We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,300
Open access books available

170,000
International authors and editors

185M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Chapter

Optimal Unmanned Aerial Vehicle Control and Designs for Load Balancing in Intelligent Wireless Communication Systems

Abhishek Mondal, Deepak Mishra, Ganesh Prasad and Ashraf Hossain

Abstract

Maintaining reliable wireless connectivity is essential for the continuing growth of mobile devices and their massive access to the Internet of Things (IoT). However, terrestrial cellular networks often fail to meet their required quality of service (QoS) demand because of the limited spectrum capacity. Although the deployment of more base stations (BSs) in a concerned area is costly and requires regular maintenance. Alternatively, unmanned aerial vehicles (UAVs) could be a potential solution due to their ability of on-demand coverage and the high likelihood of strong line-of-sight (LoS) communication links. Therefore, this chapter focuses on a UAV's deployment and movement design that supports existing BSs by reducing data traffic load and providing reliable wireless communication. Specifically, we design UAV's deployment and trajectory under an efficient resource allocation strategy, i.e., assigning devices' association indicators and transmitting power to maximize overall system's throughput and minimize the total energy consumption of all devices. For these implementations, we adopt reinforcement learning framework because it does not require all information about the system environment. The proposed methodology finds optimal policy using the Markov decision process, exploiting the previous environment interactions. Our proposed technique significantly improves the system's performance compared to the other benchmark schemes.

Keywords: unmanned aerial vehicle, reinforcement learning, energy efficiency, offloading, throughput

1. Introduction

With the proliferation of mobile electronic devices, such as smartphones, tablets, and more internet of things (IoT) gadgets, the need for high-speed wireless connectivity has been growing rapidly [1]. But, the existing cellular networks with limited

spectrum, coverage, and energy capacity fail to satisfy users' quality of service (QoS) requirements. Hence, the next generation 5G technologies, such as device-to-device (D2D) communications, ultra-dense small cell networks, and millimeter wave (mmW) communications, are emerging as potential alternatives to deal with such issues [2, 3]. However, these modern 5G cellular networks face several challenges due to resource allocation, backhaul interferences, high reliance on the line of sight (LoS) link, and signal blockage. On the other hand, integration of unmanned aerial vehicles (UAVs) into the fifth-generation (5G) and sixth-generation (6G) cellular networks as aerial base stations would be a promising aspect to achieve several goals, namely ubiquitous accessibility, robust navigation, ease of monitoring and management, etc., because they can establish LoS dominant air to ground channel in a controllable manner [4]. Notably, cellular-connected UAV-assisted system gains significant performance improvement over the existing point-to-point UAV-ground communication in terms of coverage and throughput [5]. UAV also offload temporary high-traffic demands from terrestrial BSs during huge crowd events such as festivals, concerts, and stadium games [6]. Therefore, UAVs' utility in the cellular network is directly related to the highest number of serving users. Nevertheless, many challenges related to the utilization of UAVs need to be addressed, including their deployment strategy, trajectory optimization, and resource allocation under flight time limitations which affect instantaneous LoS probability and remarkably influence the system performance.

The relevant studies [7–10] optimized the trajectory and deployment of UAVs in different circumstances. However, most of them incorporate nonlinear algorithms that rely on average spatial throughput. Thus, computational complexity grows rapidly with the higher number of users and flight time. Moreover, practically without prior knowledge about the network state, it becomes very difficult for a UAV to find its path to accomplish a given real-time task. Alternatively, machine learning (ML) techniques [11–13] intelligently support UAVs and ground users in performing mission-oriented operations with low complexity when complete network information is not available. Particularly, reinforcement learning (RL), being a part of ML, can search for the optimal policy through trial and error while interacting with the environment [14]. Hence, this chapter investigates the optimal deployment, trajectory, and resource allocation of UAVs to meet the throughput requirements of the cellular network.

2. Background

The existing literature focuses on the deployment and movement of UAV relays for numerous applications. In [15], the authors estimated the optimal UAV relay position in a multi-rate communication system using theoretical and simulated analysis. The work in [16] investigated the mission planning of UAV relays to improve the connectivity of ground users. The authors of [17, 18] maximized the lower bound of the uplink transmission rate over the link between UAV relay and ground devices using dynamic heading adjusting approaches. For throughput maximization of the mobile relaying system, an iterative algorithm was developed [19, 20], which jointly optimized the relays' trajectory and transmitting power of the sources and UAVs by satisfying the practical constraints. In [21], the authors maximized the UAV relay network's throughput by optimizing transmit power, bandwidth, transmission rate, and relay deployment. However, in these works, a model-based centralized approach

is used where all necessary system parameters are required. Additionally, the research gap still exists on enhancing network performance for source-destination device pair communication. To overcome these shortcomings, Indu et al. [22] minimized the energy consumption of UAV during its trajectory using genetic algorithm (GA). The authors in [6] proposed two meta-heuristic algorithms, such as GA and particle swarm optimization (PSO), to find the optimal UAV trajectory for satisfying users' minimum data rate requirements. They showed that PSO significantly improves the UAV's wireless coverage compared to GA. Although the meta-heuristic algorithms can deal with the complexity of UAV path planning, there are still some challenges in exchanging information between UAV and core network due to either unavailable constraints or obtaining their gradient analytically.

Another line of research studied the mobility management of UAVs for resource allocation and coverage optimization using RL techniques to deal with convergence issues. Kawamoto et al. [23] have presented a resource allocation algorithm of UAV using Q-learning techniques for allocating time slots and modulation schemes. The work in [24] presented a framework for the optimal UAV trajectory under a given data rate constraint, which relies on a state-action-reward-state-action (SARSA) algorithm. Hu et al. [25] proposed a real-time sensing and transmission protocol in UAV-aided cellular networks and designed optimal UAVs' trajectories under limited spectrum resources using RL based on a Q-learning algorithm. Furthermore, the authors of [26] transformed UAV trajectory optimization problem for maximizing cumulative collected sensors' data into a Markov decision process (MDP) and proposed two stochastic modeling RL algorithms, namely Q-learning and SARSA, to learn UAV's policy. They proved that SARSA outperforms Q-learning due to the adaptive system's state update rule. From the state-of-the-art, the coupled relationship among UAV trajectory, device association, and transmit power allocation of IoT devices for the enhancement of network lifetime has not been investigated during the data collection process of UAV-assisted IoT networks.

3. Channel characterization of UAV-operated communication system

This section proposes a multi-hop radio frequency and free space optical (RF-FSO) communication framework that analytically optimizes the UAV's altitude for performance enhancement of a relaying system. Here, we minimize the outage probability and symbol error rate based on independent and identically distributed statistical parameters i.e., pointing errors, atmospheric turbulence, and scintillation.

3.1 Channel model

Consider a multi-hop hybrid RF–FSO system as shown in **Figure 1**, where single antenna-equipped ground base stations realize periodic data exchange. Since there are significant obstacles in the LoS path, direct link cannot be established between them. Therefore, two UAVs are deployed at a certain altitude which are employed as relays between the source and destination. These UAVs operate as RF and optical link transceiver modules with single-directional apertures. Depending on various environmental conditions, three different channels categorize the source-to-destination link, i.e., Ground to UAV (G2U), UAV to UAV (U2U), and UAV to Ground (U2G) channels.
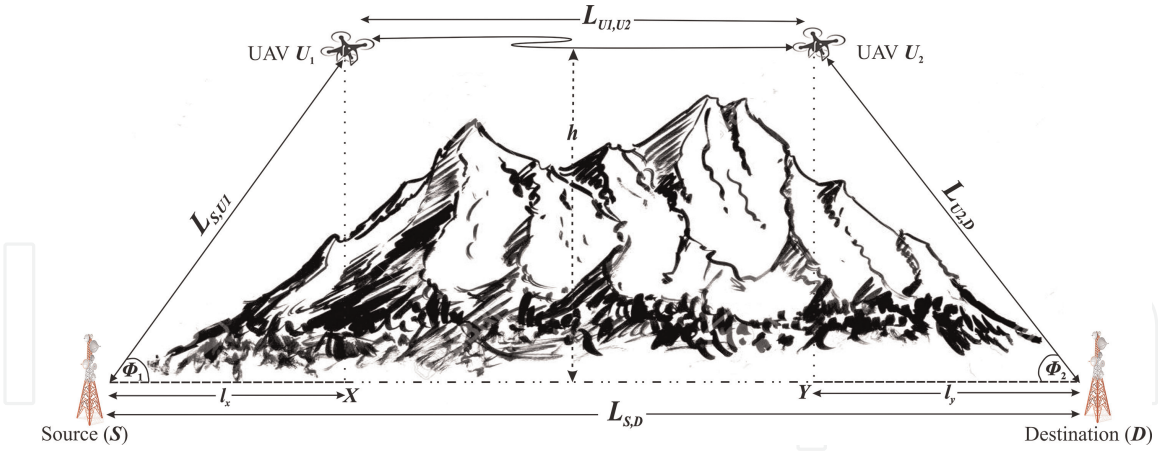
Figure 1.
UAV-assisted multihop hybrid RF–FSO system.

3.1.1 G2U channel model

As ground to UAV channel consists of RF signals, experiencing small-scale fading and large-scale path loss, the received symbol at UAV U_1 can be estimated as [27],

$$Y_{U_1} = \sqrt{P_{S,U_1}}\sqrt{a_{S,U_1}}h_{S,U_1}x_S + n_{U_1} \tag{1}$$

where, x_S is the transmitted symbol of power P_{S,U_1}, n_{U_1} represents the additive white Gaussian noise (AWGN) power of zero mean and variance N_0 at U_1, h_{S,U_1} defines the channel gain of S-U_1 link and $a_{S,U_1} = \kappa_{S,U_1}L_{S,U_1}^{-\epsilon_{S,U_1}}$ is path loss corresponding to link distance L_{S,U_1}, ϵ_{S,U_1} denotes the path loss exponent and κ_{S,U_1} is the environment-dependent constant. As multipath components govern the S-U_1 link, therefore $|h_{S,U_1}|^2 = \chi$ follows a non-central chi-square distribution, and its probability density function (PDF) is given by [28],

$$f_\chi(t) = \frac{(K_{S,U_1}+1)e^{-K_{S,U_1}}}{A_{S,U_1}} \, exp\left\{\frac{-(K_{S,U_1}+1)t}{A_{S,U_1}}\right\} \times I_0\left(2\sqrt{\frac{(K_{S,U_1}+1)K_{S,U_1}}{A_{S,U_1}}t}\right) \tag{2}$$

where $\overline{A_{S,U_1}} = E\left\{|h_{S,U_1}|^2\right\} = 1$, is average fading power, $E\{.\}$ denotes expectation operator, $I_0(.)$ defines zero order modified Bessel function, $K_{S,U_1} = |m_{S,U_1}|^2/2\sigma^2$ is Rician factor, m_{S,U_1} is the amplitude of LoS component and σ^2 is average power of multipath components. The instantaneous signal-to-noise ratio (SNR) received at UAV U_1 is expressed as [29],

$$\Upsilon_{S,U_1} = \frac{P_{S,U_1}a_{S,U_1}}{N_0}X = \overline{\Upsilon}_{S,U_1}X \tag{3}$$

where, the average SNR is given as, $\overline{\Upsilon}_{S,U_1} = \frac{P_{S,U_1}a_{S,U_1}}{N_0}$

3.1.2 U2U channel model

UAV U_1 first receive the RF signal Y_{U_1}, then convert and encode it into the optical signal and then forward it to UAV U_2 over FSO link. The received signal at UAV U_2 can be obtained as [27]

$$Y_{U_2} = \eta_{U_1} \sqrt{P_{U_1,U_2}} h_{U_1,U_2} x_{U_1} + n_{U_2} \tag{4}$$

where η_{U_1} is electrical to optical conversion coefficient of UAV U_1, x_{U_1} indicates the converted and encoded optical symbol of power P_{U_1,U_2}, n_{U_2} denotes AWGN with zero mean and variance N_0 at UAV U_2, and $h_{U_1,U_2} = h_a h_p$ is optical channel coefficient depending on atmospheric turbulence-induced fading (h_a) and pointing errors (h_p). The instantaneous SNR received at UAV U_2, can be expressed as [27]

$$\Upsilon_{U_1,U_2} = \frac{\eta_{U_1}^2 P_{U_1,U_2} h_{U_1,U_2}^2}{N_0} \tag{5}$$

Since the optical link between UAV U_1 and U_2 experience several atmospheric turbulence and corresponding optical axis misalignment, the PDF of its instantaneous SNR follows the variation of atmospheric turbulence and pointing errors, which can be expressed as [30]

$$f_{\Upsilon_{U_1,U_2}}(\Upsilon) = \frac{\xi^2}{2\Upsilon\Gamma(\alpha)\Gamma(\beta)} G_{1,3}^{3,0} \left(\alpha\beta \sqrt{\frac{\Upsilon}{\overline{\Upsilon}_{U_1,U_2}}} \Bigg|_{\xi^2,\alpha,\beta}^{\xi^2+1} \right) \tag{6}$$

where $\Gamma(.)$ is the Gamma function, α and β are scintillation parameters, ξ is the ratio between the equivalent beam radius and the misalignment displacement standard deviation at U_2, $G_{p,q}^{m,n} \left(x \Big|_{b_1,b_2,\dots,b_m,\dots,b_q}^{a_1,a_2,\dots,a_n,\dots,a_p} \right)$ is Meijer's G function and $\overline{\Upsilon}_{U_1,U_2} = P_{U_1,U_2} \eta_{U_1}^2 E\{h_{U_1,U_2}\}^2 / N_0$ is average electrical SNR.

