EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

Positioning and power optimisation for UAV-assisted networks in the presence of eavesdroppers: a multi-armed bandit approach

Xavier Alejandro Flores Cabezas*, Diana Pamela Moya Osorio and Matti Latva-aho

*Correspondence:
xavier.florescabezas@oulu.fi

Centre for Wireless Communications, University of Oulu, Oulu, Finland

Abstract

Unmanned aerial vehicles (UAVs) are becoming increasingly attractive for the ambitious expectations for 5G and beyond networks due to their several benefits. Indeed, UAV-assisted communications introduce a new range of challenges and opportunities regarding the security of these networks. Thus, in this paper we explore the opportunities that UAVs can provide for physical layer security solutions. Particularly, we analyse the secrecy performance of a ground wireless communication network assisted by *N* friendly UAV jammers in the presence of an eavesdropper. To tackle the secrecy performance of this system, we introduce a new area-based metric, the weighted secrecy coverage (WSC), that measures the improvement on the secrecy performance of a system over a certain physical area given by the introduction of friendly jamming. Herein, the optimal 3D positioning of the UAVs and the power allocation is addressed in order to maximise the WSC. For that purpose, we provide a reinforcement learning-based solution by modelling the positioning problem as a multi-armed bandit problem over three positioning variables for the UAVs: angle, height and orbit radius. Our results show that the proposed algorithm improves the secrecy of the system over time in terms of the WSC, and it converges into a stable state close to the exhaustive search solution for discretised actions, where there is a trade-off between expediency of the positioning of the UAVs to positions of better secrecy outcome and energy consumption.

Keywords: Friendly jamming, Multi-armed bandit, Physical layer security, Secrecy outage probability, Reinforcement learning, UAV-assisted communications

1 Introduction

Over the last years, the interest on the integration of unmanned aerial vehicles (UAVs) to cellular networks as new aerial nodes has exponentially increased [1, 2]. The advancements on cellular technologies with the fifth generation of wireless networks (5G) and the expected almost ubiquitous accessibility to these networks make UAVs to be considered as a crucial component of the sixth generation of wireless networks (6G). UAVs are expected to be deployed as aerial base stations (ABSs), access points (APs), or relays to assist terrestrial communications within the so-called UAV-assisted communications. In this way, the advantageous characteristics of UAV-assisted communications, such as on-demand deployment, low-cost, flexibility in network reconfiguration, and high chance of line-of-sight (LoS) links, promote the emerging of a number of novel use cases and applications in different contexts such as disaster areas, smart agriculture, traffic control, search and rescue, package delivery, among others [3, 4].

Nonetheless, with all the expected technological and architectural progress for 6G, especially with the integration of artificial intelligence (AI) into the network operation and management and the advancements of quantum computing with its potential to break pre-quantum cryptographic methods, security becomes a highly critical aspect in order to guarantee the levels of resilience and reliability planned for 6G [1]. Physical layer security (PLS) has attracted increased attention as a mechanism to provide more robust and quantum-resistant protection to wireless networks by relying on the unique physical properties of the random and noisy wireless channels to enhance confidentiality in a flexible and adaptive manner. Thus, PLS can find a new horizon in the 6G era, especially for the constrained scenarios of Internet of things (IoT) applications [5, 6].

Under these circumstances, UAVs can also be exploited for the design of secure solutions in UAV-assisted communications via PLS; thus, the challenges and opportunities for preventing passive and active attacks in wireless networks have been recently discussed in [7]. On the one hand, UAV-assisted communications are more vulnerable to eavesdropping and jamming attacks due to their strong LoS links compared to communication between ground nodes; on the other hand, UAVs can also be used to launch more effective attacks [7]. Therefore, there is a vast research area to be exploited for providing secure wireless communications in the UAV era, and some have been already reported in the literature [8–16].

Particularly, the introduction of UAV nodes acting as friendly jammers in order to improve the secrecy performance of wireless networks has recently risen special attention. For instance, in [8], an optimal three-dimensional (3D) deployment and jamming power of UAV-based jammer was proposed to improve the secrecy performance of a legitimate transmission between a pair of nodes for unknown eavesdropper location. In [9], the secrecy outage probability (SOP) of a UAV-based mmWave relay network in the presence of multiple eavesdroppers is investigated. Two scenarios are considered, with and without cooperative jamming, which is introduced via the destination and an external UAV. In [10], the authors studied the secrecy performance of a non-orthogonal multiple access (NOMA)-based scheme in a mmWave UAV-assisted wireless network by considering a protected-zone approach. In [11], the existence of an optimal UAV jammer location on a network with multiple eavesdroppers was proved, and the impact of the density of eavesdroppers, the transmission power of the UAV jammer and the density of UAV jammers on the optimal location was investigated.

Moreover, the joint optimisation of the transmit power and the trajectory of a UAV-based friendly jammer in a three-dimensional space was investigated in [12]. Therein, the problem of average achievable secrecy rate maximisation of the secondary system was investigated for a cognitive relay network by considering the imperfect location information of ground nodes, that is the eavesdropper, secondary receiver and primary receiver. Also in [13], the secrecy rate maximisation problem of a mobile user over all time slots is studied by considering a dual UAV-enabled secure communication system, where one UAV sends confidential information, while the other serves as friendly jammer. Both UAVs are considered to be energy-constrained devices, and the location information of eavesdroppers is imperfect. Therein, the optimisation problem is solved by jointly designing the 3D trajectory of UAVs and the time allocated for recharging and jamming sending under constraints such as the maximum UAV speed, UAV collision avoidance, UAV positioning error and UAV energy harvesting.

More recently, machine learning (ML) approaches have been considered in order to tackle the intricacy of the optimisation problems related to UAV-assisted scenarios, where there are a number of coupled variables and the complexity on the characteristics of the problems would lead to exhaustive searches or complex operations. Particularly, in [14], a deep reinforcement learning (RL) algorithm is proposed to jointly optimise the active beamforming of the UAV, the coefficients of a reflective intelligent surface (RIS) elements, and the UAV trajectory to maximise the sum secrecy rate of the legitimate users in the presence of multiple eavesdroppers of a mmWave UAV communication assisted by a RIS under imperfect channel state information (CSI). Besides, in [15], a deep learning method is employed to optimise a 3D beamformer for the transmission of confidential signal and friendly jamming in order to maximise the average secrecy rate by considering partial CSI of the legitimate UAV and eavesdropping UAV. Also, the authors in [16] considered UAV jammers assisting a legitimate transmission between a UAV and ground nodes in the presence of ground eavesdroppers. Therein, a multi-agent deep RL approach was used to maximise the secure capacity by jointly optimising the trajectory of UAVs, the transmit power of the UAV transmitter and the jamming power of the UAV jammers.

All in all, the employment of friendly jamming has been widely accepted as an effective manner to enable confidential transmissions in wireless networks. However, the effectiveness of friendly jamming schemes is in most cases harnessed to the perfect or partial knowledge of the CSI of the legitimate and eavesdropping links, which is hard to obtain in practice. To dive into the characterisation of the effectiveness of friendly jamming in wireless networks, the authors in [17] proposed two novel area-based metrics, the jamming coverage and the jamming efficiency, in order to provide insights into the design of optimal jamming configurations by considering different levels of CSI knowledge. Later, in [18], we considered a UAV-assisted friendly jamming scheme in a wireless network in the presence of eavesdroppers. Based on the area-based metrics in [17], a novel metric, the weighted secrecy coverage (WSC), was proposed to give a better insight into the impact of friendly jamming. Thus, the optimal positioning of two UAV jammers was tackled in order to maximise the WSC. Further in [19], we proposed a zero-forcing precoding scheme for the two friendly UAV jammers in order to enhance the efficiency of the friendly jamming, thus enhancing the WSC.

Inspired by [17] and based on [18], we will advance on the state-of-the-art by studying a UAV-assisted wireless network, where a number of UAVs assist a legitimate ground communication between a pair of ground nodes in a confined region on a fading environment, and the 3D positioning of the UAVs is optimised in order to maximise the WSC metric. For that purpose, we model our optimisation problem as a multi-armed

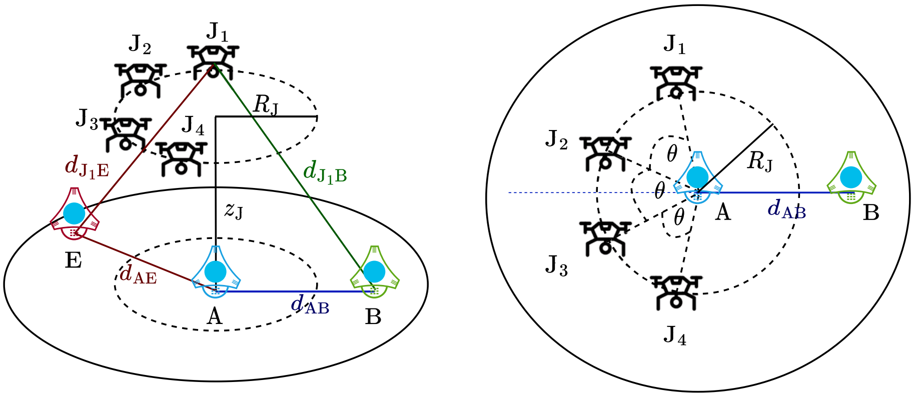

Fig. 1 System model

bandit (MAB) problem and provide a RL-based solution.[1] Thus, the contributions of the paper are the following:

- We derive an expression for the SOP of the proposed system, with Rayleigh fading ground channels and air-to-ground (A2G) channels with a Rician fading LoS component and a Rayleigh fading NLoS component, for a wireless wiretap channel with N friendly UAV jammers.
- We propose a time frame-based algorithm to optimise the WSC of the system as three independent multi-armed bandit problems, one for each positioning variable of the UAVs.
- Monte Carlo simulations are performed to validate our theoretical expressions and to evaluate the performance of the algorithm in terms of the WSC and the energy consumption of the system, which show a steady convergence to an optimal result.

The remainder of the paper is organised as follows. In Sect. 2.1, the investigated system model is presented. In Sect. 2.2, the considered secrecy metrics are introduced, namely the secrecy capacity, SOP, secrecy impact of jamming metric, and the WSC metric. In Sect. 2.3, the formulation of the WSC maximisation problem is shown. In Sect. 3, the numerical results are presented. Finally, in Sect. 4, the conclusions of this work are presented.

