# Multiscale Adaptive Scheduling and Path-Planning for Power-Constrained UAV-Relays via SMDPs

Bharath Keshavamurthy *Student Member, IEEE* and Nicolò Michelusi *Senior Member, IEEE*

*Abstract*—We describe the orchestration of a decentralized swarm of rotary-wing UAV-relays, augmenting the coverage and service capabilities of a terrestrial base station. Our goal is to minimize the time-average service latencies involved in handling transmission requests from ground users under Poisson arrivals, subject to an average UAV power constraint. Equipped with rate adaptation to efficiently leverage air-to-ground stochastics, we first derive the optimal control policy for a single relay via a semi-Markov decision process formulation, with competitive swarm optimization for UAV trajectory design. Accordingly, we detail a multiscale decomposition of this construction: outer decisions on radial wait velocities and end positions optimize the expected long-term delay-power trade-off; consequently, inner decisions on angular wait velocities, service schedules, and UAV trajectories greedily minimize the instantaneous delay-power costs. Next, generalizing to UAV swarms via replication and consensus-driven command-and-control, this policy is embedded with spread maximization and conflict resolution heuristics. We demonstrate that our framework offers superior performance vis-à-vis average service latencies and average per-UAV power consumption: $12\times$ faster data payload delivery relative to static UAV-relay deployments and $3\times$ faster than a deep-Q network solution; remarkably, $1$ relay with our scheme outclasses $3$ relays under a joint successive convex approximation policy by $80\%$.

*Index Terms*—UAV-relays, Rate Adaptation, SMDP, CSO

## I. INTRODUCTION

With sustained device proliferation, enterprises across sectors have stepped-up their adoption of Unmanned Aerial Vehicles (UAVs) to gather data, survey infrastructure, monitor operations, and automate logistics [1]. Inevitably, this has fostered varied academic research and industrial R&D on drone-augmented beyond line-of-sight connectivity and traffic offloading in cellular networks: the coverage and service capabilities of an extant terrestrial radio access network are enhanced by the mobility and maneuverability of these autonomous aerial relays [2]. Unsurprisingly, the pervasive potential of such hybrid networks brings along a plethora of challenges in real-world deployments [2]: specifically, on-board energy constraints of these aerial platforms impacting mission times, stringent Quality-of-Service (QoS) mandates for reliable connectivity, channel characteristics of Air-to-Ground (A2G) links in highly-mobile settings, and computational feasibility challenges in trajectory design brought on by the inherently large state and action spaces. Ergo, several works in the state-of-the-art have tried to tackle these challenges using tools from optimization theory, machine learning, and reinforcement learning—however, various problems remain unsolved and various challenges are left unaddressed.

**Related Work**: Perusing single UAV-relay formulations in current literature, we observe non-adaptive schemes [3]–[5] designed for applications where the IoT devices possess local storage or aggregation capabilities allowing for deterministic

arrivals of data packets. Yet, practical deployments involve dynamically-generated traffic from miscellaneous sets of users, each with varying degrees of QoS mandates and technological prowess. Unlike these works, we consider dynamic traffic generation from random deployments of ground users, thereby constructing a control strategy that is receptive to uncertain system dynamics. Furthermore, these works solve for the optimal service schedules and associated trajectories via Successive Convex Approximation (SCA) [3]–[5], which apart from being computationally infeasible to accommodate dynamic traffic due to prohibitively large convergence times, relies on first-order Taylor approximations of the optimization problem to enforce convexity, thereby introducing inaccuracies into the model. On another note, these works employ Free Space Path-Loss (FSPL) models that fail to account for the A2G channel characteristics inherent in UAV-assisted wireless networks; moreover, their approximations in the traffic delivery constraint preclude the adoption of rate adaptation which allows the transmitters in the network to leverage channel stochastics to maximize throughput. In this paper, in addition to accurately modeling A2G channel characteristics and employing rate adaptation at all the transmitters to efficiently exploit said characteristics, there are no such underlying approximations.

Pivoting to the path-planning problem for a single relay, a Competitive Swarm Optimization (CSO) [6] approach is proposed in this paper to bypass the computational infeasibility seen in [3]–[5]. Unlike SCA, which employs approximations to enforce convexity, CSO does not depend on the specific problem structure to work effectively. Contrary to the limited update scope of Particle Swarm Optimization (PSO) [7], CSO exhibits superior performance on large-scale optimization benchmarks [6], since it involves more efficient updates wherein pair-wise competition is invoked between particles— permitting the winners to advance and the loser particles to learn from the winners. Unreasonably, works that employ PSO, either optimize static hovering positions only [8], or impose impractical path and velocity restrictions [4], [9].

Next, shifting our attention to swarm orchestration frameworks, we find inefficient solutions such as centralized deployments [10], [11] in which an aggregation center coordinates the operations of the UAV-relays; or either joint multi-relay optimization methods [11], [12] or model-free formulations consisting of combined state and action spaces [13], [14]. Centralized swarm deployments bring in the need for additional capital and operational expenditure; and joint multi-UAV constructions lead to prohibitively large solution spaces resulting in unnecessary overhead in policy convergence times, which when scaled to larger swarms result in intractability. To mitigate these drawbacks, we present an orchestration framework suitable for decentralized UAV-relay swarms by embedding our single UAV-relay policy with multi-agent heuristics and replicating it across the swarm.

A previous version of this research [15] focused only on single UAV-relay deployments, assumed an FSPL channel model, and employed PSO for trajectory optimization: the challenges associated with these have been discussed above.

**Novelties** In this paper, with rate adaptation to exploit A2G channel stochastics, we first constrain our study to single relay settings, wherein the problem of minimizing the time-average service delay subject to an average UAV power constraint, is formulated as a Semi-Markov Decision Process (SMDP). We derive a multiscale decomposition to this formulation: optimizing the long-term delay-power costs yields outer decisions on radial wait velocities and service positions (via value iteration); consequently, greedily minimizing the instantaneous delay-power costs yields inner actions on angular wait velocities (via exhaustive search) and service trajectories (via competitive swarm optimization). Post single relay policy convergence, with an overlaid command-and-control network, we supplement this control strategy with multi-agent heuristics—namely, spread maximization and consensus-driven conflict resolution—and replicate it across the swarm.

**Extensions**: Further developments to this research include M/G/$x$ queue management heuristics for link-layer prescient scheduling, a hierarchical variant of CSO to facilitate efficient scalability to higher-dimensional trajectory design, and viability analyses via emulations and real-world flight-tests on NSF AERPAW (OFDM PHY radio + MAVLink vehicle control).

The rest of the paper is structured as follows: Sec. II outlines the system model; Sec. III elucidates our SMDP formulation for single relay settings; Sec. IV details our swarm extensions; Sec. V chronicles our numerical evaluations; Sec. VI lists our conclusions.

## II. System Model

In this section, we model the processes involved in the non-terrestrial augmentation of conventional radio ecosystems, aided in their coverage and service capabilities by UAV-relays.

