This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3303253

Multi-Agent Learning for Secure Wireless Access from UAVs with Limited Energy Resources

Aly Sabri Abdalla and Vuk Marojevic

***Abstract*—Terrestrial wireless network deployment challenges and the high associated costs encourage the exploration of aerial base stations (ABSs). An ABS carried by an unmanned aerial vehicle (UAV) can be dispatched at a relatively low cost to provide coverage on demand, such as in emergency situations and during temporary hot-spot events. While relatively inexpensive, battery-powered UAVs have a limited flight time and can only provide temporary service in practice. This paper therefore considers and monitors the available energy of UAVs as a constraint for the proposed communication architecture consisting of dynamically dispatched ABSs that are managed by a high-altitude platform station (HAPS) performing network optimization. We consider a fleet of UAVs for providing secure wireless service to sparsely distributed users in urban areas and propose an efficient coverage strategy to satisfy the users' data rate demands while meeting their secrecy rate requirements. Because of the complexity, dynamics, and distributed nature of the problem, we employ multiple ABSs as the agents and design a deep deterministic policy gradient (DDPG) algorithm to optimize their positions in the ABS network with time-constrained nodes. Numerical results illustrate how the DDPG-empowered HAPS is able to coordinate and leverage the ABSs fleet for wide-spread secure coverage and adjust the network deployment topology when nodes become unavailable. While the DDPG has a higher training complexity, it provides better performance over state-of-the-art solutions in terms of the number of securely served users. We discuss the practical implications of the training process and identify opportunities for research and development.**

***Index Terms*—Cellular communications, deep reinforcement learning, energy constraint, HAPS, secrecy rate, security, UAV.**

I. Introduction

THE emerging 5G and future 6G wireless network deployments will enable advanced commerce, transportation, health, science, and defense applications. Next-generation wireless technology will support the integration of seamless mobility across networks and provide an overarching architecture for delivering flexible and customizable networking and end-to-end services. Such networks are much needed for scalability with the increasing number of connected devices, such as simple sensors, actuators, user devices, sophisticated industrial control systems, medical systems, vehicles, cities, and critical infrastructure components. Unmanned aerial vehicles (UAVs) will play an important role as they can provide on-demand wireless networking support.

Manuscript received 19 April 2023; revised 7 July 2023; accepted 4 August 2023. Date of publication - - 2023; date of current version 4 August 2023. This work was supported in part by the NSF PAWR program, under grant number CNS-1939334.

Aly Sabri Abdalla and Vuk Marojevic are with the Department of Electrical and Computer Engineering, Mississippi State University, MS, USA e-mail: ($asa298$@msstate.edu; $vm602$@msstate.edu).

The integration of UAVs into cellular communication networks is commonly known as cellular-connected UAVs or network-connected UAVs which can be further classified into UAVs supporting the terrestrial network infrastructure and UAVs subscribed as user equipment (UEs) [1]. As part of the wireless communication infrastructure, the UAV can be deployed as an aerial base station (ABS) or an aerial relay (AR) [2] [3].

ABSs are useful for extending wireless coverage and enabling capacity on demand to provide seamless communications even in difficult circumstances such as during disaster recovery or crowded events. A number of ABSs are needed to provide coverage and serve ground users dispatched over dense areas without terrestrial cellular coverage. However, such wireless communication systems create a new attack surface and security vulnerabilities because the information is signaled over the air with a potentially larger radio frequency (RF) footprint than from terrestrial transmitters. Therefore, the emerging aerial wireless communication systems enabled by a network of ABSs must be protected against wireless security threats [4].

The foundation of cellular network operations and its major threats stem from the trust relationship between a UE and the network. A UE searches for well-known control signals that are broadcast from cell towers and it follows the instructions coming from the network. Many control signals are sent in the clear and can be easily reproduced for launching spoofing attacks. 5G allows null encryption and the network can decide whether to use encryption or not. Hence, eavesdropping is possible not only for capturing control information but also for user data in certain circumstances and network configurations. The network can request 5G users to provide location updates or their globally unique Subscription Permanent Identifier (SUPI); a fake base station can leverage this. 5G can encrypt the SUPI, but there may be instances where this is not implemented [5].

An eavesdropping attack occurs when an attacker attempts to capture data that is not intended for it and is being transmitted between other devices in a network. Eavesdropping attacks can compromise the confidentiality and privacy of data. An eavesdropper can be passive (receive only) or active (transmit and receive) [6]. An illegitimate user that can gain access to the network, increases its ability to eavesdrop on the network. An attacker may also eavesdrop on the control or data channels in order to launch sophisticated attacks [7].

Intelligently controlling UAV trajectories can facilitate secure wireless coverage for terrestrial users under eavesdropping attacks. UAV trajectory control is proposed for establish-

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3303253

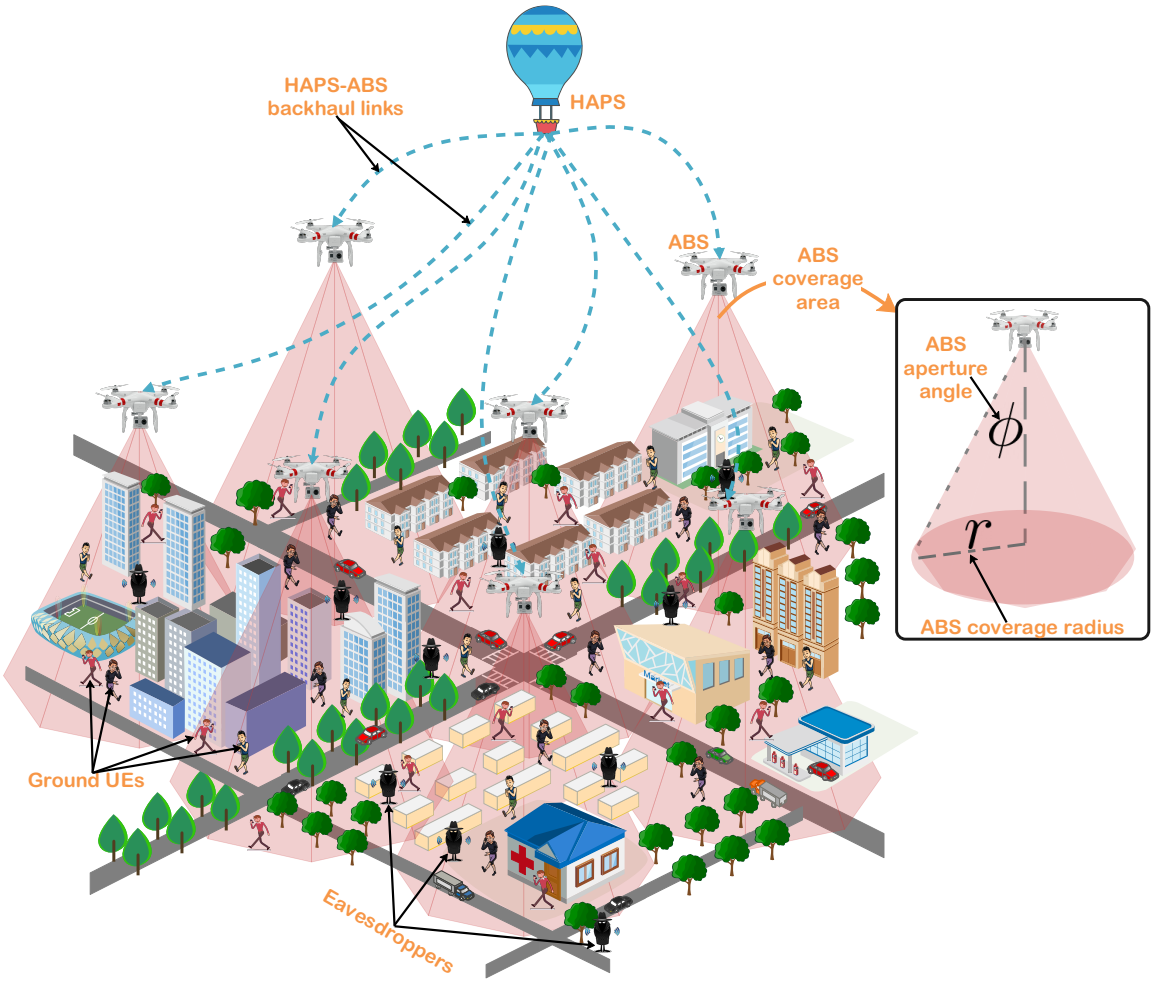


Fig. 1: Multi ABS-assisted ground secure coverage in presence of eavesdropping attack environment.

ing high data rates with low intercept rates. Multiple ABSs need to operate in concert for practical use cases with widely distributed users. Apart from the limited communication resources, it is important to also take into consideration the limited capacity of onboard batteries of untethered ABSs. The power limitations of the UAV are a critical aspect that bounds the availability of the ABS; therefore, the power consumption and remaining power levels of the ABS must be monitored and reported for effective and timely aerial network adjustments.

This paper studies the aforementioned challenges and proposes jointly optimizing the trajectories of multiple ABSs that are coordinated through a high-altitude platform station (HAPS) for establishing intelligent and secure coverage for a dense urban area with dynamic wireless service needs. The scenario of interest in this paper is illustrated in Fig. 1, which shows a HAPS coordinating multiple ABSs that are dispatched to locations for providing wireless coverage while meeting the user service demands in the presence of numerous eavesdroppers. The resource-constrained multi-UAV trajectory optimization problem is complex and using traditional modeling and optimization tools to solve it may not be feasible, especially for large-scale networks [8]–[10]. We thus propose studying the applicability of deep reinforcement learning (DRL) methods that have been shown to be effective for solving related problems that process large state spaces and time-varying environments [11]. Moreover, DRL techniques are capable of delivering high-performance solutions with a reasonable learning overhead and little or no domain knowledge [12].

The contributions of this paper are summarized in continuation:

- We devise an effective method for maximizing the coverage over sparsely connected areas, while satisfying the quality of service (QoS) and data rate requirements of ground users.
- We design a framework for UAV trajectory optimization for effectively serving users while minimizing the effectiveness of eavesdroppers.
- We develop a protocol for replacing the UAVs running our of battery to maintain the wireless connectivity during the entire mission.
- We numerical analyze the performance of the proposed aerial radio access network architecture and the optimization process.
- We study the effect of UAV speed and user mobility on the convergence performance of the DRL model and identify bottlenecks and practical considerations.

The remainder of this paper is organized as follows. Section II introduces the related work. Sections III defines the system model and Section IV formulates the problem. Section V

describes the fundamentals of DRL and presents the proposed DRL-based trajectory optimization framework. Section VI provides simulation results and analyses. Section VII discusses the practical implications of the proposed data driven method and Section VIII provides the concluding remarks.

