}}(t)\}}$
      Update $\bar{R}_{b,i}(t)$ as:
      $\bar{R}_{b,i}(t) = \frac{n_{b,i}(t-1)\bar{R}_{b,i}(t-1) + \mathbb{1}_{\{a_{b,i} = a_b^{\mathrm{UCB}}(t)\}} R_b(t)}{n_{b,i}(t)}$
    end for
  end for
end while

C. UPPER CONFIDENCE BOUND ALGORITHM

In the UCB algorithm, each player b first selects each action once. Then, at each time instant $t > |\mathcal{A}_b|$, player b selects action $a_b^{\mathrm{UCB}}(t)$ as follows [51]:

$$a_b^{\mathrm{UCB}}(t) = \underset{a_{b,i} \in \mathcal{A}_b}{\operatorname{argmax}} \left\{ \bar{R}_{b,i}(t) + \sqrt{\frac{2 \ln t}{n_{b,i}(t)}} \right\}, \tag{27}$$

where $\mathcal{A}_b$ represents the action set of player b, in which $\mathcal{A}_b = \mathcal{A}_{\mathrm{S}}$ for each SBS $b \in \mathcal{S}$, and $\mathcal{A}_b = \mathcal{A}_{\mathrm{U}}$ for each UAV $b \in \mathcal{U}$. Here, $\bar{R}_{b,i}(t)$ denotes the average reward from action $a_{b,i} \in \mathcal{A}_b$ for player b at time instant t. Parameter $n_{b,i}(t)$ is the number of times that action $a_{b,i} \in \mathcal{A}_b$ has been selected by BS b until time instant t. Note that in the case that an action $a_{b,i}$ has been selected many times, i.e. $n_{b,i}(t)$ is large, compared to the other actions, then the confidence interval $\sqrt{\frac{2 \ln t}{n_{b,i}(t)}}$ decreases. Thus, player b intends to explore the other less selected actions. Furthermore, when an action $a_{b,j} \in \mathcal{A}_b$ obtains a high reward in the past, i.e. $\bar{R}_{b,i}(t)$ is large, player b intends to exploit this action to receive the possible maximum reward. The pseudocode for the MAB based approach is presented in Algorithm 1. To initialize the locations of UAVs, the heuristic approach proposed in [52] is utilized. This approach determines the locations of UAVs iteratively. At each iteration, the location of a new UAV is selected from a predefined set of horizontal locations such that it

is placed at the furthest distances from the other BSs in the system.

V. EVALUATION

For simulations, we consider a 500m × 500m area with a set of users uniformly distributed in the area. A set of SBSs are uniformly distributed within this area with a minimum distance of 40 and 10 meters from another SBS and a user, respectively. The main system parameters used in the simulations are summarized in Table 2. All results are averaged over a large number of independent runs for various practical configurations. We evaluate the performance of our proposed scheme compared to the following benchmark algorithms:

- *Q-learning*: The Q-learning based approach for the 2D trajectory design of UAVs is selected as one of the benchmark schemes. In this approach, the altitudes of UAVs are set to 100 m, and their 2D trajectories are optimized based on the Q-learning algorithm proposed in [53] with the reward function defined in (26). Furthermore, the BSs randomly choose their channels.
- *2D-MAB*: In 2D-MAB approach, each UAV flies at a fixed altitude 150 meter, and optimizes its horizontal location according to the UCB approach with the same reward function as the proposed approach. The channels are allocated randomly among all the BSs.

TABLE 2. System-level simulation parameters.