Notation Throughout this paper, and unless stated otherwise, $|\cdot|$ denotes the absolute value, $\mathbb{E}_x[\cdot]$ denotes the expectation over random variable x, if x is missing the argument is considered the random variable, $\mathcal{N}(x, \sigma^2)$ denotes a normal distribution of mean x and variance σ^2, $\Pr[\cdot]$ denotes probability, $F_x(\cdot)$ is the cumulative density function (CDF) of random variable x, while $f_x(\cdot)$ is its probability density function (PDF) and $\iint_S$ is a double integral over surface S.

[1] This is an extended version of the paper presented at the 2021 Joint EuCNC and 6G Summit, thus part of this paper was published in [18].

2 Methods

2.1 System model

Consider the system illustrated in Fig. 1, comprised of a legitimate pair of ground nodes, the transmitter Alice (A) and the receiver Bob (B), who establish an open wireless link to send private information from A to B. They are confined on a circular area S of radius R_A around A. Within S, the presence of an illegitimate node Eve (E) is established, trying to leak the information from the legitimate transmission shared through the wireless medium. It is assumed that E is a passive eavesdropper located within the region S, but its exact position and available resources are unknown. A is located at the origin of coordinates (0, 0, 0) and B is located along the x-axis at $(d_{\mathrm{AB}}, 0, 0)$, without losing generality. To improve the secrecy performance of the system, N UAVs, $\{\mathrm{J_i}\}_{i\in\{1,\ldots,N\}}$ are deployed to act as friendly jammers by emitting pseudorandom noise isotropically in order to prevent E from leaking information. The jammers are positioned at a common height z_{J} and within a circular orbit of radius R_{J} around A, at angular positions $\theta_{\mathrm{J_i}}$ with $i \in \{1, \ldots, N\}$. We assume that the estimate of the radial position of B with respect to A is unreliable; thus, we model the distance between A and B as a random Gaussian variable with the actual distance d_{AB} being the mean of the estimate (unbiased), and a given uncertainty σ_{AB}, $\widehat{d}_{\mathrm{AB}} \sim \mathcal{N}(d_{\mathrm{AB}}, \sigma^2_{\mathrm{AB}})$, where $\widehat{d}_{\mathrm{AB}}$ is the estimate of the distance between A and B.

2.1.1 Ground channels

There are two ground channels to consider between ground nodes, one between A and B and the other between A and E. Both channels are considered to undergo Rayleigh fading and are subject to additive white Gaussian noise (AWGN) with mean power N_0. Then, the corresponding channel coefficients are h_{AB} and h_{AE}, and the respective channel gains are $|h_{\mathrm{AB}}|^2$ and $|h_{\mathrm{AE}}|^2$. For a node $\mathrm{U} \in \{\mathrm{B}, \mathrm{E}\}$, the channel coefficient h_{AU} is an independent complex circularly symmetric Gaussian random variable with a channel gain of $g_{\mathrm{AU}} = |h_{\mathrm{AU}}|^2$ with a scale parameter of $\Omega_{\mathrm{AU}} = \mathbb{E}\left[|h_{\mathrm{AU}}|^2\right] = \gamma_{\mathrm{A}} d_{\mathrm{AU}}^{-\alpha_G}$, where d_{AU} is the distance between A and node U, α_G is the path loss exponent for the ground links and γ_{A} is the transmit SNR of A given by $\gamma_{\mathrm{A}} = P_{\mathrm{A}}/N_0$ with P_{A} as the transmit power of A.

2.1.2 Air-to-ground channels

There are two air-to-ground channels for each UAV jammer, one between the UAV and B and the other between the UAV and E. The channel coefficients for those links are given by $\mathrm{h}_{\mathrm{J_iU}}$, with $\mathrm{U} \in \{\mathrm{B}, \mathrm{E}\}$ and $i \in \{1, \ldots, N\}$.

The propagation path loss for the A2G channels presents a contribution from a LoS component and a non-LoS (NLoS) component, where the contribution of each component to the overall path loss is determined by the probabilities P_{LoS} and P_{NLoS}, respectively [20]. These probabilities are functions of the UAV position with respect to the ground node of interest U and are given by [20]

$$P_{\mathrm{LoS}} = \frac{1}{1 + \psi \exp\left(-\omega\left[\frac{180}{\pi}\tan^{-1}\left(\frac{z_{\mathrm{J}}}{r_{\mathrm{J_iU}}}\right) - \psi\right]\right)}, \tag{1}$$

$$P_{\text{NLoS}} = 1 - P_{\text{LoS}}, \tag{2}$$

where ψ and ω are environmental constants [21, 22] and $r_{\text{J}_i\text{U}}$ is the distance from node U and the projection on the plane of the ith UAV.

The path loss of each component is given by

$$L_{\text{J}_i\text{U}}^{\text{LoS}} = \xi_{\text{LoS}} d_{\text{J}_i\text{U}}^{\alpha_J} \tag{3}$$

$$L_{\text{J}_i\text{U}}^{\text{NLoS}} = \xi_{\text{NLoS}} d_{\text{J}_i\text{U}}^{\alpha_J} \tag{4}$$

where α_J is the path loss exponent for the A2G links and ξ_{LoS} and ξ_{NLoS} are the attenuation factors for the LoS and the NLoS links, respectively. It is also assumed that the LoS channel undergoes Rician fading with channel coefficient $\text{h}_{\text{J}_i\text{U}}^{\text{LoS}}$ and channel gain given by $g_{\text{J}_i\text{U}}^{\text{LoS}} = |\text{h}_{\text{J}_i\text{U}}^{\text{LoS}}|^2$, with a scale parameter of $\Omega_{\text{J}_i\text{U}}^{\text{LoS}} = \mathbb{E}[|\text{h}_{\text{J}_i\text{U}}^{\text{LoS}}|^2] = \gamma_{\text{J}_i} P_{\text{LoS}} (L_{\text{J}_i\text{U}}^{\text{LoS}})^{-1}$ and shape parameter of $K_{\text{J}_i\text{U}}$, where γ_{J_i} is the transmit SNR of UAV J$_i$, $\gamma_{\text{J}_i} = P_{\text{J}_i}/N_0$ and P_{J_i} is the transmit power, with a total jamming SNR of $\gamma_{\text{T}} = \sum_i \gamma_{\text{J}_i}$. The NLoS component undergoes Rayleigh fading with channel gain $g_{\text{J}_i\text{U}}^{\text{NLoS}} = |\text{h}_{\text{J}_i\text{U}}^{\text{NLoS}}|^2$, with a scale parameter of $\Omega_{\text{J}_i\text{U}}^{\text{NLoS}} = \mathbb{E}[|\text{h}_{\text{J}_i\text{U}}^{\text{NLoS}}|^2] = \gamma_{\text{J}_i} P_{\text{NLoS}} (L_{\text{J}_i\text{U}}^{\text{NLoS}})^{-1}$. Considering that, the average channel gain can be expressed as

$$g_{\text{J}_i\text{U}} = g_{\text{J}_i\text{U}}^{\text{LoS}} + g_{\text{J}_i\text{U}}^{\text{NLoS}}. \tag{5}$$

2.1.3 *Signal analysis*

For the communication process, A sends a symbol x with mean power $\mathbb{E}\left[|x|^2\right] = 1$, while the UAVs send pseudorandom symbols s_i with mean power $\mathbb{E}\left[|s_i|^2\right] = 1$, with $i \in \{1, \ldots, N\}$. We consider a common noise level with power $\mathbb{E}\left[|w|^2\right] = N_0$ at every node in the system. Thus, the received signal at both B and E is, respectively, given by

$$y_{\text{U}} = h_{\text{AU}} x + \sum_{i=1}^{N} h_{\text{J}_i\text{U}} s_i + w, \tag{5}$$

with U $\in$ {B, E}. Then, the instantaneous received signal-to-interference-plus-noise ratio (SINR) at node U can be expressed as

$$\gamma_{\text{U}} = \frac{g_{\text{AU}}}{1 + \sum_{i=1}^{N} g_{\text{J}_i\text{U}}}, \tag{6}$$

For the particular case with no UAV jammers, the SINR values at B and E are, respectively, given by $\gamma_{\text{B}} = \gamma_{\text{A}} g_{\text{AB}}$ and $\gamma_{\text{E}} = \gamma_{\text{A}} g_{\text{AE}}$.

2.2 Performance analysis

As previously mentioned, E is located within a circular area S around A, but no further knowledge on the exact position of E is assumed, i.e. E can be in whichever point inside S. Therefore, to evaluate the secrecy performance of the proposed system, we consider the area-based secrecy metrics proposed in [17], namely jamming coverage (JC) and jamming efficiency (JE), and a new hybrid metric, the WSC, introduced in [18]. These metrics' definition is based on the SOP, which is derived for the proposed system as described below.

2.2.1 Secrecy outage probability

For the definition of the area-based secrecy metrics, we consider first the SOP [6] defined as

$$\mathrm{SOP} = \Pr\left[C_S < R_S\right], \tag{7}$$

where R_S is the chosen rate for a secrecy code and C_S is the secrecy capacity, which for our system is given by

$$C_S = [C_{\mathrm{B}} - C_{\mathrm{E}}]^{+} = \left[\log_2\left(\frac{1+\gamma_{\mathrm{B}}}{1+\gamma_{\mathrm{E}}}\right)\right]^{+}, \tag{8}$$

where C_{B} and C_{E} are the capacities of the channels between A and B and between A and E, respectively, with $[X]^{+} = \max[X, 0]$, which tells us that if the capacity of the illegitimate channel is greater than the capacity of the legitimate channel, no secrecy can be achieved.

2.2.2 Secrecy improvement metric

This metric measures the improvement on the secrecy performance of the proposed system, which is measured by the SOP, attained by the introduction of the friendly jamming sent by the UAV jammers. Thus, this metric is given by [17]

$$\Delta = \frac{\mathrm{SOP}_{\mathrm{NJ}}}{\mathrm{SOP}_{\mathrm{J}}}, \tag{9}$$

where the SOP subscript identifies if the SOP is computed with (J) or without (NJ) the presence of friendly jamming. Then, $\Delta > 1$ values imply a reduction on the SOP by the presence of the UAV jammers, while $\Delta < 1$ is the opposite.