3.1.3 U2G channel model

After receiving the optical signal Y_{U_2}, UAV U_2 first decodes and converts it to RF signal and then forwards to the destination. Hence, the channel characterization is similar as the G2U channel model, and the received signal at the destination can be expressed as [27]

$$Y_D = \eta_{U_2} \sqrt{P_{U_2,D}} \sqrt{a_{U_2,D}} h_{U_2,D} x_{U_2} + n_D \tag{7}$$

where η_{U_2} is optical to electrical conversion coefficient of UAV U_2, x_{U_2} denotes the transmitted symbol of power $P_{U_2,D}$, n_D defines AWGN of zero mean and variance N_0, $h_{U_2,D}$ is channel coefficient and $a_{U_2,D}$ is path loss attenuation factor. Instantaneous SNR received at the destination is expressed as,

$$\Upsilon_{U_2,D} = \frac{\eta_{U_2}^2 P_{U_2,D} a_{U_2,D} |h_{U_2,D}|^2}{N_0} \tag{8}$$

where $\overline{\Upsilon}_{U_2,D} = \eta_{U_2}^2 P_{U_2,D} a_{U_2,D} / N_0$ is average SNR

3.2 Performance metrics of multihop RF: FSO system

3.2.1 Outage probability

It is defined as the probability that instantaneous SNR is less than the minimum required threshold level, Υ_{th}. For decode and forward relaying mode, the equivalent SNR at destination can be expressed as [27]

$$\Upsilon_{S,D} = min\left(\Upsilon_{S,U_1}, \Upsilon_{U_1,U_2}, \Upsilon_{U_2,D}\right) \tag{9}$$

Cumulative distribution function (CDF) of equivalent SNR is expressed by,

$$
\begin{aligned}
F_{\Upsilon_{S,D}}(\Upsilon) &= Pr(\Upsilon_{S,D} \leq \Upsilon) = Pr(\,min\,(\Upsilon_{S,U_1}, \Upsilon_{U_1,U_2}, \Upsilon_{U_2,D}) \leq \Upsilon) \\
&= 1 - \left\{1 - F_{\Upsilon_{S,U_1}}(\Upsilon)\right\}\left\{1 - F_{\Upsilon_{U_1,U_2}}(\Upsilon)\right\}\left\{1 - F_{\Upsilon_{U_2,D}}(\Upsilon)\right\}
\end{aligned} \tag{10}
$$

where $F_{\Upsilon_{S,U_1}}(\Upsilon)$, $F_{\Upsilon_{U_1,U_2}}(\Upsilon)$ and $F_{\Upsilon_{U_2,D}}(\Upsilon)$ are the CDF of Υ_{S,U_1}, Υ_{U_1,U_2} and $\Upsilon_{U_2,D}$ respectively. The outage probability of the overall system is obtained in terms of Q1 (., .) i.e., the first order Marcum Q function as [31]

$$
\begin{aligned}
P_{out} &= F_{\Upsilon_{S,D}}(\Upsilon_{th}) = Pr(\Upsilon_{S,D} \leq \Upsilon_{th}) \\
&= 1 - Q_1\left(\sqrt{2K_{S,U_1}}, \sqrt{2\Upsilon_{th}L_{S,U_1}^{\epsilon_{S,U_1}}(1 + K_{S,U_1})/\tilde{\Upsilon}_{S,U_1}}\right) \\
&\quad \times Q_1\left(\sqrt{2K_{U_2,D}}, \sqrt{2\Upsilon_{th}L_{U_2,D}^{\epsilon_{U_2,D}}(1 + K_{U_2,D})/\tilde{\Upsilon}_{U_2,D}}\right) \\
&\quad \times \left[1 - \frac{\xi^2}{\Gamma(\alpha)\Gamma(\beta)} G_{2,4}^{3,1}\left(\alpha\beta\sqrt{\frac{\Upsilon_{th}}{\overline{\Upsilon}_{U_1,U_2}}}\Bigg|_{\xi^2,\alpha,\beta,0}^{1,\xi^2+1}\right)\right]
\end{aligned} \tag{11}
$$

3.2.2 Symbol error rate

It is defined as the probability of false estimation of the received symbol, which can be expressed as [32]

$$P_{M,PSK}(e) = 1 - \sum_{k=1}^{M} P_k(\Upsilon_{S,U_1})P_k(\Upsilon_{U_1,U_2})P_k(\Upsilon_{U_2,D}) \tag{12}$$

$$
P_k(\Upsilon_{s,d}) = \begin{cases}
1 - \dfrac{1}{\pi}\displaystyle\int_0^{\frac{(M-1)\pi}{M}} \mathcal{M}_{\Upsilon_{s,d}}\left(-\dfrac{sin^2\left(\frac{\pi}{M}\right)}{sin^2(\phi)}\right)d\phi, & \text{for } k = 1 \\[4mm]
\dfrac{1}{\pi}\displaystyle\int_0^{\frac{(M-1)\pi}{M}} \mathcal{M}_{\Upsilon_{s,d}}\left(-\dfrac{sin^2\left(\frac{\pi}{M}\right)}{sin^2(\phi)}\right)d\phi, & \text{for } k = \dfrac{M}{2} + 1 \\[4mm]
\left[\dfrac{1}{2\pi}\displaystyle\int_0^{\pi-a_k-1} \mathcal{M}_{\Upsilon_{s,d}}\left(-\dfrac{sin^2(a_k-1)}{sin^2(\phi)}\right)d\phi - \right. \\
\left. \dfrac{1}{2\pi}\displaystyle\int_0^{\pi-a_k} \mathcal{M}_{\Upsilon_{s,d}}\left(-\dfrac{sin^2(a_k)}{sin^2(\phi)}\right)d\phi\right], & \text{otherwise}
\end{cases} \tag{13}
$$

where, $a_k = (2k-1)\frac{\pi}{M}$. After substituting Eq. (6) in Eq. (13) and using [29], we can obtain the moment-generating function of instantaneous SNR corresponding FSO link as

$$\mathcal{M}_{\Upsilon_{U_1,U_2}}(s) = \frac{\xi^2 2^{\alpha+\beta-1}}{4\pi\Gamma(\alpha)\Gamma(\beta)} \times G_{3,6}^{6,1}\left(\frac{(\alpha\beta)^2}{16\overline{\Upsilon}_{U_1,U_2}s}\middle|\begin{matrix}1,\frac{\xi^2+1}{2},\frac{\xi^2+2}{2}\\\frac{\xi^2}{2},\frac{\xi^2+1}{2},\frac{\alpha}{2},\frac{\alpha+1}{2},\frac{\beta}{2},\frac{\beta+1}{2}\end{matrix}\right) \tag{14}$$

3.3 UAVs' optimal altitude

According to Eq. (11), outage probability is a function of UAV's altitude, distance from source to destination, and distance between the projection points of UAVs on the ground and end users. For these given parameters values, the optimal altitude is obtained as

$$\tilde{h} = l_y \tan(\tilde{\phi}_2) \tag{15}$$

where the optimal altitude must satisfy the following condition [33]

$$\tilde{h} = \arg\min_{h\epsilon[0,\infty]} P_{out}(h, l_x, l_y, L_{S,D}) \tag{16}$$

Finally, the optimal elevation angle at the receiver side $\tilde{\phi}_2$ is obtained by solving the equation,

$$[P_1.Q_1(v_2, w_2) + P_2.Q_1(v_1, w_1)].P_3 = 0 \tag{17}$$

where

$$P_1 = v_1 e^{-\frac{v_1^2+w_1^2}{2}}\left[I_1(v_1, w_1)\frac{K'_{S,U_1}(\phi_1)}{v_1} - I_0(v_1, w_1).\frac{w_1}{2}\left\{\frac{K'_{S,U_1}(\phi_1)}{1+K_{S,U_1}(\phi_1)}\right.\right. \tag{18}$$

$$\left.\left. + \epsilon'_{S,U_1}(\phi_1)\ln\left(\frac{l_x}{\cos\phi_1}\right) + \epsilon_{S,U_1}(\phi_1)\tan\phi_1\right\}\right] \times \frac{l_x l_y}{l_x^2\cos^2\phi_2 + l_y^2\sin^2\phi_2}$$

$$P_2 = v_2 e^{-\frac{v_2^2+w_2^2}{2}}\left[I_1(v_2, w_2)\frac{K'_{U_2,D}(\phi_2)}{v_2} - I_0(v_2, w_2).\frac{w_2}{2}\left\{\frac{K'_{U_2,D}(\phi_2)}{1+K_{U_2,D}(\phi_2)}\right.\right. \tag{19}$$

$$\left.\left. + \epsilon'_{U_2,D}(\phi_2)\ln\left(\frac{l_y}{\cos\phi_2}\right) + \epsilon_{U_2,D}(\phi_2)\tan\phi_2\right\}\right]$$

$$P_3 = 1 - \frac{\xi^2}{\Gamma(\alpha)\Gamma(\beta)}G_{2,4}^{3,1}\left(\alpha\beta\sqrt{\frac{\Upsilon_{th}}{\overline{\Upsilon}_{U_1,U_2}}}\middle|\begin{matrix}1,\xi^2+1\\\xi^2,\alpha,\beta,0\end{matrix}\right) \tag{20}$$

3.4 Numerical results

In this section, we provide numerical insights of optimal UAVs' altitude and corresponding performance analysis and then cross-validate the proposed methodology

using Monte-Carlo simulation. We assume that the system is operated under moderate and strong atmospheric turbulence conditions with a maximum free space optical distance 7 km, where the average SNR is set as $\overline{\Upsilon}_{S,U_1} = \overline{\Upsilon}_{U_1,U_2} = \overline{\Upsilon}_{U_2,D} = 75$ dB.

The variations of elevation angle corresponding to the optimal UAVs' altitude for the given distance between the projection points of UAVs on the ground and end users under moderate atmospheric turbulence conditions are depicted in **Figure 2**. According to this figure, the optimal elevation angles decrease with the increase in distance from the end-user location to the projection point of the UAVs on the ground because the variation of optimal elevation angle follows Eq. (15).