System Parameters		
Parameter		**Value**
Number of satellites in the orbital plane		22
Altitude of satellites		550 km
Height of SBSs		15 m
Height of users		1.5 m
Carrier frequency/channel bandwidth per BS in backhaul links		28 GHz, 100 MHz
$\lvert\mathcal{Q}\rvert$		4
ω_{ACC}		56 MHz
Noise power spectral density		-174 dBm/Hz
Number of SBSs		4
Number of UAVs		4
Total number of iterations (N)		5740
T_s		1 sec
Fixed point iterations (M)		500
ρ_b^0		0.5
α, β, γ		0.1, 750, 8
$V_{\min}, V_{\max}$		0, 1.3 m/sec
V_{U}		10 m/sec
$h_{\min}, h_{\max}$		22.5 m, 150 m
ϑ_k/ζ_k		1.8 Mbps
ϕ_b, ψ_b		0.5, 0.5
BS Parameters		
Parameter	**Terrestrial BS**	**UAV**
Transmit power	24 dBm	24 dBm
Reference path loss	LoS: 61.4 NLoS: 72 [54]	LoS: 61.4 NLoS: 61.4 [48]
Path loss exponent	LoS: 2 NLoS: 2.92	LoS: 2 NLoS: 3
Shadowing standard deviation ($\sigma^z_{b,\mathrm{SF}}$)	LoS: 5.8 NLoS: 8.7	LoS: 5.8 NLoS: 8.7

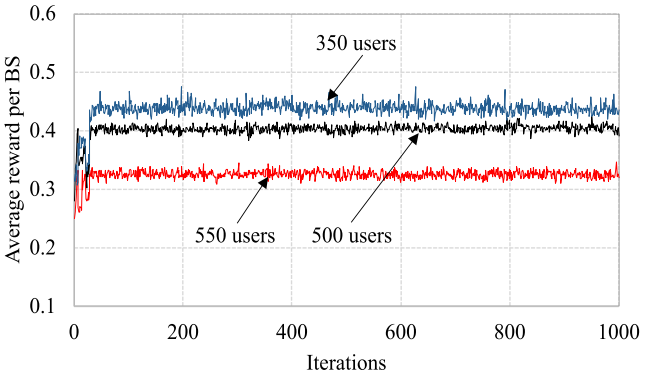

FIGURE 2. The convergence behavior of the proposed approach for 350, 500 and 550 users in a system with 4 SBSs and 4 UAVs.

- *Random*: In this scheme, each BS selects its action randomly.
- *No UAVs*: In order to demonstrate the benefits of employing the UAVs, no UAVs is used as another benchmark scheme. In particular, the MAB approach is utilized at the SBSs for channel allocation.

Fig. 2 shows the convergence behaviour of the proposed algorithm for 350, 500, and 550 users in terms of average reward per BS in a system with 4 SBSs and 4 UAVs. As can be observed, they converge fast, in which the convergence time reaches up to about 30 iterations.

In the following, the solid curves belong to the benchmark algorithms, and the dashed curve refers to the proposed approach.

A. PERFORMANCE OF PROPOSED MECHANISM VS NUMBER OF USERS

For the first set of the results, we consider a system with 4 SBSs and 4 UAVs, and vary the number of users in the system.

To verify the effectiveness of our proposed algorithm for enhancing the fairness, we illustrate the fairness index defined in (21) for all the schemes in the network shown in Fig. 3. Our proposed approach makes the system adaptable to achieve an improved spectral efficiency while ensuring the fairness. Therefore, our proposed approach can yield a significant performance in terms of the fairness under

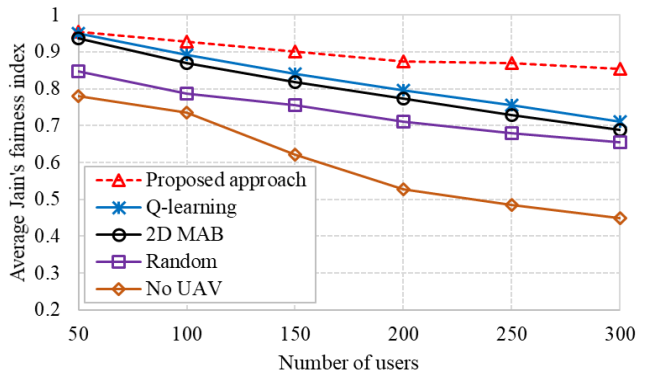

FIGURE 3. Average Jain's fairness index versus the number of users for a system with 4 SBSs and 4 UAVs.

dynamic traffic load, and achieves the highest fairness index. However, the performance gap between the proposed approach and 2D-MAB diminishes for the very low number of users (i.e. 50 users). As the number of users varies, compared to the Q-learning, 2D MAB, random, and no UAV approaches, the improvements of the fairness in the proposed approach are 20.10%, 24.22%, 30.74%, and 90.44%, respectively. Therefore, it can be confirmed that the proposed approach can achieve a remarkable improvement in the fairness. We also observe that when the number of users increases, the system starts to densify, and the average fairness index of all approaches degrade.

