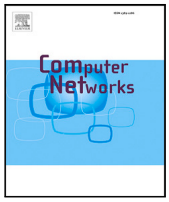
Computer Networks 247 (2024) 110462

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet



MUSCAT: Distributed multi-agent Q-learning-based minimum span channel allocation technique for UAV-enabled wireless networks

Ki-Hun Lee [a,1], Seungmin Lee [b,1], Jaedon Park [c], Howon Lee [d,*], Bang Chul Jung [a,*]

[a] Department of Electronics Engineering, Chungnam National University, Daejeon, 34134, South Korea
[b] Newratek, Inc., Seoul, 06175, South Korea
[c] Agency for Defense Development, Daejeon, 34186, South Korea
[d] Department of Electrical and Computer Engineering, Ajou University, Suwon, 16499, South Korea

ARTICLE INFO

Keywords:
Unmanned aerial system
Distributed dynamic resource allocation
Minimum span channel allocation problem
Multi-agent reinforcement learning
Stateless Q-learning

ABSTRACT

We consider a minimum span channel allocation problem (MS-CAP) to overcome spectrum scarcity and facilitate the efficiency of unmanned aerial vehicle (UAV)-enabled wireless networks. Basically, the MS-CAP minimizes the difference between the maximum and minimum used frequency, i.e., the required total bandwidth, while guaranteeing the quality-of-service (QoS) requirements for each wireless link in the network. The conventional optimal minimum span channel allocation (MS-CA) scheme is based on a centralized approach, assuming that global network information is available at the central controller. In practice, however, this may not be feasible for dynamic environments like UAV-enabled wireless networks since the real-time exchange of network information and channel allocation results with dynamically moving UAVs is formidable. Hence, we propose a novel practical MS-CA algorithm based on distributed multi-agent reinforcement learning (MARL), where each agent independently learns its best strategy with its local observations. To the best of our knowledge, the proposed technique is the first work of designing a distributed MARL for the MS-CAP for multi-UAV-enabled wireless networks in the literature. Numerical results reveal that the proposed distributed MS-CA technique can efficiently save the required total bandwidth while ensuring the QoS requirements of each link, represented by the signal-to-interference plus noise ratio (SINR) threshold, even in dynamic wireless networks. It validates the applicability of the proposed distributed MS-CA framework to dynamic networks.

1. Introduction

Future sixth-generation (6G) network envisions a three-dimensional heterogeneous network architecture with the introduction of unmanned aerial vehicles (UAVs), high altitude platform stations (HAPSs), and low earth orbit (LEO) satellites to realize cost-effective and high-capacity connectivity [1,2]. In particular, UAVs are in the spotlight as one of the indispensable platforms for 6G networks due to their versatility and maneuverability [3–6]. They can serve as additional aerial base stations (BSs) or relay nodes to provide ubiquitous coverages or configure flexible on-demand networks. As summarized in [3,6–8], extensive studies have been investigated for enhancing the performance of UAV-enabled wireless communication networks, which include UAV deployment, trajectory design, and radio resource management.

Since the dynamic environment due to the high mobility of UAVs makes network information exchanges among nodes formidable, decentralized frameworks based on multi-agent reinforcement learning (MARL) have attracted tremendous attention as a viable and practical alternative to conventional centralized optimization solvers [9–12]. In [9], the authors proposed a distributed multi-agent Q-learning-based dynamic joint power level, sub-channel, and user selection algorithm to improve system throughput while considering the power consumption of UAVs. It is worth noting that each UAV learns independently with its local observation. In [10], a joint UAVs' positions, transmit beamformers, and user association problem was investigated to maximize the achievable sum rate in cooperative multi-UAV networks, where the authors formulated a problem with mixed integer nonlinear programming and solved this with a difference of convex algorithm-assisted deep Q-learning scheme. In [11], the authors modeled a joint UAVs' power and channel allocation, user association, and trajectory design problem as a decentralized partially observable Markov decision process (MDP) to maximize both overall and fairness throughput, and solved this by exploiting a QMIX-shaped MARL framework with parameterized deep

* Corresponding authors.
E-mail addresses: kihun.h.lee@cnu.ac.kr (K.-H. Lee), sm.lee@newratek.com (S. Lee), jaedon2@add.re.kr (J. Park), howon@ajou.ac.kr (H. Lee), bcjung@cnu.ac.kr (B.C. Jung).
[1] Ki-Hun Lee and Seungmin Lee contributed equally to this work.

https://doi.org/10.1016/j.comnet.2024.110462
Received 2 January 2024; Received in revised form 21 March 2024; Accepted 24 April 2024
Available online 29 April 2024

Q-networks. In [12], deep reinforcement learning-based joint power allocation, user association, and trajectory design schemes were investigated to maximize the system utility. In summary, existing resource allocation techniques for UAV-enabled wireless networks, including the studies mentioned above, focus on maximizing the system throughput or improving the energy efficiency of UAVs.

From a different perspective, we revisit the traditional minimum span channel (frequency) allocation problem (MS-CAP) that minimizes the bandwidth used in the network while satisfying each link's quality-of-service (QoS) requirements. In other words, our resource allocation technique for UAV-enabled networks aims to strike spectral scarcity issues and alleviate spectrum usage fees by minimizing the difference between the maximum and minimum used frequency, which implies the substantially required total bandwidth in the network [5,13]. To the best of our knowledge, the MS-CAP has not been well-investigated since the literature [14], even though next-generation wireless networks are expected to introduce spectrum scarcity problems with the drastic increase of wireless terminals and data traffic [1,7]. An example scenario of the MS-CAP is a spectrum-sharing technique based on spectrum leasing, in which a network operator rents a portion of licensed frequency bands for a certain period of time [15]. It is particularly suitable for UAV-enabled wireless networks because UAVs are generally operated temporarily rather than permanently due to their limited power consumption, and their dynamic characteristics make spectrum sensing-based sharing infeasible [16]. The UAV network operators must lease a portion of the spectrum and minimize the required total bandwidth while ensuring minimum QoS requirements to reduce the price charged.

On the other hand, the conventional MS-CAP is based on a centralized approach that is no longer suitable for UAV-enabled wireless networks due to the aforementioned dynamic nature. This motivates us to design a novel distributed dynamic MS-CA technique for multi-UAV-enabled networks. Furthermore, conventional MS-CAPs are based on the *protocol* interference model for simplicity, even though they cannot capture the actual cumulative interference received at each link in the network [17]. In contrast, we exploit the *physical* interference model to reflect the more realistic received signal strength (RSS), incorporating fading and path loss for each link. The MS-CAPs based on the two interference models have the following differences:

- In the *protocol interference model*-based MS-CAP, the problem can be formulated as a directed graph in which the edges are generated depending on the distance between vertices. More specifically, when one vertex is within the guard zone radius (interference range) of another vertex, a directed edge is formed between the two vertices. Here, the guard zone radius for each vertex can be defined in proportion to the maximum tolerable interference of the vertex. Then, the MS-CAP is solved like edge coloring of a graph, allocating channels (colors) to the links (edges) such that no two incident edges have the same color while using as few colors as possible. Unfortunately, its channel allocation results cannot guarantee that the target signal-to-interference plus noise ratio (SINR) constraint for each link has been satisfied. If the distance for edge connection between two vertices is large, it will consume many channel indices; and if it is small, the SINR may deteriorate below the target SINR due to cumulate interference.
- In the *physical interference model*-based MS-CAP we formulate, all given wireless links in the network, including communication (desired) and interference links, can be transformed into a weighted complete (fully connected) directed graph, where weights represent the RSS of each link. Herein, the MS-CAP can be solved like edge coloring, which is not restricted to allocating different channels (colors) to incident links (edges). Specifically, edge coloring is performed considering weights (desired signal strength (positive), cumulative interference from the edges colored in the same color (negative)), noise power, and the target SINR; hence, some (or all) edges connected to the same vertex can be colored the same color as long as the SINR constraint is satisfied.

In this paper, we formulate a novel MS-CAP that considers the physical interference model as MDPs for all wireless links and solve them by exploiting distributed MARL framework. The main contributions of our work are summarized as follows:

- In contrast to conventional resource allocation techniques that maximize throughput or energy efficiency for UAV-enabled wireless networks, we revisit the MS-CAP to overcome the spectrum scarcity problem and facilitate the cost efficiency of the networks.
- Existing MS-CAPs are based on the *protocol* interference model for simplicity even though they cannot reflect the more realistic received signal strength (RSS) of links. In contrast, we formulate a novel *centralized* optimal MS-CAP with the *physical* interference model.
- We design a novel *decentralized dynamic* MS-CA technique based on distributed multi-agent Q-learning since the centralized MS-CA may not be feasible in UAV-enabled wireless networks due to the dynamic nature. It is worth noting that this is the first work on designing a distributed MARL framework for the dynamic MS-CAP.