For mathematical tractability purposes, in [18] we proposed an analogous secrecy improvement metric that provides the same general idea with the criteria of secrecy achievement $(1 - \mathrm{SOP})$ instead of SOP, thus given by

$$\overline{\Delta} = \frac{1-\mathrm{SOP}_{\mathrm{J}}}{1-\mathrm{SOP}_{\mathrm{NJ}}}. \tag{10}$$

The SOP without jamming term, $\mathrm{SOP}_{\mathrm{NJ}}$, is obtained in closed form in [18] as

$$\mathrm{SOP}_{\mathrm{NJ}} = 1 - e^{-\frac{1}{\Omega_{\mathrm{AB}}}\left(2^{R_S}-1\right)}\left(\frac{1}{2^{R_S}\left(\frac{\Omega_{\mathrm{AE}}}{\Omega_{\mathrm{AB}}}\right)+1}\right), \tag{11}$$

while, the SOP including jamming, $\mathrm{SOP}_{\mathrm{J}}$ is obtained as in Proposition 1.

Proposition 1 *The SOP in the presence of N UAV jammers* $\mathrm{SOP}_{\mathrm{J}}$ *for the proposed system is given by*

$$\mathrm{SOP_J} = \int_0^\infty F_{\gamma_\mathrm{B}}(2^{R_S}(1+x)-1) f_{\gamma_\mathrm{E}}(x)dx. \tag{12}$$

where $F_{\gamma_\mathrm{B}}(\cdot)$ *is the CDF of the SINR at B,* γ_B*, and* $f_{\gamma_\mathrm{E}}(\cdot)$ *is the PDF of the SINR at E,* γ_E*, which are, respectively, expressed as*

$$F_{\gamma_\mathrm{U}}(x) = 1 - e^{-\widehat{x}} e^{\sum_{i=1}^{N}\left(\frac{\eta_i}{\eta_i+\widehat{x}}-1\right)K_{\mathrm{J_iU}}} \prod_{i=1}^{N}\left(\frac{\eta_i}{\eta_i+\widehat{x}}\right) \tag{13}$$

$$\begin{aligned} f_{\gamma_\mathrm{U}}(x) = {} & \frac{1}{\Omega_\mathrm{AU}} e^{-\widehat{x}} e^{\sum_{i=1}^{N}\left(\frac{\eta_i}{\eta_i+\widehat{x}}-1\right)K_{\mathrm{J_iU}}} \left(1 + \sum_{i=1}^{N}\frac{1}{\eta_i+\widehat{x}}\left(1+\frac{\eta_i K_{\mathrm{J_iU}}}{\eta_i+\widehat{x}}\right)\right) \cdot \\ & \prod_{i=1}^{N}\left(\frac{\eta_i}{\eta_i+\widehat{x}}\right), \end{aligned} \tag{14}$$

with $\mathrm{U} \in \{\mathrm{B}, \mathrm{E}\}$, $\widehat{x} = \frac{x}{\Omega_\mathrm{AU}}$ *and*

$$\eta_i = \frac{1+K_{\mathrm{J_iU}}}{\Omega_{\mathrm{J_iU}}}. \tag{15}$$

The SOP in (12) can be extended for the channel in (5), with both LoS and NLoS components, by considering $g_{\mathrm{J_iU}} = g_{\mathrm{J_iU}}^{\mathrm{LoS}} + g_{\mathrm{J_iU}}^{\mathrm{NLoS}}$, which implies doubling the amount of terms in the sums and products in (13) and (14). The Rayleigh NLoS parameters are adapted from Rician channels by setting the shape parameters to zero, $K_{\mathrm{J_iU}}^{\mathrm{NLOS}} = 0$, making $\eta_i^{\mathrm{NLoS}} = (\Omega_{\mathrm{J_iU}})^{-1}$.

Proof Let us consider first the case with 2 UAVs and LoS connection between the UAVs and the ground nodes. Under these conditions, $g_{\mathrm{J_iU}} = g_{\mathrm{J_iU}}^{\mathrm{LoS}}$ and $\Omega_{\mathrm{J_iU}} = \Omega_{\mathrm{J_iU}}^{\mathrm{LoS}}$. For that case, the PDF and CDF of the effective A2G channel gains $g_{\mathrm{J_iU}}$ are given by [23]

$$f_{g_{\mathrm{J_iU}}}(x) = \frac{1+K_{\mathrm{J_iU}}}{\Omega_{\mathrm{J_iU}}} e^{-K_{\mathrm{J_iU}} - \frac{1+K_{\mathrm{J_iU}}}{\Omega_{\mathrm{J_iU}}}x} I_0\left(\sqrt{\frac{4K_{\mathrm{J_iU}}(K_{\mathrm{J_iU}}+1)}{\Omega_{\mathrm{J_iU}}}x}\right) \tag{16}$$

$$F_{g_{\mathrm{J_iU}}}(x) = 1 - Q_1\left[\sqrt{2K_{\mathrm{J_iU}}}, \sqrt{\frac{2(K_{\mathrm{J_iU}}+1)}{\Omega_{\mathrm{J_iU}}}x}\right] \tag{17}$$

where $I_0(\cdot)$ is the zero-order modified Bessel function of first kind and $Q_1[\cdot]$ is the Marcum-Q function of order 1. Additionally, the PDF and CDF of the ground channels g_{AU} are given by

$$f_{g_{\mathrm{AU}}}(x) = \frac{1}{\Omega_{\mathrm{AU}}} e^{-\frac{1}{\Omega_{\mathrm{AU}}}x} \tag{18}$$

$$F_{g_{\mathrm{AU}}}(x) = 1 - e^{-\frac{1}{\Omega_{\mathrm{AU}}}x} \tag{19}$$

Therefore, the CDF of γ_U is obtained as

$$\begin{aligned} F_{\gamma_{\mathrm{U}}} &= \Pr\left[\frac{g_{\mathrm{AU}}}{1+g_{\mathrm{J_1U}}+g_{\mathrm{J_2U}}} < x\right] \\ &= \Pr\left[g_{\mathrm{AU}} < x(1+g_{\mathrm{J_1U}}+g_{\mathrm{J_2U}})\right] \\ &= \int_0^\infty \int_0^\infty F_{g_{\mathrm{AU}}}\left(x(1+y+z)\right) f_{g_{\mathrm{J_1U}}}(y) f_{g_{\mathrm{J_2U}}}(z)\, dydz, \end{aligned} \tag{20}$$

while the PDF is derived from the CDF as

$$\begin{aligned} f_{\gamma_{\mathrm{U}}}(x) &= \frac{d}{dx} F_{\gamma_{\mathrm{U}}}(x) \\ &= \int_0^\infty \int_0^\infty \frac{d}{dx} F_{g_{\mathrm{AU}}}\left(x(1+y+z)\right) f_{g_{\mathrm{J_1U}}}(y) f_{g_{\mathrm{J_2U}}}(z)\, dydz \\ &= \int_0^\infty \int_0^\infty (1+y+z) f_{g_{\mathrm{AU}}}\left(x(1+y+z)\right) f_{g_{\mathrm{J_1U}}}(y) f_{g_{\mathrm{J_2U}}}(z)\, dydz. \end{aligned} \tag{21}$$

To simplify the notation, in the following steps g_{AU} is used for g_A, $g_{\mathrm{J_iU}}$ for g_i, $K_{\mathrm{J_iU}}$ for K_i and $\Omega_{\mathrm{J_iU}}$ for Ω_i. Thus, by considering [24, 8.445], the term $I_0(\cdot)$ in (16) can be rewritten as its series representation as

$$\begin{aligned} I_0\left(\sqrt{\frac{4K_i(K_i+1)}{\Omega_i}x}\right) &= \sum_{n=0}^{\infty} \frac{1}{n!\Gamma(n+1)2^{2n}} \left(\left(\frac{4K_i(1+K_i)}{\Omega_i}x\right)^{1/2}\right)^{2n} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!^2} \left(\frac{K_i(1+K_i)}{\Omega_i}\right)^n x^n, \end{aligned} \tag{22}$$

then, by defining $\eta_i \triangleq \frac{1+K_i}{\Omega_i}$, (16) can be rewritten as

$$f_{g_i}(x) = e^{-K_i} \sum_{n=0}^{\infty} \frac{K_i^n}{n!^2} \eta_i^{n+1} e^{-\eta_i x} x^n. \tag{23}$$

Then, by replacing (24) and (23) into (20) leads to

$$F_{g_A}(x(1+y+z)) = 1 - e^{-\widehat{x}} e^{-\widehat{x}y} e^{-\widehat{x}z}. \tag{24}$$

Then, by plugging (24) and (23) into (20) we obtain

$$F_{\gamma_{\mathrm{U}}} = e^{-K_1-K_2}(\mathcal{I}_1 - \mathcal{I}_2), \tag{25}$$

where $\mathcal{I}_1$ and $\mathcal{I}_2$ are given by

$$\mathcal{I}_1 = \int_0^\infty \int_0^\infty \left(\sum_{n=0}^{\infty} \frac{K_1^n}{n!^2} \eta_1^{n+1} e^{-\eta_1 y} y^n\right) \left(\sum_{m=0}^{\infty} \frac{K_2^m}{m!^2} \eta_2^{m+1} e^{-\eta_2 z} z^m\right) dydz, \tag{26}$$

and

$$\begin{aligned} \mathcal{I}_2 = e^{-\widehat{x}} \int_0^\infty \int_0^\infty & \left(\sum_{n=0}^{\infty} \frac{K_1^n}{n!^2} \eta_1^{n+1} e^{-(\eta_1+\widehat{x})y} y^n\right) \cdot \\ & \left(\sum_{m=0}^{\infty} \frac{K_2^m}{m!^2} \eta_2^{m+1} e^{-(\eta_2+\widehat{x})z} z^m\right) dydz. \end{aligned} \tag{27}$$