Fig. 4 presents the performance comparison of each approach in terms of average load per BS. Observing Fig. 4, it can be found that as the number of users increases, the average load per BS increases. Generally, a lower load means that the BSs have more capabilities to serve additional users compared to a higher load situation. We can observe that the Q-learning and proposed approach have almost the same performance in terms of BS load. Compared to the other benchmark algorithms, the proposed approach yields better performance for a low number of users. For instance, for a system with 50 users, the proposed approach can decrease the average load per BS 11.39%, 40.31%, and 50.65% compared to the 2D MAB, random, and no UAV approaches, respectively. For high number of users, the performance of all approaches will approach to 1.

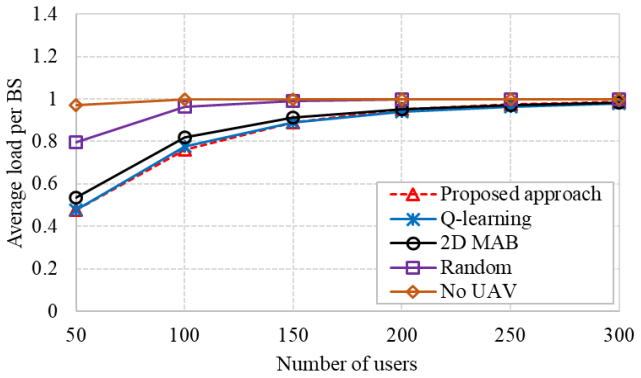

FIGURE 4. Average load per BS versus the number of users for a system with 4 SBSs and 4 UAVs.

Fig. 5 demonstrates the behavior of the reward function as the number of users increases. The result shows that our proposed approach can outperform the benchmark schemes. For the proposed approach, due to optimizing both the trajectory and resource allocation, it results better performance compared to the other approaches. Besides, the Q-learning approach can perform better than the other benchmark approaches due to optimizing the 2D trajectories of the UAVs using the Q-learning algorithm. Particularly, the proposed approach yields significant improvements over the Q-learning, 2D-MAB, random, and no UAV approaches reaching up to 18.41%, 23.49%, 46.4%, and 94.67%, respectively. Furthermore, it is observed that the

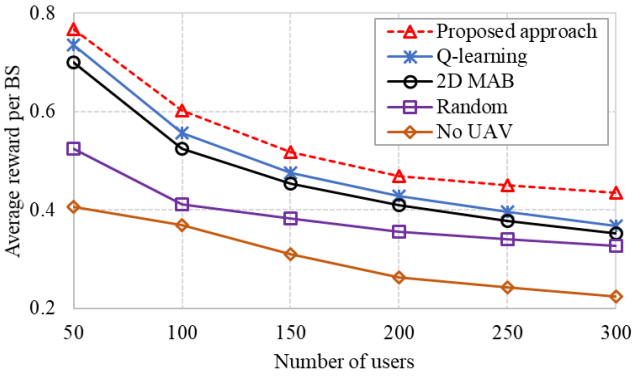

FIGURE 5. Average reward per BS versus the number of users for a system with 4 SBSs and 4 UAVs.

reward function decreases as the number of users increases. This could be explained by the fact that with densifying the system by increasing the number of users, the average load per BS increases and the fairness index per BS degrades. Thus, it yields a lower reward function per BS.

To highlight the user's QoE improvement capability of the proposed approach, we compare the average number of outage users for all the schemes in Fig. 6. The results indicate that our proposed approach has the potential to decrease the outage users by up to 87.4% with employing the UAVs by taking the advantage of compensating outage. In addition to that, it also proves the efficient way of improving the performance of the system up to 49.11%, 52.92%, and 68.37%, when compared to the Q-learning, 2D-MAB, and random approaches, respectively. Furthermore, it is clear that the average number of outage users increases proportional to the number of total users in the system. This is due to the fact that a limited amount of radio resource is available in the system.