II. RELATED WORK

In recent years, multiple comprehensive studies have been conducted that examine the deployment and trajectory optimization of UAVs [13]. Literature has proposed diverse solutions to achieve different objectives for a variety of scenarios and constraints. Machine learning (ML) techniques are gaining traction for the trajectory design and management of UAVs, among others [14]. In continuation we therefore review the recent studies that are relevant to our research. We describe them in the following categories: i) UAV deployment and coverage optimization, ii) security of terrestrial communications enabled by UAVs, iii) energy-constrained UAV communications, and iv) DRL-based solutions to UAV trajectory, deployment, security, and coverage problems. Table I discusses the technical contributions of the state-of-the-art solutions compared to the work presented in this paper.

A. UAV-Assisted Cellular Network Deployment and Coverage Extension

The optimization of UAV deployment for coverage extension is considered a complex optimization task with the potential of notably improving the performance of cellular communication networks. Noh et al. [23] propose an ellipse clustering algorithm to allow ABSs to improve the likelihood of covering a large number of ground users while preventing inter-cell interference. They establish an energy-efficient 3D deployment technique to minimize the total energy consumption of ABSs while guaranteeing a particular QoS for each user. Malandrino et al. [24] optimize the UAV path planning for wireless coverage extension in response to a disaster. An optimal separation distance between UAVs was proposed by Khuwaja et al. [15] for avoiding co-channel interference and increasing the coverage of multiple UAVs in urban areas. Wang et al. [16] introduce an iterative algorithm to obtain the minimal number of UAVs needed to enhance the communication coverage for UEs. Two fast UAV deployment solutions are studied by Zhang et al. [17] for maximizing wireless coverage. The first reduces the total deployment delay for network efficiency whereas the second minimizes the deployment delay for user fairness.

B. UAV-Assisted Secure Cellular Communications

A UAV may be deployed to support the terrestrial cellular network—as an AR or ABS—not only for extending coverage but also for improving security. Bringing the network access points closer to the users and employing physical layer security (PLS) techniques can complement traditional wireless network security procedures and improve the confidentiality, availability, and privacy of wireless communications services.

Early research has shown how UAVs can extend cellular networks and be used to improve the security of terrestrial networks against eavesdropping. Sun et al. [25] discuss how a UAV-assisted system can be leveraged for PLS enhancements applied to advanced cellular networks. Zhuo et al. [18] introduce UAV-carried friendly jammers to enhance the PLS of terrestrial cellular networks. The authors investigate the appropriate power levels for jamming to disturb eavesdropping. Wu et al. [26] and Shang et al. [27] also propose using UAVs as friendly jammers. They investigate the effects of having a strong line of sight (LoS) jamming link between the UAV and a single or multiple eavesdroppers. Sun et al. [19] suggest UAV relaying as a solution against eavesdropping in a cellular network. The authors address the secrecy rate problem by optimizing the source and AR power allocation over a finite time window. Wang et al. [28] discuss the deployment of UAVs as mobile relays to improve the results of static relays. They use the difference-of-concave (DC) program to solve the secrecy rate maximization problem and find that each DC iteration yields a closed-form solution. Hou et al. [29] propose a UAV-enabled covert federated learning architecture to enhance data security and protect against eavesdropping. A distributed proximal policy optimization-based strategy is proposed to jointly optimize the trajectory and artificial noise transmit power of the UAV, as well as the CPU frequency, transmit power, bandwidth allocation of devices, and the required local model accuracy. Bai et al. [30] address the challenge of limited computational capability and the short battery lifetime of UAVs by employing mobile-edge computing for offloading computational tasks. An energy-efficient computation task offloading technique is proposed with a focus on PLS.

C. Energy-constrained UAV communications

There have been various research projects that take into consideration the energy consumption of the UAV that is deployed for coverage extension. For example, Liu et al. [20] develop a framework for controlling multiple UAV nodes deployed for improving coverage and connectivity. They consider the user fairness while minimizing the UAV movements for energy conservation based on a DRL solution. Omoniwa et al. [21] propose a decentralized Q-learning approach that simultaneously improves the energy utilization while maximizing the number of connected ground devices served by multiple ABSs subject to interference from neighboring cells. Arani et al. [31] develop a multiarmed bandit learning algorithm for optimizing the energy and spectral efficiency for a set of rotary-wing UAVs integrated into terrestrial networks. Their solution addresses the connectivity outage by jointly optimizing the UAV trajectories and speeds.

D. DRL-Optimized UAV Deployment and Operation

We are witnessing increasing interest in applying ML schemes to operate UAV-assisted cellular networks. For example, Mozaffari et al. [8] leverage ML to optimize a 3D UAV cell association approach for a cellular network composed of UAV users and ABSs. Qi et al. [9] present a 3D UAV deployment scheme that is founded on the deep deterministic policy gradient (DDPG) for designing and scheduling the

TABLE I: Prior art and proposed research.

Category	Ref.	Metric	Strategy	Optimization
UAV-assisted deployment and coverage	[15]	Coverage area ratio	Deploy multiple UAVs in 2D formations in the presence of co-channel interference	1D and 2D UAV placement
	[16]	Fairness and load balancing	Minimize the number of deployed UAVs while balancing the load	2D UAV coordinates
	[17]	UAV deployment delay	Establish a fast deployment formation of UAVs in heterogeneous networks considering the fairness among UAVs	1D UAV placement
UAV-assisted secure communications	[18]	Intercept probability	UAV-based friendly jammer deployment for minimizing the intercept probability	UAV jamming power and 3D location
	[19]	Secrecy rate	AR for securing transmission against eavesdropping with imperfect location information	Transmit and relay powers
	[12]	Secrecy rate	ABS positioning without eavesdropper's channel state information	UAV trajectory and transmit power
Energy-constrained UAV communications	[20]	Fairness coverage	Maximize the energy efficiency with joint consideration for coverage, fairness, energy consumption and connectivity	Flight direction and distance for each UAV
	[21]	Connectivity	Maximize the connectivity while improving the UAV's energy utilization	3D trajectory of each UAV
DRL-optimized UAV deployment	[9]	Fairness coverage	Maximize the sum-rate while minimizing the energy consumption and guaranteeing user fairness	3D mobility and energy replenishment
	[22]	Age of information	Navigate an ABS under energy constraints	2D UAV trajectory
This work		Secrecy coverage	Deploy energy-constrained multi-ABSs that satisfies the users' secrecy requirements over sparsely connected areas in the presence of eavesdroppers	UAV flight distance and direction

mobility of multiple UAVs for energy restoration with the goal of providing fair coverage to ground users. Challita et al. [32] propose a DRL scheme relying on an echo state network to design an interference-aware path planning strategy for UAV-assisted networks. This strategy enables the UAV to optimize its flight direction, transmission power, and cell association. Samir et al. [10] propose a DRL to determine the trajectories for a minimum number of UAVs to support coverage for vehicles on a highway. Abedin et al. [22] tackle the problem of minimizing the energy consumption and the average age of information by optimizing the trajectory of UAVs. Seid et al. [33] propose a multi-agent DRL framework for dynamic computation offloading for IoT devices with energy harvesting in a multi-UAV-assisted IoT network. Their solution minimizes computation cost and resource price while deploying a consortium blockchain for securing the transactions among nodes.

Different from the aforementioned solutions, the novelty of this research is defining and addressing the objective of facilitating secure coverage extension with UAVs over sparsely connected areas in the presence of multiple eavesdroppers.

III. System Model

In this paper, we consider a cooperative multi-UAV framework coordinated via a HAPS for enabling secure wireless communications. Each UAV carries an ABS that is deployed for facilitating secure wireless network access to ground users in the presence of terrestrial eavesdroppers. Fig. 1 illustrates the scenario which consist of M ground users, E eavesdroppers, K ABSs, and one HAPS.

For the sake of simplicity and without loss of generality, we do not consider optimizing the ABS height in this paper. We rather assume that all UAVs are flying/hovering at a fixed altitude h. The height should be chosen to enable LoS communication links to ground users [34] while providing the necessary coverage and enabling low power transmission to minimize the eavesdropping rate [35], [36]. Each ABS has a directional antenna that focuses its radiation power to cover the region directly below it with the aperture angle ϕ. The ground coverage of each ABS is modeled as a circle region of radius $r = h\ tan(\frac{\phi}{2})$. The HAPS-ABS links are assumed to be less prone to eavesdropping and operate on a different frequency than the access links. The HAPS-ABS and the wireless backhaul link establishment and management are out of the scope of this paper.

Without loss of generality, we model and analyze the downlink transmission. While some of the insights gained through examining the downlink may carry over to the uplink situation, there exist specific aspects that require individual consideration and will be addressed in future work.

A. Air-to-Ground Channel Model

The air-to-ground (A2G) communication channel between the ABS and ground nodes features LoS, non-LoS (NLoS), and multiple reflected components which cause multipath fading [37]. The path loss of the A2G communication link between the k^{th} ABS and the i^{th} ground UE can be calculated as [38]

$$\beta_{k,i}^{t} = \begin{cases} \eta_{LoS} \left(\frac{4\pi f_C}{c}\right)^{\alpha} d_{k,i}^{-\alpha} & \text{LoS condition,} \\ \eta_{NLoS} \left(\frac{4\pi f_C}{c}\right)^{\alpha} d_{k,i}^{-\alpha} & \text{NLoS condition,} \end{cases} \quad (1)$$

where η_{LoS} and η_{NLoS} are the excessive path loss coefficients for the LoS and NLoS components, respectively. The symbol f_c represents the carrier frequency, c is the speed of light, $\alpha \approx 2$ is the pathloss exponent for LoS links, and

$$d_{k,i} = \sqrt{h^2 + (x_k - x_i)^2 + (y_k - y_i)^2} \quad (2)$$

is the distance between the k^{th} ABS and the i^{th} UE, where (x_k, y_k) and (x_i, y_i) represent the 2D coordinates of the ABS and the UE.

According to field measurements [37], the occurrence probability of an LoS link $\rho_{LoS}(\theta)$ between a UAV transmitter and a ground receiver is calculated by (3) and can be modeled as

$$\rho_{LoS}(\theta) = \frac{1}{1 + C\ \exp\left[-U\ \ (\theta - C)\right]}, \qquad \theta = \frac{180}{\pi} \arctan\left(\frac{h}{r}\right), \tag{3}$$

where C and U are variables that depend on the environment, e.g. rural, urban, or dense urban. These variables can be determined through three empirical parameters in the ITU-R model [39]: α_{ITU} is the ratio between the constructed area and the total area, β_{ITU} is the mean number of buildings per unit area, and γ_{ITU} is the building height parameter. Parameter $r = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2}$ the horizontal distance between the UAV and the ground UE. Note that the LoS and NLoS probabilities are related as $\rho_{NLoS}(\theta) = 1 - \rho_{LoS}(\theta)$. The average channel power gain between the k^{th} ABS and the i^{th} UE can then be calculated as

$$\begin{aligned} g^t_{k,i} =& \eta_{LoS}\ \rho_{LoS}(\theta)\ \left(\frac{4\pi f_C}{c}\right)^{\alpha} d^{-\alpha}_{k,i} + \\ & \eta_{NLoS}\ \rho_{NLoS}(\theta)\ \ \left(\frac{4\pi f_C}{c}\right)^{\alpha} d^{-\alpha}_{k,i}, \\ =& \hat{\rho}\ \left(\frac{4\pi f_C}{c}\right)^{\alpha} d^{-\alpha}_{k,i}, \end{aligned} \tag{4}$$

where $\hat{\rho} = \eta_{LoS}\ \rho_{LoS}(\theta) + \eta_{NLoS}\ \rho_{NLoS}(\theta)$ captures the regularized LoS probability that covers both LoS and NLoS conditions.