By considering [24, 3.326.2], each individual integral in $\mathcal{I}_1$ can be solved as

$$\begin{aligned}\int_0^{\infty}\left(\sum_{n=0}^{\infty}\frac{K_i^n}{n!^2}\eta_i^{n+1}e^{-\eta_i y}y^n\right)dy &= \sum_{n=0}^{\infty}\frac{K_i^n}{n!^2}\eta_i^{n+1}\int_0^{\infty}e^{-\eta_i y}y^n dy\\ &= \sum_{n=0}^{\infty}\frac{K_i^n}{n!^2}\eta_i^{n+1}\frac{n!}{\eta_i^{n+1}}dy\\ &= \sum_{n=0}^{\infty}\frac{K_i^n}{n!}\\ &= e^{K_i},\end{aligned} \tag{28}$$

and the same reasoning is applied for each individual integral in $\mathcal{I}_2$, which can be solved as

$$\begin{aligned}\int_0^{\infty}\left(\sum_{n=0}^{\infty}\frac{K_i^n}{n!^2}\eta_i^{n+1}e^{-(\eta_i+\widehat{x})y}y^n\right)dy &= \sum_{n=0}^{\infty}\frac{K_i^n}{n!^2}\eta_i^{n+1}\int_0^{\infty}e^{-(\eta_i+\widehat{x})y}y^n\\ &= \sum_{n=0}^{\infty}\frac{K_i^n}{n!^2}\eta_i^{n+1}\frac{n!}{(\eta_i+\widehat{x})^{n+1}}dy\\ &= \left(\frac{\eta_i}{\eta_i+\widehat{x}}\right)\sum_{n=0}^{\infty}\left(\frac{\eta_i K_i}{\eta_i\widehat{x}}\right)^n\frac{1}{n!}\\ &= \left(\frac{\eta_i}{\eta_i+\widehat{x}}\right)e^{\frac{\eta_i K_i}{\eta_i+\widehat{x}}}.\end{aligned} \tag{29}$$

Then, by replacing (28) in (26) and (29) in (27), $\mathcal{I}_1$ and $\mathcal{I}_2$ can be, respectively, expressed as

$$\mathcal{I}_1 = e^{K_1+K_2} \tag{30}$$

$$\mathcal{I}_2 = e^{-\widehat{x}}\left(\frac{\eta_1}{\eta_1+\widehat{x}}\right)\left(\frac{\eta_2}{\eta_2+\widehat{x}}\right)e^{\left(\frac{\eta_1}{\eta_1+\widehat{x}}\right)K_1+\left(\frac{\eta_2}{\eta_2+\widehat{x}}\right)K_2}. \tag{31}$$

Finally, (25) can be expressed as

$$F_{\gamma_{\mathrm{U}}}(x) = 1 - e^{-\widehat{x}}\left(\frac{\eta_1}{\eta_1+\widehat{x}}\right)\left(\frac{\eta_2}{\eta_2+\widehat{x}}\right)e^{\left(\frac{\eta_1}{\eta_1+\widehat{x}}-1\right)K_1+\left(\frac{\eta_2}{\eta_2+\widehat{x}}-1\right)K_2}. \tag{32}$$

To compute the PDF in (16), it is followed a similar process for the CDF calculation, by considering that

$$\begin{aligned}\int_0^{\infty}e^{-(\eta_i+\widehat{x})x}x^{n+1}\mathrm{d}x &= \frac{(n+1)!}{(\eta_i+\widehat{x})^{n+2}}\\ &= \left(\frac{n+1}{\eta_i+\widehat{x}}\right)\frac{n!}{(\eta_i+\widehat{x})^{n+1}},\end{aligned} \tag{33}$$

and

$$\sum_{n=0}^{\infty}\left(\frac{\eta_i K_i}{\eta_i+\widehat{x}}\right)\frac{n+1}{n!}=\left(1+\frac{\eta_i K_i}{\eta_i+\widehat{x}}\right)e^{\frac{\eta_i K_i}{\eta_i+\widehat{x}}}. \tag{34}$$

Thus, the PDF can be obtained as

$$\begin{aligned} f_{\gamma_{\mathrm{U}}}(x)=\frac{1}{\Omega_A}e^{-\widehat{x}}\left(\frac{\eta_1}{\eta_1+\widehat{x}}\right)\left(\frac{\eta_2}{\eta_2+\widehat{x}}\right)e^{\left(\frac{\eta_1}{\eta_1+\widehat{x}}-1\right)K_1+\left(\frac{\eta_2}{\eta_2+\widehat{x}}-1\right)K_2}.\\ \left(1+\frac{1}{\eta_1+\widehat{x}}\left(1+\frac{\eta_1 K_1}{\eta_1+\widehat{x}}\right)+\frac{1}{\eta_2+\widehat{x}}\left(1+\frac{\eta_2 K_2}{\eta_2+\widehat{x}}\right)\right). \end{aligned} \tag{35}$$

It is worthwhile to note that the integrals in (26) and (27) can be separated into independent terms for each UAV. Therefore, the CDF and PDF for the general case of N UAVs can be obtained as in (13) and (14), respectively.

Then, the SOP is calculated as

$$\begin{aligned} \mathrm{SOP}&=\Pr[C_S<R_S]\\ &=\Pr\left[\frac{1+\gamma_{\mathrm{B}}}{1+\gamma_{\mathrm{E}}}<2^{R_S}\right]\\ &=\Pr\left[\gamma_{\mathrm{B}}<2^{R_S}(1+\gamma_{\mathrm{E}})-1\right]\\ &=\int_0^{\infty}F_{\gamma_{\mathrm{B}}}(2^{R_S}(1+x)-1)f_{\gamma_{\mathrm{E}}}(x)\mathrm{d}x. \end{aligned} \tag{36}$$

□

2.2.3 Weighted secrecy coverage

As mentioned before, we assume no knowledge on the position of E, other than it is located inside the circular region S within a radius R_A from A, so we analyse the secrecy performance of the proposed system in terms of the area-based metrics in [17], the jamming coverage (JC) and the jamming efficiency (JE). Both of these metrics give us the notion on the effect over the secrecy performance inside S by the presence of the UAV jammers.

For the JC, consider that E is located at a single point within the area S, where a certain $\overline{\Delta}$ value can be calculated, and we are interested in such points that lead into a $\overline{\Delta}>1$ value. Then, the jamming secrecy coverage is the integral over the area where $\overline{\Delta}>1$, expressed as

$$\mathrm{JC}=\iint_{\overline{\Delta}>1}dS_{\mathrm{E}}, \tag{37}$$

where the dS_{E} term indicates an integral over the positions of E over the whole area S. To illustrate this concept, Fig. 2 shows a simplified overview of the system as a heatmap of $\overline{\Delta}$ over the whole area S. The JC would be the total area where $\overline{\Delta}>1$, which is enclosed by the yellow line surrounding the UAVs and A.

On the other hand, JE measures the average improvement in the secrecy over the whole area S:

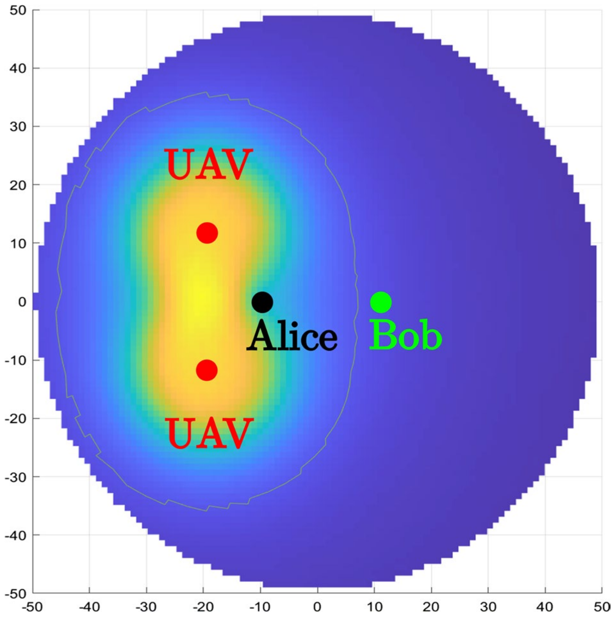

Fig. 2 Heatmap for $\overline{\Delta}$ values due to the presence of UAV jammers on a circular region around Alice

$$\mathrm{JE} = \frac{1}{|S|}\iint_{S} \overline{\Delta}\mathrm{d}S_{\mathrm{E}}, \tag{38}$$

where $|S|$ is the area of the region S.

Note that JC gives a measure of the area within S where an improvement on the secrecy performance of the system is obtained due to the UAV jammers, while JE gives a measure of the average improvement in the secrecy performance over the area S, if E were located at a random point.

To get further insights on the jamming effective coverage, in [18] we proposed a hybrid metric, the WSC, to account for both, the area over which secrecy is improved and the average secrecy improvement over the whole area S. The WSC is given by

$$\mathrm{WSC} = \left(\iint_{\overline{\Delta}>1} \mathrm{d}S_{\mathrm{E}}\right)\left(\frac{1}{|\mathrm{S}|}\iint_{\mathrm{S}} \overline{\Delta}\mathrm{d}S_{\mathrm{E}}\right). \tag{39}$$

2.3 Positioning optimisation

In this section, we consider joint optimisation of the 3D positioning of the UAVs (common height, common orbit radius and angles around A) and the power allocation between the UAVs in order to maximise the WSC, given a relative position of B with respect to A, which is characterised by d_{AB}. Thus, the optimisation problem is formulated as

$$\max_{\Omega=\{\{\theta_{\mathrm{J_i}}\}_{i\in\{1,\ldots,N\}},\{\gamma_{\mathrm{J_i}}\}_{i\in\{1,\ldots,N\}},z_{\mathrm{J}},R_{\mathrm{J}}\}} \quad \mathrm{WSC}(\Omega, d_{\mathrm{AB}}) \tag{40a}$$

$$\text{subject to} \quad 0 \le \theta_{\mathrm{J_i}} \le 2\pi \quad , \quad \forall i \in \{1,\ldots,N\} \tag{40b}$$

$$\gamma_{\text{J}_\text{i}} \geq 0 \quad , \quad \forall i \in \{1, \ldots, N\} \tag{40c}$$

$$\sum_{i=1}^{N} \gamma_{\text{J}_\text{i}} \leq \gamma_\text{T}, \tag{40d}$$

$$z_\text{MIN} \leq z_\text{J} \leq z_\text{MAX}, \tag{40e}$$

$$0 \leq R_\text{J} \leq R_\text{MAX}, \tag{40f}$$

where z_MIN is the minimum flying height, z_MAX is the maximum allowed flying height for the UAVs, R_MAX is the limit of the orbit radius around A, which is the radius of S, and γ_T is the maximum jamming transmit SNR from all UAVs.