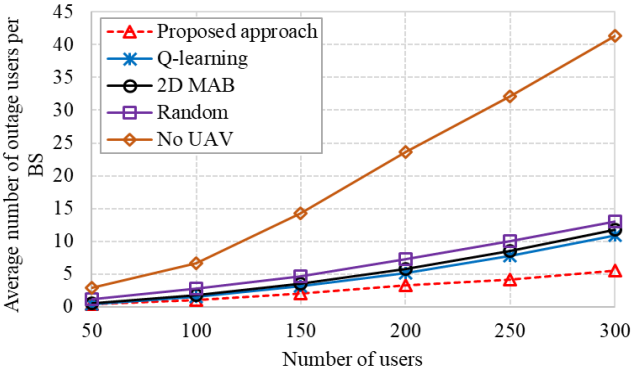

FIGURE 6. Average number of outage users versus the number of users for a system with 4 SBSs and 4 UAVs.

Next, we evaluate the effects of the number of users and the limited capacity in all the schemes in terms of the average rate. Fig. 7 shows that as we move away from the optimizing different parameters in the system, the average rate per user decreases. Accordingly, the proposed algorithm has a higher

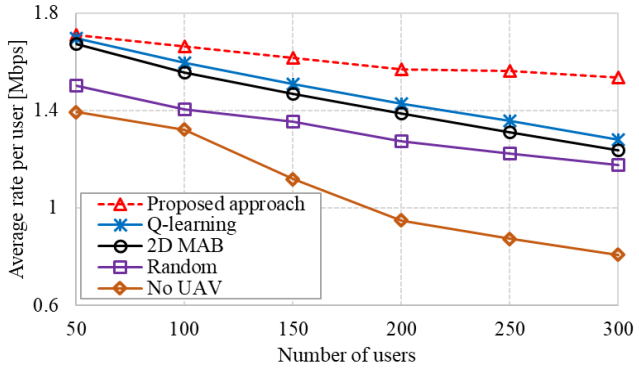

FIGURE 7. Average rate per user versus the number of users for a system with 4 SBSs and 4 UAVs.

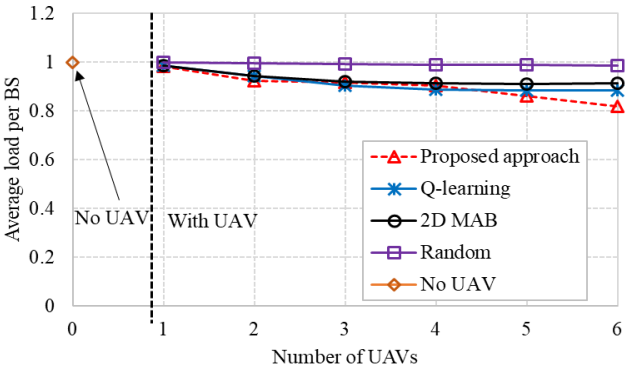

FIGURE 9. Average load per BS versus the number of UAVs for a system with 4 SBSs and 150 users.

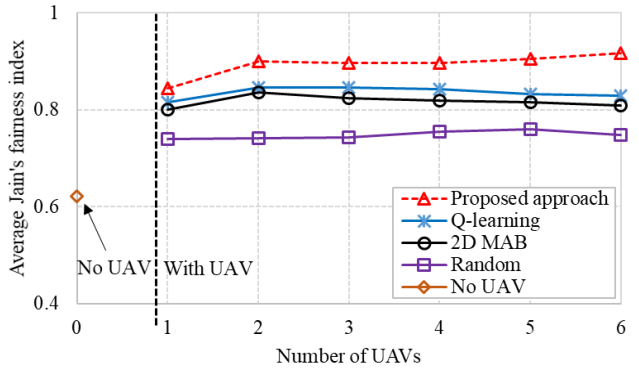

FIGURE 8. Average Jain's fairness index versus the number of UAVs for a system with 4 SBSs and 150 users.