The variation of outage probability with respect to UAVs' altitude under moderate atmospheric turbulence conditions is statistically visualized in **Figure 3** when the SNR threshold is assumed as $\Upsilon_{th} = 0.4$. Since small-scale fading and signal path loss less
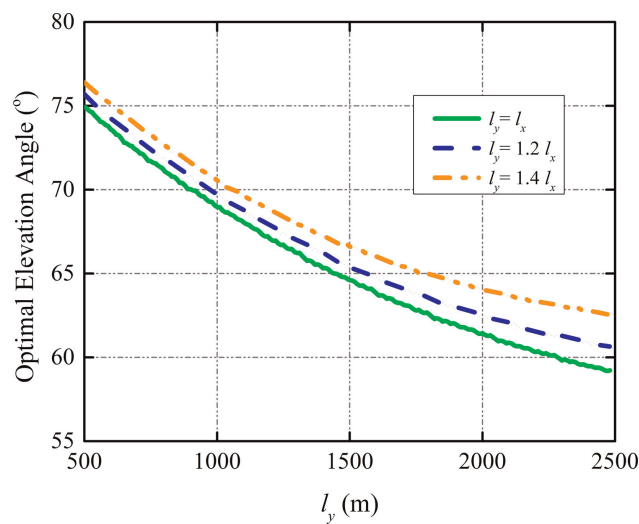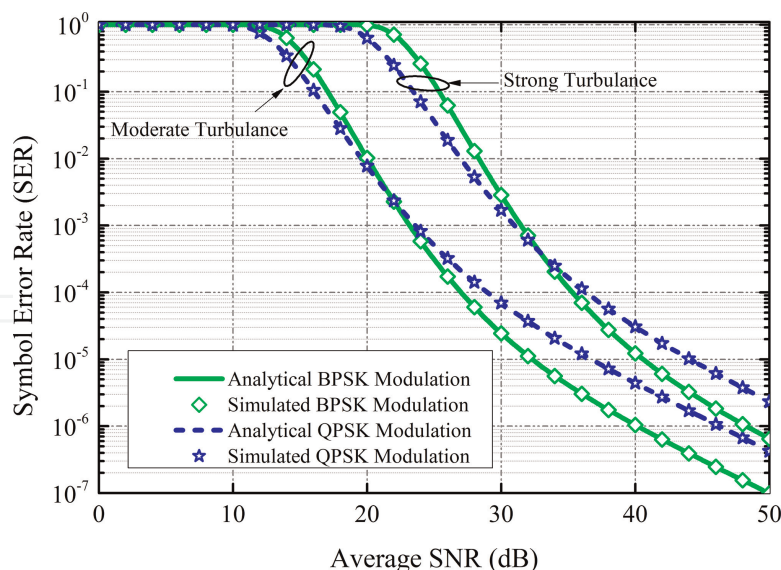

Figure 2.
Variation of optimal elevation angle while considering $\Upsilon_{th} = 0.1$.
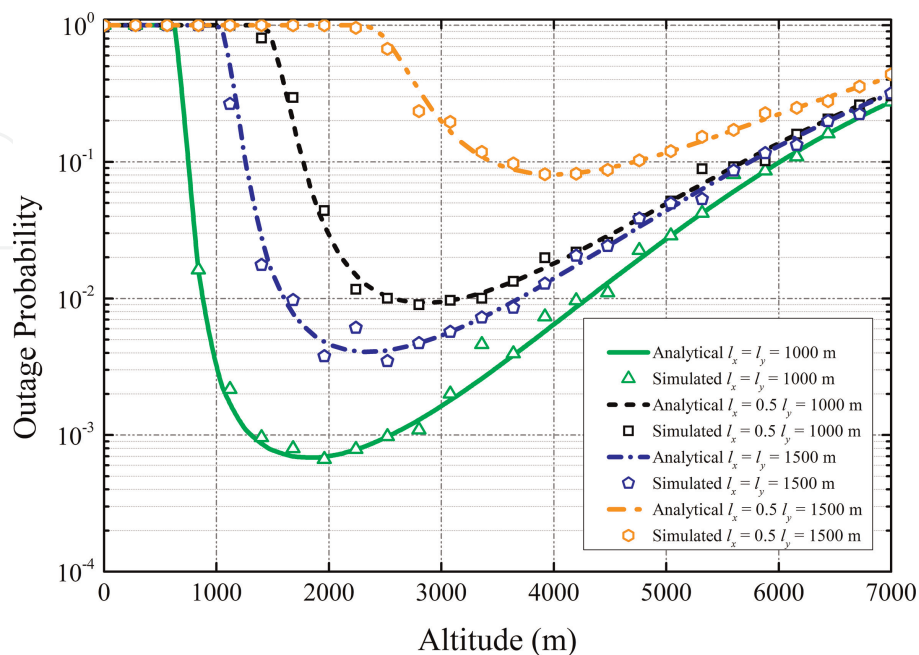
Figure 3.
Outage probability variation for different UAVs' altitude.

Figure 4.
Variation of symbol error rate for different modulation schemes.

affect the received SNR at the optimal altitude, minimum outage probability can be achieved at that altitude. On the other hand, outage probability increases if UAVs' altitude deviates from the optimal value.

Figure 4 shows the impact of various modulation schemes on symbol error rate when the distance between projection points of UAVs on the ground and end users is 2000 m under different atmospheric turbulence conditions. According to the result, it is observed that symbol error rate decreases with the average SNR value. Furtherore, binary phase shift keying (BPSK) outperforms the modulation scheme of quadrature phase shift keying (QPSK). Although higher modulation techniques offer more data rates and bandwidth efficiency, they are more complicated to implement, require a more stringent RF amplifier, and are less resilient to error. Therefore, BPSK offers more secure and errorless transmission than other modulation techniques.

4. Throughput maximization in UAVs-supported D2D network

This section proposes a UAVs-supported self-organized device-to-device (USSD2D) network containing multiple source-destination device pairs and multiple UAVs, where the objective is to find the optimal deployed location of UAVs to support reliable data transmission between source and destination device pairs. Here, we consider SNR-constrained maximization of the total instantaneous transmission rate of the USSD2D network by jointly optimizing device association, UAV's channel selection, and UAVs' deployed location at every time slot.

4.1 System model

Figure 5 depicts the UAVs-supported self-organized device-to-device (USSD2D) network where the stationary source and destination devices pairs are randomly deployed on the ground within the target area. The direct D2D pairs can establish LoS links due to good channel conditions and the short distance between them. On the other hand, UAV-assisted D2D pairs cannot establish direct links due to the presence
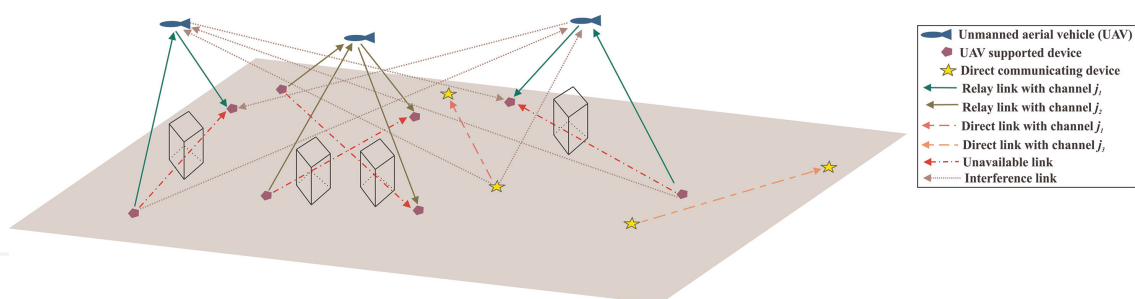
Figure 5.
UAVs-supported self-organized device-to-device network.

of significant obstacles in the signal propagation path and thereby utilize the deployed UAVs as relays.

4.1.1 Channel model

Consider M number of UAVs represented by $\mathcal{M} = \{1, 2, \dots, M\}$ at a fixed altitude of H_u acting as relays for $\overline{K}$ number of direct D2D pairs and $\tilde{K}$ number of UAV-assisted D2D pairs. There are total J number of orthogonal channels represented by $\mathcal{J} = \{1, 2, \dots, J\}$ in the USSD2D network, and each UAV selects a single orthogonal channel at a time. The set of source and destination devices of the direct D2D and UAV-assisted D2D pairs are represented as $\overline{\mathcal{K}}_S = \{1, 2, \dots, \overline{K}\}$, $\overline{\mathcal{K}}_D = \{\overline{K} + 1, \overline{K} + 2, \dots, 2\overline{K}\}$, $\tilde{\mathcal{K}}_S = \{1, 2, \dots, \tilde{K}\}$ and $\tilde{\mathcal{K}}_D = \{\tilde{K} + 1, \tilde{K} + 2, \dots, 2\tilde{K}\}$ respectively where kth device's location is $(x_k, y_k), \forall k \in \{\overline{\mathcal{K}}_S \cup \overline{\mathcal{K}}_D \cup \tilde{\mathcal{K}}_S \cup \tilde{\mathcal{K}}_D\}$. UAVs' flight period is discretized into T equally spaced time slots of duration δ each and mth UAV's location $U_m(t) = (x_m(t), y_m(t), H_u), \forall m \in \mathcal{M}, t \in \mathcal{T} = \{1, 2, \dots, T\}$ is almost unchanged within each slot. Here, we assume that one source device can only associate with a single UAV at a time slot, but multiple devices can access a single UAV simultaneously. To avoid mutual interference from nearby devices, UAVs select the orthogonal channel, and data transmission follows amplify and forward relaying (AF) protocol [34]. The association indicator of the $\tilde{k} \in \{\tilde{\mathcal{K}}_S \cup \tilde{\mathcal{K}}_D\}$ device with UAV m at time slot t is defined as

$$\overline{I}_{\tilde{k},m}(t) = \begin{cases} 1, \text{if device } \tilde{k} \text{ associates with UAV } m \\ \quad 0, \text{Otherwise} \end{cases} \tag{21}$$

Similarly, when UAV m selects an orthogonal channel j at tth time slot, the corresponding channel selection indicator is defined as

$$\tilde{I}_{m,j}(t) = \begin{cases} 1, \text{if UAV } m \text{ selects channel } j \\ \quad 0, \text{Otherwise} \end{cases} \tag{22}$$

The path loss between the device $\tilde{k}$ and UAV m can be expressed as [35]

$$L_{\tilde{k},m}(t) = \frac{\mu_{LoS} - \mu_{NLoS}}{1 + b_1 \exp\left[-b_2\left(\frac{180}{\pi}\phi_{\tilde{k},m}(t) - b_1\right)\right]} + 20\log\left(\frac{4\pi f_c D_{\tilde{k},m}(t)}{c}\right) + \mu_{NLoS} \tag{23}$$

where c is the speed of light, f_c is the carrier frequency, μ_{LoS} and μ_{NLoS} are attenuation factors corresponding to the LoS and NLoS path, respectively, b_1 and b_2 are the constant. $\phi_{\tilde{k},m}(t) = \sin^{-1}\left(H_u/D_{\tilde{k},m}(t)\right)$ is the elevation angle between the device $\tilde{k}$ and UAV m, where the instantaneous distance between them is calculated as $D_{\tilde{k},m}(t) = \sqrt{(x_m(t) - x_{\tilde{k}})^2 + (y_m(t) - y_{\tilde{k}})^2 + H_u^2}$. The instantaneous channel gain between $\tilde{k}$th device and relay UAV m can be expressed as

$$G_{\tilde{k},m}(t) = 10^{-L_{\tilde{k},m}(t)/10} \tag{24}$$

4.1.2 Transmission model

The received SNR at UAV m from the source device $\tilde{k}$ over channel j can be expressed as [34]

$$\Gamma_{\tilde{k},m}^{j}(t) = \frac{P_{\tilde{k}}^{Tx} G_{\tilde{k},m}(t) \bar{I}_{\tilde{k},m}(t) \tilde{I}_{m,j}(t)}{N_0} \tag{25}$$

where $P_{\tilde{k}}^{Tx}$ is transmit power of $\tilde{k}$ device and N_0 is noise power. The expected SNR received by the destination device $\tilde{k} + \tilde{K} \in \tilde{\mathcal{K}}_D$ from UAV m over channel j can be expressed as

$$\hat{\Gamma}_{m,\tilde{k}+\tilde{K}}^{j}(t) = \frac{P_m^{Tx} G_{m,\tilde{k}+\tilde{K}}(t) \bar{I}_{\tilde{k}+\tilde{K},m}(t) \tilde{I}_{m,j}(t)}{N_0} \tag{26}$$

where P_m^{Tx} is transmit power of UAV m. The overall SNR at the destination device of the UAV-assisted D2D pair following AF relaying protocol can be expressed as [36]

$$\hat{\Gamma}_{\tilde{k},\tilde{k}+\tilde{K}}^{j}(t) = \left[\prod_{i=1}^{N}\left(1 + \frac{1}{\Gamma_i^{j}(t)}\right) - 1\right]^{-1} \tag{27}$$

where $\Gamma_i^{j}(t)$ is the instantaneous SNR of the ith hop over jth channel, and N is the total number of hops in the link. For direct D2D pair, we consider a conventional channel model where the instantaneous channel gain between the source device $\bar{k}$ and destination device $\bar{k} + \bar{K}$ can be expressed as

$$G_{\bar{k},\bar{k}+\bar{K}}(t) = \beta_0 D_{\bar{k},\bar{k}+\bar{K}}^{-\varrho}(t) \tag{28}$$

where $\beta_0 = \left(4\pi f_c/c\right)^2$ is free space path loss at a distance of 1 m, and ϱ is the path loss exponent. The expected instantaneous SNR received by the destination device $\bar{k} + \bar{K}$ from the source device $\bar{k}$ over channel j can be expressed as

$$\hat{\Gamma}_{\bar{k},\bar{k}+\bar{K}}^{j}(t) = \frac{P_{\bar{k}}^{Tx} G_{\bar{k},\bar{k}+\bar{K}}(t)}{N_0} \tag{29}$$