To simplify the optimisation problem in (40), some trends are considered regarding the angular positioning and the allocated jamming power for the case of two UAVs provided as observed in [18]. In that work, it was found locating both UAVs symmetrically behind the line between A and B leads to the optimal performance; thus, this trend is generalised to the N UAVs case by considering a single opening angle θ_J between any pair of adjacent UAVs symmetrically located, as shown in Fig. 1. Then, it was proved that the WSC is maximised by having an equal power allocation for the friendly jammers, which is also generalised to the N UAV case.

Under these observations, the optimisation problem in (40) can be reformulated as

$$\max_{\Omega=\{\theta_\text{J}, z_\text{J}, R_\text{J}\}} \qquad \text{WSC}(\Omega, d_\text{AB}) \tag{41a}$$

$$\text{subject to} \quad 0 \leq \theta_\text{J} \leq \frac{2\pi}{N-1}, \tag{41b}$$

$$z_\text{MIN} \leq z_\text{J} \leq z_\text{MAX}, \tag{41c}$$

$$0 \leq R_\text{J} \leq R_\text{MAX}, \tag{41d}$$

where only three optimisation variables are considered, namely the opening angle θ_J, the UAV common height z_J and the UAV surveillance orbit radius R_J.

2.3.1 Reinforcement learning-based positioning

Given that the estimate of the distance from A to B is unreliable, the optimisation problem in (41) cannot be reliably solved. To account for the stochastic nature of the estimate of the distance to B, d_AB, we consider a coordinate-descent-based [25] iterative scheme to reliably solve the optimisation problem in (41) by employing an RL approach to ascertain the optimum positioning for the UAVs around A. Particularly, we model this problem as a multi-armed bandit (MAB) problem, by considering the discrete positioning variables values as the arms or actions, and the WSC reading obtained at each step as

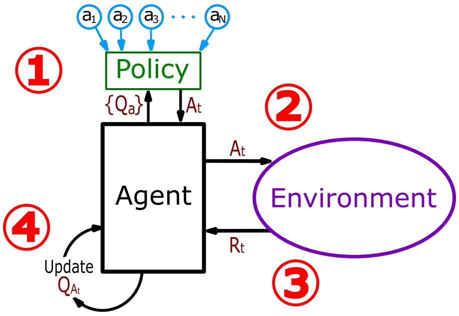

Fig. 3 Reinforcement learning process: (1) At time t the agent chooses an action A_t based on a policy and the set of estimates $\{Q_a\}$, (2) the agent applies action A_t onto the environment, (3) the agent obtains reward R_t from the environment, and (4) the agent updates the estimate for the action chosen Q_{A_t}

the values or rewards. In the following, we briefly introduce the basis of the MAB problem and some relevant RL concepts to help us explain our approach.[2]

Multi-Armed Bandit Problem [26] An MAB problem consists of an agent (bandit) which has to choose at each time step among a set of actions (arms) to obtain rewards. At each step, each chosen action provides a reward, which is a random variable with a given distribution per action. The goal of the agent is to maximise the reward obtained over the time, which could be understood as choosing the optimum action, which is the action with the highest expected reward, so-called *exploitation*. This is done by keeping estimates of each of the actions' expected rewards. Therefore, it is also of interest to keep learning more about other actions to refine the estimates for each of them, which is called *exploration*. The action chosen at each step is determined by a policy, which in part sets the exploration/exploitation balance to be taken. An illustrative example of this learning process is shown in Fig. 3.

Considering the optimisation problem in (41), we have three positioning variables, the opening angle of adjacent UAVs behind A (θ_J), the common height of the UAVs (z_J) and the orbit radius of the UAVs around A (R_J). Each variable is separated into its own RL process, independent of the other two. For each positioning variable, we define its possible actions as a range of values the variable can take, which are given by the constraints in (41), and a discretised number of actions per variable (N_θ, N_z, N_R). Each action of a variable has a reward distribution, which corresponds to the distribution of WSC values obtained by performing that action. The goal is to be able to estimate with high accuracy which of the actions has the greatest expected reward. At each step, one of the actions is chosen following a policy and the received WSC reward is processed to contribute for the estimation of the expected reward (WSC) for said action.

To simplify the computations, we perform three separated RL processes, one for each positional variable with its own action range discretisation. The RL loops for each of the variables are to be repeated back to back, alternating between the variables.

Considering that for each RL step of a given positioning variable, an assumption needs to be made regarding the other two positioning variables. The natural way of choosing which value should be considered for the other two positioning variables is to choose

[2] We refer interested readers on RL and the MAB problem to [26].

them in a *greedy* fashion, i.e. choose the values for the other two positioning variables that are estimated thus far to be the ones that lead to the highest reward. This implies that for any of the positioning variables, the RL process being carried out is non-stationary since the values for the other positioning variables, which are considered as part of the environment, change during the process, thus changing the environment. To account for the non-stationarity of the RL processes, consider the following generic estimate update rule [26]:

$$Q_{n+1} = Q_n + \alpha_n[R_n - Q_n], \tag{42}$$

where Q_n is a generic action reward estimate at time n, R_n is the observed reward at time n and α_n is the so-called step size at time n, which controls the contribution of the observed data to the estimate at time n. As we consider that all observed rewards will contribute evenly to the estimate, we set $\alpha_n = 1/n$. However, in a non-stationary environment, we may want to give a higher weight to the new observations over the past observations, so that the RL process would be more sensitive to the environmental changes. To accomplish this, we set $\alpha_n = \alpha$ for all n values to be a constant, such that $0 < \alpha < 1$ [26].

Regarding the policy to be used, we consider the upper confidence bound (UCB) policy [26] that is described next:

Upper Confidence Bound The action chosen at each step is determined by both the estimated value of the action thus far (greedy) and by the frequency of chosen that action in the past. This rule is determined by [26]

$$A_t = \arg\max_a \left[Q_t(a) + c\sqrt{\frac{\ln(t)}{N_t(a)}} \right] \tag{43}$$

where $N_t(a)$ is the number of times the action a has been chosen up to time t and c is a constant parameter that controls the degree of exploration. Then, with this policy, a continuous exploration is performed as time goes on in favour of less chosen actions over time that is controlled by the c constant, which has to be set depending on the desired degree of exploration, and the expected reward values.

2.3.2 Positioning learning block

RL loops will be employed over the positional variables of the UAVs in order to iteratively reach the optimum values in a coordinate descent fashion [25]. This processing is performed at A that has a global understanding of the system, and it transmits the positional information to the UAVs for physical adjustment. However, the transmission frequency of positional information to the UAVs is a concern, since every time this information is received, the UAVs are compelled to adjust their position, thus entailing energy consumption. If this occurs after each RL step of each variable, the movement of the UAVs may be unnecessarily erratic (given the randomness of the estimate and the discretisation level of the variable domains), consume a high amount of energy from the UAVs over time and introduce a substantial amount of delay, given that A needs to receive an acknowledgement (ACK) from the UAVs alerting that the required new position has been assumed before starting another RL step.

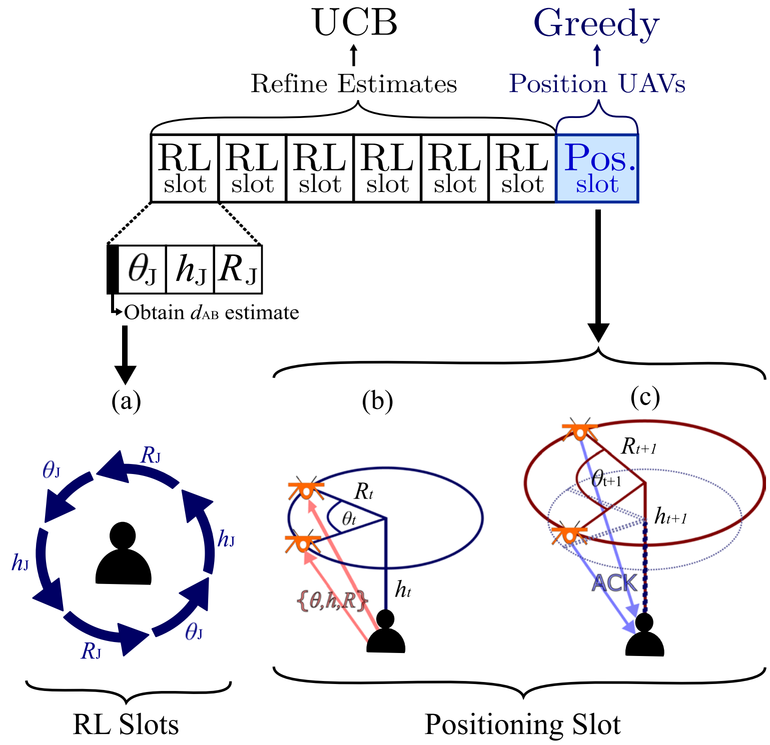

Fig. 4 PLB-based WSC improvement algorithm. **a** Iterative RL processes on the three positional variables θ_J, z_J and R_J in A, over the RLSs of the PLB, **b** signalling from A to the UAVs to adopt new positions, and **c** UAVs sending ACK signalling back to A after adopting their new positions

Thus, we propose a time frame-based scheme that splits a given time range, which we name a positioning learning block (PLB), into individual slots, namely RL slots (RLSs) and positioning slots, as shown in Fig. 4. A PLB comprises *nRLS* consecutive RLSs and a single positioning slot at the end of it. At the beginning of an RLS, a $\widehat{d}_{AB}$ estimate is obtained and used in the rest of the slot, where a single RL step is performed for each of the positioning variables (θ_J, z_J, R_J), one after another. Each RL step assumes a greedy positioning from the other variables.