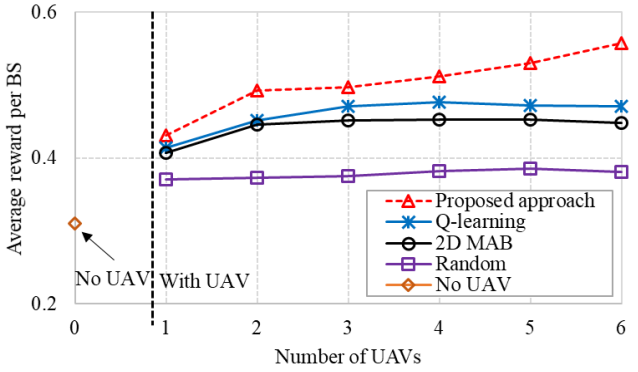

FIGURE 10. Average reward per BS versus the number of UAVs for a system with 4 SBSs and 150 users.

rate compared to the benchmark algorithms. The results indicate that the proposed approach has the potential to improve the average rate per user by up to 20.31%, 24.19%, 30.51%, and 90.59% compared to the Q-learning, 2D-MAB, random, and no UAV approaches, respectively.

B. PERFORMANCE OF PROPOSED MECHANISM VS NUMBER OF UAVs

In the further simulations, we vary the number of UAVs in the system, and observe the variation of different performance metrics. Moreover, we set the number of users to 150.

In Fig. 8, we plot Jain's fairness index defined in (21) for all the approaches. It can be seen that the proposed approach outperforms the other schemes for the different number of UAVs in the system by achieving better fairness index. Compared with the Q-learning, 2D-MAB and random approaches, the proposed approach can enhance the fairness index in a system with 150 users up to 10.56%, 13.34%, and 22.61% respectively. Furthermore, Fig. 8 shows that the proposed approach can improve the fairness up to 47.62% compared to no UAV approach, showing the benefit of deploying the UAVs in the system. Besides, increasing the number of UAVs can improve the fairness index in the proposed approach.

Fig. 9 illustrates the average load per BS by varying the number of UAVs from 0 to 6. We can observe that for the low number of UAVs, the proposed approach and the learning based benchmark algorithms have almost the same performance. While with increasing the number of UAVs, the proposed approach can balance the load among different BSs, and give more opportunities to the users to be associated to the BSs. The reductions in the average load per BS in the proposed approach are about 7.57%, 10.41%, 17.15%, and 18.25% compared to the Q-learning, 2D-MAB, random, and no UAV approaches, respectively.

Fig. 10 compares the average reward per BS for all the approaches. It shows that the higher number of UAVs yields higher reward per BS in the proposed approach. Therein, the improvements of the average reward per BS in the proposed approach compared to the Q-learning, 2D-MAB, random, and no UAV approaches are up to about 18.26%, 24.41%, 46.56%, and 79.49%, respectively.

Fig. 11 plots the average number of outage users for all the methods as the function of the number of UAVs. It can be seen that the proposed approach exhibits the lowest number of outage users compared to the other methods. The proposed approach yields about 49.01%, 54.02%, 64.33%, and 90.15% reductions in the outage users compared to the Q-learning, 2D-MAB, random and no UAV approaches, respectively.

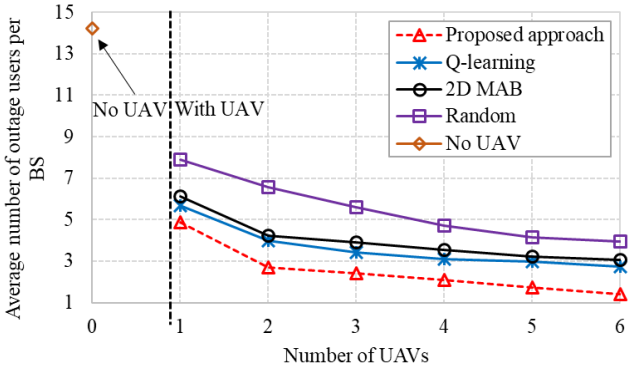

FIGURE 11. Average number of outage users versus the number of UAVs for a system with 4 SBSs and 150 users.

Furthermore, as the number of UAVs increases, the average number of outage users per BS decreases.

VI. CONCLUSION

In this paper, we have proposed a mechanism for link optimization for QoE in an STIN/SAGIN. To do that, we have decoupled our problem into three sub-problems including: the BS-satellite association problem in the backhaul links, the user-BS association and the resource management and UAV trajectory design in the access links while ensuring the fairness among users and minimizing the load of BSs. To solve our problem in the access links, we have modeled it as a MAB problem, and employed a UCB policy. Simulation results have shown that our approach substantially enhances both the user fairness and the spectral efficiency of the system compared to the benchmark algorithms.