2. System model

We consider a multi-UAV-enabled wireless network consisting of multiple UAVs and BSs, as illustrated in Fig. 1, where all communicating nodes are assumed to be equipped with a single antenna. Due to the dynamic nature of the network, all nodes operate in a decentralized fashion without the assistance of a central unit performing centralized operations, i.e., a distributed UAV-enabled network is considered. In particular, we focus on communication links between UAVs and BSs, where the BSs act as ground control stations (GCSs) for exchanging control and user plane signals with the UAVs. To model dynamic topologies according to the UAVs' mobilities, we assume that the considered network operates on a discrete-time axis partitioned into equal intervals T_s and define each interval sequentially as a time slot t ($\in \{1, 2, \dots, \lfloor T/T_s \rfloor\}$), where T is the total time.

2.1. Spatial model

Let $\mathcal{N}$ ($= \mathcal{N}_{\mathsf{BS}} \cup \mathcal{N}_{\mathsf{UAV}}$) be the set of all nodes in the network, where $\mathcal{N}_{\mathsf{BS}}$ and $\mathcal{N}_{\mathsf{UAV}}$ represent sets of BSs and UAVs, respectively; and $\mathbf{p}_n^{(t)}$ ($= [x_n^{(t)}, y_n^{(t)}, h_n^{(t)}]$) denote the three-dimensional (3D) Cartesian coordinates of node n ($\in \mathcal{N}$) at time slot t. All BSs are stationarily deployed within a circular disk area D with radius R, and UAVs fly over this plane. Specifically, the BSs are randomly distributed within D in accordance with a homogeneous Poisson point process (PPP) Φ with density λ. Each deployed BS controls a single UAV and is associated with it in a pair; hence, there are the same number of BSs and UAVs.[2] Moreover, considering the communication range of each BS-UAV pair, it is assumed that the UAV flies in a cylindrical space with radius r and height H with its BS location as the origin. Each association link is assumed to have a frequency division duplex (FDD) link. That is, each link must be assigned two different channels for transmission and reception, and we define $\mathcal{L}$ ($= \{1, 2, \dots, L\}$) as the set of all associated (desired) links where $L = |\mathcal{N}|$.

[2] We assumed this spatial model consisting of multiple BS-UAV pairs to avoid distraction from our underlying goal of the MS-CAP due to considerations such as association policy, handover mechanism, trajectory design, etc. The results obtained in this paper are not necessarily limited to such a system model, i.e., the proposed channel allocation technique can be straightforwardly applied to any network topology for the MS-CA.

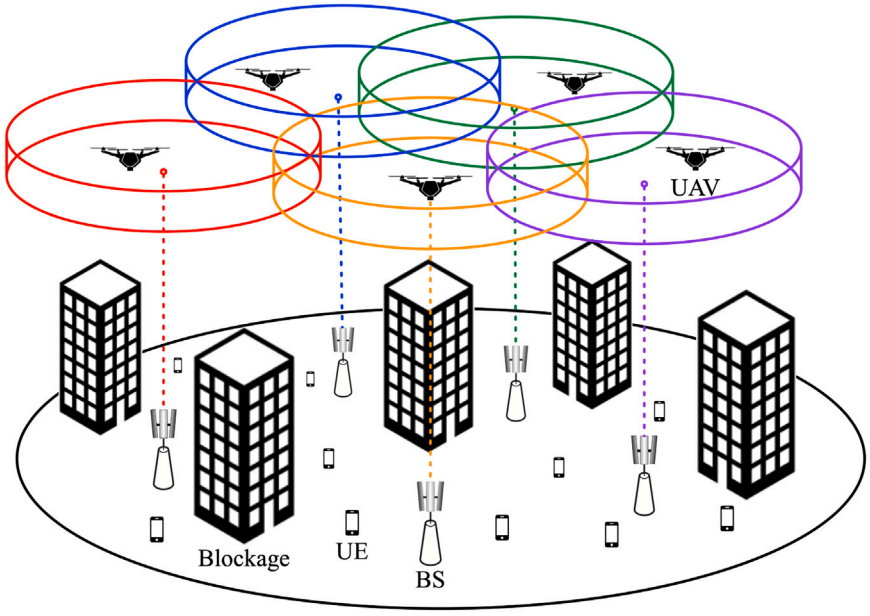


Fig. 1. An example of the multi-UAV-enabled wireless network with multiple UAV-BS pairs.

2.2. UAV mobility model

We focus on the resource (channel) allocation problem given the UAVs' trajectories as in the literature [9]. Hence, we assume that UAVs fly according to their predefined trajectories and adopt a random mobility model for the UAVs' trajectories to configure various UAV-enabled wireless networks. Such random trajectories can also be considered in practical UAV application scenarios, such as reconnaissance, random arrival of UAVs for package delivery, etc [18].

Specifically, the movement of UAVs is modeled to follow the mixed mobility (MM) model, which is a mixture of the one-dimensional random waypoint mobility (RWPM) and two-dimensional random work (RW) models for the vertical and horizontal directions, respectively [18]. Without loss of generality, we consider an arbitrary UAV in the network. The UAV is initially located at a random point within its cylindrical flight area and moves by repeating the following two steps.

2.2.1. Vertical movement

First, the UAV selects a next waypoint H' from a uniform distribution $\mathcal{U}[H_{\min}, H_{\max}]$, and moves towards it with a constant speed v_{v} which is also chosen from $\mathcal{U}[v_{\mathsf{v},\min}, v_{\mathsf{v},\max}]$, where $H_{\min}$ and $H_{\max}$ denote the minimum and maximum altitude of the UAV; $v_{\mathsf{v},\min}$ and $v_{\mathsf{v},\max}$ represent the minimum and maximum speed of the UAV in the vertical direction, respectively. Upon reaching the waypoint H', the UAV stays at this altitude for a dwell time T_{d} drawn from a uniform distribution $\mathcal{U}[\tau_{\min}, \tau_{\max}]$, during which it moves horizontally.

2.2.2. Horizontal movement

Unlike [18], which employs an RW model, we design a more realistic horizontal mobility model. First, the speed v_{h} and direction φ of the UAV are randomly initialized from $\mathcal{U}[v_{\mathsf{h},\min}, v_{\mathsf{h},\max}]$ and $\mathcal{U}[0, 2\pi)$, respectively, where $v_{\mathsf{h},\min}$ and $v_{\mathsf{h},\max}$ denote the minimum and maximum speed of the UAV in the horizontal direction, respectively. The UAV then moves from its previous position with changes in speed $v'_{\mathsf{h}} = v_{\mathsf{h}} \pm \Delta v_{\mathsf{h}}$ and direction $\varphi' = \varphi \pm \Delta\varphi$ for each time slot until the end of the dwell time T_{d}. When the UAV reaches the boundary of its operating circle during horizontal movement, it changes direction to the origin, i.e., bounce-back.

2.3. Propagation model

Let $\bar{r}_{j,i}^{(t)}$ be the average received signal strength (RSS) that the receiver (RX) of link j ($\in \mathcal{L}$) receives from the transmitter (TX) of link i ($\in \mathcal{L}$) at time slot t. We can define air-to-air (A2A), ground-to-ground (G2G), air-to-ground (A2G), and ground-to-air (G2A) channels depending on each link's TX and RX nodes, then the average RSS of each link can be calculated as follows:

$$\bar{r}_{j,i} = \begin{cases} P_i L_{(1\ m)} \|\mathbf{p}_j - \mathbf{p}_i\|^{-2}, & \text{for A2A link,} \\ P_i L_{(1\ m)} \|\mathbf{p}_j - \mathbf{p}_i\|^{-3}, & \text{for G2G link,} \\ \prod_{\xi} \left[P_i L_{(1\ m)} \|\mathbf{p}_j - \mathbf{p}_i\|^{-2} \eta_\xi\right]^{\Pr(\xi|\overline{\mathbf{p}_j,\mathbf{p}_i})}, & \text{for A2G link,} \end{cases} \tag{1}$$

where the time index t is omitted for the sake of brevity, i.e., $\bar{r}_{j,i} := \bar{r}_{j,i}^{(t)}$ and $\mathbf{p}_j := \mathbf{p}_j^{(t)}$; $\mathbf{p}_j^{(t)}$ and $\mathbf{p}_i^{(t)}$ denote the locations of the RX of link j and the TX of link i corresponding to the time slot t, respectively; P_i is the transmission power of link i; and $L_{(1\ m)} = (\frac{c}{4\pi f_c})^2$ represents the propagation loss at a reference distance (1 m), where c and f_c are the light speed (m/s) and carrier frequency (Hz), respectively. For A2A and G2G links, path loss models are adopted taking into account free space and urban scenario path loss exponents, respectively [19]. Since the G2A and A2G channels have the same characteristics, we describe the A2G channel model in this subsection. Using the probabilistic A2G path loss model in [20], the average RSS reading for each A2G (G2A) link can be calculated as shown in (1) in linear scale, where ξ ($\in \{\mathsf{LoS}, \mathsf{NLoS}\}$) denotes the link scenario group and η_ξ is the mean value of the excessive path loss for the line-of-sight (LoS) and non-LoS (NLoS) scenarios. In this propagation model, the sigmoid function-shaped LoS probability denoted as $\Pr(\mathsf{LoS}|\overline{\mathbf{p}_j,\mathbf{p}_i})$ is given by

$$\Pr(\mathsf{LoS}|\overline{\mathbf{p}_j,\mathbf{p}_i}) = \frac{1}{1 + \mathfrak{a}\exp\left(-\mathfrak{b}\sin^{-1}\left(\frac{|h_j - h_i|}{\|\mathbf{p}_j - \mathbf{p}_i\|}\right) - \mathfrak{a}\right)},$$

where $\overline{\mathbf{p}_j,\mathbf{p}_i}$ denotes the link between the TX of link i and the RX of link j; $\mathfrak{a}$ and $\mathfrak{b}$ are parameters that depend on the propagation environment, such as rural, urban, dense urban, etc; and $\sin^{-1}(\cdot)$ is the elevation angle of link $\overline{\mathbf{p}_j,\mathbf{p}_i}$ in degrees. Accordingly, the NLoS probability becomes $\Pr(\mathsf{NLoS}|\overline{\mathbf{p}_j,\mathbf{p}_i}) = 1 - \Pr(\mathsf{LoS}|\overline{\mathbf{p}_j,\mathbf{p}_i})$. Note that $\overline{\mathbf{p}_j,\mathbf{p}_i}$ is the desired link if $j = i$, otherwise the interference link.