For the duration of the RLSs, A performs internal processing of the RL steps, and at the positioning slot, A chooses the greedy actions from the three positioning variables and transmits this information to the UAVs. Then, the UAVs assume their new positions based on this information and send an ACK signal to A, which, upon reception, starts another PLB as shown in Fig. 4. Therefore, we define an off-policy scheme, where we employ a greedy policy at the positioning slots, and a UCB policy at the RLSs.

Given this approach, each UAV incurs in energy consumption at each positioning slot that is simply given by: the energy needed to receive the positioning instructions from A (E_{RX}), the energy needed to manoeuvre to its new position (E_{Mov}) and the energy needed to send an ACK back to A (E_{ACK}). This energy term is given by

$$E = E_{\mathrm{RX}} + E_{\mathrm{ACK}} + E_{\mathrm{Mov}} \tag{44}$$

$$= E_{\mathrm{RX}} + E_{\mathrm{ACK}} + \Delta t_v P_{\mathrm{Mov}}, \tag{45}$$

where P_{Mov} is the power needed by the UAV to manoeuvre and Δt_v is the time it takes the UAV to perform this change in position. Assuming that the UAV changes its position by assuming its new angle, height and radius in that order, Δt_v is given by

$$\Delta t_v = \frac{1}{v_{\mathrm{J}}}\left(|\Delta s| + |\Delta z_{\mathrm{J}}| + |\Delta R_{\mathrm{J}}|\right) \tag{46}$$

$$= \frac{1}{v_{\mathrm{J}}}\left(\frac{1}{2}R_{\mathrm{J}_0}|\Delta\theta_{\mathrm{J}}| + |\Delta z_{\mathrm{J}}| + |\Delta R_{\mathrm{J}}|\right), \tag{47}$$

where v_{J} is the manoeuvring speed of the UAV (assumed constant throughout the flight), $\Delta\theta_{\mathrm{J}}$, Δz_{J} and ΔR_{J} are the angle, height and radius variations, and R_{J_0} is the initial UAV radius value.

2.3.3 MAB-based WSC improvement UAV positioning algorithm

The concepts defined so far have the main goal of establishing the optimal position for the N UAV jammers in order to maximise the WSC, while A sends out information to B over the wireless medium. In Algorithm 1, we present the process followed by the proposed algorithm, where the variables in brackets ($[\theta_{\mathrm{J}}]$, $[z_{\mathrm{J}}]$, $[R_{\mathrm{J}}]$) represent the action values estimates array for each of the variables.

Algorithm 1 provides a description of the processes depicted in Fig. 4 over time. In this algorithm, MAB processes are carried out, once for each RLS, for every positioning variable sequentially with the UCB action-choosing policy, over a number of PLBs. This algorithm refines its action estimates for each of the positioning variables over time in each RLS, adapting to the changes in the other positioning variables and allowing the UAVs to take positions that increase their WSC at the end of each PLB. Thus, the WSC of the system increases closer to the optimum at every PLB.

Algorithm 1: WSC improvement UAV positioning algorithm

```
$[\theta_{\mathrm{J}}] \leftarrow N_\theta$ values from $0$ to $\pi$;
$[z_{\mathrm{J}}] \leftarrow N_z$ values from $0$ to $z_{\mathrm{MAX}}$;
$[R_{\mathrm{J}}] \leftarrow N_R$ values from $0$ to $R_{MAX}$;
$Q_0(\theta_{\mathrm{J}}) \leftarrow 0 \;\; \forall \theta_{\mathrm{J}} \in [\theta_{\mathrm{J}}]$;
$Q_0(z_{\mathrm{J}}) \leftarrow 0 \;\; \forall z_{\mathrm{J}} \in [z_{\mathrm{J}}]$;
$Q_0(R_{\mathrm{J}}) \leftarrow 0 \;\; \forall R_{\mathrm{J}} \in [R_{\mathrm{J}}]$;
while Alice requires secrecy improvement aid do
    PLB starts;
    for $i \leftarrow 0$ to $nRLS$ do
        RLS starts;
        $\widehat{d}_{\mathrm{AB}} \leftarrow$ estimate of $d_{\mathrm{AB}}$ ;
        RL process for $\theta_{\mathrm{J}}$;
        $\theta_t \leftarrow$ choose action based on $(43)$;
        $z_t \leftarrow \arg\max_{z_{\mathrm{J}} \in [z_{\mathrm{J}}]} Q_t\left(z_{\mathrm{J}}\right)$ ;
        $R_t \leftarrow \arg\max_{R_{\mathrm{J}} \in [R_{\mathrm{J}}]} Q_t\left(R_{\mathrm{J}}\right)$ ;
        $\mathrm{WSC}_t \leftarrow$ compute $\mathrm{WSC}(\theta_t, z_t, R_t)$ from $(39)$ with $\widehat{d}_{\mathrm{AB}}$ estimate;
        $Q_t(\theta_t) \leftarrow$ update as in $(42)$ with $R_t = \mathrm{WSC}_t$;
        RL process for $z_{\mathrm{J}}$;
        $z_t \leftarrow$ choose action based on $(43)$;
        $\theta_t \leftarrow \arg\max_{\theta_{\mathrm{J}} \in [\theta_{\mathrm{J}}]} Q_t\left(\theta_{\mathrm{J}}\right)$ ;
        $R_t \leftarrow \arg\max_{R_{\mathrm{J}} \in [R_{\mathrm{J}}]} Q_t\left(R_{\mathrm{J}}\right)$ ;
        $\mathrm{WSC}_t \leftarrow$ compute $\mathrm{WSC}(\theta_t, z_t, R_t)$ from $(39)$ with $\widehat{d}_{\mathrm{AB}}$ estimate;
        $Q_t(z_t) \leftarrow$ update as in $(42)$ with $R_t = \mathrm{WSC}_t$;
        RL process for $R_{\mathrm{J}}$;
        $R_t \leftarrow$ choose action based on $(43)$;
        $z_t \leftarrow \arg\max_{z_{\mathrm{J}} \in [z_{\mathrm{J}}]} Q_t\left(z_{\mathrm{J}}\right)$ ;
        $\theta_t \leftarrow \arg\max_{\theta_{\mathrm{J}} \in [\theta_{\mathrm{J}}]} Q_t\left(\theta_{\mathrm{J}}\right)$ ;
        $\mathrm{WSC}_t \leftarrow$ compute $\mathrm{WSC}(\theta_t, z_t, R_t)$ from $(39)$ with $\widehat{d}_{\mathrm{AB}}$ estimate;
        $Q_t(R_t) \leftarrow$ update as in $(42)$ with $R_t = \mathrm{WSC}_t$;
    end
    Positioning Slot starts;
    Alice sends out greedy actions for $\theta_{\mathrm{J}}$, $z_{\mathrm{J}}$ and $R_{\mathrm{J}}$ to UAVs;
    UAVs receive this information and position themselves accordingly;
    Alice receives ACK sent by the UAVs after reaching their final position;
end
```

3 Results and discussion

In this section, we evaluate the secrecy performance of the proposed system, in terms of the WSC, and the proposed RL algorithm for certain illustrative cases. The parameters used for the evaluation, unless stated otherwise, are shown in Table 1, channel-specific parameters chosen for the urban environment taken from [21, 22]. For UAV-specific parameters, such as the energy for receiving a data frame E_{RX}, for sending an ACK E_{ACK}, and the power spent on manoeuvring from one point to another P_{Mov}, we refer to values based on common transceiver energy consumption values [27] and manoeuvring power values [28]. Also, we consider a UAV movement speed of v_J and a processing time for an RLS of Δt_{RL}. The actual practical values of these parameters depend on the specific UAVs used, so the values considered here are simplified for comparison purposes.

To validate the expression for the SOP in (12), Fig. 5 shows a comparison of theoretical results and results obtained from Monte Carlo simulations for different configurations of parameters.

Note that the simulation results perfectly match with the analytical results, thus validating our expressions. As it is expected, the SOP increases as R_S increases, but it converges more rapidly for larger numbers of UAVs. A better performance in terms of SOP

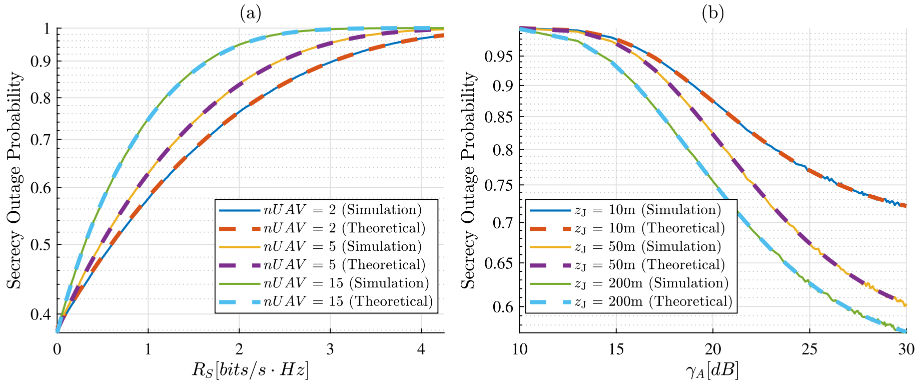

Fig. 5 **a** SOP over R_S values varying the number of UAVs (*nUAV*), **b** SOP over γ_A values varying the common UAV height z_J

is obtained with higher transmit SNR values, tending to a floor in the performance. However, as the height of the UAV increases, this floor of the SOP decreases and it is reached more slowly.

To evaluate the performance of our algorithm, in the following figures, results from Monte Carlo simulations are presented by considering two performance metrics, the energy spent over time and the secrecy performance in terms of the WSC obtained over time. The energy spent is presented as a cumulative metric over a given time step (PLB). The secrecy performance is illustrated as the WSC obtained at each time step and normalised by the secrecy area. These results are also compared to the WSC resulting from an exhaustive search over discretised positioning values. Figure 6 presents the normalised WSC and the cumulative energy consumed obtained over time (PLBs) by the proposed algorithm for different values of σ_{AB}. Note that the WSC increases until it reaches a convergence level, which is higher as σ_{AB} decreases, obtaining a better secrecy performance. This behaviour occurs because, as σ_{AB} decreases, the variance of the estimates of the action rewards also decreases; thus, more reliable action reward estimates are obtained, and it is more likely to choose the optimal actions from the discretised sets.