B. Secrecy Capacity

The extent to which the confidentiality of the system is compromised can be measured through the secrecy capacity, which is defined as the transmission capacity at which no information will be decoded by the eavesdropper. The secrecy capacity can thus be calculated as the difference between the legitimate and the wiretap channel capacities [40]:

$$SC = \max\big((C_L - C_W), 0\big). \tag{5}$$

Parameters C_L and C_W represent the legitimate channel capacity between an ABS and a ground user and the wiretap channel capacity between the ABS and a ground eavesdropper, respectively.

For the legitimate transmission capacity C_L, we obtain the channel capacity between the k^{th} ABS ($k \in K = [1, .., k, .., K]$) and the i^{th} UE ($i \in M = [1, .., i, .., M]$) as

$$C^{k,i}_L = B\ \log_2\left(1 + SINR^{k,i}\right), \tag{6}$$

where B is the system bandwidth and

$$SINR^{k,i} = \frac{P\ g_{k,i}}{\sigma^2 + \sum\limits_{j \neq k}^{k_c} P\ g_{k,j}}, \tag{7}$$

is the signal-to-interference plus noise ratio (SINR) between the k^{th} ABS and the i^{th} UE. Parameter P is the transmit power of the ABS and $\sigma^2 = BN_0$ the noise variance with N_0 being the power spectral density of the additive white Gaussian noise (AWGN). The sum term in (7) represents the interference at the i^{th} user cased by other ABSs in the set k_c of ABSs that have the i^{th} user within their coverage areas. The purposes of this paper is studying resource dependencies including UAV energy, convergence time, secrecy capacity, and flight speed. Spectrum availability is another critical resource for real-world ABS deployments. We will consider spectrum-related constraints as part of the problem formulation in our future research.

The capacity between the k^{th} ABS and the e^{th} ground eavesdropper is

$$C^{k,e}_W = B\ \log_2\left(1 + SINR^{k,e}\right), \tag{8}$$

where

$$SINR^{k,e} = \frac{P\ g_{k,e}}{\sigma^2 + \sum\limits_{m \neq k}^{k_e} P\ g_{k,j}} \tag{9}$$

is the SINR between these. The set k_e contains the ABSs that currently provide downlink transmissions to ground users with the e^{th} eavesdropper being within their combined coverage area. In other words, k_e captures the actively transmitting ABSs that eavesdropper e sees. Parameter $g_{k,e}$ denotes the A2G channel gain for the wiretap communication channel between the k^{th} ABS and the e^{th} eavesdropper. It is obtained after calculating the distance $d_{k,e}$ between them as

$$d_{k,e} = \sqrt{h^2 + (x_k - x_e)^2 + (y_k - y_e)^2}, \tag{10}$$

where (x_e, y_e) are the 2D coordinates of the e^{th} eavesdropper. Here we assume that the channel state information of all eavesdropping channels is available at the HAPS and this information can be used to localize the eavesdroppers [41]–[43]. Knowing the CSI and locations of eavesdroppers is the information theoretic ideal case and allows calculating the theoretically achievable secrecy capacity. Future work will extend the models and analyses to account for imperfect CSI conditions.

C. Energy Consumption Model

The on-board battery on each UAV powers the vehicle and the ABS with an initial energy level that is known before the start of the task. The total energy of the ABS's on-board battery is consumed by two main parts: the propulsion unit and the communication unit. The propulsion unit draws power during the UAV flight for enabling mobility and hovering of the UAV. The communication unit consumes energy for signal transmission and acquisition through RF, baseband, and protocol processing operations. It has been shown that the communication energy (a few Watts) can be ignored when compared to the propulsion energy (a few hundred Watts). We can thus assume that the on-board battery is drained by the propulsion energy requirement [44].

The propulsion energy has two phases: the mobile phase and the hovering phase. The mobility phase of the UAV includes horizontal and vertical movement. In the considered scenario, the UAVs carrying the ABSs take off from a building or

tower and fly in the horizontal direction without changing the altitude. The ABS operation is divided into R time slots with a maximum time slot duration of T. For each time slot t, the k^{th} ABS flies in horizontal direction $\chi_k^t \in [0, 2\pi]$ with a constant speed v for a period of time $t_{\mathbb{M}} \leq t_{\mathbb{M}}^{max}$ and traverses a distance of $d_k^t \in [0, d_{max}]$. Parameter $t_{\mathbb{M}}^{max} < T$ is the maximum allowed time for the mobile phase and d_{max} is the maximum traversed distance time slot t. After reaching its desired position, the ABS hovers at that location until the end of the time slot period $t_{\mathbb{H}} \geq t_{\mathbb{H}}^{min}$ while serving a wireless user. Parameter $t_{\mathbb{H}}^{min} = T - t_{\mathbb{M}}^{max}$ is the minimum required time for providing wireless access to a ground UE.

The total energy consumption of the UAV operation per time slot has two components, corresponding to the mobility and hovering phases:

$$PC^t = PC_{\mathbb{M}}^t + PC_{\mathbb{H}}^t. \tag{11}$$

The UAV power consumption during the mobility phase for a given velocity in the horizontal direction χ_t^k in the t^{th} time slot is given by

$$PC_{\mathbb{M}}^t = W\ v_i, \tag{12}$$

where $W = m_u\ g_a$ is the weight of the ABS (UAV with payload) in Newton, m_u is the mass of the ABS, g_a is the gravitational acceleration, and v_i is the induced velocity calculated based on (7.10) of [45]. Specifically,

$$v_i = \frac{W}{\sqrt{2}\ \gamma\ a}\ \frac{1}{\sqrt{v^2 + \sqrt{v^4 + (\frac{W}{\gamma\ a})^2}}}, \tag{13}$$

where γ is the air density a in the total area of the UAV rotor disks and v the UAV speed.

The horizontal speed of the ABS is zero while hovering and providing wireless access. The energy consumption for hovering during time slot t can be simplified from (13) as follows:

$$\begin{aligned} PC_{\mathbb{H}}^t &= W\ v_i^h, \\ v_i^h &= \frac{W}{\sqrt{2}\ \gamma\ a}\ \frac{1}{\sqrt{\frac{W}{\gamma\ a}}}, \\ PC_{\mathbb{H}}^t &= W\ \frac{W}{\sqrt{2}\ \gamma\ a}\ \frac{1}{\sqrt{\frac{W}{\gamma\ a}}} = \sqrt{\frac{W^3}{2\ \gamma\ a}}. \end{aligned} \tag{14}$$

Equations (12) to (14) evince that the ABS's on-board battery is drained more during hovering than during horizontal flight movement.

By the end of time slot t, the residual on-board battery energy of the k^{th} ABS is calculated as

$$J_{k_f}^t = J_{k_0}^t - PC^t, \tag{15}$$

where $J_{k_0}^t$ is the initial energy level of the k^{th} ABS at the start of time slot t. The remaining energy at the end of each time slot is compared against threshold J_{th} to determine whether the k^{th} ABS is able to continue providing wireless access to ground users or if it needs to quit its operation and move back to the recharging station. The same process applies to all operating ABSs.

IV. Problem Formulation

The objective of this paper is to maximize the accumulative number of served ground users with the available ABSs with practical energy constraints while establishing secure communication links that minimize the effect of eavesdropping. This requires determining the 3D trajectories for the deployed ABSs and dynamically adjusting the planned 3D trajectory when any of the deployed ABS needs to quit its operation for recharging. Each ABS is constrained by its on-board battery of limited capacity. Higher-capacity batteries are heavier and require more propulsion energy [46]. The ABS's time of operation is therefore bound and it may quit before completing the task. One or multiple ABSs in the deployed ABS network may quit before the overall mission is completed.

The optimization problem for serving users within the coverage area of K ABSs in R time slots is formulated as

$$\mathrm{P}: \max_{\boldsymbol{x_k^t}, \boldsymbol{y_k^t}} \sum_{t=1}^{R} \boldsymbol{\xi} = \sum_{t=1}^{R} \Big(\sum_{i=1}^{M} \eta_i^t \Big),$$

where (16.a)

$$\eta_i^t = \begin{cases} 1 & SC_i^t > Sc_{th}, \\ 0 & SC_i^t < Sc_{th}, \end{cases}$$

$$s.t. \qquad 0 \leq \boldsymbol{x_k^t} \leq L, \qquad \forall k = 1, \cdots, K, \tag{16.b}$$

$$0 \leq \boldsymbol{y_k^t} \leq L, \qquad \forall k = 1, \cdots, K. \tag{16.c}$$

The $\sum_{t=1}^{R} \boldsymbol{\xi}$ term denotes the accumulative number of served users that satisfy the QoS requirement over the period of R time slots. The QoS requirement is defined here as a secrecy rate threshold. We define $\eta_i^t \in \{0, 1\}$ as a binary variable that indicates whether the i^{th} currently served user in time slot t has achieved the QoS requirement ($\eta_i^t = 1$) or not ($\eta_i^t = 0$). The optimization of the trajectory of the k^{th} ABS is bounded within a 2D plane.

V. Proposed Solution

The optimization problem involves the trajectory optimization for multiple UAVs with energy constraints. The binary variable η of the optimization problem (16.a) makes the problem hard to solve because it involves integer constraints. In addition, η is a non-convex constraint with respect to the UAV trajectory. Applying traditional optimization tools to solve this problem would incur a high computational complexity. Most of the traditional solutions to equivalent multi-parameter optimization problems are iterative and alternately optimize the parameters to reach suboptimal results [47].

It is important that the solution to this problem be of low complexity and scalable to more UEs, ABSs, and larger areas for practical reasons. We therefore propose a multi-trajectory design algorithm by defining a transition process based on the current state of the system. Since the next system state is independent of the previous states and actions, the process can be modeled as a Markov decision process (MDP). This facilitates applying a reinforcement learning (RL) algorithm without requiring the knowledge of the system model to find

the optimal coordinates $\boldsymbol{x_k^t}$ and $\boldsymbol{y_k^t}$ for each deployed ABS in the t^{th} time slot.

In what follows we first describe the DRL model and define the states, the actions, and the reward. Then we introduce the proposed DDPG algorithm for the DRL model to maximize the accumulative number of UEs that are served while satisfying their security-driven QoS requirements. The reason for choosing DDPG is that it employs two DNNs—the actor network and the critic network—that can efficiently process the highly dimensional state-action space.