Furthermore, the instantaneous RSS of each link considering small-scale fading, denoted by $r_{j,i}^{(t)}$, can be modeled as

$$r_{j,i}=\begin{cases}\bar{r}_{j,i}, & \text{for A2A link,}\\ \bar{r}_{j,i}|\widetilde{g}_{j,i}|^2, & \text{for G2G link,}\\ \bar{r}_{j,i}\left|\sqrt{\frac{\kappa_{j,i}}{\kappa_{j,i}+1}}\bar{g}_{j,i}+\sqrt{\frac{1}{\kappa_{j,i}+1}}\widetilde{g}_{j,i}\right|^2, & \text{for A2G link,}\end{cases} \tag{2}$$

where the time index t is still omitted for the sake of brevity, i.e., $r_{j,i} := r_{j,i}^{(t)}$, $\kappa_{j,i} := \kappa_{j,i}^{(t)}$, $\widetilde{g}_{j,i} := \widetilde{g}_{j,i}^{(t)}$, and $\bar{g}_{j,i} := \bar{g}_{j,i}^{(t)}$, $\widetilde{g}_{j,i}$ denotes the random scattered small-scale fading component which is assumed to follow an independent and identically distributed complex standard normal distribution, i.e., $\widetilde{g}_{j,i} \sim \mathcal{CN}(0,1)$ and $|\widetilde{g}_{j,i}|^2 \sim \exp(1)$; $\kappa_{j,i}$ denotes the angle-dependent Rician K-factor of link $\overline{\mathbf{p}_j,\mathbf{p}_i}$ among A2G and G2A links, given by

$$\kappa_{j,i} = \mathfrak{c}\exp\left(\mathfrak{e}\sin^{-1}\left(\frac{|h_j-h_i|}{\|\mathbf{p}_j-\mathbf{p}_i\|}\right)\right),$$

where $\mathfrak{c}$ and $\mathfrak{e}$ are constant coefficients determined by specific environments [10] and $\sin^{-1}(\cdot)$ is the elevation angle of the link in radians; and $\bar{g}_{j,i}$ denotes the deterministic LoS component with $|\bar{g}_{j,i}|=1$. Intuitively, Rayleigh and Rician fading channel gains are included for G2G and A2G links, respectively [10,21]. Note that the proposed MS-CA technique does not depend on such a certain channel model; hence, it can be applied to any given channel model.

It is also worth noting that the instantaneous RSS (2) is only available for the distributed MS-CA in Section 3.2, but not for the centralized one in Section 3.1. This is because the centralized resource allocation problem, considering the instantaneous small-scale fading effects for all links, may not be feasible in practical wireless networks because of the signaling overhead to obtain all RSS readings and feedback channel allocation results. In contrast, there is no such signaling overhead for the distributed scheme [21]. Hence, we leverage the average RSS for the centralized CAP by averaging out the small-scale fading effects as $\bar{r}_{j,i}^{(t)} = \mathbb{E}[r_{j,i}^{(t)}]$.

3. Minimum span channel allocation

We first formulate a *centralized* MS-CAP considering the physical interference model, and then design a *decentralized* MS-CA algorithm based on MARL, especially distributed multi-agent stateless Q-learning. As aforementioned, our main objective is to minimize the total bandwidth required in the network while ensuring the QoS requirements of each link, where the required total bandwidth is defined as the difference between the maximum and minimum used frequency. To design a more realistic dynamic resource allocation scheme, we assume that the network employs a dual-band connection frequency plan and focus on the channel allocation for the payload communication links. For high reliability, orthogonal channels are assigned to the command and control (C2) links, which are also exploited to convey messages containing information for the RL process in Section 3.2 between UAVs and BSs. This plan is reasonable enough as C2 links require low data rates of 60–100 kbps in general, which could be covered with 50 kHz bandwidth according to Shannon's theoretical channel capacity. Meanwhile, we will allocate a channel with 20 MHz bandwidth to each payload link because it requires a data rate of up to 50 Mbps [22].

3.1. Problem formulation

We introduce binary optimization variables to formulate the centralized optimal MS-CAP: $z_k^{(t)}$ ($\in\{0,1\}$, $\forall k \in \mathcal{K}$) and $z_{j,k}^{(t)}$ ($\in\{0,1\}$, $\forall j \in \mathcal{L}$, $k \in \mathcal{K}$), where $\mathcal{K}$ ($=\{1,2,\dots,K\}$) denotes the set of available channel indices in the network. To be specific,

$$\begin{cases} z_k^{(t)}=1, & \text{If channel } k \text{ is used on any link at time slot } t,\\ z_k^{(t)}=0, & \text{Otherwise;}\end{cases}$$

and

$$\begin{cases} z_{j,k}^{(t)}=1, & \text{If channel } k \text{ is allocated to link } j \text{ at time slot } t,\\ z_{j,k}^{(t)}=0, & \text{Otherwise.}\end{cases}$$

Then, for each time slot t, the *centralized* optimal MS-CAP with the physical interference model can be formulated based on [14] as follows:

$$\min_{z_k} \sum_{k\in\mathcal{K}} z_k \tag{3a}$$

$$\text{s.t.}\quad z_k \ge z_{k+1}, \qquad \forall k \in \mathcal{K}\backslash K, \tag{3b}$$

$$z_k \ge z_{j,k}, \qquad \forall j \in \mathcal{L}, \forall k \in \mathcal{K}, \tag{3c}$$

$$\sum_{k\in\mathcal{K}} z_{j,k} = 1, \qquad \forall j \in \mathcal{L}, \tag{3d}$$

$$\sum_{k\in\mathcal{K}} \frac{\bar{r}_{j,j} z_{j,k}}{\sum_{i\in\mathcal{L}\backslash j} \bar{r}_{j,i} z_{i,k} + N_0 W_j} \ge \underline{\gamma}_j, \ \forall j \in \mathcal{L}, \tag{3e}$$

where $\underline{\gamma}_j$ denotes the target QoS requirement represented by the signal-to-interference plus noise ratio (SINR) threshold of link j. We omitted the time index $(\cdot)^{(t)}$ for notational convenience. The objective function (3a) minimizes the number of channel indices required in the network. Constraint (3b) implies that a channel with a lower index should be used preferentially over a higher one. Constraint (3c) indicates that channel k is used if it is allocated to any link j. Constraint (3d) states that a single channel is allocated to each link. Finally, constraint (3e) includes the QoS requirements of each link j represented by SINR, where N_0 and W represent the noise power and channel bandwidth, respectively. The formulated optimum MS-CAP takes the average RSSs of all links and the channel bandwidth and SINR threshold of each desired link as the main inputs and returns the channel index assigned to each link.

Although various efficient solvers have been developed for optimization problems, this formulation may be practically infeasible in UAV-enabled wireless networks because global network information, such as RSSs for all links, is unavailable on the central controller due to UAVs' high mobilities and dynamic nature. Moreover, this problem is basically *NP-hard*, as proved in Remark 1, even if global information is given. Hence, it is desirable to design a decentralized algorithm with low-complexity.

Remark 1. Formulation (3a)–(3e) is an NP-hard problem.

Proof. Constraint (3e) can be equivalently transformed into linear form as $\alpha_j z_{j,k} + B(1-z_{j,k}) \ge \sum_{i\in\mathcal{L}\backslash j} r_{j,i} z_{i,k}$, $\forall j \in \mathcal{L}, \forall k \in \mathcal{K}$, where $\alpha_j = r_{j,j}/\underline{\gamma}_j - N_0 W_j$, and B is a sufficiently big number. Moreover, this can be rearranged as

$$\sum_{i\in\mathcal{L}} r_{j,i} z_{i,k} \le B, \ \forall j \in \mathcal{L}, \forall k \in \mathcal{K},$$

where we have redefined $r_{j,j} := B - \alpha_j$. Then, our problem becomes a vector bin packing problem stated as follows:

- bin k = channel k; item j = link j
- each bin has L resources, denoted by $b_1,\dots,b_L$
- the capacity of each resource in bin k is B
- resources consumed by item j if it is put into bin k are

  $$b_1 : r_{1,j},\ b_2 : r_{2,j},\dots,\ b_j : B-\alpha_j,\dots,\ b_L : r_{L,j}$$

- the goal is to pack all the items (links) into the minimum number of bins (channels) while the consumption of each resource b_l in each bin remains below the capacity (or SINR requirements are satisfied).