**Deployment Model**: Consider a generalized deployment in which a swarm of $N_U$ rotary-wing Unmanned Aerial Vehicles (UAVs)—each equipped with an on-board transceiver chain—operate as cellular relays to supplement the coverage and service capabilities of a terrestrial Base Station (BS) by relaying data traffic dynamically-generated by Ground Nodes (GNs). The BS is located at the center of the circular cell (of radius $a$), at height $H_B$, while the UAVs operate at a fixed height of $H_U$. The GNs are distributed uniformly at random throughout the cell, with a density of $\lambda_G$ [GNs per unit area]. The BS utilizes $k$ orthogonal channels to serve the GNs simultaneously via an Orthogonal Frequency Division Multiple Access (OFDMA) strategy; on the other hand, the UAV-relays are restricted to serve one GN at a time through a decode-and-forward scheme. All channels are assumed to have a bandwidth of $B$. Without any loss of generality, studying uplink transmissions only, these GNs generate random data traffic that is to be transmitted to the BS, either directly or by using one of the UAV-relays.

**Communication Model**: Each GN generates uplink transmission requests of $L$ bits, according to a Poisson process with rate $\lambda_{R|G}$ [requests per GN per unit time]. Coupled with the random deployment of GNs, uplink requests arrive in time according to a Poisson process with rate $\lambda_R = \lambda_G \cdot \lambda_{R|G}$

[requests per unit time per unit area]. Thus, $\Lambda \triangleq \lambda_R \pi a^2$ [requests per unit time] is the overall request arrival rate over the circular cell. Since a new request is uniformly distributed in the cell area, its angular coordinate $\theta$ is uniform in $[0, 2\pi)$ and the probability density function of its radial coordinate is $f_R(r) = \frac{2r}{a^2} \mathbb{I}(r \leq a)$, where $\mathbb{I}(\cdot)$ is the indicator function. Allocating the band-edges of the spectrum under use as control channels, a fully-connected mesh network is overlaid on the BS and the UAVs to establish a command-and-control network. Since the packets exchanged among the mesh nodes over the control channel constitute short frames relative to the large data payloads generated by the GNs (and communicated over orthogonal data channels), it is reasonable to neglect the latencies involved in these control operations. When a GN decides to upload its data, it informs the BS—over the control channel—about the need for an uplink transmission of $L$ bits, and includes its physical location in this preliminary request for service. Considering potential delay-power costs for this request, the BS and the UAVs coordinate over the control network to arrive at a consensus on the best scheduling decision: if direct transmission is chosen, the BS assigns a data channel $k \in \{1, 2, \ldots, N_B\}$ to the GN and instructs it to begin transmission; else, if relaying the data payload through UAV $i$ is determined to be the most efficient choice, the UAV instructs the GN to begin transmission over its designated pre-determined data channel $k_{U_i} \in \{1, 2, \ldots, N_U\}$. A Decode-and-Forward (D&F) strategy underlies the communication process encountered in the latter case: while moving along a designed trajectory (a sequence of way-points and velocities), UAV $i$ first receives the entire data payload from the GN over channel $k_{U_i}$ (decode) and subsequently transmits it to the BS over the same channel (forward). Crucially, inherent in these scheduling decisions is the A2G channel model underlying the GN→BS, GN→UAV, and the UAV→BS links.

**A2G Channel Model**: For a generic link, we denote the flat-fading channel coefficient as $h \triangleq \sqrt{\beta} g$, where $\beta$ captures the large-scale channel variations, and $g$ with $\mathbb{E}[|g|^2] = 1$ is the small-scale fading component. We model the large-scale component as $\beta = \beta_{\text{LoS}}(d) \triangleq \beta_0 d^{-\alpha}$ for line-of-sight (LoS) and $\beta = \beta_{\text{NLoS}}(d) \triangleq \kappa \beta_0 d^{-\tilde{\alpha}}$ for non-LoS (NLoS) links, where $\beta_0$ is the pathloss referenced at a distance of 1 m, $2 \leq \alpha \leq \tilde{\alpha}$ are the LoS and NLoS path-loss exponents, $\kappa \in (0, 1]$ captures the additional NLoS attenuation, and $d$ is the Euclidean distance between the transmitter (Tx) and the receiver (Rx) [3]. We model the LoS and NLoS probability as a function of the elevation angle $\varphi \in (0, 90^o]$ as

$$P_{\text{LoS}}(\varphi) = \frac{1}{1 + z_1 e^{-z_2[\varphi - z_1]}}; P_{\text{NLoS}}(\varphi) = 1 - P_{\text{LoS}}(\varphi), \quad (1)$$

where $z_1$ and $z_2$ are environment-specific parameters. The distribution of the small-scale fading component $g$ also depends on the LoS or NLoS link state—specifically, for LoS, we model $g$ as Rician fading with a $\varphi$-dependent $K$-factor, i.e., $K(\varphi) = k_1 \exp(k_2 \varphi)$, where coefficients $k_1$ and $k_2$ are determined by the propagation environment [5]; for NLoS, we model $g$ as Rayleigh fading (Rician with $K = 0$) [5]. Given $h$, the link capacity is $C(h) = B \cdot \log_2\left(1 + \frac{|h|^2 P_T}{N_0 B \Gamma}\right)$, where $P_T$ is the transmission power, $N_0$ is the noise power spectral density at the receiver, $B$ is the channel bandwidth, and

$\Gamma$ is the Signal-to-Noise Ratio (SNR) gap between practical modulation-and-coding schemes and theoretical Gaussian signaling [5]. We assume that other sources of signal degradation, such as the Doppler effect, are well-compensated at the receiver. Since the large-scale components typically vary slowly relative to the rate of acquisition of Channel State Information (CSI), we assume that the current large-scale parameters $(\beta, K)$ are known at the transmitter's side throughout the communication process, which enables rate control at the transmitter; on the other hand, small-scale fading conditions vary on a much faster timescale, hence cannot be tracked at the transmitter, which may result in outages when the selected rate exceeds the channel capacity $C(h)$. Thus, given $(\beta, K)$ and a transmission rate of $\Upsilon$ [bits per second], we define the outage probability $P_{\text{out}}(\Upsilon, \beta, K) \triangleq \mathbb{P}(C(\sqrt{\beta}g) < \Upsilon | \beta, K) = \mathbb{P}(|g|^2 < u(\Upsilon, \beta))$, where $u(\Upsilon, \beta) \triangleq N_0 B \Gamma (2^{\Upsilon/B} - 1)/(\beta P)$. Since $2(K+1)|g|^2$ has a non-central $\chi^2$ distribution with 2 degrees of freedom and a non-centrality parameter $2K$, we find

$$P_{\text{out}}(\Upsilon, \beta, K) = 1 - Q_1\left(\sqrt{2K}, \sqrt{2(K+1)u(\Upsilon, \beta)}\right), \quad (2)$$

where $Q_1(\cdot, \cdot)$ is the standard Marcum $Q$-function [5]. Note that when $K=0$ (Rayleigh fading NLoS link), the function specializes to $Q_1\left(0, \sqrt{2u(\Upsilon, \beta)}\right) = \exp(-u(\Upsilon, \beta))$. We assume that the small-scale fading is averaged out across time and space, yielding the expected throughput

$$R(\Upsilon, \beta, K) = \Upsilon \cdot Q_1\left(\sqrt{2K}, \sqrt{2(K+1)u(\Upsilon, \beta)}\right). \quad (3)$$