A. Deep Reinforcement Learning Model

The main components of the MDP of any sequential problem for the RL agent are the state space $\mathcal{S}$, the action space $\mathcal{A}$, the reward space $\mathcal{R}$, and the transition probability space $\mathcal{T}$, i.e., $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{T})$. In each time slot t, the current state $s^t \in \mathcal{S}$ is monitored by the RL agent, which takes a specific action $a^t \in \mathcal{A}$ based on the current policy parameters. After the action is performed, the agent transitions to the new state s^{t+1} following the transition probability $\mathcal{T}(s^{t+1}|s^t, a^t)$. Because it is difficult and often impractical to accurately model the deployment environment, we adopt RL to attain an optimal action based on the received instantaneous reward and the state transitions. This allows us to neglect $\mathcal{T}$ and focus on carefully identifying the states, the actions, and the reward as follows.

State: The set of states holds the various observations characterizing the environment. It can be modeled as $\mathcal{S} = \{s^1, s^2, ..., s^t, .., s^T\}$, where each state s^t encapsulates two parts for each ABS: the current location and the residual energy level of the on-board battery. The instantaneous location of the ABS is captured by coordinates $\boldsymbol{x_k^t}$ and $\boldsymbol{y_k^t}$. These are used by the learning agent to find the number of ground UEs that are located within the coverage area of the ABS in time slot t. The residual energy level $J_{k_f}^t$ in time slot t informs the agent of the current status of the ABS node lineup to be able to adjust the tasks for the remaining ABSs. The individual state observation for the k^{th} ABS in time slot t is thus given by $s^t = \{\boldsymbol{x_k^t}, \boldsymbol{y_k^t}, J_{k_f}^t\}$. The state space for all ABSs in any time slot can then be represented as

$$\mathcal{S}^t = \{\boldsymbol{x_1^t}, \boldsymbol{x_2^t}, .., \boldsymbol{x_K^t}, \boldsymbol{y_1^t}, \boldsymbol{y_2^t}, .., \boldsymbol{y_K^t}, J_{1_f}^t, J_{2_f}^t, .., J_{K_f}^t\}. \quad (17)$$

Action: The decisions that the DRL agent can take to initiate the changeover from the current state to the following state are provided by the action space. We define a set of actions that the agent takes as $\mathcal{A} = \{a^1, a^2, ..., a^t, .., a^T\}$. Any action a_k^t accommodates two parts: the flight distance of the k^{th} ABS $\boldsymbol{d_k^t}$ and the flight direction $\boldsymbol{\chi_k^t}$. Therefore, action a^t for the k^{th} ABS can be expressed as $\{\boldsymbol{d_k^t}, \boldsymbol{\chi_k^t}\}$ and the action space for all ABSs in any time slot as

$$\mathcal{A}^t = \{\boldsymbol{d_1^t}, \boldsymbol{d_2^t}, .., \boldsymbol{d_K^t}, \boldsymbol{\chi_1^t}, \boldsymbol{\chi_2^t}, .., \boldsymbol{\chi_k^t}\}. \quad (18)$$

Reward: After the agent takes action a^t in state s^t at time t, the agent gets reward $r_t(s^t, a^t)$. For our system, we define the reward function as the accumulative number of served UEs whose secrecy capacity goals are met over time period R:

$$r^t(s^t, a^t) = \sum_{t=1}^{R} \boldsymbol{\xi}. \quad (19)$$

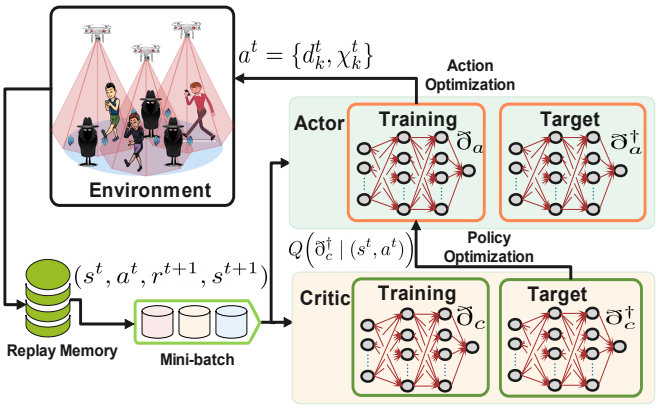


Fig. 2: Block diagram of the proposed DDPG architecture.

B. Deep Deterministic Policy Gradient

For more complex dynamic environments with underlying continuous actions, deep Q-network (DQN) agents may not be the most suitable approximators for achieving the optimal policy. Unlike DQN agents, DDPG agents can model continuous actions with high-dimensional continuous action spaces [48]. The DDPG is based on the deterministic policy gradient (DPG) algorithm [49], which is capable of handling continuous action spaces with discrete state spaces. It is designed by coupling the DPG with neural networks for enabling it to efficiently handle continuous state and action spaces.

The DDPG algorithm is particularly well suited for problems with continuous action spaces. It uses a deterministic policy, which maps states to specific actions rather than mapping probability distributions to actions. This allows for a more efficient exploration of the action space, which can be especially useful for effective learning and decision making in complex and high-dimensional continuous environments.

The DDPG architecture features two DNNs, the actor and the critic networks, which jointly estimate a deterministic policy that maximizes the long-term reward in a DRL setting where the environment is modeled as a continuous state space [50]. Fig. 2 illustrates this.

The DDPG uses a method called off-policy learning, which means it can learn from the past as well as from the present actions. In other words, it can learn from a set of past experiences as well as from the current state. This can help stabilize the learning process and improve the overall performance of the agent. The actor network learns a policy parameterized by the action value function while the critic network estimates the state-action value function, which is then used to update the actor's policy parameters. The joint estimation of the policy and action-selection function achieved by the actor and the critic network is called the actor-critic policy network. This scheme is different from traditional DQN approaches in which the policy is learned independently of the action selection function or actor network.

The advantage that the DDPG has over traditional DQN methods is that it provides a more principled way of learning policies that capture the dynamics within a complex system without requiring explicit knowledge of the system dynamics while avoiding locally optimal solutions. Similar to a DQN, both the actor and critic networks encompass two DNNs—the

training and the target networks—for the purpose of improving the stability [51]. However, the manner in which the target networks are updated is slightly different from the DQN approach. Rather than directly copying the parameters from the learning network every fixed number of steps, the parameters of the target network are updated softly by gradually updating the parameters of the target networks towards the parameters of the learning networks.

Algorithm 1: DDPG-empowered multi-ABS deployment for providing secure wireless access to terrestrial users.

Input $(x_i, y_i) \forall i, J_{k_0}^0 \forall k, J_{th}, Sc_{th}$
Initialize $\mathcal{M}$ with capacity $\aleph$, ζ, $\eth_c$, $\eth_c^\dagger = \eth_c$, ϱ_c, $\eth_a$, $\eth_a^\dagger = \eth_a$, ϱ_a
for *episode = 1, 2, ..., N* **do**
  Obtain the initial state s^0;
  for *t = 1, 2, ..., R* **do**
    From the training actor network, acquire $a^t = \{\boldsymbol{d_k^t}, \boldsymbol{\chi_k^t}\}$;
    Observe next state s^{t+1} *given* a^t;
    for ABS_k, *k= 1, 2, ..., K* **do**
      if $J_{k_f}^t > J_{th}$ **then**
        Obtain s_k^{t+1} *based on the* s_k^t *and* a^t;
      else
        Define $s_k^{t+1} = s_k^t$;
        Exclude the ABS_k *at the calculations of the instant reward* r^{t+1};
      end
      if $(x_i^{t+1}, y_i^{t+1}) > (L, L)$ **then**
        Define $(x_i^{t+1}, y_i^{t+1}) = (x_i^t, y_i^t)$;
      end
    end
    Calculate r^{t+1};
    Store experience $e^t = (s^t, a^t, r^t, s^{t+1})$ *in* $\mathcal{M}$
    Obtain $Q\Big(\eth_c \mid (s^t, a^t)\Big)$ *from the training critic network*
    Calculate $\ell(\eth_c)$ *via eq. (26)*
    Calculate $\Delta_{\eth_c}\ell(\eth_c)$, $\Delta_a Q\Big(\eth_c^\dagger \mid (s^t, a^t)\Big)$, $\Delta_{\eth_a}\mho(\eth_a \mid s^t)$
    Update critic and actor training networks $\eth_c$ *and* $\eth_a$
    Update critic and actor target networks $\eth_c^\dagger$ *and* $\eth_a^\dagger$ *after* ε *steps*
    Train the DNN network with s^{t+1} *as input*
  end
end
Result: Optimal $\boldsymbol{d_k}$ and $\boldsymbol{\chi_k}$, $\forall k$.

The updates of the training critic network are obtained as

$$\eth_c = \eth_c - \varrho_c \ \Delta_{\eth_c}\ell(\eth_c), \tag{25}$$

where $\eth_c$ captures the weights of the network, ϱ_c the learning rate, and $\Delta_{\eth_c^t}$ the gradient. Parameter $\ell(\eth_c)$ is the loss function of the training critic network. It is calculated as

$$\ell(\eth_c) = \mathbb{E}\Bigg[\Bigg(\Big[r^t + \Pi \times \ Q\Big(\eth_c^\dagger \mid (s^{t+1}, \tilde{a})\Big)\Big] - \Big[Q\Big(\eth_c \mid (s^t, a^t)\Big)\Big]\Bigg)^2\Bigg], \tag{26}$$

where $\tilde{a}$ is the action of the agent that follows the deterministic policy drafted by the target actor network and $\eth_c^\dagger$ captures the network's weights.

It is important to note that the update of the training network is more frequent than the update of the target network. The update of the training actor network is obtained as

$$\eth_a = \eth_a - \varrho_a \ \Delta_a Q\Big(\eth_c^\dagger \mid (s^t, a^t)\Big) \ \Delta_{\eth_a}\mho(\eth_a \mid s^t), \tag{27}$$

where $\eth_a$ symbolizes the weights of the network $\mho(\eth_a \mid s^t)$, ϱ_a the learning rate, $\Delta_a Q\Big(\eth_c^\dagger \mid (s^t, a^t)\Big)$ the gradient of the target critic network output with reference to the taken action, and $\Delta_{\eth_a}\ \mho(\eth_a \mid s^t)$ the gradient of the training actor network with respect to $\eth_a$.

Following the updates of the training critic and training actor networks, the target critic and target actor network updates are obtained as

$$\eth_c^\dagger \leftarrow \tau_c \ \eth_c + (1 - \eta_c) \ \eth_c, \tag{28}$$

$$\eth_a^\dagger \leftarrow \tau_a \ \eth_a + (1 - \eta_a) \ \eth_a, \tag{29}$$

where τ_c and τ_a are the learning rates for updating the critic and actor networks, respectively. Algorithm 1 provides the pseudocode for the proposed DDPG algorithm.