The vector bin packing problem is a well-known NP-hard problem [23], so formulation (3a)–(3e) is NP-hard as well. □

Algorithm 1 Distributed MARL framework of MUSCAT

Input: $\mathcal{L}$, $\mathcal{K}$, $\varepsilon_{\text{init}}$, ζ, α, β, W_j, $\underline{\gamma}_j$.
Initialization: $\mathcal{A}_j = \mathcal{K}$, $Q_j^{(1)} = \mathbf{0}^{|\mathcal{A}_j|\times 1}$, $\forall j \in \mathcal{L}$, $t = 1$
while $t \le T/T_{\text{s}}$ **do**
  Update $\varepsilon^{(t)} = \varepsilon_{\text{init}}(1-\varepsilon_{\text{init}})^{\frac{t-1}{\zeta\times|\mathcal{A}_j|}}$
  Select actions $A_j^{(t)}$ according to (5), $\forall j \in \mathcal{L}$
  for all $j \in \mathcal{L}$ **do**
    Measure the SINR at the RX according to (7)
    Obtain the reward $R_j^{(t)}$ according to (6)
    Update the Q-value $Q_j^{(t+1)}(A_j^{(t)})$ according to (4)
    Update $t \leftarrow t+1$
  end for
end while

3.2. MARL-based multi-UAS MS-CA: MUSCAT

We design a distributed MARL framework for solving the dynamic MS-CAP in a *decentralized* manner. In particular, we leverage stateless Q-learning [21], also known as single-state Q-learning.[3] Hence, each agent (link) generates a Q-table consisting of its action space and recursively learns its Q-value in the Q-table to find an optimal strategy (*channel allocation*) only through its local information. Specifically, in time slot $t+1$, agent j ($\in \mathcal{L}$)'s Q-function is updated according to the following update rule:

$$Q_j^{(t+1)}(A_j^{(t)}) = (1-\alpha)Q_j^{(t)}(A_j^{(t)}) + \alpha\left(R_j^{(t)} + \beta \max Q_j^{(t)}\right), \tag{4}$$

where $A_j^{(t)}$ and $R_j^{(t)}$ denote the selected action and obtained reward of agent j at time slot t which will be discussed in Sections 3.2.2 and 3.2.4, respectively. Note that the Q-function represents the expected long-term reward of the action that agent j decides, which is given by

$$Q_j^{(t)}(A_j^{(t)}) = \mathbb{E}\left[\sum_{\tau=0}^{\infty}\beta^{\tau}R_j^{(t+\tau+1)}\middle|A_j^{(t)}\right],$$

where $\sum_{\tau=0}^{\infty}\beta^{\tau}R_j^{(t+\tau+1)}$ is the long-term reward of agent j at a certain time slot t. The potential meaning of the Q-values are the expected reward that the agent ultimately can be obtained when following the action, rather than the true reward $R_j^{(t)}$. Therefore, it is important to learn optimal Q-function guaranteeing the best reward.

We formulate an MS-CAP as Markov decision processes (MDPs) for all links and solve them through a stateless Q-learning-based distributed MARL framework. Algorithm 1 represents the pseudo-code of the proposed multi-UAS MS-CA technique, named MUSCAT (Multi-Unmanned aerial vehicle minimum Span Channel Allocation Technique), where α ($\in (0,1]$) and β ($\in (0,1]$) denote the learning rate and discount factor, respectively. From the system model, the decision period is defined within a time interval T_{s} [9].[4] Each link runs the MUSCAT algorithm independently and simultaneously determines an optimal strategy for the MDP, which consists of:

[3] Stateless Q-learning is also called multi-armed bandit (MAB) in the sense that there is no state space [24]. In this paper, however, we adopt the name of stateless Q-learning because the update rule for the Q-function follows the Q-learning framework.

[4] Due to the limited decision period, traditional alternative distributed optimization methods with iterative algorithms, such as the alternating direction method of multipliers (ADMM), are not applicable here. It is worth noting in Algorithm 1 that each agent has no iterations to update its Q-table in each time slot.

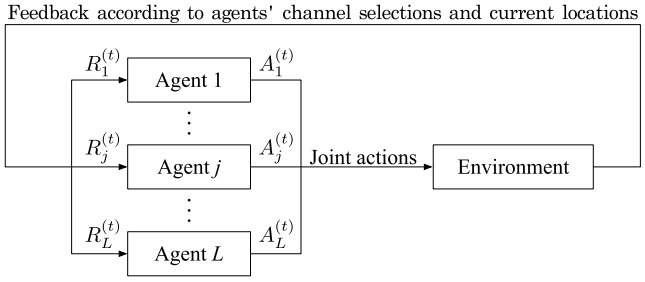


Fig. 2. Distributed stateless Q-learning-based MARL architecture.

3.2.1. Agent

We can take each link in the network as an independent learning agent, so there are L agents corresponding to L different links. It is noteworthy that each agent cannot observe information about other agents, such as their actions (channel selections), rewards, locations, and interferences; i.e., each agent runs the Q-learning procedure independently with only its local information.

3.2.2. Action

In each time slot, each agent decides on channel allocation. We define the set of all possible actions performed by agent j as $\mathcal{A}_j$ ($= \mathcal{K}$), which consists of available channel indices in the network. Moreover, we adopt a decaying ε-greedy exploration policy as the action selection mechanism, which is used to select the agent's action in each time slot. The action of agent j at time slot t, denoted by $A_j^{(t)}$, is then conducted as follows:

$$A_j^{(t)} = \begin{cases} k\ (\sim \mathcal{U}\{1,K\}), & \text{with probability } \varepsilon^{(t)}, \\ \arg\max\limits_{a_j\in\mathcal{A}_j} Q_j^{(t)}, & \text{with probability } 1-\varepsilon^{(t)}, \end{cases} \tag{5}$$

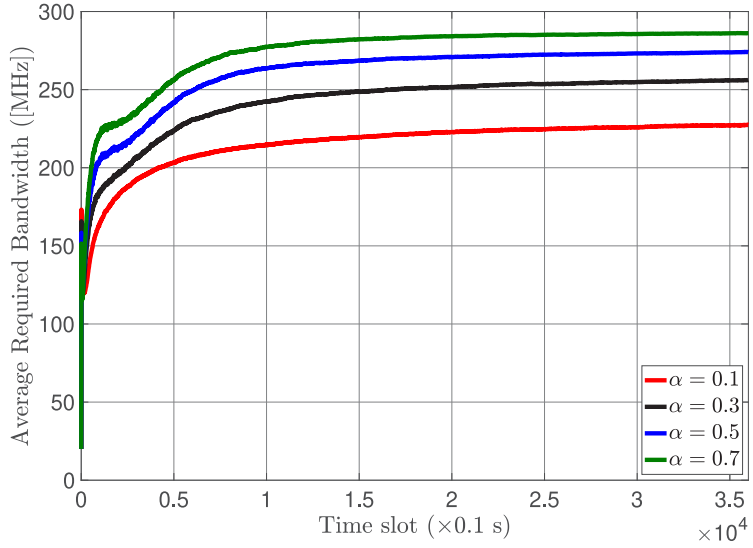
where $\varepsilon^{(t)} = \varepsilon_{\text{init}}(1-\varepsilon_{\text{init}})^{\frac{t-1}{\zeta\times|\mathcal{A}_j|}}$ and $\mathcal{U}\{\cdot\}$ represents a discrete uniform distribution. Also, $\varepsilon_{\text{init}}$ and ζ denote the initial value of ε and the exploration parameter, respectively. That is, the agent performs the learning process by exploring a new action through random channel allocation with probability $\varepsilon^{(t)}$ or exploiting an action known as the best so far, which corresponds to the highest Q-value at the moment, with probability $1-\varepsilon^{(t)}$ to strike a trade-off between exploitation and exploration.