The instantaneous transmission rate achieved by the destination device $\overline{k} + \overline{K}$ can be expressed as

$$\overline{R}^{j}_{\overline{k},\overline{k}+\overline{K}}(t) = B \log_2 \left[1 + \hat{\Gamma}^{j}_{\overline{k},\overline{k}+\overline{K}}(t) \right] \tag{30}$$

The total instantaneous transmission rate achieved by all direct D2D pairs can be calculated as

$$\overline{R}_{Sum}(t) = \sum_{j=1}^{J} \sum_{\overline{k}=1}^{\overline{K}} \overline{R}^{j}_{\overline{k},\overline{k}+\overline{K}}(t) \tag{31}$$

Similarly, $\tilde{k} + \tilde{K}$th device obtains the instantaneous transmission rate over channel j as

$$\tilde{R}^{j}_{\tilde{k},\tilde{k}+\tilde{K}}(t) = B \log_2 \left[1 + \hat{\Gamma}^{j}_{\tilde{k},\tilde{k}+\tilde{K}}(t) \right] \tag{32}$$

The total instantaneous transmission rate of all UAV-assisted D2D pairs can be expressed as

$$\tilde{R}_{Sum}(t) = \sum_{j=1}^{J} \sum_{m=1}^{M} \sum_{\tilde{k}=1}^{\tilde{K}} \tilde{R}^{j}_{\tilde{k},\tilde{k}+\tilde{K}}(t) \tag{33}$$

The overall instantaneous transmission rate of the USSD2D network is formulated as

$$R_{Sum}(t) = \overline{R}_{Sum}(t) + \tilde{R}_{Sum}(t) \tag{34}$$

4.1.3 Problem formulation

From the practical scenario, it is observed that when UAVs fly toward a group of devices to obtain better channel conditions, the remaining devices of the network cannot receive adequate services from the UAV, and consequently, UAVs cannot allocate network resources fairly. Hence, we jointly optimize UAVs' location, device association, and channel selection indicators at every time slot to maximize the total instantaneous transmission rate of the USSD2D network while assuring that each device should achieve a minimum SNR of ς to maintain the required QoS. The corresponding optimization problem is formulated as

$$\text{P1}: \left\{ \begin{array}{c} \text{Maximize} \\ (x_m(t), y_m(t)), \overline{I}_{k,m}(t), \tilde{I}_{m,j}(t) \\ \forall k \in \left\{ \overline{\mathcal{K}}_S \cup \overline{\mathcal{K}}_D \cup \tilde{\mathcal{K}}_S \cup \tilde{\mathcal{K}}_D \right\}, m \in \mathcal{M}, j \in \mathcal{J} \end{array} \right\} R_{Sum}(t) \tag{35}$$

Subject to the constraints

$$C1 : \Gamma^{j}_{k,k+K}(t) > \varsigma, \forall k \in \left\{ \overline{\mathcal{K}}_S \cup \overline{\mathcal{K}}_D \cup \tilde{\mathcal{K}}_S \cup \tilde{\mathcal{K}}_D \right\} \tag{36}$$

$$C2 : \overline{I}_{k,m}(t), \overline{I}_{k+K,m}(t) = \{0,1\}, \tilde{I}_{m,j}(t) = \{0,1\}, \forall k \in \left\{ \overline{\mathcal{K}}_S \cup \overline{\mathcal{K}}_D \cup \tilde{\mathcal{K}}_S \cup \tilde{\mathcal{K}}_D \right\}, m \in \mathcal{M}, j \in \mathcal{J} \tag{37}$$

$$C3 : \sum_{m \in \mathcal{M}} \overline{I}_{k,m}(t) \leq 1, \sum_{m \in \mathcal{M}} \overline{I}_{k+K,m}(t) \leq 1, \forall k \in \left\{ \overline{\mathcal{K}}_S \cup \overline{\mathcal{K}}_D \cup \tilde{\mathcal{K}}_S \cup \tilde{\mathcal{K}}_D \right\}, \tag{38}$$

$$C4 : \sum_{j \in \mathcal{J}} \tilde{I}_{m,j}(t) \leq 1, \forall m \in \mathcal{M} \tag{39}$$

C1 indicates that a device should achieve a minimum SNR threshold to maintain the required QoS. C2 defines the instantaneous device association indicator and UAVs' channel selection indicator. C3 assures that each device can be associated with a single UAV at a time slot, and C4 implies UAVs' channel selection conditions at each time slot. The optimization variables $(x_m(t), y_m(t))$, $\overline{I}_{k,m}(t)$ and $\tilde{I}_{m,j}(t)$ are coupled and interactable, where the deflection of one variable impacts the optimization of other variables and the objective value. Hence, this optimization problem becomes complicated using standard optimization tools. In order to tackle this situation, we adopt an RL-based UAV deployment strategy to find their optimal position by estimating the required system parameters using real-time measurements and statistics of collected information.

4.2 RL-based solution methodology

UAVs acting as RL agents select the action depending on their current positions, which are only related to their previous states. Hence, the proposed framework follows Markovian properties composed of state, action, reward, state transition probability, and the flying time periods. In the next sub-section, we explain each of those elements elaborately.

4.2.1 State space

The state of the mth UAV at t-th time slot is the vector of two elements which represent its current position as $\delta_m(t) = (x_m(t), y_m(t)), \forall \delta_m(t) \in \mathcal{S}$. Here, $\mathcal{S}$ is the state space, whose elements are independent and identically distributed random variables arranged by combining all possible values across the time horizon.

4.2.2 Action space

UAV's action $a_m(t) \in \mathcal{A}$ in the current state is the change of its position, which is measured with respect to its immediate X and Y coordinates. Here, we consider a benchmark RL gridworld environment where UAVs have maximum of eight possible moving directions at each state, i.e., NORTH, NORTH-WEST, WEST, SOUTH-WEST, SOUTH, SOUTH-EAST, EAST, and NORTH-EAST. After selecting an action, the X and Y coordinate changes of UAV m at t-th time slot are represented as $\delta^m_x(t) \in \{-\vartheta(t)\delta, 0, \vartheta(t)\delta\}$ and $\delta^m_y(t) \in \{-\vartheta(t)\delta, 0, \vartheta(t)\delta\}$ respectively, $\forall a_m(t) = \left\{ \delta^m_x(t), \delta^m_y(t) \right\} \in \mathcal{A}, t \in \mathcal{T}$, where $\vartheta(t)$ is the velocity of UAVs at time slot t and

$\mathcal{A}$ is the action set containing all possible actions. The obtained X and Y coordinate of UAV m for next time slot is measured as

$$x_m(t+1) = x_m(t) + \delta_x^m(t) \tag{40}$$

$$y_m(t+1) = y_m(t) + \delta_y^m(t) \tag{41}$$

4.2.3 Reward formulation

RL agents choose their actions in such a manner that maximizes long-term cumulative reward. Since our objective is to maximize the total instantaneous transmission rate of the USSD2D network, we need to find such locations of UAVs that impacts immediate objective value. Hence, we model the instantaneous reward function contributed by UAV m as

$$\mathcal{R}(\mathbf{s}_m(t), \mathbf{a}_m(t)) = \sum_{j=1}^{J} \sum_{\tilde{k}=1}^{\tilde{K}} \tilde{R}_{\tilde{k},m,\tilde{k}+\tilde{K}}^{j}(t) + \sum_{j=1}^{J} \sum_{\overline{k}=1}^{\overline{K}} \overline{R}_{\overline{k},\overline{k}+\overline{K}}^{j}(t), \forall m \in \mathcal{M} \tag{42}$$

4.2.4 State transition probability

It is the probability that UAV m changes its state from $\mathbf{s}_m(t)$ to $\mathbf{s}_m(t+1)$ after selecting an action $\mathbf{a}_m(t)$, denoted as $P_{tr}\{\mathbf{s}_m(t+1) \in \mathcal{S} | \mathbf{s}_m(t), \mathbf{a}_m(t)\}$. Let us consider the probability vectors of device association and UAVs' channel selection at time slot t as $P_{\tilde{k}}^{DA}(t) = \left[\tilde{P}_{\tilde{k},1}(t), \tilde{P}_{\tilde{k},2}(t), \ldots, \tilde{P}_{\tilde{k},M}(t)\right], \forall \tilde{k} \in \{\tilde{\mathcal{K}}_S \cup \tilde{\mathcal{K}}_D\}$ and $P_m^{CS}(t) = \left[\overline{P}_{m,1}(t), \overline{P}_{m,2}(t), \ldots, \overline{P}_{m,J}(t)\right], \forall m \in \mathcal{M}$ respectively where $\tilde{P}_{\tilde{k},m}(t)$ indicates the association probability of device $\tilde{k}$ with UAV m at time slot t and $\overline{P}_{m,j}(t)$ is the probability that UAV m selects channel j at time slot t. In each time slot, source and destination devices associated with a single UAV according to probability vectors $P_{\tilde{k}}^{UA}(t)$ and UAV selects a single orthogonal channel with a probability vector of $P_m^{CS}(t)$. The probabilities of device association and UAV's channel selections are updated for the next time slot as follows:

$$\tilde{P}_{\tilde{k},m}(t+1) = \begin{cases} \tilde{P}_{\tilde{k},m}(t) + w_1 \tilde{r}_{\tilde{k},m}(t)\left(1 - \tilde{P}_{\tilde{k},m}(t)\right), m = U_{\tilde{k}}^{Max}(t) \\ \tilde{P}_{\tilde{k},m}(t) - w_1 \tilde{r}_{\tilde{k},m}(t)\tilde{P}_{\tilde{k},m}(t), m \neq U_{\tilde{k}}^{Max}(t) \end{cases} \tag{43}$$

$$\overline{P}_{m,j}(t+1) = \begin{cases} \overline{P}_{m,j}(t) + w_2 \overline{r}_{m,j}(t)\left(1 - \overline{P}_{m,j}(t)\right), j = C_m^{Max}(t) \\ \overline{P}_{m,j}(t) - w_2 \overline{r}_{m,j}(t)\overline{P}_{m,j}(t), j \neq C_m^{Max}(t) \end{cases} \tag{44}$$

where w_1 and w_2 are the learning step sizes. $U_{\tilde{k}}^{Max}(t)$ is the current best UAV for device $\tilde{k}$ for a fixed selected channel and $C_m^{Max}(t)$ is the current best channel of UAV m for associated devices at that time slot respectively, which can be expressed as

$$U_{\tilde{k}}^{Max}(t) = \arg\max_{m \in \mathcal{M}} \tilde{R}_{\tilde{k},m,\tilde{k}+\tilde{K}}(t), \forall \tilde{k} \in \{\tilde{\mathcal{K}}_S \cup \tilde{\mathcal{K}}_D\} \tag{45}$$

$$C_m^{Max}(t) = \arg\max_{j \in \mathcal{J}} \tilde{R}_{m,j}(t), \forall m \in \mathcal{M} \tag{46}$$

where $\tilde{r}_{\tilde{k},m}(t)$ and $\bar{r}_{m,j}(t)$ are the normalized reward achieved by the source device $\tilde{k}$ and UAV m at time slot t respectively, which are defined as

$$\tilde{r}_{\tilde{k},m}(t) = \frac{\tilde{R}_{\tilde{k},m,\tilde{k}+\tilde{K}}(t)}{\max_{m \in \mathcal{M}} \tilde{R}_{\tilde{k},m,\tilde{k}+\tilde{K}}(t)} \tag{47}$$

$$\bar{r}_{m,j}(t) = \frac{\tilde{R}_{m,j}(t)}{\max_{j \in \mathcal{J}} \tilde{R}_{m,j}(t)} \tag{48}$$

From (43) and (44), it is observed that the update of selection probability vectors depends on the instantaneous transmission rate, which does not need any prior information. Thus, device association and UAVs' channel selection at each time slot is entirely model-free.