ACKNOWLEDGMENT

The authors would like to thank anonymous reviewers who took the precious time on providing invaluable comments.

REFERENCES

[1] M. D. Sanctis, E. Cianca, G. Araniti, I. Bisio, and R. Prasad, "Satellite communications supporting Internet of remote things," *IEEE Internet Things J.*, vol. 3, no. 1, pp. 113–123, Feb. 2016.

[2] *Study on Management and Orchestration Aspects With Integrated Satellite Components in a 5G Network*, document TR 28.808, 3rd Generation Partnership Project, 2020. [Online]. Available: https://tinyurl.com/yxjptlwa

[3] *Architectural Framework for Machine Learning in Future Networks Including IMT-2020*, document ITU-T Y.3172, Telecommunication Standardization Sector of ITU, 2019. [Online]. Available: https://tinyurl.com/y66ppbuy

[4] 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. Surveys Tuts.*, vol. 21, no. 3, pp. 2334–2360, 3rd Quart., 2019.

[5] J. Saldana, A. Arcia-Moret, B. Braem, E. Pietrosemoli, A. Sathiasee, and M. Zennaro, *Alternative Network Deployments: Taxonomy, Characterization, Technologies, and Architectures*, document RFC 7962, Internet Requests for Comments, RFC Editor, Aug. 2016. [Online]. Available: https://www.rfc-editor.org/rfc/rfc7962.txt

[6] T. Kärkkäinen, R. Baig, A. Sathiaseelan, A. Ali, M. Isah, F. Arnal, R. Sallantin, D. Trossen, C. Theodorou, L. Pajevic, and J. Benseny, "RIFE (architectuRe for an Internet For everybody) D3.3: Final platform design and set of dissemination strategies," RIFE Consortium, Tech. Rep., 2017. [Online]. Available: https://rife-project.eu/wp-content/uploads/sites/31/2017/08/RIFE_D3.3_final.pdf

[7] 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 Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1688–1691, Dec. 2019.

[8] Y. Sun, D. Xu, D. W. K. Ng, L. Dai, and R. Schober, "Optimal 3D-trajectory design and resource allocation for solar-powered UAV communication systems," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4281–4298, Jun. 2019.

[9] Z. Wang, L. Duan, and R. Zhang, "Adaptive deployment for UAV-aided communication networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4531–4543, Sep. 2019.

[10] M. M. Azari, F. Rosas, K.-C. Chen, and S. Pollin, "Ultra reliable UAV communication using altitude and cooperation diversity," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 330–344, Jan. 2018.

[11] M. Alzenad and H. Yanikomeroglu, "Coverage and rate analysis for unmanned aerial vehicle base stations with LoS/NLoS propagation," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2018, pp. 1–7.

[12] J. Fan, M. Cui, G. Zhang, and Y. Chen, "Throughput improvement for multi-hop UAV relaying," *IEEE Access*, vol. 7, pp. 147732–147742, 2019.

[13] A. A. Khuwaja, G. Zheng, Y. Chen, and W. Feng, "Optimum deployment of multiple UAVs for coverage area maximization in the presence of co-channel interference," *IEEE Access*, vol. 7, pp. 85203–85212, 2019.

[14] A. H. Arani, M. M. Azari, W. Melek, and S. Safavi-Naeini, "Learning in the sky: Towards efficient 3D placement of UAVs," in *Proc. IEEE 31st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, Aug. 2020, pp. 1–7.

[15] H. E. 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. Wireless Commun.*, vol. 18, no. 12, pp. 5831–5844, Dec. 2019.

[16] J. Plachy, Z. Becvar, P. Mach, R. Marik, and M. Vondra, "Joint positioning of flying base stations and association of users: Evolutionary-based approach," *IEEE Access*, vol. 7, pp. 11454–11463, 2019.

[17] J. Cui, Y. Liu, and A. Nallanathan, "Multi-agent reinforcement learning-based resource allocation for UAV networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 729–743, Feb. 2020.