3.2.3. State

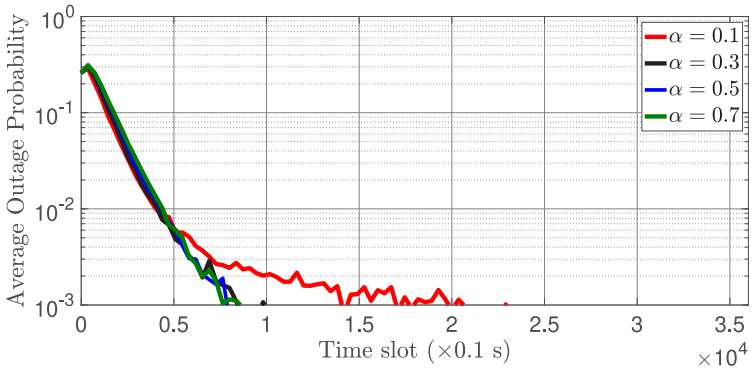
In canonical Q-learning-based RL frameworks, defining a state–action pair for an agent is fundamental for solving an MDP. However, if the state is meaningless, using a stateless approach is preferable [21, 24]. Fig. 2 illustrates the stateless Q-learning-based MARL architecture. There is no state space here, but the agents interact with the environment, which includes the result of all agents' joint actions. More specifically, since each agent's instantaneous SINR depends on all agents' channel selections and current locations, each agent can still adapt to the dynamic environment and learn the best strategy through iterative decision-making with location changes. In particular, such a stateless Q-learning method is known to allow the agent to adapt more effectively to the dynamic environment than standard Q-learning because the Q-value can be quickly changed by the reward value [21,25]. Moreover, this RL method has the advantages of low computational complexity and fast convergence [26].

In our MARL framework, the state for each agent j at time slot t, denoted by $S_j^{(t)}$, can be defined as in [9, (18)] as follows:

$$S_j^{(t)} = \begin{cases} 1, & \text{if } \gamma_j^{(t)} \ge \underline{\gamma}_j, \\ 0, & \text{otherwise}, \end{cases}$$



(a) Average required bandwidth per time slot



(b) Average outage probability per time slot

Fig. 3. Learning results per time slot with different learning rates α.

but the optimal policy of each link is not closely related to a certain state. Specifically, since all agents act simultaneously, the following channel selection of each agent does not depend on whether the current SINR is satisfied (state). From this perspective, we adopt stateless Q-learning for each agent.

3.2.4. Reward

After taking an action, the agent obtains an instantaneous reward from the environment that includes the goal of the learning problem. Recall that the proposed MUSCAT aims to minimize the required bandwidth in the dynamic network while ensuring the QoS requirements represented by the SINR for all links. From this perspective, we define the instantaneous reward function of the environment to agent j at time slot t, denoted by $R_j^{(t)}$, as follows:

$$R_j^{(t)} = \begin{cases} K/(|K/2 - A_j^{(t)}|^{\mu} + K), & \text{if } \gamma_j^{(t)} \geq \underline{\gamma}_j, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where μ is the outage tolerant parameter that controls the trade-off between required bandwidth and link outages, which will be discussed in Section 4. Intuitively, this reward implies minimizing the required bandwidth by encouraging agents to use a channel index close to the center frequency if the SINR constraint is satisfied. It is worth noting that each agent's reward relies only on its current action and SINR without information about other agents, such as their actions, locations, interferences, and rewards [9]. Although the reward depends on the joint strategy, the required information, whether the SINR is satisfied or not, can be obtained through the C2 link. In particular, since each agent knows its action, it can obtain the reward value by receiving binary feedback from its receiver on whether the SINR is satisfied.

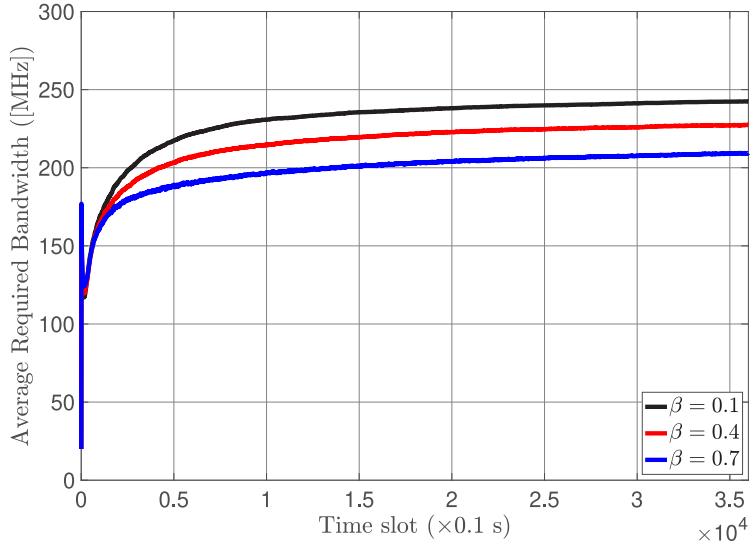
In Algorithm 1, the initial Q-values are set to zero for all agents. At any time slot t, the RX of link j measures the current SINR level $\gamma_j^{(t)}$ and provides feedback to the TX via the C2 link. The observed SINR from the RX of link j at time slot t is given by

$$\gamma_j^{(t)} = r_{j,j}^{(t)} \Bigg/ \left(\sum_{i \in \mathcal{L} \setminus j} r_{j,i}^{(t)} \mathbb{1}_{\{A_j^{(t)} = A_i^{(t)}\}} + N_0 W_j \right), \tag{7}$$

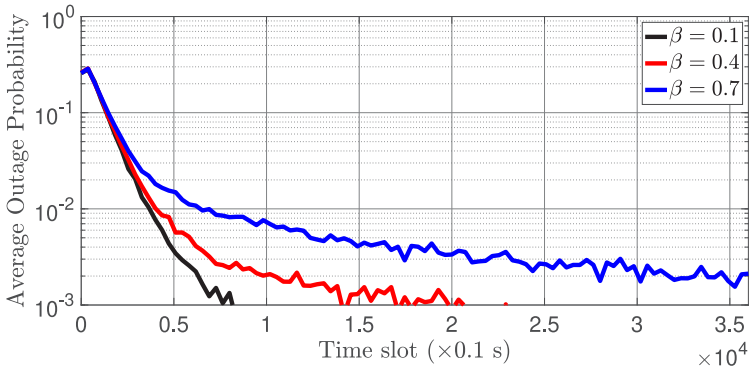
where $\mathbb{1}_{\{\cdot\}}$ is an indicator function. For each time slot, as shown in Fig. 2, each agent takes the selected channel index (action) and the current SINR as the primary inputs and updates its Q-table through (4) and (6). Each agent then selects either a best channel index corresponding to the highest Q-value in the updated Q-table or a random one, according to the ϵ-greedy exploration policy. The chosen channel index is mapped to its new action, which affects the environment, and an instantaneous SINR is fed back to the agent as an output. Note that agents selfishly and rationally execute Algorithm 1, aiming to minimize the required total bandwidth in the network.

3.2.5. Computational complexity

We theoretically analyze the computational complexity of the proposed MUSCAT (Algorithm 1) based on the worst-case complexity analysis framework. Since each link runs the MUSCAT algorithm as an independent agent, the computational complexity can be derived from single-agent Q-learning in each iteration [21]. First, line 4 performs



(a) Average required bandwidth per time slot



(b) Average outage probability per time slot

Fig. 4. Learning results per time slot with different discount factors β.

some constant calculations and one assignment operation, so it has a computational complexity of $\mathcal{O}(1)$. Line 5 requires a computational complexity of $\mathcal{O}(K) = \mathcal{O}(L)$ because K comparison operations are performed when the agent selects an exploitation as an action. Here, $K = L$ corresponds to the worst cases where all links use orthogonal channels to each other. Also, line 7 has $\mathcal{O}(L)$ computational complexity because $L - 1$ comparison, multiplication, and summation operations for interferences in the denominator, one multiplication, addition, and division operations are performed. With the same approach, lines 8 and 9 require computational complexity of $\mathcal{O}(1)$ and $\mathcal{O}(L)$, respectively. Finally, there are L agents, the overall computational complexity of the proposed MUSCAT algorithm in each iteration is $\mathcal{O}(L^2)$. This implies that the proposed MUSCAT has polynomial computational complexity and can be feasible and viable even for large-scale UAV-enabled networks.

4. Simulation results

In this section, we validate the effectiveness of the proposed multi-agent stateless Q-learning-based dynamic distributed MS-CA technique, named MUSCAT, through extensive computer simulations using MATLAB. IBM ILOG CPLEX Optimizer has been employed to solve the optimal MS-CA in the simulations, and no built-in modules have been used for the proposed MUSCAT. Moreover, the simulation server used has the specifications of AMD Ryzen Threadripper 3960X 24-core CPU and 256 GB memory. We consider two scenarios: a static UAV-enabled wireless network where UAVs hover at stationary positions and a dynamic one where UAVs travel according to the mobility model described in Section 2.2. Here, the static network scenario is considered to compare the proposed distributed MS-CA technique with the centralized one that derives an optimal solution. Recall that when the UAVs are stationary, there is enough time to exchange network information and channel allocation results between the central unit and the nodes, so that the centralized optimal MS-CA technique is applicable. Conversely, when the UAVs are moving, the RSS of each link varies instantaneously, making it infeasible to apply the centralized method.

The simulation parameters are summarized in Table 1. Without loss of generality, the same channel bandwidth and SINR threshold are assumed for all links drawn from the required data rate of 50 Mbps for the A2G data link [22].[5] The A2G channel parameters, $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$, $\mathfrak{e}$ and η_ξ, are set by assuming the urban environment [20] and considering the center frequency of 2 GHz [27]. In addition, we set $K = |\mathcal{L}|$ for each simulation, which corresponds to the worst cases where all links use orthogonal channels with each other. With these parameters, we generated 2000 different UAV-enabled wireless network topologies for the same parameters and extracted the average performance.