4.2.5 Updating the action value function

During the operation period, each UAV acts as an RL agent where UAV m takes an action $a_m(t)$ at current state $\imath_m(t)$. Then it generates an immediate reward $\mathcal{R}(\imath_m(t), a_m(t))$, and computes corresponding $Q(\imath_m(t), a_m(t))$ value. Finally, the current state $\imath_m(t)$ is updated to the next state $\imath_m(t+1)$ and UAV m selects the next action $a_m(t+1)$ using the same policy where the action-value function is updated as [37]

$$Q(\imath_m(t), a_m(t)) \leftarrow (1-\alpha)Q(\imath_m(t), a_m(t)) + \alpha[\mathcal{R}(\imath_m(t), a_m(t)) + \gamma\, Q(\imath_m(t+1), a_m(t+1))] \tag{49}$$

UAVs consider all the possible actions from the action space and select an action with a certain probability that provides maximum long-term reward. ϵ-greedy action selection policy is adopted under which the probability that UAV m takes action $a_m(t) \in \mathcal{A}$ corresponding to a state $\imath_m(t) \in \mathcal{S}$ at time slot t can be expressed as [37]

$$\pi_m^\epsilon = \begin{cases} \arg\max_{a_m(t)=\{\delta_x^m(t), \delta_y^m(t)\}} Q(\imath_m(t), a_m(t)), \text{with probability } 1-\epsilon \\ \text{Random Selection, with probability } \epsilon \end{cases} \tag{50}$$

UAVs execute state-action pairs repeatedly to gain experience of interacting with the environment. These interaction results are recorded in Q-table and updated the learning policy in each episode until convergence. Algorithm 1 summarizes the optimal deployment strategy using the adaptive State-Action-Reward-State-Action (SARSA) technique.

4.3 Simulation results

In this sub-section, we validate the proposed analysis and provide various numerical insights on key system parameters to improve the system's performance. Later, we compare the obtained results corresponding to the proposed SARSA algorithm with the existing works [34], such as random selection with fixed optimal relay deployment (RS-FORD), an exhaustive search for relay assignment and channel allocation with fixed initial relay deployment (ES-FIRD), and alternative optimization for the
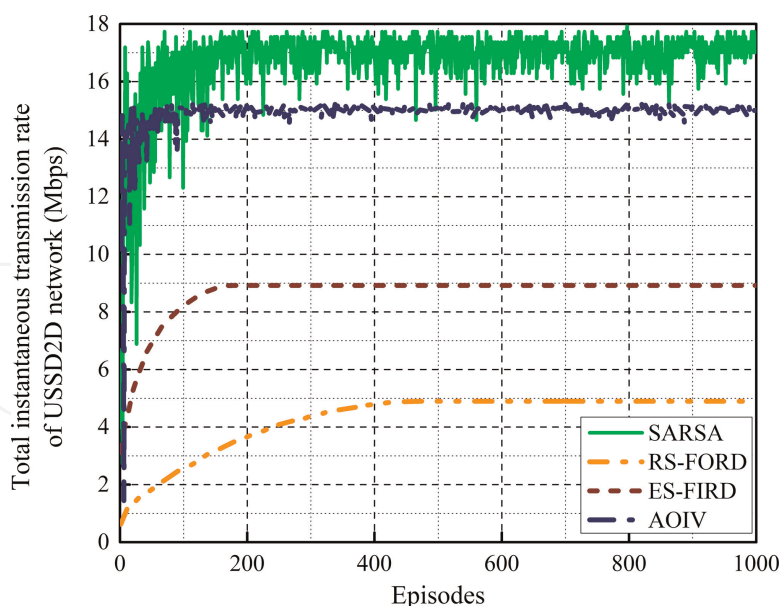
Figure 6.
The variation of the total transmission rate of the USSD2D network corresponding to each episode.

individual variable (AOIV). Here, we consider that direct D2D pair and UAV-assisted D2D pair devices are uniformly distributed in a 4 km×4 km square area where the primary simulation parameters are adopted from [38].

The iterative evolutions of the proposed and benchmark schemes are depicted in **Figure 6**, where the number of UAVs, UAV-assisted D2D pairs, direct D2D pairs, orthogonal channels, and transmit power are set as 5, 10, 2, 7, and 10 mW respectively. From this figure, it is clear that the proposed algorithm outperforms the benchmark scheme with respect to the converged value because it utilizes ϵ-greedy action policy to obtain the large search space by exploring the target region more efficiently. Furthermore, UAV acting as an RL agent learns to improve the cumulative reward, i.e., the total instantaneous transmission rate, from its past learning experiences. Hence, according to this figure, the SARSA algorithm enhances the overall transmission rate by 75.37%, 49.74%, and 11.01%, compared with RS-FORD, ES-FIRD, and AOIV schemes, respectively.

Algorithm 1: Optimal UAV deployment strategy using adaptive SARSA technique

---

Input: $N_0, B, \mu_{LoS}, \mu_{NLoS}, f_c, b_1, b_2, H_u, \vartheta_0, \overline{K}, \overline{\mathcal{K}}_S, \overline{\mathcal{K}}_D, \tilde{K}, \tilde{\mathcal{K}}_S, \tilde{\mathcal{K}}_D, P_k^{Tx}, M, \mathcal{M}, P_m^{Tx}, J, \mathcal{J},$
$w_1, w_2, \gamma, \alpha, \epsilon, \varsigma, \forall s_m(t) \in \mathcal{S}, a_m(t) \in \mathcal{A}, k \in \mathcal{K} = \{\overline{\mathcal{K}}_S \cup \overline{\mathcal{K}}_D \cup \tilde{\mathcal{K}}_S \cup \tilde{\mathcal{K}}_D\}, m \in \mathcal{M}$

Output: Instantaneous reward generated by all UAVs as $\mathcal{R}(t)$

1: Initialize $Q(s_m(t), a_m(t)) = 0, \forall s_m(t) \in \mathcal{S}, a_m(t) \in \mathcal{A}, m \in \mathcal{M}$

2: Set initial device association probability as $\tilde{P}_{\tilde{k},m}(1) = \frac{1}{M}, \forall \tilde{k} \in \{\tilde{\mathcal{K}}_S \cup \tilde{\mathcal{K}}_D\}, m \in \mathcal{M}$

3: Set initial channel selection probability of UAVs as $\overline{P}_{m,j}(1) = \frac{1}{J}, \forall m \in \mathcal{M}, j \in \mathcal{J}$

4: Initially deploy UAV m at the random position as $s_m(1) = (x_m(1), y_m(1), H_u), \forall m \in \mathcal{M}$

5: **for** $t = 1, 2, \ldots, T$ **do**

6: **for** $\tilde{k} = 1, 2, \ldots, \tilde{K}$ **do**

7: **for** $m = 1, 2, \ldots, M$ **do**

8: Obtain the association probability of device $\tilde{k}$ with UAV m as $\tilde{P}_{\tilde{k},m}(t)$

9: Calculate $\hat{\Gamma}_{\tilde{k},m}^j(t)$ and $\hat{\Gamma}_{m,\tilde{k}+\tilde{K}}^j(t)$ by (25) and (26), respectively, for a fixed assigned channel

10: **if** $\hat{\Gamma}_{\tilde{k},m}^j(t), \hat{\Gamma}_{m,\tilde{k}+\tilde{K}}^j(t) \geq \varsigma$ **then**


11: Calculate $\tilde{R}^j_{\tilde{k},\tilde{k}+\tilde{K}}(t)$ using (32) for a fixed assigned channel

12: **else**

13: $\tilde{R}^j_{\tilde{k},\tilde{k}+\tilde{K}}(t) = 0$

14: **for** $\tilde{k} = 1, 2, \ldots, \tilde{K}$ **do**

15: **for** $m = 1, 2, \ldots, M$ **do**

16: Set $\bar{I}_{\tilde{k},m}(t) = 1$ when $m = \arg \max_{m \in \mathcal{M}} \tilde{P}_{\tilde{k},m}(t)$, otherwise $\bar{I}_{\tilde{k},m}(t) = 0$

17: According to (43), update the association probability as $\tilde{P}_{\tilde{k},m}(t) \leftarrow \tilde{P}_{\tilde{k},m}(t+1)$

18: **for** $m = 1, 2, \ldots, M$ **do**

19: **for** $j = 1, 2, \ldots, J$ **do**

20: UAV m obtains the jth channel selection probability as $\overline{P}_{m,j}(t)$

21: Calculate $\hat{\Gamma}^j_m(t)$ according to (25) for the fixed associated devices

22: **if** $\hat{\Gamma}^j_m(t) \geq \varsigma$ **then**

23: $\tilde{R}^j_m(t) = \sum_{\tilde{k}=1}^{\tilde{K}} B \log_2 \left[1 + \hat{\Gamma}^j_{\tilde{k},m}(t) \right]$

24: **else**

25: $\tilde{R}^j_m(t) = 0$

26: **for** $m = 1, 2, \ldots, M$ **do**

27: **for** $j = 1, 2, \ldots, J$ **do**

28: Set $\tilde{I}_{m,j}(t) = 1$ when $j = \arg \max_{j \in \mathcal{J}} \overline{P}_{m,j}(t)$, otherwise $\tilde{I}_{m,j}(t) = 0$

29: According to (44), update channel selection probability as $\overline{P}_{m,j}(t) \leftarrow \overline{P}_{m,j}(t+1)$

30: **for** $m = 1, 2, \ldots, M$ **do**

31: Choose the action values $a_m(t) = \left\{ \delta^m_x(t), \delta^m_y(t) \right\}$ by (50)

32: Find next state as $s_m(t+1) = \left(x_m(t+1), y_m(t+1), H_u \right)$ by (40) and (41)

33: Calculate the immediate reward $\mathcal{R}(s_m(t), a_m(t))$ of UAV m by (42)

34: Choose the action $a_m(t+1) = \left\{ \delta^m_x(t+1), \delta^m_y(t+1) \right\}$ by (50) and obtain $Q(s_m(t+1), a_m(t+1))$ value

35: Update $Q(s_m(t), a_m(t))$ value according to (49) and store it in Q-table

36: Update the state and action for the next time slot as $s_m(t) \leftarrow s_m(t+1)$ and $a_m(t) \leftarrow a_m(t+1)$ respectively

37: Calculate the instantaneous reward generated by all UAVs as $\mathcal{R}(t) = \sum_{m=1}^{M} \mathcal{R}(s_m(t), a_m(t))$

Figure 7a shows the variation of instantaneous transmission rate for different number of UAVs while the other3 network parameters are the same, as mentioned in **Figure 6**. It can be observed in this figure that the performance metric value increases with the number of UAVs because all UAVs utilize the available channels efficiently at their deployed location. But when the number of UAVs exceeds 7, the total instantaneous transmission rate does not increase significantly because all UAVs reuse the limited spectrum, which increases mutual interferences among UAVs and source-destination device pairs.

Figure 7b plots the objective value corresponding to the different number of available orthogonal channels. From this figure, we can say that the instantaneous transmission rate increases with the number of channels because all the communication nodes select individual channels according to the channel selection probability vectors. But when the number of channels exceeds 7, no such variation in objective value is found because this is a sufficient resource to avoid mutual interferences completely.