In our model, we permit rate adaptation at the transmitter based on the large-scale parameters $(\beta, K)$, coordinated through the control channel via CSI feedback. The transmission rate $\Upsilon$ is chosen to maximize the expected throughput given $(\beta, K)$, i.e., $\Upsilon^*(\beta, K) \triangleq \arg\max_{\Upsilon \geq 0} R(\Upsilon, \beta, K)$. Let $Z \triangleq \sqrt{\frac{2\beta P_T}{N_0 B \Gamma} u(\Upsilon, \beta)}$, so $\Upsilon = B \log_2\left(1 + \frac{1}{2}Z^2\right) \triangleq f(Z)$,

$$\Upsilon^*(\beta, K) = f(Z^*(\beta, K)), \quad Z^*(\beta, K) \triangleq \arg\min_{Z \geq 0} g(Z),$$

$$g(Z) \triangleq -\ln f(Z) - \ln Q_1\left(\sqrt{2K}, \sqrt{\frac{(K+1)\beta P}{N_0 B \Gamma}} Z\right). \quad (4)$$

Since the function $g(Z)$ is convex, $Z^*(\beta, K)$ can be found efficiently using a bisection method. Upon determining the optimal transmission rate $\Upsilon^*(\beta, K)$, we define the optimized throughput, as a function of the large-scale conditions, as $R^*(\beta, K) \triangleq R(\Upsilon^*(\beta, K), \beta, K)$. Assuming that the LoS and NLoS conditions are averaged out in the temporal and spatial dimensions, we compute the average link throughput coupled with rate adaptation as

$$\bar{R}(d, \varphi) \triangleq P_{\text{LoS}}(\varphi) \cdot R^*(\beta_{\text{LoS}}(d), K(\varphi)) +$$
$$P_{\text{NLoS}}(\varphi) \cdot R^*(\beta_{\text{NLoS}}(d), 0), \quad (5)$$

which is then specialized to the three distinct communication links by expressing the transmission powers, the environment-specific parameters $z_1$, $z_2$, $k_1$, and $k_2$, the large-scale parameters $(\beta, K)$, and the LoS or NLoS probabilities (1) based on the spatial configuration, i.e., Tx-Rx distance and
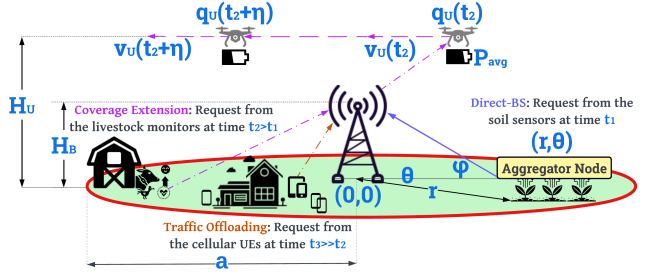


Fig. 1: Single UAV-relay specialization of our generalized deployment setting.

elevation angle. Specifically, for the GN→BS link, we let $\bar{R}_{GB}(r)$ be the throughput with the GN in position $(r, \theta)$, computed by setting the GN-BS distance as $d = \sqrt{H_B^2 + r^2}$ and the elevation angle as $\varphi = \sin^{-1}(H_B/d)$ in (5). Similarly, for the GN→UAV link, we let $\bar{R}_{GU}(r_{GU})$ be the throughput when the GN-UAV distance (projected onto the $x-y$ plane) is $r_{GU}$, computed by setting the GN-UAV Euclidean distance as $d = \sqrt{r_{GU}^2 + H_U^2}$ and the elevation angle as $\varphi = \sin^{-1}(H_U/d)$ in (5). Finally, for the UAV→BS link, we let $\bar{R}_{UB}(r_{UB})$ be the throughput when the $x-y$ projected UAV-BS distance is $r_{UB}$, computed by setting the GN-UAV Euclidean distance as $d = \sqrt{r_{UB}^2 + (H_U - H_B)^2}$ and $\varphi = \sin^{-1}\left(\frac{(H_U - H_B)}{d}\right)$ in (5).

## III. THE SMDP FORMULATION

In this section, we specialize the generalized deployment, communication, and channel models detailed in Sec. II to single UAV-relay settings. Accordingly, we describe the mathematical constructions involved in the design of our solution framework to minimize the time-average service delay experienced by the GNs in the cell, subject to an average UAV mobility power constraint via a Semi-Markov Decision Process (SMDP) formulation. The effective traffic rate experienced by a single UAV is $\frac{\Lambda}{N_U}$ [requests per unit time], which is assumed in this section in place of the overall rate $\Lambda$. Let $\mathbf{q}_U(t) = (r_U(t), \theta_U(t))$ be the polar coordinate of the UAV at time $t$, projected onto the $x-y$ plane, where $r_U(t) \in \mathbb{R}_+$ and $\theta_U(t) \in [0, 2\pi)$ denote the UAV's radius and angle with respect to the BS (cell center). This setup is depicted in Fig. 1.

We note that the operations of the UAV-relay can be split into the following phases. In the waiting phase, no GN requests are being served by the UAV, which thus moves according to a waiting policy, until a new request is received. When a new GN request is received, say from position $(r, \theta)$, the system transitions to the request scheduling phase, where the system decides if the GN should transmit its data payload directly to the BS, or relay it through the UAV. If direct transmission is selected, the system immediately re-enters the waiting phase, as the UAV remains free to serve other requests; else, the system enters the UAV-relay phase, in which the GN relays its data payload through the UAV using the D&F protocol; upon the completion of this relay service, the system re-enters the waiting phase. Note that the BS can accommodate simultaneous transmissions (see Sec. II): new requests received during the UAV-relay phase are directly served by the BS. Under a given policy $\mu$, averaged out over $M_t$ decision time-slots, we denote the expected service delay for scheduled requests as $\bar{W}_\mu^{(s)}$, the expected UAV energy expenditure as $\bar{E}_\mu$, and the expected average operational duration as $\bar{T}_\mu$. With this setup, the optimization problem given by $\min_\mu \bar{W}_\mu^{(s)}$ s.t.

$\bar{E}_\mu - P_{\text{avg}}\bar{T}_\mu \leq 0$, will be the focus of the our analyses in this section. To solve it, consider

$$g(\nu) = \min_\mu \bar{W}_\mu^{(s)} + \nu(\bar{E}_\mu - P_{\text{avg}}\bar{T}_\mu), \qquad (6)$$

where $\nu$ is the dual variable, optimized by solving $\max_{\nu \geq 0} g(\nu)$. We now demonstrate that for a given $\nu \geq 0$, (6) can be cast as an SMDP and solved via dynamic programming. **States**: The state is defined by the UAV position $\mathbf{q}_U$, taking value from the set $\mathcal{Q}_{\text{UAV}} \triangleq \mathbb{R}_+ \times [0, 2\pi)$ (in polar coordinates), and the position of an uplink transmission request $\mathbf{q}_G$, taking values from $\mathcal{Q}_{\text{GN}} \triangleq [0, a] \times [0, 2\pi)$ (polar coordinates). The state space is then $\mathcal{S} = \mathcal{S}_{\text{wait}} \cup \mathcal{S}_{\text{comm}}$, where $\mathcal{S}_{\text{wait}} = \mathcal{Q}_{\text{UAV}}$ is the set of waiting states and $\mathcal{S}_{\text{comm}} = \mathcal{Q}_{\text{UAV}} \times \mathcal{Q}_{\text{GN}}$ is the set of communication states. Crucial to the definition of the SMDP is how the system is sampled in time to define Markov dynamics in the evolution of the sampled states. Accordingly, we define the actions available in each state $\mathbf{s} \in \mathcal{S}$ and the transition probabilities, along with the duration $T(\mathbf{s}; \mathbf{a})$, the UAV energy usage $E(\mathbf{s}; \mathbf{a})$, and the communication delay $\Delta(\mathbf{s}; \mathbf{a})$ metrics accrued in state $\mathbf{s}$ under action $\mathbf{a}$.