[5] In this paper, the common channel bandwidth and SINR threshold were assumed for all links, but they can be readily extended to more practical UAV-enabled networks with different values for each link. Specifically, we can reflect the different channel conditions and QoS requirements of each link j by adjusting the channel bandwidth W_j and SINR threshold $\underline{\gamma}_j$ in (3e), (6), and (7).

Table 1
Key system parameters of simulations.

Parameter	Notation	Value
Radius of the area	R	250 m
GCS (BS) density	λ	20~100/km^2
GCS (BS) height	H_{BS}	20 m
UAV operating radius	r	100 m
UAV altitude	H	100~120 m
UAV ascent/descent speed	$v_{\mathrm{v,min}}$, $v_{\mathrm{v,max}}$	5, 10 m/s
UAV cruise speed	$v_{\mathrm{h,min}}$, $v_{\mathrm{h,max}}$	30, 40 m/s
UAV dwell time	$\tau_{\min}$, $\tau_{\max}$	2, 4 s
UAV speed variance	Δv_{h}	−1~+1 m/s
UAV self-rotation angle	$\Delta\varphi$	−5~+5°/s
Transmission powers	P_{UAV}, P_{GCS}	23 dBm (200 mW), 30 dBm (1 W)
Time interval	T_{s}	0.1 s
Carrier frequency	f_c	2 GHz
Channel bandwidth	W	20 MHz
A2G channel constants	$\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$, $\mathfrak{e}$	9.6117, 0.1581, 0.9028, 1.8637
Excessive path losses	$\eta_{\mathrm{LoS}}, \eta_{\mathrm{NLoS}}$	0.7943, 0.0100
Noise power	$N_0 W$	-174 dBm/Hz $+ 10\log_{10} W$ Hz $+ 3$ dB (noise-figure)
SINR threshold	$\underline{\gamma}$	7 dB

Table 2
Ratio of used bandwidth to total bandwidth of MUSCAT ($\mu = 4$) and optimal MS-CA (centralized) in static UAV-enabled networks ([%]).

λ [No. of GCSs/km^2]	20	40	60	80	100
Total BW [MHz]	169.76	313.18	467.62	630.26	786.56
MUSCAT ($\mu = 4$)	53.85	45.82	**42.67**	**41.11**	**41.30**
Optimum ((3a)–(3e))	48.22	40.51	Cannot solve		

Table 3
Ratio of used bandwidth to total bandwidth of MUSCAT and conventional CA (rate maximization) in dynamic UAV-enabled networks ([%]).

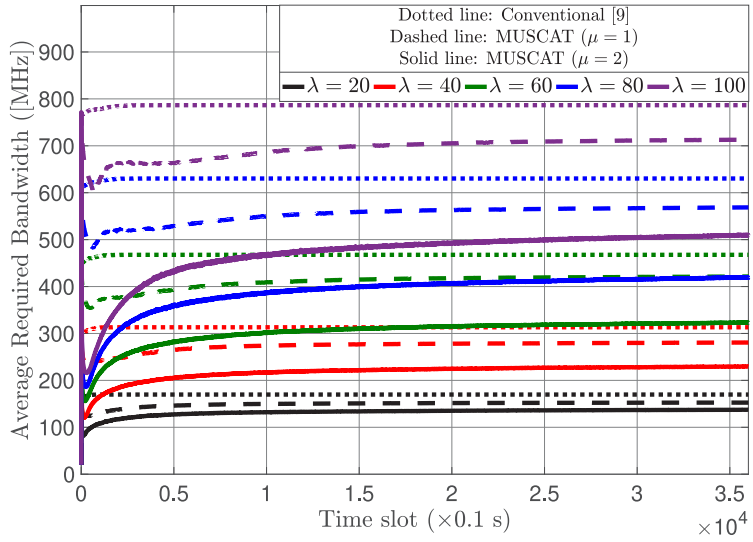
λ [No. of GCSs/km^2]	20	40	60	80	100
Total BW [MHz]	169.76	313.18	467.62	630.26	786.56
Conv. (Rate max.) [9]	100	100	100	100	100
MUSCAT ($\mu = 1$)	89.88	89.61	90.12	90.22	90.63
MUSCAT ($\mu = 2$)	80.90	73.25	69.08	66.50	64.76

We first simulate the MUSCAT algorithm for different MARL hyperparameters in dynamic UAV-enabled wireless networks with $T = 3600$ s to investigate its sensitivity depending on the hyperparameters and obtain the best parameters. Here, we set the GCS density λ, the exploration parameter ζ in (5), and the outage tolerant parameter μ in (6) as $\lambda = 40$, $\zeta = 200$, and $\mu = 2$, respectively. Fig. 3 shows the average required total bandwidth (Fig. 3(a)) and the average outage probability (Fig. 3(b)) over time for different learning rates α ($\in \{0.1, 0.3, 0.5, 0.7\}$) and a discount factor $\beta = 0.4$. Since each link performs the learning process for each time slot, the required bandwidth progressively converges to be as narrow as possible while alleviating the outage probability. Although the outage probability for $\alpha \geq 0.3$ decreases more rapidly than for $\alpha = 0.1$, the required total bandwidth tends to increase as the learning rate α increases. Moreover, Fig. 4 shows the simulation results with different discount factors β ($\in \{0.1, 0.4, 0.7\}$) and a learning rate $\alpha = 0.1$. Similar to Fig. 3, the outage probability decreases over time, and the required total bandwidth gradually converges. In particular, we can observe that the higher the discount factor, the lower the total bandwidth required (Fig. 4(a)), but the higher the outage probability (Fig. 4(b)). Considering such trade-offs, we set $\alpha = 0.1$, $\beta = 0.4$, $\zeta = 200$, and $\epsilon_{\mathrm{init}} = 0.99$ for the MARL parameters of MUSCAT.

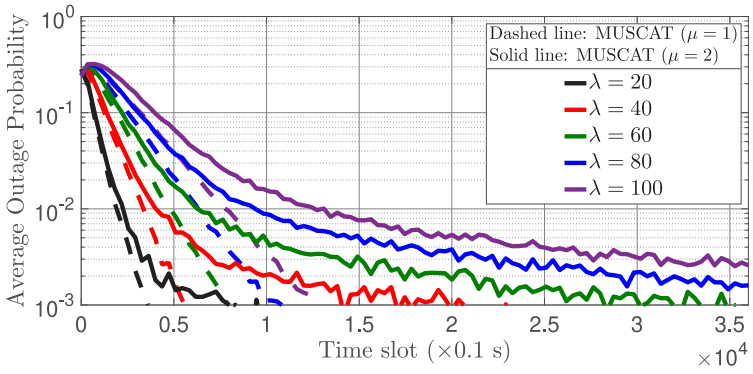
Table 2 shows the required average total bandwidth with respect to the GCS (BS) density λ of the proposed MUSCAT (Algorithm 1) and the centralized optimal MS-CA ((3a)–(3e)) in *static* UAV-enabled wireless networks with $T = 1200$ s. Note that as the GCS density increases, the network density increases, and the complexity of MS-CAP increases accordingly. We observe that the proposed MUSCAT requires slightly more bandwidth than the optimal MS-CA, but it sufficiently minimizes the required bandwidth even though it is in a decentralized manner. In particular, the centralized optimal MS-CA cannot obtain a solution when the GCS density λ is greater than 40 due to its NP-hardness, while the MUSCAT still operates well. This demonstrates the effectiveness of the proposed MARL-based dynamic distributed MS-CA framework.

From now on, we only consider the proposed MUSCAT in dynamic UAV-enabled wireless networks with $T = 3600$ s, as shown in Figs. 3 and 4. To the best of our knowledge, there is no distributed MS-CA technique, so we compared the MUSCAT with a conventional CA scheme (Conv.) that maximizes the network throughput [9] as a benchmark. Fig. 5 shows the average required total bandwidth (Fig. 5(a)) and the average outage probability (Fig. 5(b)) per time slot for various GCS densities λ. And, Table 3 presents the ratio of the converged value of the average required bandwidth in Fig. 5(a) to the given total bandwidth. The conventional CA technique allocates different channels to each link in our setup to maximize network throughput by eliminating inter-link interference. As a result, it consumes all of the given spectrum, as shown in Table 3 and Fig. 5(a). It is noteworthy that most existing resource allocation techniques for UAV-enabled networks have the same results because they consider the same underlying goal of maximizing throughput. Meanwhile, Fig. 5 shows that the MUSCAT incrementally alleviates the average outage probability of the links while appropriately minimizing the required total bandwidth, even in the dynamic topologies where UAVs move. In particular, we can observe that the MUSCAT with $\mu = 2$ results in a significant reduction in the required bandwidth with acceptable outage probability (<0.5%), and the MUSCAT with $\mu = 1$ results in a robust allocation in which the outage probability converges to near zero, although it requires a broader bandwidth. It is expected that the UAVs will be able to operate robustly in their operating area through learned channel allocation results.