The energy consumption of the UAVs remains the same over the first time steps, but increases more rapidly for lower values of σ_{AB}. This is expected as at lower σ_{AB} values, the estimates of the action rewards are more reliably found earlier, and any new sample taken to adjust the estimates will not cause a big deviation from its current value (low variance). As the same actions are more reliable chosen, UAVs move less between PLBs, thus consuming less energy. In general, a smaller uncertainty of the distance between A and B will achieve greater secrecy performance and, at the same time, reduce the power consumption of the UAVs.

Figure 7 shows the impact of the shape parameter K of the A2G channels on the normalised WSC and the cumulative energy consumed obtained over time (PLBs) for different values of K. Note that a strong LoS component, higher K, leads to a significant loss on the WSC. However, the convergence for lower K values is slower, thus involving more movement between actions that may be further apart, which increases the energy consumption.

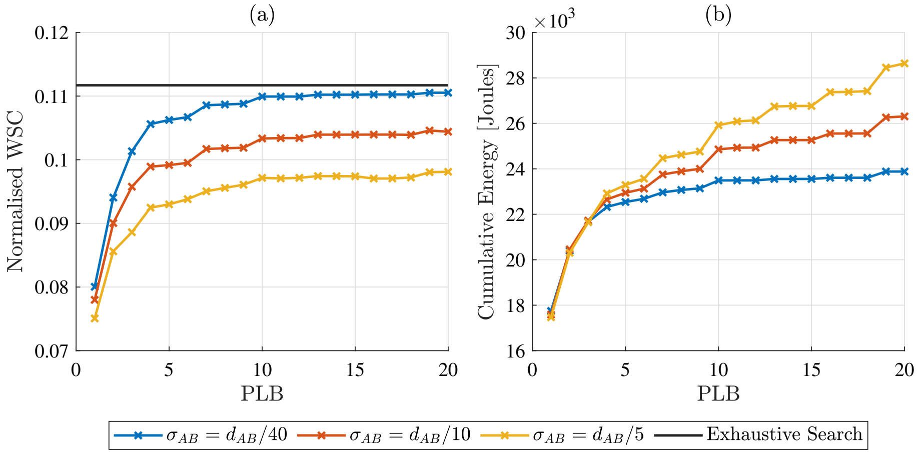

Fig. 6 **a** Normalised WSC mean values obtained over time, and **b** cumulative energy consumed in kilo-Joules by all the UAVs over time. Both measured over PLBs with varying uncertainty of the distance between A and B σ_{AB}, compared to the exhaustive search results

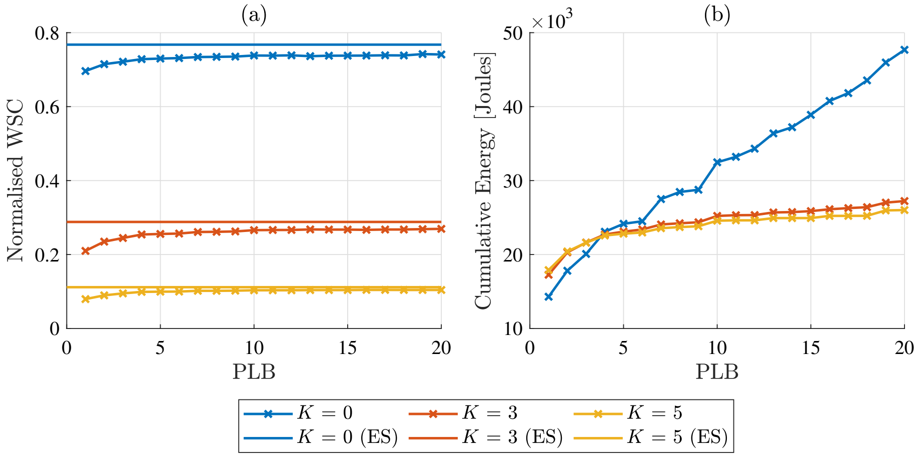

Fig. 7 **a** Normalised WSC mean values obtained over time, and **b** cumulative energy consumed in kilo-Joules by all the UAVs over time. Both measured over PLBs with varying number of shape parameter *K* of the A2G links, compared to the exhaustive search (ES) results

Figure 8 shows the impact of the number of UAVs in the system over the normalised WSC and the cumulative energy consumed obtained over time (PLBs) for different numbers of *nUAV*. The transmit SNR at each UAV is considered as $\gamma_J = \gamma_T/nUAV$. The results obtained by exhaustive search are also illustrated.

Note that as more UAVs are introduced in the system, while maintaining the total jamming power constant, the secrecy performance decreases. It can mean that having more UAVs affect more the legitimate node B than the illegitimate node E, as it is considered that E can be anywhere in the region S. It is also observed that a good level of convergence is reached up to 3 UAVs within 10 PLBs; thus, the energy consumption over time is maintained low. However, the energy consumption increases drastically for four UAVs. This can be explained due to the late convergence of the case with four UAVs, suggesting a more erratic, less stable movement as the number of UAVs increases. This result

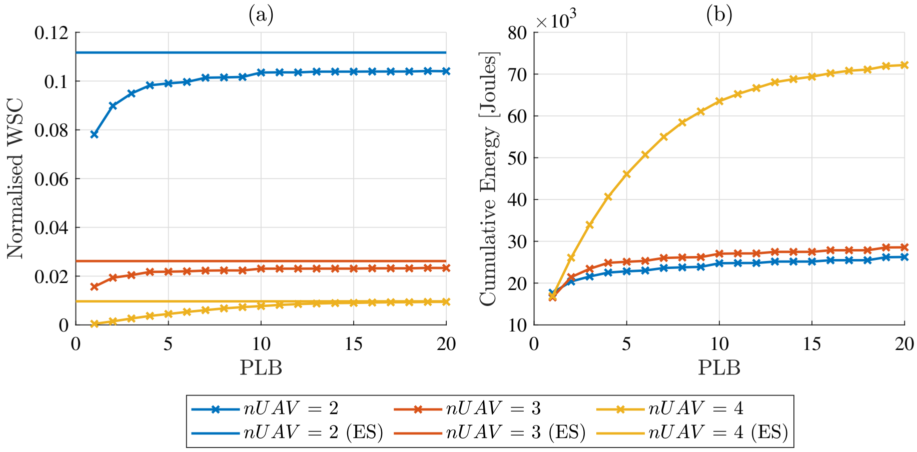

Fig. 8 **a** Normalised WSC mean values obtained over time, and **b** cumulative energy consumed in kilo-Joules by all the UAVs over time. Both measured over PLBs with varying number of UAVs maintaining the total jamming power constant and compared to the exhaustive search (ES) results

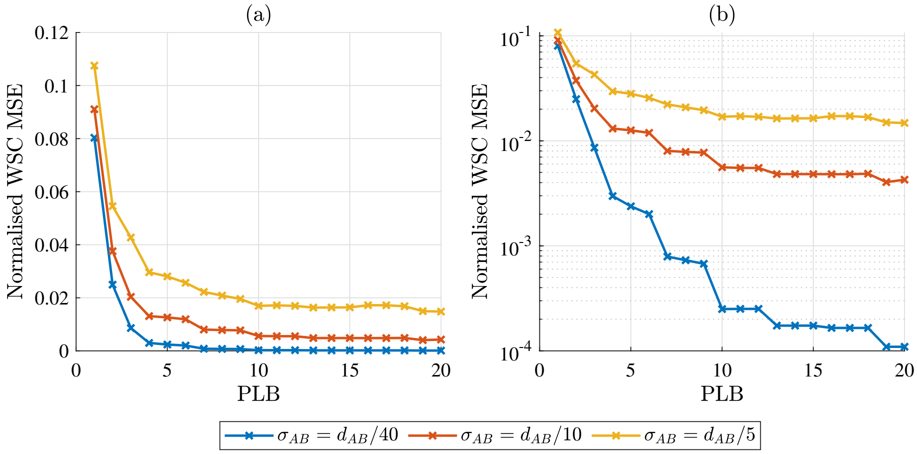

Fig. 9 Normalised MSE of WSC obtained with the algorithm, compared to exhaustive search results over time for varying values of in **a** linear scale and **b** semilogarithmic scale. Both measured over PLBs with varying uncertainty of the distance between A and B σ_{AB}

suggests that the inclusion of two UAVs may be enough and efficient to provide secret transmissions to a single legitimate pair.

Finally, to analyse the convergence of the algorithm, Fig. 9 shows the normalised minimum squared error (MSE) of the WSC, which is obtained by comparing to the exhaustive search results over time and then normalised to the exhaustive search value. The results are shown for different values of σ_{AB}. Note that the algorithm quickly converges to low values of MSE as it reaches a steady low level within 10 PLBs. As σ_{AB} increases, the MSE converges to a higher level, which occurs because a higher uncertainty introduces a larger variance in the action estimates, allowing for the optimal actions to be chosen less reliably, thus increasing the MSE.

It is also worth noting that simulations with $nRLS = 10$ within 10 PLBs (100 RLSs in total), where the UCB algorithm has been applied to the three positioning variables 100 times each, proved to be enough to reach a good level of convergence with very low

Table 1 Common simulation parameters

Parameter	Value	Parameter	Value
γ_{A}	23 (dB)	α	0.5
γ_{T}	13 (dB)	N_θ	90
d_{AB}	20 (m)	N_z	30
R_A	50 (m)	N_R	30
Δt_{RL}	0.3 (s)	z_{MIN}	100 (m)
v_J	2 (m/s)	z_{MAX}	200 (m)
E_{RX}	0.1 (J)	R_{MAX}	50 (m)
E_{ACK}	0.1 (J)	ψ (Urban)	9.61
P_{Mov}	80 (W)	ω (Urban)	0.16
$nPLB$	20	ξ_{LoS} (Urban)	1.0
$nRLS$	10	ξ_{NLoS} (Urban)	20
σ_{AB}	1 (m)	α_G (Urban)	0.3
c	0.3	α_J (Urban)	0.3
$K_{\mathrm{J_iU}}$	5	$nUAV$	2

MSE. This convergence speed is possible because the number of actions for each of the positioning variables is kept relatively low. Importantly, the treatment of the three MAB processes as independent favours the convergence speed, compared to a joint action space or state-action pairs that would greatly increase the amount of actions or state-action values to be considered.