**Waiting actions and transitions**: If the UAV is in the waiting state $\mathbf{s}_n = \mathbf{q}_U \in \mathcal{S}_{\text{wait}}$ at time $t$, then the actions available are to move the UAV with radial (referred outward) and angular (referred counter-clockwise) velocity components $(v_r, \theta_c)$, over an arbitrarily small duration $\Delta_0 \ll 1/\Lambda$. Under a maximum velocity constraint $V_{\max}$, the action space is then $\mathcal{A}_{\text{wait}}(r_U) \triangleq \left\{ (v_r, \theta_c) \in \mathbb{R}^2 \middle| \sqrt{v_r^2 + r_U^2 \cdot \theta_c^2} \leq V_{\max} \right\}$, where $v_U = \sqrt{v_r^2 + r_U^2 \theta_c^2}$ is the velocity expressed with respect to polar coordinates. Upon choosing action $\mathbf{a} = (v_r, \theta_c) \in \mathcal{A}_{\text{wait}}(r_U)$, the communication delay is $\Delta(\mathbf{s}; \mathbf{a}) = 0$, since there is no ongoing communication; the duration of a waiting state visit is $T(\mathbf{s}; \mathbf{a}) = \Delta_0$, during which the UAV uses an amount of energy $E(\mathbf{s}; \mathbf{a}) = \Delta_0 P_{\text{mob}}(v_U)$ to move at velocity $v_U$. The new state is then sampled at time $t + \Delta_0$, with the UAV moved to the new position $\mathbf{q}_U(t + \Delta_0) \approx (r_U, \theta_U) + (v_r, \theta_c)\Delta_0$. With probability $e^{-\Lambda\Delta_0}$, no new request is received in the time interval $[t, t+\Delta_0]$, so that the new state is a waiting state. Otherwise, a new request is received from a GN in position $(r, \theta)$, so that the new state is a communication state. Thus, the transition probability from the waiting state $\mathbf{s}_n = (\mathbf{q}_U, \mathbf{w}_0)$ under action $\mathbf{a}_n = (v_r, \theta_c) \in \mathcal{A}_{\text{wait}}(r_U)$ is

$$\mathbb{P}(\mathbf{s}_{n+1} = \mathbf{q}_U + \mathbf{a}_n\Delta_0 | \mathbf{s}_n, \mathbf{a}_n) = e^{-\Lambda\Delta_0} \text{ and} \qquad (7)$$

$$\mathbb{P}(\mathbf{s}_{n+1} = (\mathbf{q}_U + \mathbf{a}_n\Delta_0, \mathbf{q}_G') | \mathbf{s}_n, \mathbf{a}_n) = \frac{A(\mathcal{F})}{\pi a^2} \cdot (1 - e^{-\Lambda\Delta_0}),$$

where $\mathbf{q}_G' \in \mathcal{F}$ and $A(\mathcal{F})$ is the area of region $\mathcal{F}$, $\forall \mathcal{F} \subseteq \mathcal{Q}_{\text{GN}}$, since requests are uniformly distributed in the cell.

**Communication actions and transitions**: Upon reaching a communication state $\mathbf{s}_n = (\mathbf{q}_U, \mathbf{q}_G) \in \mathcal{S}_{\text{comm}}$ at time $t$, the system must serve a GN request at position $\mathbf{q}_G = (r, \theta)$. The system first determines the best scheduling decision $\xi \in \{0, 1\}$. If $\xi = 0$, the GN transmits directly to the BS, and the next state immediately after this decision is the waiting state $\mathbf{s}_{n+1} = \mathbf{q}_U \in \mathcal{S}_{\text{wait}}$ with probability 1. In this case, the cost metrics under action $\mathbf{a}_n = \mathbf{a} = (0, [\ ])$ are computed as $\Delta(\mathbf{s}_n; \mathbf{a}) = \frac{L}{\bar{R}_{GB}(r)}$, $E(\mathbf{s}_n; \mathbf{a}) = 0$, $T(\mathbf{s}_n; \mathbf{a}) = 0$, since direct transmissions occur at throughput $\bar{R}_{GB}(r)$ and the system

moves immediately to the waiting state $\mathbf{q}_U \in \mathcal{S}_{\text{wait}}$ resulting in the action duration and energy expenditure being 0. On the other hand, if $\xi = 1$, the UAV uses the D&F protocol described next, while following a trajectory starting from its current position $\mathbf{q}_U$ and ending in position $\mathbf{q}_U'$. We denote this action as $\mathbf{a}_n = \tilde{\mathbf{a}} = (1, \mathbf{q}_U \to \mathbf{q}_U')$. In the first phase (of duration $t_p$) of the D&F protocol, the GN transmits its payload to the UAV; in the second phase (of duration $\Delta - t_p$), the UAV-relays the data payload to the BS. Assuming a move-and-transmit implementation [3], the trajectory $(\mathbf{q}_U \to \mathbf{q}_U')$ and the time periods ($T_1 \triangleq [0, t_p]$ and $T_2 \triangleq [t_p, \Delta]$) must satisfy

$$\int_{T_1} \bar{R}_{GU}(r_{GU}(t+\eta))\mathrm{d}\eta, \int_{T_2} \bar{R}_{UB}(r_{UB}(t+\eta))\mathrm{d}\eta \geq L, \quad (\text{C.1})$$

i.e., the entire payload of $L$ bits is first transmitted to the UAV with rate $\bar{R}_{GU}(r_{GU}(t+\eta))$, where $r_{GU}(t+\eta)$ is the GN-UAV distance (projected onto the $x-y$ plane) at time $t+\eta$; then, the UAV transmits the payload to the BS with rate $\bar{R}_{UB}(r_{UB}(t+\eta))$, where $r_{UB}(t+\eta)$ is the radial position of the UAV at time $t+\eta$, so that the total communication delay is $\Delta$. In this case, the cost metrics under action $\tilde{\mathbf{a}}$ are $\Delta(\mathbf{s}_n; \tilde{\mathbf{a}}) = \Delta$, $E(\mathbf{s}_n; \tilde{\mathbf{a}}) = \int_0^\Delta P_{\text{mob}}(v_U(t+\eta))\mathrm{d}\eta$, $T(\mathbf{s}_n; \tilde{\mathbf{a}}) = \Delta$. Upon completing the D&F protocol, the UAV enters the waiting phase again, so that $\mathbf{s}_{n+1} = \mathbf{q}_U'$ becomes the new SMDP state, sampled at time $t + \Delta$. Let $\mathcal{Q}_{\mathbf{q}_G}(\mathbf{q}_U \to \mathbf{q}_U')$ be the set of feasible UAV trajectories starting in $\mathbf{q}_U$, terminating in $\mathbf{q}_U'$, to serve a GN located at $\mathbf{q}_G$ using D&F:

$$\mathcal{Q}_{\mathbf{q}_G}(\mathbf{q}_U \to \mathbf{q}_{U'}) = \left\{ \mathbf{p}_U : [0, \Delta] \mapsto \mathbb{R}_+ \times [0, 2\pi) \text{ s.t. C.1,} \right.$$

$$\mathbf{p}_U(0) = \mathbf{q}_U, \mathbf{p}_U(\Delta) = \mathbf{q}_U', \exists \Delta \geq 0, \exists 0 \leq t_p \leq \Delta, \qquad (\text{C.2})$$

$$\left. v_U(\eta) \leq V_{\max}, \forall \eta \in [0, \Delta] \right\} \qquad (\text{C.3})$$

where C.2 reflects the trajectory constraints and C.3 reflects the maximum velocity constraint. Then, the action space in state $(\mathbf{q}_U, \mathbf{q}_G) \in \mathcal{S}_{\text{comm}}$ when $\xi = 1$ is the set $\mathcal{Q}_{\mathbf{q}_G}(\mathbf{q}_U) \triangleq \cup_{\mathbf{q}_U' \in \mathcal{Q}_{\text{UAV}}} \mathcal{Q}_{\mathbf{q}_G}(\mathbf{q}_U \to \mathbf{q}_U')$ of feasible trajectories starting in $\mathbf{q}_U$ that serve the GN at $\mathbf{q}_G$ via the D&F protocol. The overall communication action space is then $\mathcal{A}_{\text{comm}}(\mathbf{q}_U, \mathbf{q}_G) \triangleq \{0, [\ ]\} \cup \{\{1\} \times \mathcal{Q}_{r,\theta}(r_U, \theta_U)\}$. Here the set $\{0, [\ ]\}$ is associated with $\xi = 0$ (no trajectory design space); while the set $\{\{1\} \times \mathcal{Q}_{r,\theta}(r_U, \theta_U)\}$ is associated with $\xi = 1$, whose trajectory design space is $\mathcal{Q}_{r,\theta}(r_U, \theta_U)$.

**Policy** $\mu$: For *waiting* states $\mathbf{q}_U \in \mathcal{S}_{\text{wait}}$, the policy selects a velocity $(v_r, \theta_c)$ from the waiting action space, i.e., $\mu(\mathbf{q}_U) \in \mathcal{A}_{\text{wait}}(r_U)$. Likewise, for *communication* states $(\mathbf{q}_U, \mathbf{q}_G) \in \mathcal{S}_{\text{comm}}$, the policy selects the scheduling decision $\xi \in \{0, 1\}$ and if $\xi = 1$, the trajectory followed in the D&F protocol, i.e., $\mu(\mathbf{q}_U, \mathbf{q}_G) \in \mathcal{A}_{\text{comm}}(\mathbf{q}_U, \mathbf{q}_G)$. With a stationary policy $\mu$ defined, the Lagrangian metric $L_\mu^{(\nu)} \triangleq \bar{W}_\mu^{(s)} + \nu(\bar{E}_\mu - P_{\text{avg}}\bar{T}_\mu)$ in (6) is reformulated using Little's Law as

$$L_\mu^{(\nu)} = \lim_{N \to \infty} \mathbb{E}_\mu \left[ \frac{\frac{1}{N}\sum_{n=0}^{N-1} \ell_\nu(\mathbf{s}_n; \mu(\mathbf{s}_n))}{\frac{1}{N}\sum_{n=0}^{N-1} \mathbb{I}(\mathbf{s}_n \in \mathcal{S}_{\text{comm}})} \right]$$

$$= \frac{1}{\pi_{\text{comm}}} \int_\mathcal{S} \Pi_\mu(\mathbf{s})\ell_\nu(\mathbf{s}; \mu(\mathbf{s}))\mathrm{d}\mathbf{s}, \qquad (8)$$

where $\Pi_\mu(\mathbf{s})$ is the steady-state probability density function of the SMDP being in a state $\mathbf{s}$ under policy $\mu$,

$\pi_{\text{comm}} = \int_{\mathcal{S}_{\text{comm}}} \Pi_\mu(\mathbf{s}) \mathrm{d}\mathbf{s}$ is the steady-state probability that the UAV is in the communication phase in the SMDP, and $\ell_\nu(\mathbf{s}; \mathbf{a}) \triangleq \Delta(\mathbf{s}; \mathbf{a}) + \nu\big(E(\mathbf{s}; \mathbf{a}) - P_{\text{avg}} T(\mathbf{s}; \mathbf{a})\big)$ is the overall Lagrangian metric in state $\mathbf{s}$ under action $\mathbf{a}$. In (8), note that $\sum_{n=0}^{N-1} \ell_\nu(\mathbf{s}_n; \mu(\mathbf{s}_n))$ is the total Lagrangian cost accrued during the first $N$ SMDP stages, and $\sum_{n=0}^{N-1} \mathbb{I}(\mathbf{s}_n \in \mathcal{S}_{\text{comm}})$ is the number of communication states encountered in the SMDP; since a new decision interval is initiated after a communication state, this in turn equals the number of decision intervals. Therefore, after taking the expectation and the limit $N \to \infty$, $L_\mu^{(\nu)}$ represents the expected Lagrangian cost per decision interval, as expressed in (6). The right hand expression in (8) follows by noticing that the SMDP achieves a steady-state behavior when $N \to \infty$. We now specialize the Lagrangian metric $\ell_\nu(s; \mathbf{a})$. Specifically, for waiting states,

$$\ell_\nu(r_U, \theta_U; v_r, \theta_c) = \nu\Big(P_{\text{mob}}\Big(\sqrt{v_r^2 + r_U^2 \theta_c^2}\Big) - P_{\text{avg}}\Big)\Delta_0; \quad (9)$$

for communication states under $\xi = 0$, $\ell_\nu(r_U, \theta_U, r, \theta; 0, [\,]) = \frac{L}{\bar{R}_{GB}(r)}$; and for communication states under $\xi = 1$ with trajectory $\mathbf{p}_U$ of duration $\Delta$, we obtain

$$\ell_\nu(r_U, \theta_U, r, \theta; 1, \mathbf{p}_U) = (1 - \nu P_{\text{avg}})\Delta + \nu \int_0^\Delta P_{\text{mob}}(v_U(\eta))\, \mathrm{d}\eta.$$