Finally, we stated in Section 2.1 that the formulated MS-CAP and the proposed MS-CA technique, MUSCAT, can be directly applied to any wireless network by defining a network topology (graph) consisting of nodes and links. In other words, the proposed MS-CA techniques are not limited to the system model described in Section 2. Fig. 6 illustrates examples of different UAV-enabled wireless network scenarios.



(a) Average required bandwidth per time slot



(b) Average outage probability per time slot

Fig. 5. Learning results per time slot with different BS densities λ.

Specifically, Fig. 6(a) shows two multi-UAV-enabled wireless networks with a centralized architecture for each cell, and Fig. 6(b) shows a UAV ad-hoc network architecture with a GCS [28]. The figures also show the link association between nodes and the channel index assigned to each link according to the proposed MS-CA scheme under the same parameter values as in Table 1. In particular, the reused channel index is marked in green: $\{f_4, f_5, f_6, f_7, f_8, f_9\}$ in 6(a) and $\{f_5, f_6, f_7\}$ in 6(b), respectively. With the same approach, the designed system model and the proposed MS-CA techniques can be readily extended to any wireless network.

5. Conclusion

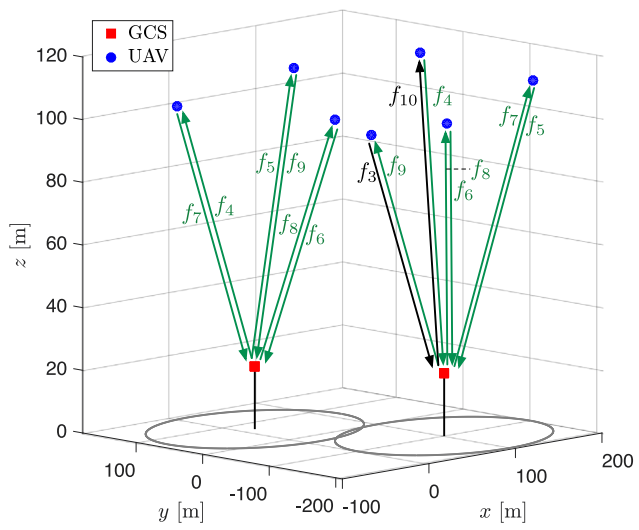
We have revisited the traditional MS-CAP and proposed a novel MS-CA technique for multi-UAV-enabled wireless networks. The inability to exploit global network information due to the dynamic nature of UAV-enabled wireless networks motivated us to design a novel distributed dynamic MS-CA technique called MUSCAT. Specifically, the MS-CAP is to minimize the required bandwidth in the network while guaranteeing the QoS requirements of links. We have formulated this MS-CAP as MDPs for all links and solved them through a distributed multi-agent stateless Q-learning framework in which each link learns its best strategy independently without any observation of the other links. It is worth noting that the links selfishly and rationally allocate their channel in a dynamic environment with the common goal of minimizing the total bandwidth required in the network. Simulation results revealed that the proposed MUSCAT can effectively reduce the required bandwidth while drastically reducing the outage probability of the links, even in multi-UAV-enabled dynamic wireless networks. As a further study, we will jointly optimize the resource allocation, user association, and UAVs' trajectories based on distributed deep reinforcement learning algorithms to improve the objective of MS-CAP.

CRediT authorship contribution statement

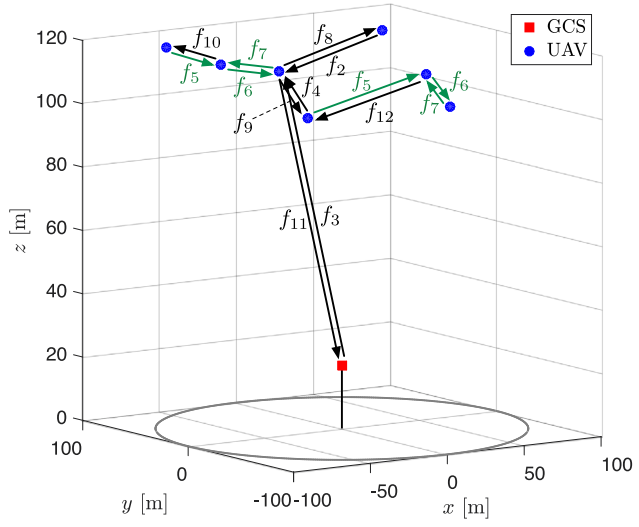
Ki-Hun Lee: Writing – original draft, Software, Methodology, Investigation, Formal analysis. **Seungmin Lee:** Software, Methodology, Investigation, Formal analysis. **Jaedon Park:** Project administration, Funding acquisition, Conceptualization. **Howon Lee:** Writing – review & editing, Validation, Supervision, Resources, Investigation, Conceptualization. **Bang Chul Jung:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

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.



(a) A centralized architecture for each cell



(b) A UAV ad-hoc network architecture with a GCS

Fig. 6. Examples of different multi-UAV-enabled wireless network architectures with channel indices f_k allocated to the links through the proposed MUSCAT. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Data availability

No data was used for the research described in the article.

Acknowledgments

This research was financially supported by the Institute of Civil Military Technology Cooperation funded by the Defense Acquisition Program Administration (DAPA) and Ministry of National Defense (MND) of Korean Government under Grant 22-CM-TN-39.

References

[1] J. Liu, Y. Shi, Z.M. Fadlullah, N. Kato, Space-air-ground integrated network: a survey, IEEE Commun. Surv. Tutor. 20 (4) (2018) 2714–2741, http://dx.doi.org/10.1109/COMST.2018.2841996.

[2] M. Giordani, M. Zorzi, Non-terrestrial networks in the 6G era: challenges and opportunities, IEEE Netw. 35 (2) (2021) 244–251, http://dx.doi.org/10.1109/MNET.011.2000493.

[3] Z. Ullah, F. Al-Turjman, U. Moatasim, L. Mostarda, R. Gagliardi, UAVs joint optimization problems and machine learning to improve the 5G and beyond communication, Comput. Netw. 182 (2020) 107478, http://dx.doi.org/10.1016/j.comnet.2020.107478.

[4] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, M. Zorzi, Toward 6G networks: use cases and technologies, IEEE Commun. Mag. 58 (3) (2020) 55–61, http://dx.doi.org/10.1109/MCOM.001.1900411.

[5] D. Liu, Y. Xu, J. Wang, J. Chen, K. Yao, et al., Opportunistic UAV utilization in wireless networks: motivations, applications, and challenges, IEEE Commun. Mag. 58 (5) (2020) 62–68, http://dx.doi.org/10.1109/MCOM.001.1900687.

[6] M.-A. Lahmeri, M.A. Kishk, M.-S. Alouini, Artificial intelligence for UAV-enabled wireless networks: a survey, IEEE Open J. Commun. Soc. 2 (2021) 1015–1040, http://dx.doi.org/10.1109/OJCOMS.2021.3075201.

[7] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, M. Debbah, A tutorial on UAVs for wireless networks: applications, challenges, and open problems, IEEE Commun. Surv. Tutor. 21 (3) (2019) 2334–2360, http://dx.doi.org/10.1109/COMST.2019.2902862.

[8] N. Cheng, S. Wu, X. Wang, Z. Yin, C. Li, et al., AI for UAV-assisted IoT applications: a comprehensive review, IEEE Int. Things J. 10 (16) (2023) 14438–14461, http://dx.doi.org/10.1109/JIOT.2023.3268316.

[9] J. Cui, Y. Liu, A. Nallanathan, Multi-agent reinforcement learning-based resource allocation for UAV networks, IEEE Trans. Wireless Commun. 19 (2) (2020) 729–743, http://dx.doi.org/10.1109/TWC.2019.2935201.

[10] P. Luong, F. Gagnon, L.-N. Tran, F. Labeau, Deep reinforcement learning-based resource allocation in cooperative UAV-assisted wireless networks, IEEE Trans. Wireless Commun. 20 (11) (2021) 7610–7625, http://dx.doi.org/10.1109/TWC.2021.3086503.

[11] S. Yin, F.R. Yu, Resource allocation and trajectory design in UAV-aided cellular networks based on multiagent reinforcement learning, IEEE Int. Things J. 9 (4) (2022) 2933–2943, http://dx.doi.org/10.1109/JIOT.2021.3094651.

[12] Z. Chang, H. Deng, L. You, G. Min, S. Garg, G. Kaddoum, Trajectory design and resource allocation for multi-UAV networks: deep reinforcement learning approaches, IEEE Trans. Netw. Sci. Eng. 10 (5) (2023) 2940–2951, http://dx.doi.org/10.1109/TNSE.2022.3171600.

[13] K.I. Aardal, S.P.M. van Hoesel, A.M.C.A. Koster, C. Mannino, A. Sassano, Models and solution techniques for frequency assignment problems, Ann. Oper. Res. 153 (2007) 79–129, http://dx.doi.org/10.1007/s10479-007-0178-0.