C. Computational Complexity

The main factor that determines the computational complexity per learning episode defined in Line 3 of Algorithm 1 is the composition of the employed DQN architecture and the retrieval of the corresponding experience (s^t, a^t, r^t, s^{t+1}). The proposed DDPG is composed of two DQNs for the actor and critic networks and both of them have almost the same structure except for the number of states of the input layers.

For any given DQN structure of the actor-critic network, let us denote the total number of fully connected hidden layers as $\hbar^f$, where the ℓ^{th} hidden layer is composed of N'_ℓ neurons. Based on (17), the size of the state space is $3 \times K$, which corresponds to the input layer neurons. The size of the action space is $2 \times K$ as defined in (18), and the output layer has therefore $2 \times K$ neurons. Based on the number of neurons of the input and output layers, we can define the weights of the first and last hidden layer as $(3 \times K) \times N'_i$ and $(2 \times K) \times N'_o$, respectively. The number of weights in the hidden layers can be calculated as $\big(\sum_{\ell=2}^{\hbar^f} N'_{\ell-1} \times N'_\ell\big)$. Consequently, the total number of weights that need to be updated for either the actor or the critic network can be expressed as

$$(3 \times K) \times N'_i + (2 \times K) \times N'_o + \big(\sum_{\ell=2}^{\hbar^f} N'_{\ell-1} \times N'_\ell\big).$$

By defining the computational complexity of one training iteration for a single neuron weight as $\mho$, a training iteration for all weights of the actor or critic DQN becomes

$$\mho\big((3 \times K) \times N'_i + (2 \times K) \times N'_o + \big(\sum_{\ell=2}^{\hbar^f} N'_{\ell-1} \times N'_\ell\big)\big).$$

The other factor that defines the computational complexity of the proposed DDPG algorithm is the acquisition of (s^t, a^t, r^t, s^{t+1}). The acquisition of states s^t and s^{t+1}, and actions a^t, as opposed to the reward r^t, can be directly obtained from the environment and the agents without incurring additional computational complexity. The computational complexity of acquiring r^t based on (16.a) is (RM). The overall computational complexity of the actor or critic DQN then becomes

$$RM + \mho\big((3 \times K) \times N'_i + (2 \times K) \times N'_o + \big(\sum_{\ell=2}^{\hbar^f} N'_{\ell-1} \times N'_\ell\big)\big).$$

Since the structure of the actor DQN is similar to the critic DQN, the total computational complexity of the proposed

TABLE II: Simulation ar

Symbol	Definition	
M	Number of ground users	
E	Number of ground eavesd	
K	Number of ABS nodes	
f_c	Center frequency	
β_0	Path loss at reference dist	
h	Height of the ABS nodes	
ϕ	Aperture angel of the AB	
C	A2G path loss parameter	
U	A2G path loss parameter	
η_{LoS}	Attenuation loss for LoS	
η_{NLoS}	Attenuation loss for NLoS	
B	System bandwidth	
P	Transmit power	
N_0	Spectral power density of	
m_u	Mass of ABS node	
g_a	Gravitational acceleration	9.8 m/s^2
γ	Air density	1.225 kg/m^3
a	Total area of rotor disks	0.18 m^2
v	ABS's velocity	40 km/h
T	Max. duration per time slot	10 s
$t_{\mathbb{M}}^{max}$	Max. duration of mobile phase	9 s
$t_{\mathbb{H}}^{min}$	Min. duration of hovering phase	1 s
d_{max}	Max. distance to fly per time slot	10 m
Sc_{th}	Secrecy capacity threshold	3.55 Mbits
J_{th}	On-board battery threshold to quit	4500 mAh
$\aleph$	Replay memory capacity	10^6
ϱ_c	Learning rate of training critic network	10^{-3}
ϱ_a	Learning rate of training actor network	10^{-3}
τ_c	Learning rate of target critic network	10^{-3}
τ_a	Learning rate of target actor network	10^{-3}
δ_c	Decaying rate of critic network	10^{-3}
δ_a	Decaying rate of actor network	10^{-5}
N	Number of episodes	1000
R	Number of time slots	100
BC	Batch size	64
Π	Discount factor	0.9
ϵ	Number of steps before updating critic and actor target networks	1

DDPG algorithm for optimizing the trajectory of multiple ABSs serving a number of ground users and satisfying their security requirements can be approximated as

$$2(RM+\mho((3\times K)\times N_i'+(2\times K)\times N_o'+(\sum_{\ell=2}^{\hbar^f} N_{\ell-1}'\times N_\ell'))).$$

VI. Simulations and Analyses

The performance of Algorithm 1 is numerically evaluated. The objective of the DDPG scheme is maximizing the accumulative number of ground users that are securely served by the ABS lineup that dynamically changes based on practical limitations of the on-board battery capacity. The simulation is performed in Python with Pytorch 1.3 and a custom-designed environment using OpenAI Gym.

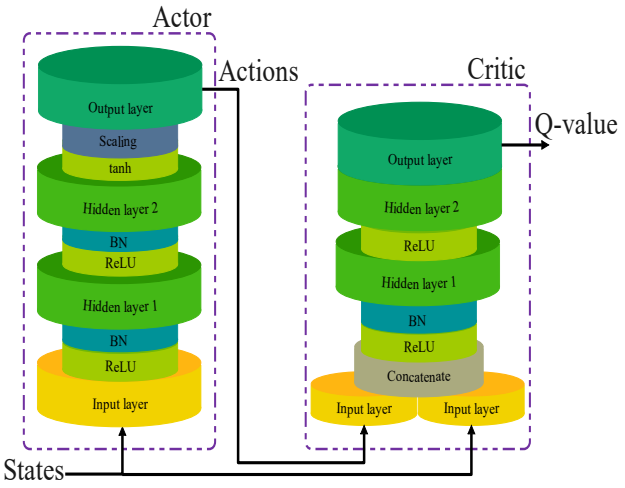


Fig. 3: The DNN design for the actor and critic networks used in the proposed DDPG algorithm (BN–batch normalization).

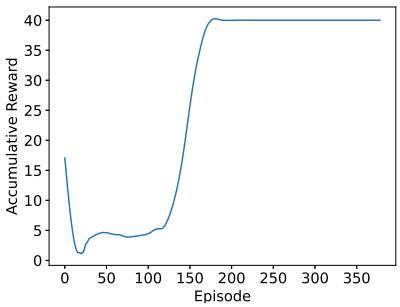
A. Environment Setting

The simulation environment is characterized by $M = 100$ ground users that are randomly distributed within a square area $A = 1000 \times 1000\ m^2$.

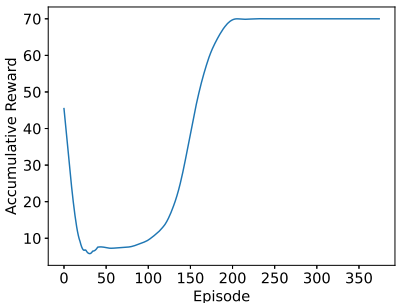
Figure 3 illustrates the structure of the critic and actor DNN designs for the proposed DDPG. The critic and actor DNNs employ the same structure and consist of four layers: the input layer with N_i' neurons, the output layer with N_o' neurons, and two fully connected hidden layers with 564 and 432 neurons. Parameter $N_i' = 3K$ corresponds to the size of the state space and $N_o' = 2K$ corresponds to the size of the action space. All DNNs employ the $ReLU$ activation function across all layers and $Adam$ with adaptive learning as the optimizer, where $\varrho_c^{t+1} = \delta_c \varrho_c^t$ and $\varrho_a^{t+1} = \delta_a \varrho_a^t$ with the decaying rates δ_c an δ_a for the critic and actor networks, respectively. Both $tanh$ and $Scaling$ layers are implemented at the output layer of the actor network while the L_2 regularization is implemented in both networks for suppressing the overfitting problem. Batch normalization helps achieve fast convergence. The input states and produced actions of the actor network are concatenated in one input stream for the critic network.

The wireless communication and energy parameters follow the models presented in Section III. The minimum user rate and secrecy rate requirements—the QoS requirements—are defined as 3.55 Mbps for each user. The hyperparameters of the DDPG algorithm are critical to its performance; therefore, we have handcrafted them to optimize the DDPG performance with respect to the specific objective. Table II summarizes the simulation and DDPG parameters.

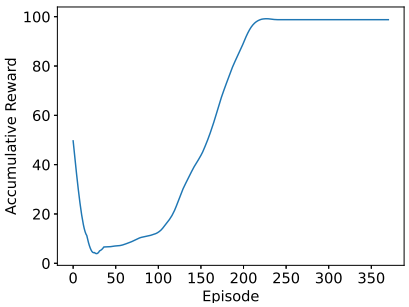
This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3303253



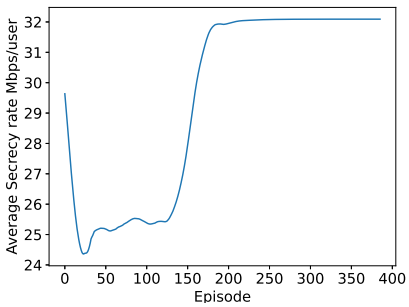
(a) Scenario 1: Accumulative reward.



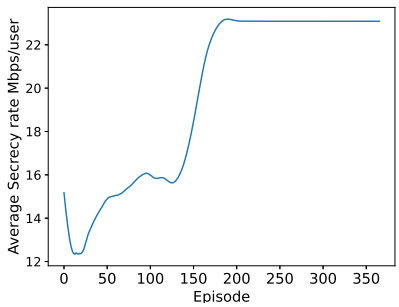
(b) Scenario 2: Accumulative reward.



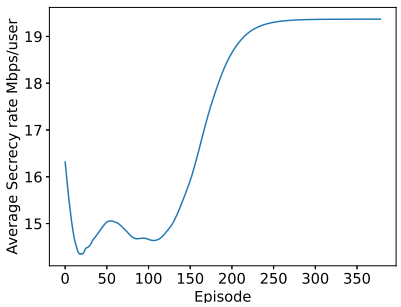
(c) Scenario 3: Accumulative reward.



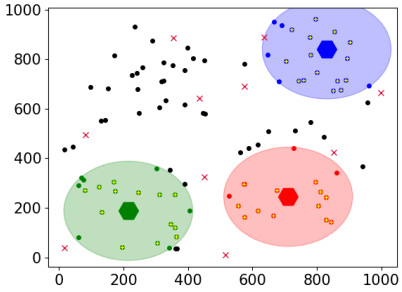
(d) Scenario 1: Average secrecy capacity.



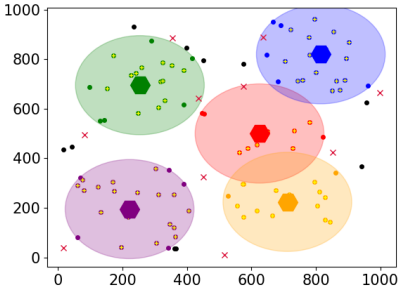
(e) Scenario 2: Average secrecy capacity.