The minimization problem of (6) can then be expressed as the average cost-per-stage problem

$$g(\nu) = \frac{1}{\pi_{\text{comm}}} \min_\mu \int_{\mathcal{S}} \Pi_\mu(s) \ell_\nu(s; \mu(s)) \mathrm{d}s, \quad (10)$$

solvable via standard dynamic programming approaches, after discretization of the state space, followed by the dual maximization $\max_{\nu \geq 0} g(\nu)$. Since GN transmission requests are uniformly distributed in the circular cell with the BS in the center, the UAV radius information is a sufficient statistic in decision making for a waiting state $(r_U, \theta_U)$, which can be thus expressed as $\mathbf{s} = r_U \in \mathcal{S}_{\text{wait}}$; likewise, for a communication state $(r_U, \theta_U, r, \theta)$, only the UAV radius, GN request radius, and the angle $\psi \in [0, 2\pi)$ between them suffice to characterize the state—thus, communication states can be compactly represented as $\mathbf{s} = (r_U, r, \psi) \in \mathcal{S}_{\text{comm}}$. A consequence of these sufficient statistics for decision-making is that the policy affects the SMDP state transitions (hence, steady-state behavior) only through the UAV radial velocity $v_r$ in the waiting states and the UAV trajectory's target end radius position $\hat{r}_U$ in the communication states. On the other hand, the angular velocity $\theta_c$ in the waiting states and the UAV trajectory's target end angular coordinate $\hat{\theta}_U$ in the communication states do not influence state dynamics, but only the instantaneous Lagrangian metric $\ell_\nu$. With this observation, let $O(\mathbf{s}) \triangleq v_r \in [-V_{\text{max}}, V_{\text{max}}]$ define the radial velocity policy of the waiting states $\mathbf{s} \in \mathcal{S}_{\text{wait}}$, specifying the radial velocity component of a waiting action $\mathbf{a} = (v_r, \theta_c) \in \mathcal{A}_{\text{wait}}(\mathbf{s})$; let $U(\mathbf{s}) \triangleq \hat{r}_U \in [0, a]$ define the next radius position policy of the communication states $\mathbf{s} \in \mathcal{S}_{\text{comm}}$, specifying the end radius position of a scheduling and communication action $\mathbf{a} \in \mathcal{A}_{\text{comm}}(\mathbf{s})$. Under this decomposition, $O$ and $U$ constitute the outer decisions made by the SMDP and are the only actions to affect the steady-state distribution, denoted as $\Pi_{O,U}$ under the outer policy $(O, U)$. Thus, the optimization problem (10)
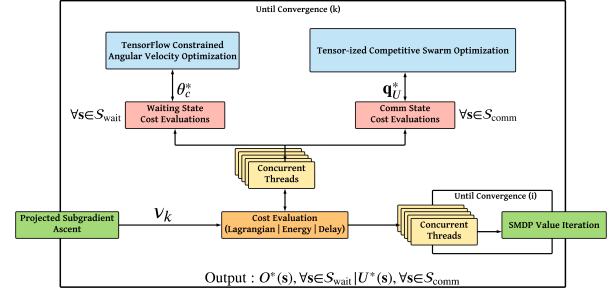


Fig. 2: Algorithmic flow inherent in our single UAV policy optimization.

can be restated as

$$g(\nu) = \frac{1}{\pi_{\text{comm}}} \min_{O,U}\Big[\int_{\mathcal{S}_{\text{wait}}} \Pi_{O,U}(\mathbf{s}) \ell_\nu^*(\mathbf{s}; O(\mathbf{s})) \mathrm{d}s +$$
$$\int_{\mathcal{S}_{\text{comm}}} \Pi_{O,U}(\mathbf{s}) \ell_\nu^*(\mathbf{s}; U(\mathbf{s})) \mathrm{d}s\Big], \quad (11)$$

where $\ell_\nu^*$ is the Lagrangian metric optimized with respect to the *inner action* components not specified by $O$ and $U$. In particular, for waiting states $\mathbf{s} = r_U$ and radial velocity $O(\mathbf{s}) = v_r$, the inner optimization is performed with respect to the UAV angular velocity $\theta_c$, i.e.,

$$\ell_\nu^*(\mathbf{s}; v_r) = \min_{\theta_c} \nu\left(P_{\text{mob}}\left(\sqrt{v_r^2 + r_U^2 \theta_c^2}\right) - P_{\text{avg}}\right)\Delta_0$$
$$\text{s.t. } \sqrt{v_r^2 + r_U^2 \theta_c^2} \leq V_{\text{max}}. \quad (12)$$

Since $\nu \geq 0$, $\Delta_0 > 0$, and $P_{\text{avg}}$ are constant, the optimizer $\theta_c^*$ is the angular velocity minimizing the UAV power consumption for a given UAV radial velocity $v_r$ and radius $r_U$, solvable through exhaustive search. For communication states $\mathbf{s} = (r_U, r, \psi)$, $\ell_\nu^*(\mathbf{s}; \hat{r}_U)$ is determined by optimizing over $\xi \in \{0, 1\}$ and if $\xi = 1$, the trajectory $\mathbf{p}_U$ followed by the UAV, terminating in $\hat{r}_U$. Let $\ell_\nu^*(\mathbf{s}; \hat{r}_U, \xi)$ denote the optimized metric as a function of $\xi \in \{0, 1\}$. For $\xi = 1$ (D&F protocol),

$$\ell_\nu^*(\mathbf{s}; \hat{r}_U, 1) = \min_{\Delta, \mathbf{p}_U, t_p} (1 - \nu P_{\text{avg}})\Delta + \quad (13)$$
$$\nu \int_0^\Delta P_{\text{mob}}\left(\sqrt{r_U'(\eta)^2 + r_U^2(\eta)\theta_U'(\eta)^2}\right) \mathrm{d}\eta \text{ s.t.}$$
$$\text{C.1, C.3, } \mathbf{p}_U(0) = (r_U, 0), \|\mathbf{p}_U(\Delta)\|_2 = \hat{r}_U \quad (\hat{\text{C}}.2)$$

where $\hat{\text{C}}.2$ enforces the trajectory constraints. For $\xi = 0$ ($\hat{r}_U = r_U$), $\ell_\nu^*(\mathbf{s}; r_U, 0) = \frac{L}{R_{GB}(r)}$. Hence $\ell_\nu^*(\mathbf{s}; r_U)$ is obtained by further minimizing over $\xi \in \{0, 1\}$, yielding

$$\ell_\nu^*(\mathbf{s}; \hat{r}_U) = \min_{\xi \in \{0,1\}} \ell_\nu^*(\mathbf{s}; r_U, \xi)\mathbb{I}(\hat{r}_U = r_U) + \ell_\nu^*(\mathbf{s}; \hat{r}_U, 1)\mathbb{I}(\hat{r}_U \neq r_U).$$

Thus, if the outer decision selects $U(\mathbf{s}) = r_U$, the inner scheduling decision $\xi \in \{0, 1\}$ is obtained by greedily minimizing a cost metric that trades off communication delay and energy consumption, i.e., direct transmission to the BS occurs if $\ell_\nu^*(\mathbf{s}; r_U, 0) < \ell_\nu^*(\mathbf{s}; r_U, 1)$. Otherwise, the UAV handles the GN request using the D&F protocol, and the inner decision on UAV trajectory—designed via Competitive Swarm Optimization (CSO) [6]—greedily minimizes the instantaneous delay-power trade-off, terminating at the target radius of the outer decision, $U(\mathbf{s}) = \hat{r}_U$. Discretizing the trajectory between the UAV's initial and final service positions, and using a modified version of (13) (incorporating time and energy penalties for

data transmission failures during D&F) as the cost function, CSO invokes pair-wise cost comparisons among a randomly initialized set of way-points and velocities: specifically, in each iteration, the winner particles (way-points and velocities with lower cost function values) advance to the next iteration while the loser particle values are updated in relation to these winners; this update process continues until a maximum number of cost evaluations are performed. We design the outer policy and compute the average cost-per-stage metric $g(\nu)$, along with the average energy- and time-per-stage metrics for a given $\nu$, by solving problem (11) via value iteration; also the dual maximization problem $\max_{\nu \geq 0} g(\nu)$ is solved via projected sub-gradient ascent [16]. Fig. 2 illustrates the sequence of operations involved in solving for the optimal policy.