4 Conclusions

This paper investigated the secrecy performance of a legitimate transmission between a pair of ground nodes aided by N friendly UAV-based jammers , in terms of the secrecy metric WSC, that measures the efficiency of friendly jamming over an area and is obtained from the SOP; thus, the exact position of the eavesdropper is not assumed. For that purpose, we first derived an integral-form expression for the SOP of the proposed system, which was validated via Monte Carlo simulations. Additionally, we proposed an RL-based algorithm to optimise the 3D positioning of the UAVs in order to maximise the WSC. The time frame-based algorithm periodically updates the positioning information of the UAVs and allows a control of energy consumption for UAV positioning.

Extensive simulations showed that the proposed algorithm improved the secrecy of the system over time and converged to the exhaustive search upper bound, as the uncertainty of the position of B decreases. The proposed time frame structure of the algorithm proved to be efficient to lead to optimal values of WSC while being flexible with the trade-off between secrecy and energy consumption. Furthermore, the algorithm can be explored for solving different problems in novel wireless communications networks that require periodic parameter updates to be learnt over time in a non-stationary environment.

Abbreviations

AWGN	Additive white Gaussian noise
ACK	Acknowledgement
CDF	Cumulative distribution function
CSI	Channel state information

Los	Line-of-sight
MAB	Multi-armed bandits
MSE	Mean squared error
NLoS	Non-line-of-sight
PLS	Physical layer security
RL	Reinforcement learning
PLB	Positioning learning block improvement block
RLS	RL slots
SINR	Signal-to-interference plus noise ratio
SNR	Signal-to-noise ratio
SOP	Secrecy outage probability
UAV	Unmanned aerial vehicle
UCB	Upper confidence bound
WSC	Weighted secrecy coverage

Author Contributions

ML-a contributed heavily to the conception of the study. DPMO contributed to the conception of the study, revised and verified the methods, results and the entire article, and wrote the Introduction section. XAFC wrote the entire article except for the Introduction section, developed the proposed algorithm, carried out the simulations and prepared the graphs. All authors read and approved the final manuscript.

Funding

This work was partially supported by the Academy of Finland, 6Genesis Flagship, under Grant 318927 and FAITH project under Grant 334280.

Data Availability

The scripts used for data gathering during the current study are available from the following repository: https://github.com/xflorescStaff/MAB-based-UAV-positioning.git.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 26 October 2021 Accepted: 1 September 2022
Published online: 14 September 2022

References

1. P. Porambage, G. Gür, D.P.M. Osorio, M. Liyanage, A. Gurtov, M. Ylianttila, The roadmap to 6G security and privacy. IEEE Open J. Commun. Soc. **2**, 1094–1122 (2021). https://doi.org/10.1109/OJCOMS.2021.3078081
2. D.P. Moya Osorio, I. Ahmad, J.D.V. Sánchez, A. Gurtov, J. Scholliers, M. Kutila, P. Porambage, Towards 6g-enabled internet of vehicles: security and privacy. IEEE Open J. Commun. Soc. **3**, 82–105 (2022)
3. W. Jiang, B. Han, M.A. Habibi, H.D. Schotten, The road towards 6G: a comprehensive survey. IEEE Open J. Commun. Soc. **2**, 334–366 (2021). https://doi.org/10.1109/OJCOMS.2021.3057679
4. Y. Zeng, Q. Wu, R. Zhang, Accessing from the sky: a tutorial on UAV communications for 5G and beyond. Proc. IEEE **107**(12), 2327–2375 (2019). https://doi.org/10.1109/JPROC.2019.2952892
5. L. Mucchi, S. Jayousi, S. Caputo, E. Panayirci, S. Shahabuddin, J. Bechtold, I. Morales, R.-A. Stoica, G. Abreu, H. Haas, Physical-layer security in 6G networks. IEEE Open J. Commun. Soc. **2**, 1901–1914 (2021). https://doi.org/10.1109/OJCOMS.2021.3103735
6. D.P. Moya Osorio, J. Vega Sanchez, H. Alves, Physical-Layer Security for 5G and Beyond (2019), pp. 1–19. https://doi.org/10.1002/9781119471509.w5GRef152
7. X. Sun, D.W.K. Ng, Z. Ding, Y. Xu, Z. Zhong, Physical layer security in UAV systems: Challenges and opportunities. IEEE Wirel. Commun. **26**(5), 40–47 (2019). https://doi.org/10.1109/MWC.001.1900028
8. Y. Zhou, P.L. Yeoh, H. Chen, Y. Li, R. Schober, L. Zhuo, B. Vucetic, Improving physical layer security via a UAV friendly jammer for unknown eavesdropper location. IEEE Trans. Veh. Technol. **67**(11), 11280–11284 (2018). https://doi.org/10.1109/TVT.2018.2868944
9. X. Pang, M. Liu, N. Zhao, Y. Chen, Y. Li, F.R. Yu, Secrecy analysis of UAV-based mmWave relaying networks. IEEE Trans. Wireless Commun. **20**(8), 4990–5002 (2021). https://doi.org/10.1109/TWC.2021.3064365
10. Y. Yapici, N. Rupasinghe, I. Güvenç, H. Dai, A. Bhuyan, Physical layer security for NOMA transmission in mmWave drone networks. IEEE Trans. Veh. Technol. **70**(4), 3568–3582 (2021). https://doi.org/10.1109/TVT.2021.3066350
11. M. Kim, S. Kim, J. Lee, Securing communications with friendly unmanned aerial vehicle jammers. IEEE Trans. Veh. Technol. **70**(2), 1972–1977 (2021). https://doi.org/10.1109/TVT.2021.3052503
12. P.X. Nguyen, V.-D. Nguyen, H.V. Nguyen, O.-S. Shin, UAV-assisted secure communications in terrestrial cognitive radio networks: joint power control and 3D trajectory optimization. IEEE Trans. Veh. Technol. **70**(4), 3298–3313 (2021). https://doi.org/10.1109/TVT.2021.3062283
13. W. Wang, X. Li, R. Wang, K. Cumanan, W. Feng, Z. Ding, O.A. Dobre, Robust 3D-trajectory and time switching optimization for dual-UAV-enabled secure communications. IEEE J. Sel. Areas Commun. (2021). https://doi.org/10.1109/JSAC.2021.3088628

14. X. Guo, Y. Chen, Y. Wang, Learning-based robust and secure transmission for reconfigurable intelligent surface aided millimeter wave UAV communications. IEEE Wirel. Commun. Lett. **10**(8), 1795–1799 (2021). https://doi.org/10.1109/LWC.2021.3081464
15. R. Dong, B. Wang, K. Cao, Deep learning driven 3D robust beamforming for secure communication of UAV systems. IEEE Wirel. Commun. Lett. **10**(8), 1643–1647 (2021). https://doi.org/10.1109/LWC.2021.3075996
16. Y. Zhang, Z. Mou, F. Gao, J. Jiang, R. Ding, Z. Han, UAV-enabled secure communications by multi-agent deep reinforcement learning. IEEE Trans. Veh. Technol. **69**(10), 11599–11611 (2020). https://doi.org/10.1109/TVT.2020.3014788
17. J.P. Vilela, M. Bloch, J. Barros, S.W. McLaughlin, Wireless secrecy regions with friendly jamming. IEEE Trans. Inf. Forensics Secur. **6**(2), 256–266 (2011). https://doi.org/10.1109/TIFS.2011.2111370
18. X.A.F. Cabezas, D.P.M. Osorio, M. Latva-aho, Weighted secrecy coverage analysis and the impact of friendly jamming over UAV-enabled networks, in *2021 Joint European Conference on Networks and Communications 6G Summit (EuCNC/6G Summit)* (2021), pp. 124–129. https://doi.org/10.1109/EuCNC/6GSummit51104.2021.9482493
19. X.A.F. Cabezas, D.P.M. Osorio, M. Latva-aho, Distributed UAV-enabled zero-forcing cooperative jamming scheme for safeguarding future wireless networks, in *2021 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC2021)*
20. Y. Zhou, P.L. Yeoh, H. Chen, Y. Li, R. Schober, L. Zhuo, B. Vucetic, Improving physical layer security via a UAV friendly jammer for unknown eavesdropper location. IEEE Trans. Veh. Technol. **67**(11), 11280–11284 (2018). https://doi.org/10.1109/TVT.2018.2868944
21. V. Dao, H. Tran, S. Girs, E. Uhlemann, Reliability and fairness for UAV communication based on non-orthogonal multiple access, in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)* (2019), pp. 1–6. https://doi.org/10.1109/ICCW.2019.8757160
22. A. Al-Hourani, S. Kandeepan, S. Lardner, Optimal LAP altitude for maximum coverage. IEEE Wirel. Commun. Lett. **3**(6), 569–572 (2014). https://doi.org/10.1109/LWC.2014.2342736
23. N. Bhargav, S.L. Cotton, D.E. Simmons, Secrecy capacity analysis over κ-μfading channels: theory and applications. IEEE Trans. Commun. **64**(7), 3011–3024 (2016)
24. I.S. Gradshteyn, I.M. Ryzhik, D. Zwillinger, V. Moll, *Table of Integrals, Series, and Products*, 8th edn. (Academic Press, Amsterdam, 2014)
25. S.J. Wright, Coordinate descent algorithms. Math. Program. **151**(1), 3–34 (2015)
26. R.S. Sutton, A.G. Barto, *Reinforcement Learning, Second Edition: An Introduction. Adaptive Computation and Machine Learning series* (MIT Press, Cambridge, 2018)
27. D.P. Moya Osorio, E.E. Benítez Olivo, H. Alves, J.C.S. Santos Filho, M. Latva-aho, An adaptive transmission scheme for amplify-and-forward relaying networks. IEEE Trans. Commun. **65**(1), 66–78 (2017). https://doi.org/10.1109/TCOMM.2016.2616136
28. C.W. Chan, T.Y. Kam, A procedure for power consumption estimation of multi-rotor unmanned aerial vehicle. J. Phys. Conf. Ser. **1509**, 012015 (2020). https://doi.org/10.1088/1742-6596/1509/1/012015

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.