Figure 7c represents the network throughput variation for different UAV-assisted D2D pairs when their transmitting power is 10 mW. Since all the devices and UAVs share the fixed amount of orthogonal channels, the network's performance is
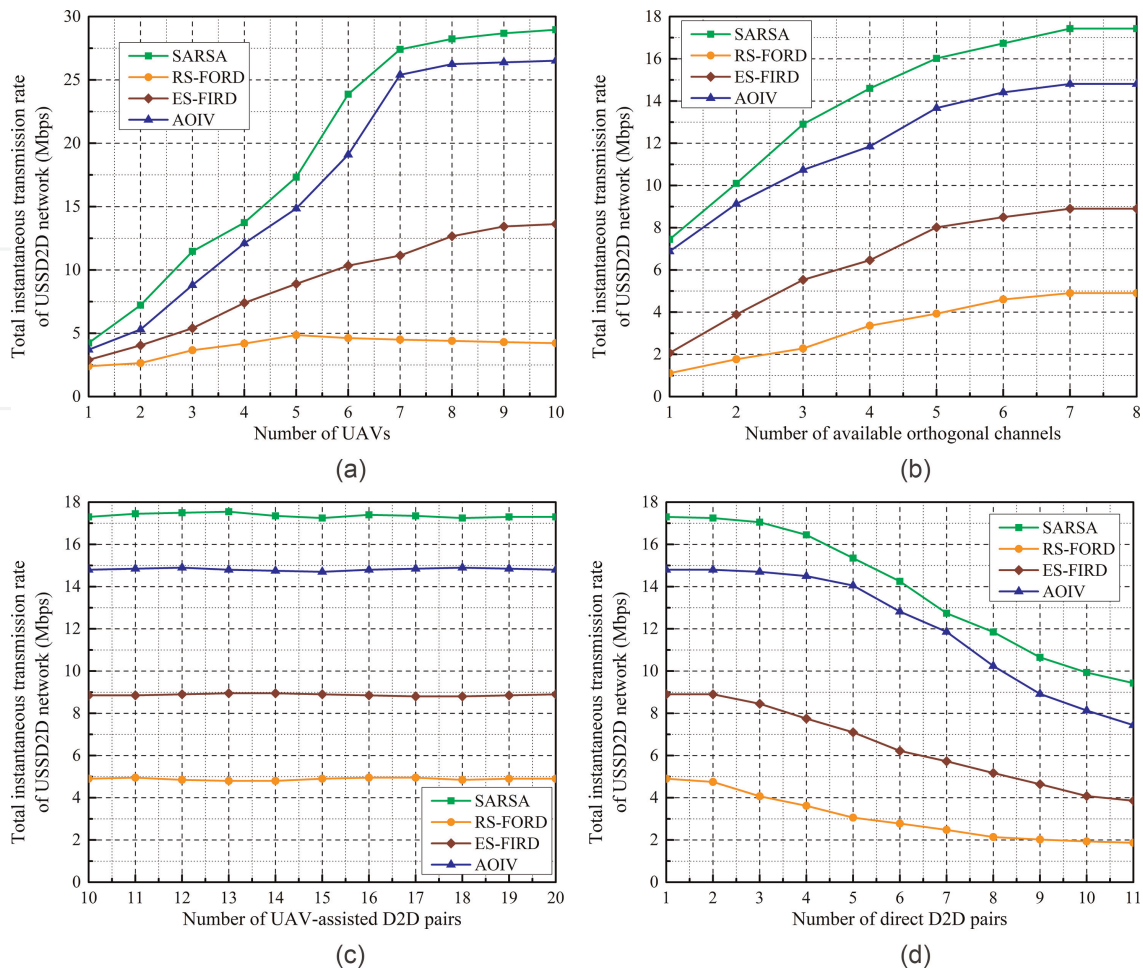
Figure 7.
Total overall network performance of the USSD2D network for different network parameters value. (a) Network throughput for different number of UAVs. (b) Network throughput for different number of channels. (c) Network throughput corresponding to the different number of UAV-assisted D2D pairs. (d) Network throughput corresponding to the different number of direct D2D pairs.

independent with respect to the number of UAV-assisted D2D pairs, and the performance metric value is almost constant for variation of the key system parameters.

The performance metric variations for different number of direct D2D pairs are illustrated in **Figure 7d** when their transmitting power is set as 10 mW. It is observed that the instantaneous transmission rate decreases with the number of direct D2D pairs because they utilize more orthogonal channels. As a result, mutual interference among UAV-assisted D2D pairs increases since they share limited network resources. Furthermore, our proposed scheme has the capabilities for adaptive action selection, which significantly outperforms the benchmark techniques. From **Figure 7**, we can say that the overall network throughput can be improved by 77.58%, 52.51%, and 12.14% compared to the RS-FORD, ES-FIRD, and AOIV schemes, respectively.

5. Minimization of devices' energy consumption in UAV-assisted IoT network

The devices at the cell edge consume high energy to achieve the required data rate when transmitting data to the nearest BS because of the large LoS distance between
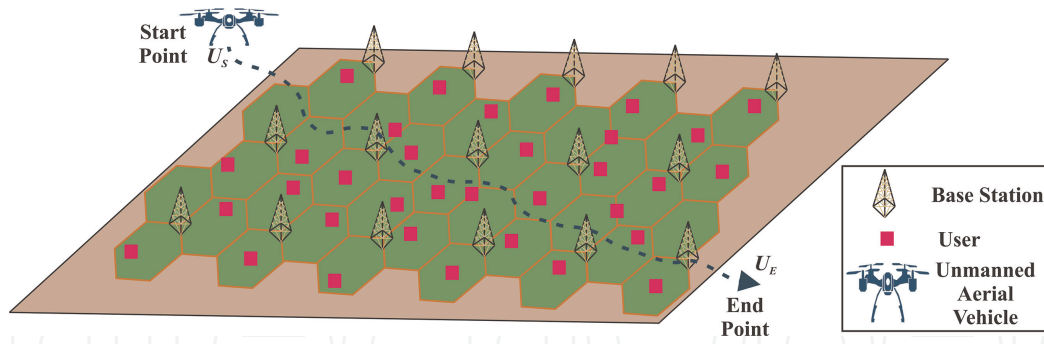
Figure 8.
Illustration of UAV-assisted IoT network.

BSs and those devices. Alternatively, a quad-rotor UAV-assisted IoT network could provide reliable communication compared to fixed terrestrial BSs. Therefore, in this section, we aim to find the optimal trajectory of UAV and the association of IoT devices that simultaneously support energy-efficient data collection.

5.1 System model

Figure 8 illustrates the UAV-assisted IoT network, in which M terrestrial BSs with fixed height of H_B and a single UAV collect data from K stationary uniformly distributed IoT devices. The UAV flies at a fixed altitude H_u with the constant speed of ϑ_u where its start and end locations are represented by $U_S = (x_s, y_s, H_u)$ and $U_E = (x_e, y_e, H_u)$ respectively. To track the UAV's location at each instance, we discretize its flight period into N equally spaced time slots, each of duration T_s, and assume that UAV's location at nth time slot $U[n] = (x[n], y[n], H_u), \forall n \in \mathcal{N} = \{1, 2, ..., N\}$ is constant. All devices transmit atleast $\mathcal{D}_{Min}$ bits data to the core network to maintain reliable QoS.

5.1.1 Data collection of core network

The transmission environment is categorized into two scenarios, i.e., ground to ground (G2G) and ground to air (G2A) channels. G2G channel establishes the links between BS and IoT devices, whereas G2A channel connects the IoT devices with the UAV platform. We generalize the wireless channel gain between each device and its destination (either UAV or BS) at each time slot as the combination of large-scale path loss and small-scale fading. The channel gain between each device and its destination can be modeled as [39]

$$h_k^i[n] = g_k^i[n] \sqrt{L_k^i[n]}, \forall k \in \mathcal{K} = \{1, 2, ..., K\} \tag{51}$$

where $i \in \{B \text{ or } U\}$ is the destination indicator in which B and U represent nearest BS and UAV, respectively, $L_k^i[n]$ is the large scale path loss, $g_k^i[n]$ is the small scale fading coefficient. The achievable instantaneous transmission rate of the kth IoT device can be formulated as [40]

$$R_k^i[n] = C_B \log_2 \left[1 + \frac{\left| h_k^i[n] \right|^2 P_k^i[n]}{\eta_0} \right] \tag{52}$$

where C_B is channel bandwidth, $P_k^i[n]$ is transmitting power of the kth device, and η_0 is noise power. Instantaneous data transmitted by the kth device over G2G and G2A channel is measured as $\mathcal{D}_k^B[n] = R_k^B[n]T_s$ and $\mathcal{D}_k^U[n] = R_k^U[n]T_s$ respectively. The energy consumption of device k at nth time slot can be calculated as

$$E_k[n] = \left(I_k^U[n]P_k^U[n] + I_k^B[n]P_k^B[n]\right)T_s, \forall k \in \mathcal{K} \tag{53}$$

where $P_k^U[n]$ and $P_k^B[n]$ are the instantaneous transmit powers of kth device when connecting with UAV and BS, respectively and $I_k^U[n], I_k^B[n] \in \{0,1\}$ are the binary device association indicators with UAV and BS respectively. The k th device transmits data to the core network during each time slot is measured as

$$\mathcal{D}_k[n] = I_k^U[n]\mathcal{D}_k^U[n] + I_k^B[n]\mathcal{D}_k^B[n], \forall k \in \mathcal{K}, n \in \mathcal{N} \tag{54}$$

5.1.2 Problem formulation

We aim for energy-efficient data collection that jointly exploit reliable data transmission, optimal instantaneous position of UAV and transmit power control. The fluctuation of channel gain causes unstable network performance, leading to quickly drain out devices' on-board battery energy. Thus, to minimize total energy consumption of all devices we jointly optimize UAV' trajectory, device association indicators and their transmit power allocation, while ensuring that each device should transmit a minimum data to the destination and UAV chooses a constant speed during its trajectory between the initial and final locations. Therefore the optimization problem is formulated as

$$P1 : \underset{\left\{\begin{array}{c}\{(x[n],y[n])\}, I_k^U[n], I_k^B[n], P_k^U[n], \text{ and } P_k^B[n]\\ \forall k \in \mathcal{K}, n \in \mathcal{N}\end{array}\right\}}{\text{Minimize}} \sum_{n=1}^{N}\sum_{k=1}^{K}\left[\left(I_k^u[n]P_k^u[n] + I_k^B[n]P_k^B[n]\right)T_s\right] \tag{55}$$

Subject to the constraints

$$\text{C1} : I_k^U[n]\mathcal{D}_k^U[n] + I_k^B[n]\mathcal{D}_k^B[n] \geq \mathcal{D}_{Min}, \forall k \in \mathcal{K}, n \in \mathcal{N} \tag{56}$$

$$\text{C2} : I_k^U[n] \in \{0,1\}, I_k^B[n] \in \{0,1\}, \forall k \in \mathcal{K}, n \in \mathcal{N} \tag{57}$$

$$\text{C3} : I_k^U[n] + I_k^B[n] \leq 1, \forall k \in \mathcal{K}, n \in \mathcal{N} \tag{58}$$

$$\text{C4} : \sum_{k=1}^{K} I_k^U[n] \leq K, \forall n \in \mathcal{N} \tag{59}$$

$$\text{C5} : U[1] = U_S, U[N] = U_E \tag{60}$$

Here, C1 ensures that each device transmits atleast $\mathcal{D}_{Min}$ bits data to either UAV or nearest BS at a time slot. C2 defines the device association indicators. C3 verifies that each device associates with either UAV or the nearest BS at each time slot. C4 implies that UAV can associate with maximum K number of devices instantaneously; and C5 guarantees that UAV starts its trajectory from an initial given position and ends to the final predefined location. The optimization problem contains multiple interactive and coupled variables, and they have a complex relationship by which changing one's value may impact to others. Furthermore, these discrete optimizing variables make the problem highly non-convex to find a limited time trajectory between the start and end points.

Hence, standard optimization methods face difficulties in obtaining exact solutions. In order to tackle this situation, we propose RL framework and adaptive decision-making policy to find UAV's successive locations, and device association along with their transmit power allocation. We adopt the SARSA algorithm to control the UAV, which acts as an RL-agent for taking the optimal action at each step to maximize its reward.

5.2 Reinforcement learning based on SARSA algorithm

As discussed earlier in Section 4.3, the RL framework follows MDP, where the current state only depends on the immediate past state, and the UAV acting as RL agent chooses an action according to the ϵ-greedy policy. Here, the generated reward depends on UAV's current state and taken action at each time slot. The expected trajectory is obtained more precisely when the reward generated by the UAV at the current time slot is beneficial for the long term. To reflect this property, we model the instantaneous reward for every time slot as UAV's instantaneous objective value, which is expressed as

$$\mathcal{R}(\mathit{s}[n], \mathit{a}[n]) = \left[\sum_{k=1}^{K} \left(I_k^U[n] P_k^U[n] + I_k^B[n] P_k^B[n] \right) T_s \right]^{-1} \tag{61}$$

Algorithm 2 summarizes the optimal trajectory learning procedure using the improved SARSA technique. In this framework, we first calculate UAV's current state, channel gain, and distances from all devices to UAV and the nearest BS at every time slot. Then, all devices select the destination (either UAV or nearest BS) by estimating the instantaneous device association indicator and the required transmit power while satisfying the data rate constraint value. This process is repeated at each step, and UAV obtains optimal policy at the final episode. Since the number of episodes is T and each episode goes through N time slots, the computation complexity depends on total steps TN, including state space and action space in RL. In our scenario, there are $L_1 L_2$ possible state locations and eight possible actions for each time slot. Therefore, the computational complexity of algorithm 1 is $\mathcal{O}(8TNL_1L_2)$, including the complexity of the action selection scheme in each step.