[18] E. Kalantari, M. Z. Shakir, H. Yanikomeroglu, and A. Yongacoglu, "Backhaul-aware robust 3D drone placement in 5G+ wireless networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2017, pp. 109–114.

[19] T. M. Nguyen, W. Ajib, and C. Assi, "A novel cooperative NOMA for designing UAV-assisted wireless backhaul networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2497–2507, Nov. 2018.

[20] M. D. Nguyen, T. M. Ho, L. B. Le, and A. Girard, "UAV placement and bandwidth allocation for UAV based wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[21] A. Fotouhi, M. Ding, L. Galati Giordano, M. Hassan, J. Li, and Z. Lin, "Joint optimization of access and backhaul links for UAVs based on reinforcement learning," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2019, pp. 1–6.

[22] M. Gapeyenko, V. Petrov, D. Moltchanov, S. Andreev, N. Himayat, and Y. Koucheryavy, "Flexible and reliable UAV-assisted backhaul operation in 5G mmWave cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2486–2496, Nov. 2018.

[23] C. T. Cicek, H. Gultekin, B. Tavli, and H. Yanikomeroglu, "Backhaul-aware optimization of UAV base station location and bandwidth allocation for profit maximization," *IEEE Access*, vol. 8, pp. 154573–154588, 2020.

[24] Y. Hu, M. Chen, and W. Saad, "Joint access and backhaul resource management in satellite-drone networks: A competitive market approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 3908–3923, Jun. 2020.

[25] A. Fouda, A. S. Ibrahim, I. Güvenç, and M. Ghosh, "Interference management in UAV-assisted integrated access and backhaul cellular networks," *IEEE Access*, vol. 7, pp. 104553–104566, 2019.

[26] A. H. Arani, P. Hu, and Y. Zhu, "Re-envisioning space-air-ground integrated networks: Reinforcement learning for link optimization," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jan. 2021, pp. 1–6.

[27] E. L. Hahne, "Round-robin scheduling for max-min fairness in data networks," *IEEE J. Sel. Areas Commun.*, vol. 9, no. 7, pp. 1024–1039, Sep. 1991.

[28] F. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. Telecommun.*, vol. 8, no. 1, pp. 33–37, Jan. 1997.

[29] R. K. Jain, D.-M. W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Eastern Res. Lab., Digit. Equip. Corp., Hudson, MA, USA, Tech. Rep. DEC-TR-301, 1984.