[14] Z.Á. Mann, A. Szajkó, Complexity of different ILP models of the frequency assignment problem, in: Z.Á. Mann (Ed.), Linear Programming: New Frontiers in Theory and Applications, Nova Science Pub Inc., UK, 2012, pp. 305–326.

[15] M. Thakur, Y. Kortesniemi, D. Lagutin, Streamlining 5G spectrum leasing, IEEE Access 11 (2023) 136179–136194, http://dx.doi.org/10.1109/ACCESS.2023.3337880.

[16] R.I. Ansari, N. Ashraf, S.A. Hassan, D.G.C.H. Pervaiz, C. Politis, Spectrum on demand: A competitive open market model for spectrum sharing for UAV-assisted communications, IEEE Netw. 34 (6) (2020) 318–324, http://dx.doi.org/10.1109/MNET.011.2000253.

[17] J. Wildman, S. Weber, On protocol and physical interference models in Poisson wireless networks, IEEE Trans. Wireless Commun. 17 (2) (2018) 808–821, http://dx.doi.org/10.1109/TWC.2017.2771773.

[18] P.K. Sharma, D.I. Kim, Random 3D mobile UAV networks: mobility modeling and coverage probability, IEEE Trans. Wireless Commun. 18 (5) (2019) 2527–2538, http://dx.doi.org/10.1109/TWC.2019.2904564.

[19] N.T.T. Docomo, White Paper on 5G Channel Model for Bands Up To 100 GHz, Tech. Rep., 2016, http://www.5gworkshops.com/5GCMSIG_White%20Paper_r2dot3.pdf (Accessed 20 March 2024).

[20] A. Al-Hourani, S. Kandeepan, S. Lardner, Optimal LAP altitude for maximum coverage, IEEE Wireless Commun. Lett. 3 (6) (2014) 569–572, http://dx.doi.org/10.1109/LWC.2014.2342736.

[21] S. Lv, X. Xu, S. Han, X. Tao, P. Zhang, Energy-efficient secure short-packet transmission in NOMA-assisted mMTC networks with relaying, IEEE Trans. Veh. Technol. 71 (2) (2022) 1699–1712, http://dx.doi.org/10.1109/TVT.2021.3133907.

[22] 3GPP TR 36.777, Enhanced LTE support for aerial vehicles, Release 15 2018.

[23] R. Panigrahy, K. Talwar, L. Uyeda, U. Wieder, Heuristics for Vector Bin Packing. Microsoft Research, Tech. Rep., 2011, https://www.microsoft.com/en-us/research/wp-content/uploads/2011/01/VBPackingESA11.pdf (Accessed 20 March 2024).

[24] S. Barrachina-Muñoz, A. Chiumento, B. Bellalta, Multi-armed bandits for spectrum allocation in multi-agent channel bonding WLANs, IEEE Access 9 (2021) 133472–133490, http://dx.doi.org/10.1109/ACCESS.2021.3114430.

[25] T. Jiang, Q. Zhao, D. Grace, A.G. Burr, T. Clarke, Single-state Q-learning for self-organised radio resource management in dual-hop 5G high capacity density networks, Trans. Emerg. Telecommun. Technol. 27 (12) (2016) 1628–1640, http://dx.doi.org/10.1002/ett.3019.

[26] Y. Zhang, R.W. Heath, Multi-armed bandit for link configuration in millimeter-wave networks: An approach for solving sequential decision-making problems, IEEE Veh. Technol. Mag. 18 (2) (2023) 39–49, http://dx.doi.org/10.1109/MVT.2023.3237940.

[27] Iskandar, S. Shimamoto, Channel characterization and performance evaluation of mobile communication employing stratospheric platforms, IEICE Trans. Commun. E89-B (3) (2006) 937–944, http://dx.doi.org/10.1093/ietcom/e89-b.3.937.

[28] M.A. Khan, I.M. Qureshi, F. Khanzada, A hybrid communication scheme for efficient and low-cost deployment of future flying ad-hoc network (FANET), Drones 3 (1) (2019) 16, http://dx.doi.org/10.3390/drones3010016.

Ki-Hun Lee received the B.S. degree in electronics engineering from Chungnam National University, Daejeon, South Korea, in 2018, where he is currently pursuing the Ph.D. candidate with the Department of Electronics Engineering.

He has experience as a Visiting Scholar with the William Paterson University of New Jersey (WPUNJ), Wayne, NJ, USA, in 2023. His research interests include wireless communications, radio network planning and optimization, statistical signal processing, information theory, interference management, spectrum-sharing techniques, and machine learning.

Mr. Lee was a recipient of the Best Paper Award at the Korean Institute of Communications and Information Sciences (KICS) Fall Conference in 2018; the Best Paper Award at the KICS Summer Conference in 2019; the Best Paper Award at the KICS Summer Conference in 2021; the Journal of KICS (J-KICS) Best Paper Award in 2022; and the Best Paper Award at the KICS Summer Conference in 2023. He was the Web Chair of IEEE CCNC 2023.

Seungmin Lee received the B.S. and M.S. degrees from the School of Electronic and Electrical Engineering, Hankyong National University, Anseong, South Korea, in 2021 and 2023, respectively.

He is currently an Engineer with Newratek, Inc., Seoul, South Korea. His current research interests include next generation 802.11 networks, wireless communications, machine learning for wireless communications.

Mr. Lee was a recipient of the Best Paper Award at KICS 2020 Winter Conference in 2020.

Jaedon Park received the B.S. degree in electronics engineering from Hanyang University, Seoul, South Korea, in 2000, and the M.S. and Ph.D. degrees from the School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2002 and 2016, respectively.

He is currently a Senior Researcher with the Agency for Defense Development (ADD), Daejeon. His research interests include electromagnetic spectrum operations (EMSO) technology, MIMO and relay systems, and FSO systems.

Howon Lee received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2003, 2005, and 2009, respectively.

From 2009 to 2012, he was a Senior Research Staff/Team Leader of the Knowledge Convergence Team, KAIST Institute for Information Technology Convergence (KI-ITC). From 2012 to 2024, he was with the School of Electronic and Electrical Engineering and the Institute for IT Convergence (IITC), Hankyong National University (HKNU), Anseong, South Korea. Since 2024, he has been with the Department of Electrical and Computer Engineering, Ajou University, Suwon, South Korea. He has also experienced as a Visiting Scholar with the University of California at San Diego (UCSD), La Jolla, CA, USA, in 2018. His current research interests include B5G/6G wireless communications, ultra-dense distributed networks, innetwork computations for 3D images, cross-layer radio resource management, reinforcement learning for UAV networks, unsupervised learning for wireless communication networks, and the Internet of Things.

Prof. Lee was a recipient of the 2006 Joint Conference on Communications and Information (JCCI) Best Paper Award and the Bronze Prize at the Intel Student Paper Contest in 2006. He was a recipient of the Telecommunications Technology Association (TTA) Paper Contest Encouragement Award in 2011; the Best Paper Award at the Korean Institute of Communications and Information Sciences (KICS) Summer Conference in 2015; the Best Paper Award at the KICS Fall Conference in 2015; the Honorable Achievement Award from 5G Forum Korea in 2016; the Best Paper Award at the KICS Summer Conference in 2017; the Best Paper Award at the KICS Winter Conference in 2018; the Best Paper Award at the KICS Summer Conference in 2018; the Best Paper Award at the KICS Winter Conference in 2020; and the Best Paper Award at the KICS Winter Conference in 2022. He received the Minister's Commendation by the Minister of Science and ICT in 2017. He is the TPC Chair of IEEE CCNC 2025.

Bang Chul Jung received the B.S. degree in electronics engineering from Ajou University, Suwon, South Korea, in 2002, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Korea Advanced Institute for Science and Technology (KAIST), Daejeon, South Korea, in 2004 and 2008, respectively.

He was a Senior Researcher/Research Professor with the KAIST Institute for Information Technology Convergence, Daejeon, from 2009 to 2010. From 2010 to 2015, he was a Faculty Member of Gyeongsang National University, Tongyeong, South Korea. He is currently a Professor with the Department of Electronics Engineering, Chungnam National University, Daejeon, South Korea. His research interests include wireless communication systems, Internet-of-Things (IoT) communications, statistical signal processing, information theory, interference management, radio resource management, spectrum-sharing techniques, and machine learning.

Prof. Jung was a recipient of the Fifth IEEE Communication Society Asia Pacific Outstanding Young Researcher Award in 2011, the Bronze Prize of Intel Student Paper Contest in 2005, the First Prize of KAISTs Invention Idea Contest in 2008, and the Bronze Prize of Samsung Humantech Paper Contest in 2009. He has been selected as a winner of Haedong Young Scholar Award in 2015, which is sponsored by the Haedong Foundation and given by KICS. He has been selected as a winner of the 29th Science and Technology Best Paper Award in 2019, which is sponsored by the Korean Federation of Science and Technology Societies. He was the Associate Editor of IEEE Vehicular Technology Magazine from 2020 to 2022, and is now the Senior Editor of IEEE Vehicular Technology Magazine. He was the TPC Chair of IEEE CCNC 2023, and is the General Chair of IEEE CCNC 2025.