Algorithm 2: UAV trajectory learning process using SARSA

Input: $\gamma, \alpha, \hat{\epsilon}, \zeta, T, (x_s, y_s, H_u), (x_e, y_e, H_u), T_s, \beta_0, \vartheta_u, H_u, \mathcal{D}_{Min}, IoT_k, K, BS_m, M, N, e_{Max},$

$h_k^i[n], \mathit{s}[n], \mathit{a}[n], \forall \mathit{s}[n] \in \mathcal{S}, \mathit{a}[n] \in \mathcal{A}, k \in \mathcal{K}, m \in \mathcal{M}, n \in \mathcal{N}, i \in \{U \text{ or } B\}$

Output: Optimal policy π_h^*

1: Initialize $Q(\mathit{s}[n], \mathit{a}[n]) = 0, \forall s \in \mathcal{S}, a \in \mathcal{A},$ and $e[1] = e_{Max}$

2: **for** $t = 1, 2, \dots, T$ **do**

3: Set the starting point as $\mathit{s}[1] = (x[1], y[1]) = (x_s, y_s)$

4: **for** $n = 1, 2, \dots, N$ **do**

5: **if** $n \leq N - 2$ and $\sqrt{(x_e - x_u[n])^2 + (y_e - y_u[n])^2} \leq \vartheta_u(N-n)T_s$ **then**

6: Choose the action values $\mathit{a}[n] = \left\{ \delta_x^u[n], \delta_y^u[n] \right\}$ by (50)

7: Find next state by (40) and (41) as $\mathit{s}[n+1] = (x[n+1], y[n+1])$

8: Calculate reward $\mathcal{R}(\mathit{s}[n], \mathit{a}[n])$ by (61)

9: Choose the next action $\mathit{a}[n+1] = \left\{ \delta_x^u[n+1], \delta_y^u[n+1] \right\}$ by (50) and obtain $Q(\mathit{s}[n+1], \mathit{a}[n+1])$ value

10: Update $Q(\mathit{s}[n], \mathit{a}[n])$ value according to (49)

11: Update the respective state and action as $\mathit{s}[n] \leftarrow \mathit{s}[n+1]$ and $\mathit{a}[n] \leftarrow \mathit{a}[n+1]$

12: **else if** $n = N - 1$ and $\sqrt{(x_e - x_u[n])^2 + (y_e - y_u[n])^2} \leq \vartheta_u T_s$ **then**

13: Obtain the next state as $\mathfrak{s}[n + 1] = [x_e, y_e]$

14: Calculate reward $\mathcal{R}(\mathfrak{s}[n], \mathfrak{a}[n])$ by (61)

15: Choose the next action $\mathfrak{a}[n + 1] = \left\{ \delta_x^u[n + 1], \delta_y^u[n + 1] \right\}$ by (50) and obtain $Q(\mathfrak{s}[n + 1], \mathfrak{a}[n + 1])$ value

16: Update $Q(\mathfrak{s}[n], \mathfrak{a}[n])$ value according to (49)

17: Update the respective state and action as $\mathfrak{s}[n] \leftarrow \mathfrak{s}[n + 1]$ and $\mathfrak{a}[n] \leftarrow \mathfrak{a}[n + 1]$

18: **else**

19: **Break**

20: Find an optimal policy as $\pi_h^* = arg \max\limits_{\mathfrak{a}[n] = \left\{ \delta_x^u[n], \delta_y^u[n] \right\}} Q(\mathfrak{s}[n], \mathfrak{a}[n]), \forall \mathfrak{s}[n] \in \mathcal{S}, \mathfrak{a}[n] \in \mathcal{A}, n \in \mathcal{N}$

5.3 Simulation results

This sub-section presents the training outcomes corresponding to the proposed SARSA algorithm for optimal trajectory and subsequently evaluates the energy-efficient data collection. Here, we compare the effectiveness and superiority of the proposed design with the benchmark PSO technique [41], where 100 IoT devices are uniformly distributed within a square field of size 2000 × 2000 m. Moreover, we adopt the required simulation parameters from [40] and [24] to implement the proposed algorithm.

### 5.3.1 Convergence analysis

The agents' training evaluations using RL-based SARSA algorithm are illustrated in **Figure 9a**, when all IoT devices maintain the data rate constraint of 10 Mbps. In this figure, we have found that the convergence rate varies for flying time because UAV explores the target area more efficiently with the available time slots. As a result more devices associate with UAV and the convergence occurs before 10,000 episodes.

Figure 9b shows the episode-wise objective value evaluation using PSO algorithm. From this figure, it is visible that PSO takes more time to converge, and its final convergence value is less than the SARSA algorithm. This is because PSO updates particles' position and velocity according to the random inertial weight which causes less exact regulation of particles' moving directions and speed. Hence, its
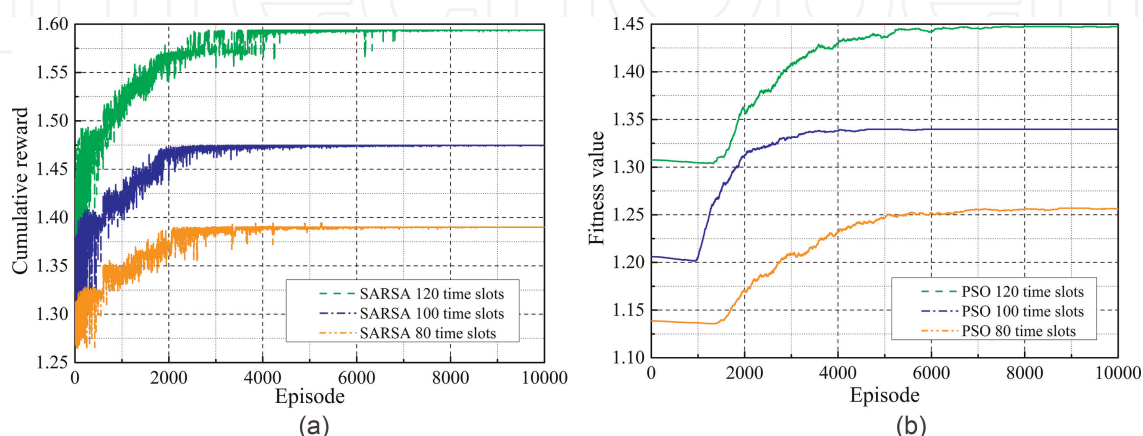
Figure 9.
Training results corresponding to the proposed and benchmark algorithms. (a) Cumulative reward generated by proposed SARSA. (b) Fitness value generated by benchmark PSO.
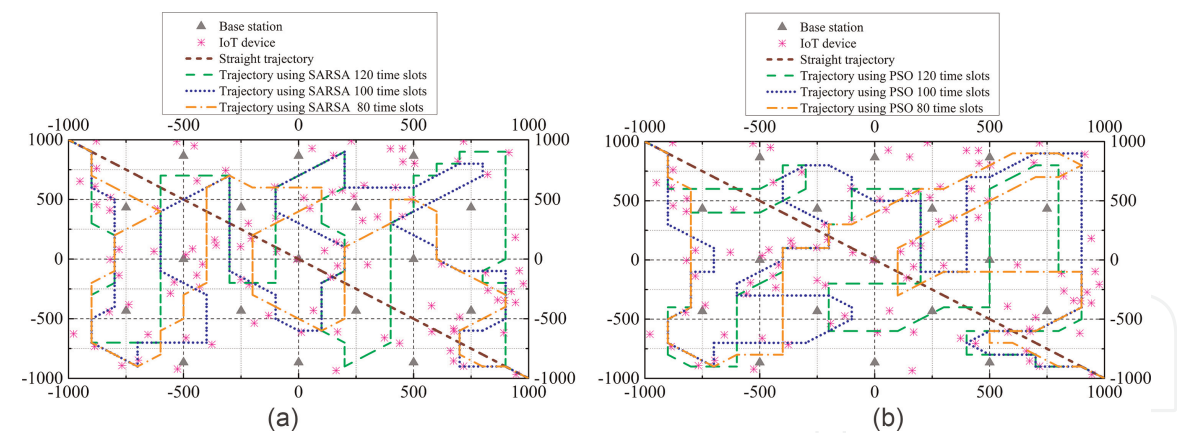
Figure 10.
Optimal trajectories corresponding to the proposed and benchmark algorithms. (a) Optimal UAV trajectory using SARSA. (b) Optimal UAV trajectory using PSO.

computational complexity increase due to the high dimensions of decision variables. Therefore, the proposed SARSA algorithm improves the cumulative reward by 10.26% with respect to the PSO.

5.3.2 Optimal trajectory

Using the same parameters mentioned in **Figure 9**, UAV finds its optimal trajectories with the help of SARSA and PSO algorithms, depicted in **Figure 10**. These figures indicate that UAV moves toward the devices, far away from the BS, and within the flight period, it reaches the final destination point. Since devices consume more energy while transmitting data to BS, UAV fly toward those devices to improve their channel conditions. as we mentioned earlier, device association with UAV increases with the flying time, more devices transmit their data to the UAV instead of BS, reducing their energy consumption.

5.3.3 Performance comparison of proposed SARSA with benchmark PSO

The variation of devices' average transmit power to achieve 10 Mbps data rate with the index value is demonstrated in **Figure 11a** where a device's index indicates its
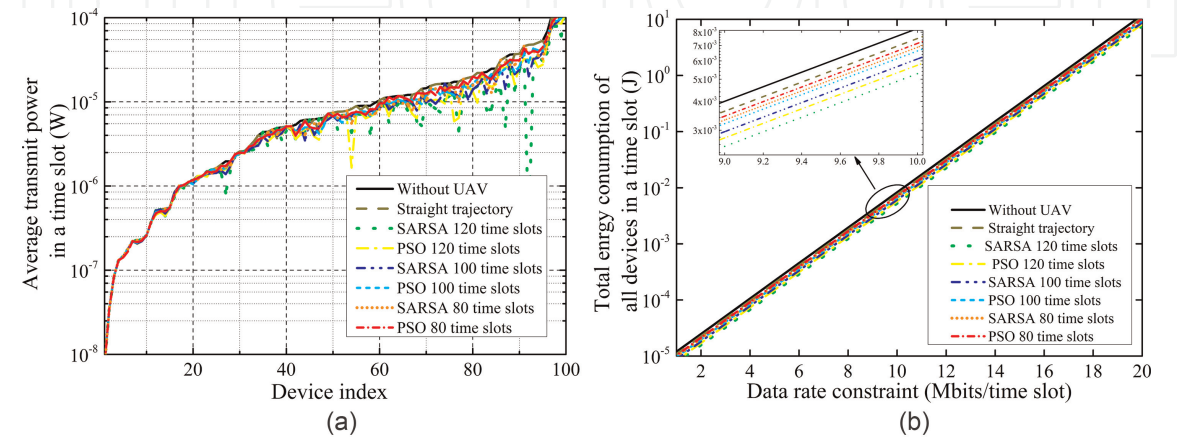
Figure 11.
Performance comparison of the proposed and benchmark algorithms. (a) Devices' transmit power corresponding to their index value. (b) Devices' energy consumption versus data rate constraint.

distance from the nearest BS. It is observed that, when there is no UAV support, average transmit power increases with the index value because, according to (52) devices far away from BS utilize more power to obtain the given data rate. But when UAV is employed, its optimal trajectory focuses the devices which are consuming more power and associates with them for data collection. Furthermore, since UAV's straight trajectory cannot improve all devices' channel conditions, the corresponding energy-efficient data collection would not be possible.