## IV. MULTI-AGENT EXTENSIONS

In this section, we remove the specializations considered for our single relay construction in Sec. III, and extend the SMDP-based control strategy to our generalized hybrid wireless network model of $N_U$ UAV-relays in the swarm. To this end, the single relay optimal policy is embedded with supplementary heuristics and replicated across the swarm.

**Command-and-Control**: Truly decentralized and coordinated operations of the UAV-relays in the swarm necessitates the need for a control network over which the server nodes can collaborate to ensure collision-free movements among the UAVs, facilitate consensus-driven decision-making on the best server to handle a GN uplink transmission request, and guarantee resilient fault-tolerant operations by setting-up fallback mechanisms to handle UAV failures. We designate the band-edges of the allocated spectrum as control channels, over which the server nodes in the cell exchange short collaboration messages: the structure of a control frame in our system is shown in Fig. 3. A fully-connected distributed mesh topology (employing these designated control channels) overlaid over the BS and the UAV swarm constitutes the design of our command-and-control network. Since each UAV-relay in the swarm possesses the same optimal waiting and communication state policies, we embed spread maximization (in the waiting states) and conflict resolution (in the communication states) to cooperatively handle their operations.

**Spread Maximization**: To efficiently position and prime the idle UAVs for a potential new GN request, in the waiting states, a UAV-relay in the swarm, in addition to executing the optimal action, determines the direction of angular motion (clockwise or counter-clockwise) based on our spread maximization heuristic—wherein each UAV-relay in the waiting state and in a radius level with zero radial velocity, executes either positive (counter-clockwise) or negative (clockwise) angular movements in order to maximize the minimum distance among them. These coordinated movements among the UAVs is made possible through periodic exchanges of control frames over the command-and-control network. Studying this frame structure in Fig. 3, we note that for waiting states, the *state flag* is set to 0, and the *GN position* and *cost-of-service* fields are empty; UAVs in the waiting state extract the positional information of their peers from the *GPS event* field and apply a maximize-the-minimum-distance heuristic over other UAVs which are waiting in the same radius level. This methodology
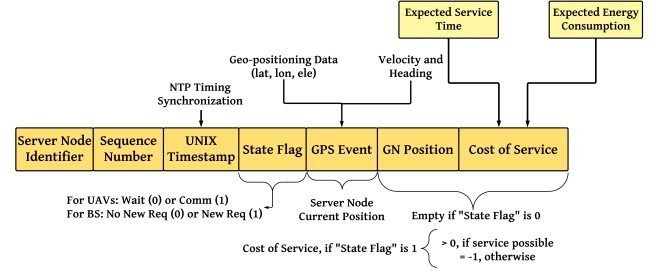


Fig. 3: Structure of the control frames exchanged among the BS and UAVs.

ensures that a suitable spread is maintained among the idle UAVs, in order to facilitate faster response times when a new uplink request is generated.

**Consensus-driven Conflict Resolution**: When a new uplink transmission request originates in the cell, the UAVs already serving a GN continue to do so, i.e., they do no participate in the consensus-driven conflict resolution process. These relays, termed unavailable under this context, transition into their corresponding waiting states upon service completion. On the other hand, UAVs in the waiting states transition into their respective communication states. The BS along with these relays are deemed to be available. Studying the control frame exchanged by these available server nodes: the *state flag* is set to 1, the *GN position field* is populated with the originating position of the request under consideration, and the *cost-of-service* field constitutes a weighted sum of the request service time and the energy required to serve it. These values on communication service times and energy consumption are obtained from the policy lookup table. Note that the BS' energy consumption for serving a GN request is set to $0$. Upon sharing these metrics with each other, these nodes arrive at a consensus on the best choice (delay and energy) for serving the request.

## V. NUMERICAL EVALUATIONS

We use a channel bandwidth of $B=5$ MHz; for all links, NLoS attenuation constant $\kappa=0.2$, 1-meter reference SNR $\frac{\beta_0 P_T}{\sigma^2 \Gamma}=40$ dB, LoS path-loss exponent $\alpha=2$, NLoS path-loss exponent $\tilde{\alpha}=2.8$, Rician $K$-factor parameters $k_1=1$ and $k_2=0.05$, and LoS probability parameters $z_1=9.61$ and $z_2=0.16$; UAV height $H_U=200$ m; BS antenna height $H_B=80$ m; maximum UAV velocity $V_{\max}=55$ m/s; and cell radius $a=1000$ m. Our UAV mobility power model uses the relationship and parameters detailed in [3]. We solve an approximation of problem (11) by discretizing the SMDP state and action spaces, and applying value iteration; additionally, the optimal dual variable value is attained via projected sub-gradient ascent. We discretize the states with $N_{\mathrm{sp}}=25$ equispaced radii values; similarly, $R_{\mathrm{sp}}=25$ corresponds to the equispaced radial velocity actions.