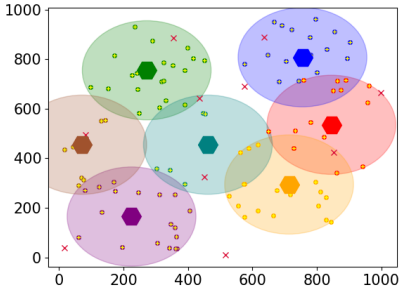
(f) Scenario 3: Average secrecy capacity.



(g) Scenario 1: Final ABS positions and UE distributions.



(h) Scenario 2: Final ABS positions and UE distributions.



(i) Scenario 3: Final ABS positions and UE distributions.

Fig. 4: The accumulative reward (a)-(c), average secrecy capacity (d)-(f), and node deployments (g)-(i) for three scenarios with three, five, and seven remaining ABSs. Circles with a plus in (g)-(i) represent users whose QoS requirements are met, filled circles represent users where the QoS requirement is not met, and crosses represent the eavesdroppers.

B. Overall Performance Evaluation

We evaluate the performance of the proposed solution in three simulation scenarios to gain insights into the performance and capabilities of the proposed multi-ABS communication framework. We assume that there are seven ABSs and that the numbers of active ABSs for completing the task are three, five, and seven, respectively, for the three scenarios. The initial energy levels control the ABS endurance and characterize each scenario. The four and two abandoning ABSs of the first and second scenarios are each initialized with an energy level of $J_0 \leq 1200$, which forces these nodes to stop providing coverage to ground UEs at some point and have the remaining nodes cover for them. On the other hand, the remaining active ABSs in all scenarios are each initialized with an energy level of $30000 \geq J_0 \leq 34000$ to remain active all the time of the task.

Figure 4 presents the accumulative reward of the system over the number of episodes, the average secrecy capacity of the served users by all active ABSs over the number of episodes, and the final deployment positions of the remaining ABSs along with their coverage areas and the user distributions. The accumulative reward functions plotted in Figs. 4 (a), (b), (c) show that as the number of operational aerial nodes increases, the coverage improves, which allows serving more UEs. The accumulated number of served users whose secrecy requirements are met are approximately 40, 70, and 100 for the three scenarios. We obverse that the aerial network can meet the QoS requirements for several UEs for the initial ABS positions. The ABSs aim to increase this number by adjusting its position until reaching the optimal location that enables secure coverage for the highest number of ground users.

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3303253

During the training, the number of served users decreases in the early phase of the learning process and then gradually increases over time as the agents get more experienced by interacting with the environment and adapting to the user and eavesdropper distributions.

As for the security performance of the system, the average secrecy rates across all served users whose secrecy requirements are met by the active ABSs are shown in Fig. 4 (d), (e), (f) for the three scenarios. We observe that in case of fewer ABSs, the algorithm forces the ABSs to focus on the denser groups of ground users that are far from the ground eavesdroppers (Fig. 4 (g)), increasing the average secrecy rate (Fig. 4 (d)) at the cost of serving fewer users (Fig. 4 (a)). In the case where the full ABS lineup is available for supporting the mission during the entire time, the distribution of the ABSs becomes denser and the coverage areas partially overlap (Fig. 4 (i)). The average secrecy rate per served user drops with respect to the previous two scenarios because several eavesdroppers fall within the cell coverage areas and the users also suffer increased inter-cell interference because of limited spectrum availability in practice. However, all users can be served and their QoS requirements satisfied, which is the objective and which shows the value of having all ABSs available. The approach we presented in this study represents a balance between realistic deployment conditions and simulation complexity. Our methodology provides a comprehensive evaluation of both the best-case (with eavesdroppers closer to the cell edges or nonexistent) and worst-case scenarios (with eavesdroppers at locations with favorable channel conditions).

Figures 4 (g), (h), (i) illustrate the final positions of the ABSs with their corresponding coverage areas. Dots symbolize users and crosses eavesdroppers. Black dots identify users that are outside of cellular coverage. The users that are located within the coverage area of an ABS are of the same color as its associated ABS. A link that satisfies the security requirement is identified with a yellow plus sign. Intuitively, an ABS tends to center its position as close as possible to the largest group of UEs without cellular coverage. This facilitates offering higher data rates and meeting the secrecy requirement, provided that the legitimate user channel is better than the wiretap channel of any eavesdropper that is in the coverage area of the same cell.

Figure 5 illustrates the trajectories of five active ABSs that are dispatched in Scenario 2. These trajectories were optimized employing Algorithm 1 to ensure efficient wireless coverage, maximizing the number of served users while satisfying their security-driven QoS requirements in the presence of multiple eavesdroppers.

C. Learning Performance Evaluation

For the learning performance evaluation, we compare the resulting reward of the DDPG algorithm for optimizing the ABS trajectories against a deep DQN-based approach which uses the same objective and the same reward function. The DQN-based solution has the same DNN structure as the critic and the actor networks. We also compare our solution to a baseline technique that randomly chooses the trajectories of the ABSs.

Figure 6 illustrates the accumulative reward functions. It corresponds to the accumulative number of ground users whose QoS requirements are met. Scenario 2 is considered here where seven ABSs are initially dispatched but only five of them are able to complete the mission until the end. Overall we notice a significant increase in the accumulative number of UEs when employing the proposed DDPG solution over the DQN-based algorithm and the baseline. The superiority of DDPG over DQN for this problem comes from the unique capabilities of the actor-critic model, as mentioned in Section V.B, to deal with the continuous state and action spaces that are employed for optimizing the policies that reflect on the agents' chosen actions. On the other hand, the DQN algorithm partitions the continuous input state space into discrete input and output state spaces based on the probability of each action. Thus, for the underlying case with high-dimensional action and state spaces, the training effect of the DQN algorithm

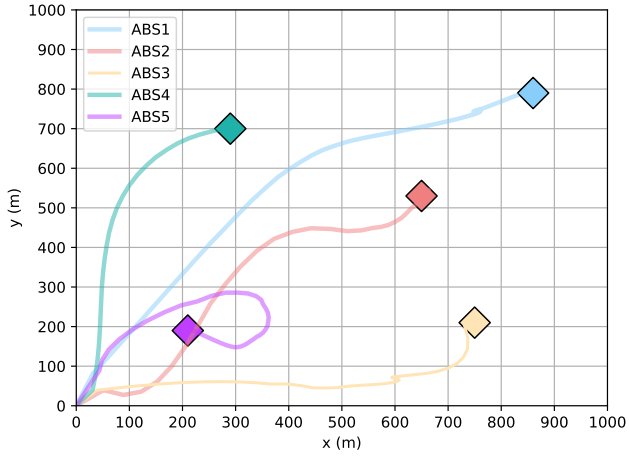


Fig. 5: The trajectories of the active multiple ABS dispatched in scenario 2.

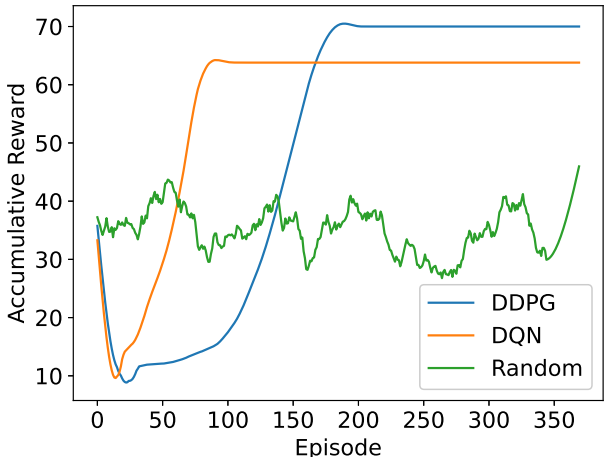


Fig. 6: The accumulative reward comparison.

is less promising than that of the DDPG. Another factor that is contributing to the superiority of DDPG over DQN is the absence of the quantization error because operating over continuous action and state spaces. However, it can be observed that the DQN is faster than the DDPG because of the DDPG's more complex architecture where four DNNs are employed compared to two DNNs of the DQN and because of the fewer hyperparameters required for DQN training.

D. Performance Evaluation As a Function of ABS Configuration and Environment Setting

We finally examine the performance of the proposed framework and DDPG algorithm in different environmental settings, which are characterized by different RF channel conditions and LoS and NLoS probabilities. We consider two environments, urban and dense urban, and two ABS aperture angles, $\phi = 30$ and 60. The environmental conditions are based on two densities of buildings and populations. These environmental conditions affect the A2G communication channel through the parameters C, U, η_{LoS}, and η_{NLoS}. We consider an urban environment, where $C = 9.61$, $U = 0.16$, $\eta_{LoS} = 1$, and $\eta_{NLoS} = 20$, and a dense urban environment, where $C = 12.08$, $U = 0.11$, $\eta_{LoS} = 1.6$, and $\eta_{NLoS} = 23$.

Figure 7 shows the accumulative reward function corresponding to the total number of users whose QoS requirements are met over the UAV flight height for different ABS aperture angles ϕ. The number of securely served users increases linearly with UAV height. Higher altitudes result in broader ground coverage and higher LoS probability. This increases the number of randomly distributed users that fall within the coverage area of each ABS while maximizing the coverage quality to satisfy the secrecy capacity requirement. As expected, better LoS conditions and higher $SINR$ values are observed for the urban than for the dense urban environment.

E. Mobile Multi-User Secure Coverage Performance Evaluation

Here we evaluate the proposed DDPG algorithm's performance under non-static conditions, simulating an environment with continuously changing user positions. More precisely, we simulate the dynamic nature of the environment by updating the UE positions at regular intervals with an approximate average UE speed of 0.8 m/s (pedestrian mobility). We assume the deployment case of five ABSs with the reward function being defined as the accumulative number of users that have their security-driven QoS requirements satisfied. Figure 4(b) shows the result of the equivalent stationary user case. We consider two UAV speeds: fast and slow. In the first simulation, we consider the case where the UAV moves at a high speed of 75 m/time slot toward the location of the UEs. Since we simulate the ground UEs as pedestrians and their speeds are very low compared to the speed of the UAVs, the UE locations do not significantly change while the UAVs are positioning themselves.

Figure 8 shows the reward and convergence behavior of the proposed DDPG solution for 5 ABSs with pedestrian UEs for the fast UAV case. The achieved reward satisfies the security-driven QoS requirements for approximately 47 users after convergence at around 350 episodes. The reward of the proposed DDPG solution for the same deployment case with stationary UEs satisfies approximately 70 users after only 200 episodes (Fig. 4(b)). This is the result of different location-reward mappings between mobile and stationary UEs.