[30] M. F. Sohail and C. Y. Leow, "Maximized fairness for NOMA based drone communication system," in *Proc. IEEE 13th Malaysia Int. Conf. Commun. (MICC)*, Nov. 2017, pp. 119–123.
[31] C. H. Liu, Z. Chen, J. Tang, J. Xu, and C. Piao, "Energy-efficient UAV control for effective and fair communication coverage: A deep reinforcement learning approach," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 2059–2070, Sep. 2018.
[32] O. Abbasi, H. Yanikomeroglu, A. Ebrahimi, and N. M. Yamchi, "Trajectory design and power allocation for drone-assisted NR-V2X network with dynamic NOMA/OMA," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7153–7168, Nov. 2020.
[33] R. Ghanavi, M. Sabbaghian, and H. Yanikomeroglu, "Q-learning based aerial base station placement for fairness enhancement in mobile networks," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2019, pp. 1–5.
[34] J.-H. Lee, J. Park, M. Bennis, and Y.-C. Ko, "Integrating LEO satellite and UAV relaying via reinforcement learning for non-terrestrial networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.
[35] I. Leyva-Mayorga, B. Soret, and P. Popovski, "Inter-plane inter-satellite connectivity in dense LEO constellations," 2020, *arXiv:2005.07965*. [Online]. Available: http://arxiv.org/abs/2005.07965
[36] *Technical Specification Group Radio Access Network; Study on New Radio (NR) to Support Non-Terrestrial Networks*, document TR 38.811, 3rd Generation Partnership Project (3GPP), Version V15.3.0, 2017.
[37] N. Okati, T. Riihonen, D. Korpi, I. Angervuori, and R. Wichman, "Downlink coverage and rate analysis of low Earth orbit satellite constellations using stochastic geometry," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5120–5134, Aug. 2020.
[38] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Commun. Mobile Comput.*, vol. 2, no. 5, pp. 483–502, 2002.
[39] A. H. Arani, A. Mehbodniya, M. J. Omidi, and F. Adachi, "Distributed load balancing user association and self-organizing resource allocation in HetNets," in *Proc. IEEE 84th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–5.
[40] A. H. Arani, A. Mehbodniya, M. J. Omidi, F. Adachi, W. Saad, and I. Güvenç, "Distributed learning for energy-efficient resource management in self-organizing heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9287–9303, Oct. 2017.
[41] A. Hajijamali Arani, M. J. Omidi, A. Mehbodniya, and F. Adachi, "Minimizing base stations' ON/OFF switchings in self-organizing heterogeneous networks: A distributed satisfactory framework," *IEEE Access*, vol. 5, pp. 26267–26278, 2017.
[42] I. Viering, M. Dottling, and A. Lobinger, "A mathematical perspective of self-optimizing wireless networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2009, pp. 1–6.
[43] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.
[44] R. L. G. Cavalcante, S. Stanczak, J. Zhang, and H. Zhuang, "Low complexity iterative algorithms for power estimation in ultra-dense load coupled networks," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 6058–6070, Nov. 2016.
[45] A. J. Fehske and G. P. Fettweis, "Aggregation of variables in load models for interference-coupled cellular data networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 5102–5107.
[46] *Propagation Data and Prediction Methods Required for the Design of Terrestrial Broadband Radio Access Systems Operating in a Frequency Range From 3 to 60 GHz*, document ITU-R P.1410-5, Feb. 2012.
[47] J. Holis and P. Pechac, "Elevation dependent shadowing model for mobile communications via high altitude platforms in built-up areas," *IEEE Trans. Antennas Propag.*, vol. 56, no. 4, pp. 1078–1084, Apr. 2008.
[48] G. Fontanesi, A. Zhu, and H. Ahmadi, "Outage analysis for millimeter-wave fronthaul link of UAV-aided wireless networks," *IEEE Access*, vol. 8, pp. 111693–111706, 2020.
[49] H. Shi, R. V. Prasad, E. Onur, and I. G. M. M. Niemegeers, "Fairness in wireless networks: Issues, measures and challenges," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 5–24, 1st Quart., 2014.
[50] J. Wang, C. Jiang, H. Zhang, Y. Ren, K.-C. Chen, and L. Hanzo, "Thirty years of machine learning: The road to Pareto-optimal wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1472–1514, 3rd Quart., 2020.
[51] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An introduction*. Cambridge, MA, USA: MIT Press, 2018.
[52] F. Lagum, I. Bor-Yaliniz, and H. Yanikomeroglu, "Strategic densification with UAV-BSs in cellular networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 384–387, Jun. 2018.
[53] B. Khamidehi and E. S. Sousa, "Reinforcement learning-based trajectory design for the aerial base stations," in *Proc. IEEE 30th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2019, pp. 1–6.
[54] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

ATEFEH HAJIJAMALI ARANI received the Ph.D. degree in electrical engineering communication systems from the Isfahan University of Technology (IUT), Iran, in 2018. She is currently a Postdoctoral Fellow with the University of Waterloo, ON, Canada. Her research interests include machine learning, resource management, heterogeneous, and aerial networks.

PENG HU (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Queen's University, Canada. He is currently a Research Officer with the National Research Council of Canada, and an Adjunct Professor with the University of Waterloo. His current research interests include autonomous networking, AI-enabled edge computing, and the Internet of Things systems. He has served as a member for the IEEE Sensors Standards Committee and the organizing and technical boards/committees for industry consortia and international conferences, including AllSeen Alliance, DASH7, IEEE PIMRC'17, and IEEE AINA'15. He serves as an Associate Editor for the *Canadian Journal of Electrical and Computer Engineering*.

YEYING ZHU (Member, IEEE) received the Ph.D. degree in statistics from Pennsylvania State University, State College, PA, USA. She is currently an Associate Professor with the Department of Statistics and Actuarial Science, University of Waterloo. Her research interests include causal inference and machine learning methods.