The total energy consumption of all devices for various data rate constraint values is illustrated in **Figure 11b**. It is clear that devices' energy consumption increases with data rate constraint because, according to (49), devices allocate more power to achieve the given rate constraint. Furthermore, from **Figure 11a**, UAV's optimal trajectory corresponding to the proposed SARSA algorithm reduces devices' transmit power with its available flying time as compared to PSO algorithm, because PSO achieves low convergence rate in an iterative process and could not identify the local optimal in high-dimension space. Hence, the proposed SARSA methodology significantly reduces the total energy consumption of all devices by 8.15%, 7.72%, and 5.67% for UAV's flying time of 80, 100, and 120 timeslots, respectively as compared to PSO.

6. Conclusion

This chapter proposes deployment and trajectory designs of UAVs for efficient resource allocation to achieve reliable wireless communication. The main features of this structure are three folded. In the first part, we optimize UAVs altitude to minimize outage probability and symbol error rate, considering pointing errors, atmospheric turbulence, and scintillation parameters where a hybrid RF-FSO channel governs the transmission environment. The second part finds the optimal deployed locations of UAVs to maximize the total instantaneous transmission rate of the devices in USSD2D network under SNR constraint. Finally, the last feature focuses on energy-efficient data collection where devices' total energy consumption is minimized by jointly optimizing their association with the nearest BS or UAV, their transmitting power, and UAV trajectory while satisfying a given data rate requirements. Numerical results validate the analysis and provide insights on the optimal UAV control design for various key system parameters. Our proposed methodology significantly improves system performance compared with the benchmark techniques. This work would be extended toward a multi UAVs-assisted energy-efficient data collection system considering the age of information aspect where the users follow a certain mobility model.

Author details

Abhishek Mondal[1*], Deepak Mishra[2], Ganesh Prasad[1] and Ashraf Hossain[1]

1 Department of Electronics and Communication Engineering, National Institute of Technology, Silchar, India

2 School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia

*Address all correspondence to: abhishekmondal532@gmail.com

IntechOpen

References

[1] Zanella A, Bui N, Castellani A, Vangelista L, Zorzi M. Internet of things for smart cities. IEEE Internet of Things Journal. 2014;**1**(1):22-32. DOI: 10.1109/JIOT.2014.2306328

[2] Samarakoon S, Bennis M, Saad W, Debbah M, Latva-Aho M. (2016, may). Ultra-dense small cell networks: Turning density into energy efficiency. IEEE Journal on Selected Areas in Communications. 2016;**34**(5):1267-1280. DOI: 10.1109/JSAC.2016.2545539

[3] Semiari O, Saad W, Bennis M, Dawy Z. Inter-operator resource management for millimeter wave multi-hop backhaul networks. IEEE Transactions on Wireless Communications. 2017;**16**(8):5258-5272. DOI: 10.1109/TWC.2017.2707410

[4] Mozaffari M, Saad W, Bennis M, Nam YH, Debbah M. A tutorial on UAVs for wireless networks: Applications challenges and open problems. IEEE Communications Surveys and Tutorials. 2019;**21**(3):2334-2360. DOI: 10.1109/COMST.2019.2902862

[5] Esrafilian O, Gangula R, Gesbert D. 3D map-based trajectory design in UAV-aided wireless localization systems. IEEE Internet of Things Journal. 2021;**8**(12):9894-9904. DOI: 10.1109/JIOT.2020.3021611

[6] Sawalmeh AH, Othman NS, Shakhatreh H, Khreishah A. Wireless coverage for mobile users in dynamic environments using UAV. IEEE Access. 2019;**7**:126376-126390. DOI: 10.1109/ACCESS.2019.2938272

[7] Lyu J, Zeng Y, Zhang R, Lim TJ. Placement optimization of UAV-mounted mobile base stations. IEEE Communications Letters. 2017;**21**(3):604-607. DOI: 10.1109/LCOMM.2016.2633248

[8] Wang Z, Duan L, Zhang R. Adaptive deployment for UAV-aided communication networks. IEEE Transactions on Wireless Communications. 2019;**18**(9):4531-4543. DOI: 10.1109/TWC.2019.2926279

[9] Alzenad M, El-Keyi A, Yanikomeroglu H. 3-D placement of an unmanned aerial vehicle base station for maximum coverage of users with different QoS requirements. IEEE Wireless Communications Letters. 2018;**7**(1):38-41. DOI: 10.1109/LWC.2017.2752161

[10] El-Hammouti H, Benjillali M, Shihada B, Alouini M. Learn-as-you-fly: A distributed algorithm for joint 3D placement and user association in multi-UAVs networks. IEEE Transactions on Wireless Communications. 2019;**18**(12):5831-5844. DOI: 10.1109/TWC.2019.2939315

[11] Zhang H, Hanzo L. Federated learning assisted multi-UAV networks. IEEE Transactions on Vehicular Technology. 2020;**69**(11):14104-14109. DOI: 10.1109/TVT.2020.3028011

[12] Liu X, Liu Y, Chen Y, Hanzo L. Trajectory design and power control for multi-UAV assisted wireless networks: A machine learning approach. IEEE Transactions on Vehicular Technology. 2019;**68**(8):7957-7969. DOI: 10.1109/TVT.2019.2920284

[13] Duong TQ, Nguyen LD, Tuan HD, Hanzo L. Learning-aided real-time performance optimization of cognitive UAV-assisted disaster communication. In: Proceeding IEEE Global

Communications Conference (GLOBECOM); 09–13 December 2019. Waikoloa, HI, USA: IEEE; 2020. pp. 1-6

[14] Liu X, Liu Y, Chen Y. Reinforcement learning in multiple UAV networks: Deployment and movement design. IEEE Transactions on Vehicular Technology. 2019;**68**(8):8036-8049. DOI: 10.1109/TVT.2019.2922849

[15] Larsen E, Landmark L, Kure O. Optimal UAV relay positions in multi-rate networks. In: Proceedings Wireless Days; 29–31 March 2017. Porto, Portugal: IEEE; 2017. pp. 8-14

[16] Han Z, Swindlehurst AL, Liu KJR. Optimization of MANET connectivity via smart deployment/movement of unmanned air vehicles. IEEE Transactions on Vehicular Technology. 2009;**58**(7):3533-3546. DOI: 10.1109/TVT.2009.2015953

[17] Jiang F, Swindlehurst AL. Dynamic UAV relay positioning for the ground-to-air uplink. In: Proceedings IEEE Globecom Workshop; 06–10 December 2010. Miami, FL, USA: IEEE; 2011. pp. 1766-1770

[18] Zhan P, Yu K, Swindlehurst AL. Wireless relay communications with unmanned aerial vehicles: Performance and optimization. IEEE Transactions on Aerospace and Electronic Systems. 2011; **47**(3):2068-2085. DOI: 10.1109/TAES.2011.5937283

[19] Zeng Y, Zhang R, Lim TJ. Throughput maximization for UAV-enabled mobile relaying systems. IEEE Transactions on Communications. 2016; **64**(12):4983-4996. DOI: 10.1109/TCOMM.2016.2611512

[20] Ono F, Ochiai H, Miura R. A wireless relay network based on unmanned aircraft system with rate optimization.

IEEE Transactions on Wireless Communications. 2016;**15**(11): 7699-7708. DOI: 10.1109/TWC.2016.2606388

[21] Fan R, Cui J, Jin S, Yang K. Optimal node placement and resource allocation for UAV relaying network. IEEE Communications Letters. 2018;**22**(4): 808-811. DOI: 10.1109/LCOMM.2018.2800737

[22] Indu SRP, Choudhary HR, Dubey AK. Trajectory design for UAV-to-ground communication with energy optimization using genetic algorithm for agriculture application. IEEE Sensors Journal. 2021;**21**(16):17548-17555. DOI: 10.1109/JSEN.2020.3046463

[23] Kawamoto Y, Takagi H, Nishiyama H, Kato N. Efficient resource allocation utilizing Q-learning in multiple UA communications. IEEE Transaction on Network Science and Engineering. 2019;**6**(3):293-302. DOI: 10.1109/TNSE.2018.2842246

[24] Mondal A, Mishra D, Prasad G, Hossain A. Joint optimization framework for minimization of device energy consumption in transmission rate constrained UAV-assisted IoT network. IEEE Internet of Things Journal. 2021; **9**(12):9591-9607. DOI: 10.1109/JIOT.2021.3128883

[25] Hu J, Zhang H, Song L. Reinforcement learning for decentralized trajectory design in cellular UAV networks with the sense-and-send protocol. IEEE Internet of Things Journal. 2019;**6**(4):6177-6189. DOI: 10.1109/JIOT.2018.2876513

[26] Cui J, Ding Z, Deng Y, Nallanathan A, Hanzo L. Adaptive UAV trajectory optimization under the quality of service constraints: A model-free solution. IEEE Access. 2020;**8**:

112253-112265. DOI: 10.1109/ACCESS.2020.3001752

[27] Yang L, Yuan J, Liu X, Hasna MO. On the performance of LAP-based multiple-hop RF/FSO systems. IEEE Transactions on Aerospace and Electronic Systems. 2019;**55**(1):499-505. DOI: 10.1109/TAES.2018.2852399

[28] Azari MM, Rosas F, Chen KC, Pollin S. Ultra reliable UAV communication using altitude and cooperation diversity. IEEE Transactions on Communications. 2018;**66**(1): 330-344. DOI: 10.1109/TCOMM.2017.2746105

[29] Puri P, Garg P, Aggarwal M. Outage and error rate analysis of network-coded coherent TWR-FSO systems. IEEE Photonics Technology Letters. 2014; **26**(18):1797-1800. DOI: 10.1109/LPT.2014.2333032

[30] Gappmair W. Further results on the capacity of free-space optical channels in turbulent atmosphere. IET Communications. 2011;**5**(9):1262-1267. DOI: 10.1049/iet-com.2010.0172

[31] Gil A, Segura J, Temme NM. Computation of the Marcum Q-function. ACM Transactions on Mathematical Software. 2013;**40**(3):280-295. DOI: 10.48550/arXiv.1311.0681

[32] Muller A, Speidel J. Exact symbol error probability of M-PSK for multihop transmission with regenerative relays. IEEE Communications Letters. 2007; **11**(12):952-954. DOI: 10.1109/LCOMM.2007.070820

[33] Mondal A, Hossain A. Channel characterization and performance analysis of UAV operated communication system with multihop RF–FSO link in dynamic environment. International Journal of Communication Systems.

2020;**33**(16):e4568. DOI: 10.1002/dac.4568

[34] Zhong X, Guo Y, Li N, Chen Y. Joint optimization of relay deployment, channel allocation, and relay assignment for UAVs-aided D2D networks. IEEE/ACM Transactions on Networking. 2020;**28**(2):804-817. DOI: 10.1109/TNET.2020.2970744

[35] Al-Hourani A, Kandeepan S, Lardner S. Optimal LAP altitude for maximum coverage. IEEE Wireless Communications Letters. 2014;**3**(6): 569-572. DOI: 10.1109/LWC.2014.2342736

[36] Hasna MO, Alouini MS. Outage probability of multihop transmission over Nakagami fading channels. IEEE Communications Letters. 2003;**7**(5): 216-218. DOI: 10.1109/LCOMM.2003.812178

[37] Mu X, Zhao X, Liang H. Power allocation based on reinforcement learning for MIMO system with energy harvesting. IEEE Transactions on Vehicular Technology. 2020;**69**(7): 7622-7633. DOI: 10.1109/TVT.2020.2993275

[38] Mondal A, Hossain A. Maximization of instantaneous transmission rate in UAVs-supported self-organized device-to-device network. International Journal of Communication Systems. 2022;**35**(6): e5064. DOI: 10.1002/dac.5064

[39] You C, Zhang R. 3D trajectory optimization in Rician fading for UAV-enabled data harvesting. IEEE Transactions on Wireless Communications. 2019;**18**(6):3192-3207. DOI: 10.1109/TWC.2019.2911939

[40] Ho TM, Nguyen KK, Cheriet M. UAV control for wireless service provisioning in critical demand areas:

A deep reinforcement learning approach.
IEEE Transactions on Vehicular
Technology. 2021;**70**(7):7138-7152.
DOI: 10.1109/TVT.2021.3088129

[41] Milner S, Davis C, Zhang H, Llorca J.
Nature-inspired self-organization,
control, and optimization in
heterogeneous wireless networks. IEEE
Transactions on Mobile Computing.
2012;**11**(7):1207-1222. DOI: 10.1109/
TMC.2011.141