In the next simulation, the UAVs carrying the ABSs travel at a lower speed of 40 m/time slot. The ABSs will thus take longer to reach their target locations and during that time the UE locations and channel states will significantly change. Figure 9 shows the accumulated reward. We notice that the number of served users is around 30 after more than 500 learning episodes. The proposed DDPG solution struggles to fully converge in this scenario even with an increased number of iterations. Its performance is lower with more more fluctuations than for the fast UAV case.

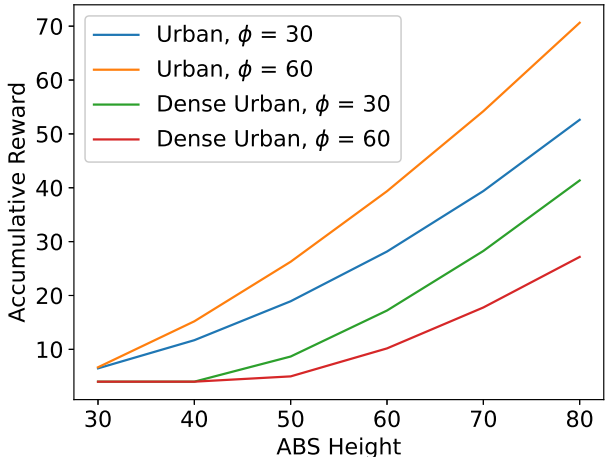


Fig. 7: The achieved accumulative reward for different environments and cell coverage areas as a function of ABS height.

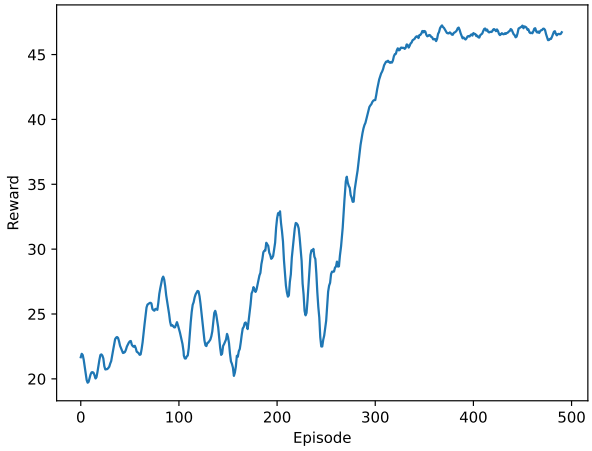


Fig. 8: The achieved accumulative reward for mobile UEs environment for higher ABSs speed.

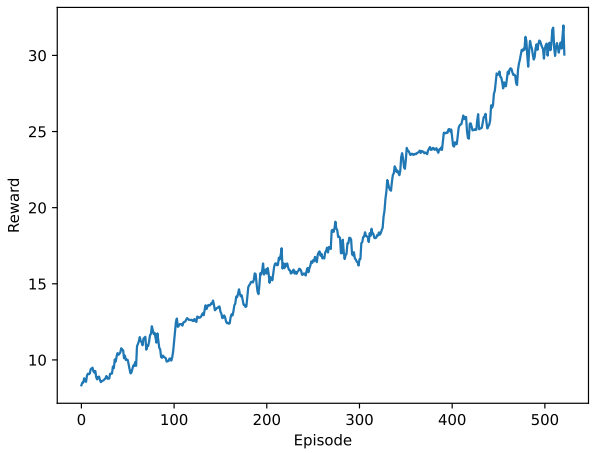


Fig. 9: The achieved accumulative reward for mobile UEs environment for lower ABSs speed.

The outcomes of these additional simulations in terms of convergence speed, solution optimality, and training time show that the proposed centralized DDPG is capable of achieving acceptable performance with slow UE mobility and fast UAV speeds. We plan to further analyze the learning efficiency versus UAV update speed bottleneck in future research to address the practical implications of implementing the proposed system in a variety of communication contexts. We therefore suggest researching a task-driven ABS network that is self-organized and relies on distributed decision-making. Moreover, we will study a distributed and local information based decision-making DDPG solution to solve the problem of providing secure wireless coverage with relatively short channel coherence times in mobile scenarios.

VII. Practical Considerations

In the context of this paper, the training of an agent is based on the data obtained through radio signaling. Modern wireless protocols operate on millisecond subframes, but sub-millisecond levels are envisaged for next generation networks. The inference phase of a trained ML model has a low complexity because it batches a much smaller number of inputs than the training phase [53]; it operates in real time.

Effective model training is essential for the proposed aerial communication network. In continuation we discuss the practical considerations, challenges, and opportunities for realizing it.

A. Training Time

The training time of the DDPG model proposed in this paper depends on a number of factors, including the input data size, complexity and features of the model itself, and the available computational hardware resources. For practical applications, it is necessary to ensure that timely results are produced. At the same time, the accuracy of the proposed model also needs to be guaranteed. It is common to apply data preprocessing techniques to reduce the dimension of the input data and reduce the training time. There are various methods for reducing the dimension of the input data, but among all the techniques that have been proposed, the auto-encoder model is the most widely used [54].

We have implemented the DDPG algorithm in a host client with multiple parallel workers. The parallel workers obtain the same set of policy parameters from the host client and independently update the global model at the host client according to their own experiences by exploring different versions of the environment. The host client then immediately updates the policy parameters for each parallel worker. It takes approximately 4 GB of memory and 633 s to train the model on a standard eight-core Intel i9 general-purpose processor. This time can be considerably reduced by optimizing the implementation for multiple processors or by applying data preprocessing techniques.

We conclude that while the DDPG has a higher training complexity than other ML solutions, it provides better performance for the given problem. Meeting the timing requirements is a matter of implementation. Practical systems will deploy a highly-optimized software implementation and leverage the latest processing technology. A HAPS can carry heavy payloads and offer high computational power. The Stratobus airship, for instance, can accommodate a payload of 450 kg with a power rating of 8 kW [55].

B. Training Convergence

DDPG is an off-policy RL algorithm that uses a critic network to learn an approximate value function and a deterministic policy to select the actions. Ensuring that the critic and actor networks converge to good solutions can be a challenge, particularly if the environment or task is complex or has high-dimensional state and action spaces. If the critic network fails to converge to a good solution, the actor network will become increasingly irrelevant and unstable, and the behavior of the trained agent will be erratic. For this reason, it is important to perform hyperparameter optimization to ensure that the critic and actor networks converge quickly.

A good method for hyperparameter optimization is choosing an initial set of parameters by performing a grid search over a large parameter space and then using the empirical mean of the best-performing models as the initial values for subsequent training iterations. Performing regular hyperparameter optimization during the training process will ensure that the critic and actor networks converge to the global maximum.

Another way to facilitate convergence is to incorporate regularization terms into the loss function of the critic or the actor network. Doing so will ensure that the learned value of the critic network converges to a desirable value and that the actor network learns a policy that is sufficiently generalizable. Regularization terms can also be used to prevent model overfitting and to improve the stability of the learned policies during training. An overfitted DDPG algorithm may rely on the same experiences in each iteration rather than exploring new parts of the state space. We must therefore incorporate a method to reduce the correlations between the sampled experiences within one mini-batch. The asynchronous parallel computing method that is adopted from the asynchronous

advantage actor-critic technique can be employed to stabilize the state exploration of the DDPG.

C. Centralized versus Decentralized Learning

In this paper, multiple agents are trained centrally at the HAPS which delivers the optimized control policy to each ABS. This centralized approach is computational resource intensive, has a high signaling overhead, and is of low resiliency since relying on a single node that may fail or become compromised. Identifying malicious agents from gaining access to the training data will be critical and mechanisms for authentication of nodes and data are need to be in place. A decentralized DDPG architecture leveraging the ABSs would remove the single node of failure and facilitate scalability. An intermediate approach that is worthwhile exploring is leveraging federated learning where different parts of the model training happen at the distributed agents and only the parameters, as opposed to the entire model and data sets, are exchanged with the HAPS.

Decentralized learning can also employ parallel computing at each agent, where a set of workers explore different versions of the environment, inform the host client, which then updates the policy parameters for each parallel worker. The process or merging distributed and parallel learning can be accomplished by dividing the training across a number of ABSs, or agents, then further splitting the data into smaller chunks, and processing each chunk in parallel on multiple processors that are available to each agent.

VIII. Conclusions

The ABS can be employed to provide network coverage on demand, such as in emergency situations and during temporary events. We formulate and optimize the deployment of multiple ABSs that are intelligently dispatched to provide wireless service to as many ground UEs as possible while maximizing the secrecy rate of each link. We employ multiple agents, the ABSs, and introduce a centralized DDPG algorithm on a HAPS for solving the trajectory optimization problem with limited UAV on-board energy resources. We analyze the convergence, complexity, and performance compared to an alternative ML approach and a baseline and show the superiority of the proposed optimization framework. The results show that the available UAV energy has a significant impact on providing broad coverage and meeting the security-based QoS demands. Future work will study the further scalability and decentralized management of ABSs and the coordination of ABSs and ground base stations. One can prototype and validate the here presented techniques on the Aerial Experimentation and Research Platform for Advanced Wireless (AERPAW) [56], which facilitates implementing ABSs with software radios and conducting static and mobile multi-user experiments with unmanned ground vehicles in AERPAW's development environment and outdoor testbed.

References

[1] A. S. Abdalla and V. Marojevic, "Communications standards for unmanned aircraft systems: The 3GPP perspective and research drivers," *IEEE Commun. Standards Mag.*, vol. 5, no. 1, pp. 70–77, March 2021.

[2] B. Li, Z. Fei, and Y. Zhang, "UAV communications for 5G and beyond: recent advances and future trends," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2241–2263, 2019.

[3] F. Qi, X. Zhu, G. Mang, M. Kadoch, and W. Li, "UAV network and IoT in the sky for future smart cities," *IEEE Network*, vol. 33, no. 2, pp. 96–101, 2019.

[4] A. S. Abdalla and V. Marojevic, "Security threats and cellular network procedures for unmanned aircraft systems: Challenges and opportunities," *IEEE Communications Standards Magazine*, vol. 6, no. 4, pp. 104–111, 2022.

[5] R. P. Jover and V. Marojevic, "Security and protocol exploit analysis of the 5G specifications," *IEEE Access*, vol. 7, pp. 24 956–24 963, 2019.

[6] D. Kapetanovic, G. Zheng, and F. Rusek, "Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 21–27, 2015.

[7] A. S. Abdalla, K. Powell, V. Marojevic, and G. Geraci, "UAV-assisted attack prevention, detection, and recovery of 5G networks," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 40–47, 2020.

[8] M. Mozaffari, A. Taleb Zadeh Kasgari, W. Saad, M. Bennis, and M. Debbah, "Beyond 5G with UAVs: Foundations of a 3D wireless cellular network," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 357–372, 2019.

[9] H. Qi, Z. Hu, H. Huang, X. Wen, and Z. Lu, "Energy efficient 3-D UAV control for persistent communication service and fairness: A deep reinforcement learning approach," *IEEE Access*, vol. 8, pp. 53 172–53 184, 2020.