To analyze the optimal waiting behavior of a UAV-relay in our control framework, we fix $P_{\mathrm{avg}}=1.2$ kW and $L=1$ Mb. Fig. 4a shows that the UAV moves to and waits by flying around an optimal radius level ($\approx 95$ m) to address two considerations: to be well-positioned for future requests and to fly at the power-minimizing velocity so as to reduce its energy consumption. Moreover, we note that the angular velocity optimization process via exhaustive search adheres to the observations in [3] about optimizing towards the minimum, i.e., $P_{\min}$ corresponding to $V=22$ m/s.
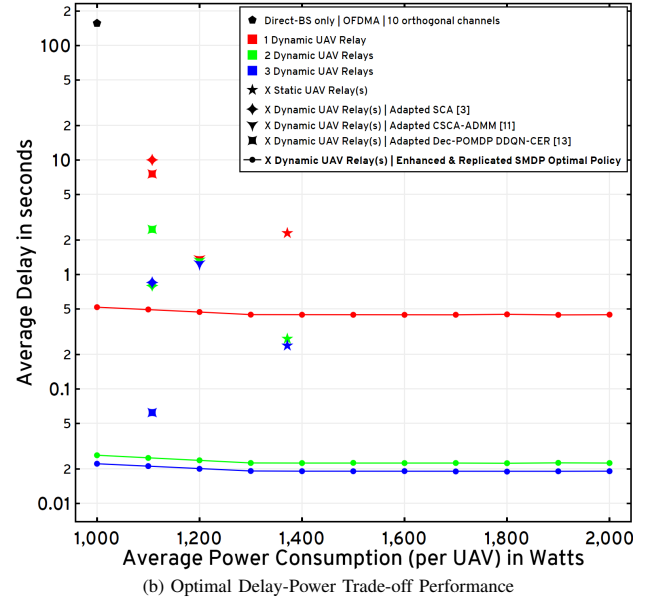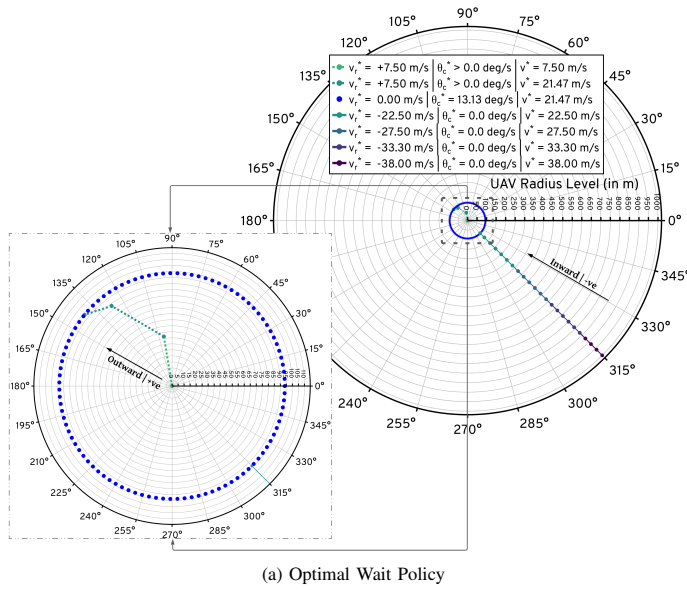
(a) Optimal Wait Policy

(b) Optimal Delay-Power Trade-off Performance

Fig. 4: (a) Optimal waiting state policy for $L$=1 Mb with $P_{\text{avg}}$=1.2 kW; (b) Average service latencies vs average power constraints for $L$=1 Mb.

As depicted in Fig. 4b, for uplink transmission requests of size $L$=1 Mb from the 300 GNs in the cell, we observe the following improvements in performance over custom network deployment heuristics and state-of-the-art frameworks, averaged over $10,000$ requests with a Poisson arrival rate of 1 request every 60 s. Considering only direct transmissions to a 10-channel OFDMA BS at the cell center, we find a significant reduction in the average communication delay experienced by the GNs in the cell, by employing UAVs to relay data traffic. Also, we observe that employing dynamic UAVs with optimized trajectories demonstrate lower service delays compared to static relay deployments—specifically, for a swarm of 3 UAV-relays, with $L$=1 Mb and a per-UAV power consumption of 1.37 kW, our solution services uplink transmission requests from the GNs $12\times$ faster than a static deployment of 3 UAVs positioned equidistant from the cell center. Furthermore, with CVXPY implementations of joint multi-agent SCA strategies [3], [11] involving split conic solvers with $10^6$ iterations and an accuracy of $10^{-6}$, we note that our control system with 1 UAV-relay exceeds the QoS performance offered by 3 relays under these SCA approaches. Finally, with an average per-UAV power constraint of 1.1 kW, our solution demonstrates $3\times$ faster service times relative to DDQNs [13].

## VI. Conclusion

In this paper, we detail the development of an adaptive framework for the decentralized orchestration of a swarm of rotary-wing UAV-relays in next-generation non-terrestrial networks. First, employing rate adaptation to leverage A2G channel dynamics, we specialize our system model to single UAV-relay settings and design the optimal request scheduling and trajectory optimization policy under an SMDP formulation (via value iteration and CSO). Next, we extend this single relay policy to distributed deployments of two or more UAVs by supplementing it with multi-UAV coordination heuristics and replicating it across the swarm. Numerical evaluations demonstrate that our solution delivers significant performance improvements over BS-only strategies, static UAV deployments, SCA approaches, and DDQN frameworks.

## References

[1] A. Fotouhi *et al.*, "Survey on UAV Cellular Communications: Practical Aspects, Standardization Advancements, Regulation, and Security Challenges," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2019.

[2] Q. Wu *et al.*, "Fundamental Trade-offs in Communication and Trajectory Design for UAV-Enabled Wireless Network," *IEEE Wireless Communications*, vol. 26, pp. 36–44, 02 2019.

[3] Y. Zeng *et al.*, "Energy Minimization for Wireless Communication With Rotary-Wing UAV," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2329–2345, April 2019.

[4] M. A. Abd-Elmagid *et al.*, "Average Peak Age-of-Information Minimization in UAV-Assisted IoT Networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 2003–2008, 2019.

[5] C. You *et al.*, "3D Trajectory Optimization in Rician Fading for UAV-Enabled Data Harvesting," *IEEE Transactions on Wireless Communications*, vol. 18, no. 6, pp. 3192–3207, 2019.

[6] R. Cheng *et al.*, "A Competitive Swarm Optimizer for Large Scale Optimization," *IEEE Transactions on Cybernetics*, vol. 45, no. 2, pp. 191–204, 2015.

[7] M. Clerc, *Particle Swarm Optimization*, ser. ISTE. London ; Newport Beach: ISTE, 2006.

[8] H. Shakhatreh *et al.*, "Efficient 3D placement of a UAV using particle swarm optimization," in *2017 8th International Conference on Information and Communication Systems (ICICS)*, 2017, pp. 258–263.

[9] R. K. Patra *et al.*, "Optimisation of Spectrum and Energy Efficiency in UAV-Enabled Mobile Relaying Using Bisection and PSO Method," in *2018 3rd International Conference for Convergence in Technology (I2CT)*, 2018, pp. 1–7.

[10] M. Mozaffari *et al.*, "Efficient Deployment of Multiple Unmanned Aerial Vehicles for Optimal Wireless Coverage," *IEEE Communications Letters*, vol. 20, no. 8, pp. 1647–1650, 2016.

[11] Q. Hu *et al.*, "Low-Complexity Joint Resource Allocation and Trajectory Design for UAV-Aided Relay Networks With the Segmented Ray-Tracing Channel Model," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 6179–6195, 2020.

[12] Y. Li *et al.*, "UAV-Assisted Dynamic Coverage in a Heterogeneous Cellular System," *IEEE Network*, vol. 31, no. 4, pp. 56–61, July 2017.

[13] H. Bayerlein *et al.*, "Multi-UAV Path Planning for Wireless Data Harvesting With Deep Reinforcement Learning," *IEEE Open Journal of the Communications Society*, vol. 2, p. 1171–1187, 2021.

[14] L. Wang *et al.*, "Multi-Agent Deep Reinforcement Learning-Based Trajectory Planning for Multi-UAV Assisted Mobile Edge Computing," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 1, pp. 73–84, 2021.

[15] M. Bliss *et al.*, "Power-Constrained Trajectory optimization for Wireless UAV Relays with Random Requests," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[16] S. Boyd *et al.*, "Subgradient methods," *Lecture notes, Stanford University, Autumn Quarter*, vol. 2004, pp. 2004–2005, 2003.