[10] M. Samir, D. Ebrahimi, C. Assi, S. Sharafeddine, and A. Ghrayeb, "Leveraging UAVs for coverage in cell-free vehicular networks: A deep reinforcement learning approach," *IEEE Transactions on Mobile Computing*, vol. 20, no. 9, pp. 2835–2847, 2021.

[11] A. S. Abdalla and V. Marojevic, "DDPG learning for aerial RIS-assisted MU-MISO communications," in *2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2022, pp. 701–706.

[12] A. S. Abdalla, A. Behfarnia, and V. Marojevic, "Aerial base station positioning and power control for securing communications: A deep Q-network approach," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 2470–2475.

[13] W. Wei, J. Wang, Z. Fang, J. Chen, Y. Ren, and Y. Dong, "3U: Joint design of UAV-USV-UUV networks for cooperative target hunting," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 3, pp. 4085–4090, March 2023.

[14] A. S. Abdalla, A. Behfarnia, and V. Marojevic, "UAV trajectory and multi-user beamforming optimization for clustered users against passive eavesdropping attacks with unknown CSI," *IEEE Transactions on Vehicular Technology*, vol. 72, pp. 1–16, 2023.

[15] A. A. Khuwaja, G. Zheng, Y. Chen, and W. Feng, "Optimum deployment of multiple UAVs for coverage area maximization in the presence of co-channel interference," *IEEE Access*, vol. 7, pp. 85 203–85 212, 2019.

[16] H. Wang, H. Zhao, W. Wu, J. Xiong, D. Ma, and J. Wei, "Deployment algorithms of flying base stations: 5G and beyond with UAVs," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 009–10 027, 2019.

[17] X. Zhang and L. Duan, "Fast deployment of UAV networks for optimal wireless coverage," *IEEE Transactions on Mobile Computing*, vol. 18, no. 3, pp. 588–601, 2019.

[18] Y. Zhou, et al., "Improving physical layer security via a UAV friendly jammer for unknown eavesdropper location," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 11 280–11 284, Nov 2018.

[19] G. Sun, N. Li, X. Tao, and H. Wu, "Power allocation in UAV-enabled relaying systems for secure communications," *IEEE Access*, vol. 7, pp. 119 009–119 017, 2019.

[20] C. H. Liu, Z. Chen, J. Tang, J. Xu, and C. Piao, "Energy-efficient UAV control for effective and fair communication coverage: A deep reinforcement learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 9, pp. 2059–2070, 2018.

[21] B. Omoniwa, B. Galkin, and I. Dusparic, "Energy-aware optimization of UAV base stations placement via decentralized multi-agent Q-learning," in *2022 IEEE 19th Annual Consumer Communications Networking Conference (CCNC)*, 2022, pp. 216–222.

[22] S. F. Abedin, M. S. Munir, N. H. Tran, Z. Han, and C. S. Hong, "Data freshness and energy-efficient UAV navigation optimization: A deep reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 9, pp. 5994–6006, 2021.

[23] S.-C. Noh, H.-B. Jeon, and C.-B. Chae, "Energy-efficient deployment of multiple UAVs using ellipse clustering to establish base stations," *IEEE Wireless Communications Letters*, vol. 9, no. 8, pp. 1155–1159, 2020.

[24] F. Malandrino, C.-F. Chiasserini, C. Casetti, L. Chiaraviglio, and A. Senacheribbe, "Planning UAV activities for efficient user coverage in disaster areas," *Elsevier Ad Hoc Networks*, vol. 89, pp. 177–185, 2019.
[25] X. Sun, D. W. K. Ng, Z. Ding, Y. Xu, and Z. Zhong, "Physical layer security in UAV systems: Challenges and opportunities," in *IEEE Wireless Communications*, vol. 26, no. 5, October 2019, pp. 40–47.
[26] Q. Wu, W. Mei, and R. Zhang, "Safeguarding wireless network with UAVs: A physical layer security perspective," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 12–18, October 2019.
[27] B. Shang, V. Marojevic, Y. Yi, A. S. Abdalla, and L. Liu, "Spectrum sharing for UAV communications: Spatial spectrum sensing and open issues," *IEEE Vehicular Technology Magazine*, vol. 15, no. 2, pp. 104–112, 2020.
[28] Q. Wang, Z. Chen, W. Mei, and J. Fang, "Improving physical layer security using UAV-enabled mobile relaying," *IEEE Wireless Communications Letters*, vol. 6, no. 3, pp. 310–313, June 2017.
[29] X. Hou, J. Wang, C. Jiang, X. Zhang, Y. Ren, and M. Debbah, "UAV-enabled covert federated learning," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.
[30] T. Bai, J. Wang, Y. Ren, and L. Hanzo, "Energy-efficient computation offloading for secure UAV-Edge-Computing systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 6074–6087, 2019.
[31] A. Hajijamali Arani, M. M. Azari, P. Hu, Y. Zhu, H. Yanikomeroglu, and S. Safavi-Naeini, "Reinforcement learning for energy-efficient trajectory design of UAVs," *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 9060–9070, 2022.
[32] U. Challita, W. Saad, and C. Bettstetter, "Interference management for cellular-connected UAVs: A deep reinforcement learning approach," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2125–2140, 2019.
[33] A. M. Seid, J. Lu, H. N. Abishu, and T. A. Ayall, "Blockchain-enabled task offloading with energy harvesting in multi-UAV-assisted IoT networks: A multi-agent DRL approach," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 12, pp. 3517–3532, 2022.
[34] B. Galkin, J. Kibilda, and L. A. DaSilva, "Coverage analysis for low-altitude UAV networks in urban environments," in *2017 IEEE Global Communications Conference*, 2017, pp. 1–6.
[35] A. S. Abdalla, B. Shang, V. Marojevic, and L. Liu, "Performance evaluation of aerial relaying systems for improving secrecy in cellular networks," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, 2020, pp. 1–5.
[36] A. Sabri Abdalla, B. Shang, V. Marojevic, and L. Liu, "Securing mobile IoT with unmanned aerial systems," in *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, 2020, pp. 1–6.
[37] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569–572, Dec 2014.
[38] J. Lu, S. Wan, X. Chen, Z. Chen, P. Fan, and K. B. Letaief, "Beyond empirical models: Pattern formation driven placement of UAV base stations," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3641–3655, 2018.
[39] ITU-R, "Rec. P.1410-2 propagation data and prediction methods for the design of terrestrial broadband millimetric radio access systems," *P Series, Radiowave propagation*, 2003.
[40] N. Yang, L. Wang, G. Geraci, M. Elkashlan, J. Yuan, and M. D. Renzo, "Safeguarding 5G wireless communication networks using physical layer security," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 20–27, 2015.
[41] C.-H. Hsieh, J.-Y. Chen, and B.-H. Nien, "Deep learning-based indoor localization using received signal strength and channel state information," *IEEE Access*, vol. 7, pp. 33 256–33 267, 2019.
[42] W. Kui, S. Mao, X. Hei, and F. Li, "Towards accurate indoor localization using channel state information," in *2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, 2018, pp. 1–2.
[43] J. Yan, L. Wan, W. Wei, X. Wu, W.-P. Zhu, and D. P.-K. Lun, "Device-free activity detection and wireless localization based on CNN using channel state information measurement," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24 482–24 494, 2021.
[44] Y. Zeng and R. Zhang, "Energy-efficient UAV communication with trajectory optimization," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3747–3760, 2017.
[45] J. M. Seddon and S. Newman, *Basic helicopter aerodynamics*. John Wiley & Sons, 2011.
[46] A. S. Abdalla, A. Yingst, K. Powell, and V. Marojevic, "Open-source software radio performance for cellular communications research with UAV users," in *94th IEEE Vehicular Technology Conference (VTC2021-Fall)*, 2021, pp. 1–6.
[47] A. S. Abdalla and V. Marojevic, "Securing mobile multiuser transmissions with UAVs in the presence of multiple eavesdroppers," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 11 011–11 016, 2021.
[48] D. Silver, *et al.*, "Deterministic policy gradient algorithms," in *31st International Conference on Machine Learning*, Bejing, China, 22–24 Jun. 2014, pp. 387–395.
[49] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *Proceedings of the 31st International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, E. P. Xing and T. Jebara, Eds., vol. 32, no. 1. Bejing, China: PMLR, 22–24 Jun 2014, pp. 387–395.
[50] A. S. Abdalla and V. Marojevic, "ARIS for safeguarding MISO wireless communications: A deep reinforcement learning approach," in *2022 5th International Conference on Advanced Communication Technologies and Networking (CommNet)*, 2022, pp. 1–6.
[51] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint 1509.02971*, pp. 1–14, 2015.
[52] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," *Proceedings of the IEEE*, vol. 107, no. 12, pp. 2327–2375, 2019.
[53] Y. LeCun, Y. LeCun, Y. Bengio, G. E. Hinton, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[54] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.
[55] Thales, "What's Up With Stratobus?" https://www.thalesgroup.com/en/worldwide/space/news/whats-stratobus, accessed: February 2023.
[56] A. S. Abdalla, A. Yingst, K. Powell, A. Gelonch-Bosch, and V. Marojevic, "Open source software radio platform for research on cellular networked UAVs: It works!" *IEEE Communications Magazine*, vol. 60, no. 2, pp. 60–66, 2022.

Aly Sabri Abdalla received the Ph.D. degree in Electrical and Computer Engineering at Mississippi State University, MS, USA in 2023. He received the B.S. and M.S. degrees in Electronics and Communications Engineering from the Arab Academy for Science Technology and Maritime Transport, Egypt, in 2014 and 2019, respectively. Currently, he is a Postdoctoral Researcher in Electrical and Computer Engineering from Mississippi State University, Mississippi State, MS. Since 2019 he has been a Research Assistant in the Department of Electrical and Computer Engineering at Mississippi State University. His research interests include wireless communication and networking, software radio, spectrum sharing, wireless testbeds and testing, and wireless security with application to mission-critical communications, open radio access network (O-RAN), unmanned aerial vehicles (UAVs), and reconfigurable intelligent surfaces (RISs).

Vuk Marojevic (Senior Member, IEEE) is an associate professor in Electrical and Computer Engineering at Mississippi State University. He received the M.S. degree in electrical engineering from Leibniz University Hannover, Germany, in 2003, and the Ph.D. degree in electrical engineering from Barcelona Tech-UPC, Barcelona, Spain, in 2009. His research interests include mobile communications, software radio, spectrum sharing, wireless testbeds and testing, and wireless security with application to mission-critical communications, O-RAN, and unmanned aircraft systems. Prof. Marojevic is an Editor of the IEEE Transactions on Vehicular Technology, an Associate Editor of IEEE Vehicular Technology Magazine, the Vice Chair of the IEEE VTS AdHoc Committee on Drones, and an Officer of the IEEE ComSoc Aerial Communications Emerging Technology Initiative.