sensors

MDPI

Article

UAV Trajectory Optimization in a Post-Disaster Area Using Dual Energy-Aware Bandits †

Amr Amrallah [1,2,*], Ehab Mahmoud Mohamed [3,4], Gia Khanh Tran [1,2] and Kei Sakaguchi [1,2]

[1] Department of Electrical and Electronic Engineering, School of Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan
[2] Academy for Super Smart Society, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan
[3] Department of Electrical Engineering, College of Engineering in Wadi Addawasir, Prince Sattam Bin Abdulaziz University, Wadi Addawasir 11991, Saudi Arabia
[4] Department of Electrical Engineering, Faculty of Engineering, Aswan University, Aswan 81542, Egypt
* Correspondence: amrallah@mobile.ee.titech.ac.jp
† This paper is an extended version of our paper published in Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN), "Dual Energy-Aware based Trajectory Optimization for UAV Emergency Wireless Communication Network: A Multi-armed Bandit Approach", Barcelona, Spain, 5–8 July 2022.

Abstract: Over the past few years, with the rapid increase in the number of natural disasters, the need to provide smart emergency wireless communication services has become crucial. Unmanned aerial Vehicles (UAVs) have gained much attention as promising candidates due to their unprecedented capabilities and broad flexibility. In this paper, we investigate a UAV-based emergency wireless communication network for a post-disaster area. Our optimization problem aims to optimize the UAV's flight trajectory to maximize the number of visited ground users during the flight period. Then, a dual cost-aware multi-armed bandit algorithm is adopted to tackle this problem under the limited available energy for both the UAV and ground users. Simulation results show that the proposed algorithm could solve the optimization problem and maximize the achievable throughput under these energy constraints.

Keywords: unmanned aerial vehicle; trajectory optimization; reinforcement learning; multi-armed bandit; cost subsidy; post-disaster

Citation: Amrallah, A.; Mohamed, E.M.; Tran, G.K.; Sakaguchi, K. UAV Trajectory Optimization in a Post-Disaster Area Using Dual Energy-Aware Bandits. *Sensors* **2023**, *23*, 1402. https://doi.org/10.3390/s23031402

Academic Editors: Andrzej Łukaszewicz, Carlos Tavares Calafate and Wojciech Giernacki

Received: 20 December 2022
Revised: 15 January 2023
Accepted: 20 January 2023
Published: 26 January 2023

1. Introduction

Across the globe, large-scale natural disasters are known for their severe casualties damage to property. Besides thousands of deaths and injuries resulting from various types of natural disasters around the world, there has been additional increase in material losses of about 100–150% [1]. The first few hours after a catastrophe are regarded as the "golden hours" of relief because rescue workers have a high probability of evacuating people from the damaged region during this period. Keep in mind that the wireless infrastructure in the disaster area might not be functional or even might be ravaged after the disaster. What makes the situation even more complicated is the paralysis of the power transmission lines after the disaster. The most powerful earthquake ever recorded in Japan, with a magnitude of 9.1, triggered a tsunami on the northeastern shore in March 2011. In the region of the catastrophe, around 6000 base stations (BSs) were wrecked, and the remaining BSs were highly overloaded with tremendous amounts of voice and data traffic. As a result of the high call block rate, communication services were suspended for roughly four days [2]. As a result, it is critical to develop an emergency wireless network that is completely independent of the conventional broadband network as soon as possible in order to preserve those valuable human lives. Unmanned aerial vehicles (UAVs) are well-known for their distinct characteristics, such as flexible deployment and rapid reaction. Thus, they can be deployed

as temporary mobile BSs to establish this type of temporary emergency wireless network [3]. UAVs are now employed for a variety of emergency wireless communication applications, such as disaster management, surveillance, early warnings, post-disaster fusion centers, damage assessment, and supply-aid drop, in addition to temporary emergency wireless networks [4].

Notwithstanding the advantages of utilizing UAVs for establishing emergency wireless communication networks in a post-disaster area, there are a number of issues that need to be neutralized. In this tough environment induced by a natural disaster, the UAV must first design and optimize its flying route. This necessitates a quick online optimization procedure to accommodate the dramatic shift in the geographical field [5]. Secondly, the available energy for victims is ephemeral due to the limited battery capacity of their UEs and the destruction of the power supply infrastructure as a result of the natural disaster [6]. Thirdly, the UAV's operating duration is restricted by the onboard battery's capacity. The UAV should return to its base for recharging before it is completely depleted [7]. Therefore, while constructing an emergency wireless communication network, all of these concerns should be addressed. In addition, since this is a crucial mission, the UAV must assist as many people as possible in the disaster zone before its battery dies. Consequently, it is vital to seek out a robust mathematical tool capable of tackling such novel challenges.

Machine learning (ML) algorithms, and more precisely, reinforcement learning (RL) algorithms, are leveraged to tackle these kinds of optimization problems. Since RL algorithms are capable of achieving superb results in terms of efficiency and generalization, and due to their ability to deal with optimization problems with conflicting parameters, researchers have been inspired to utilize them in dealing with real-time issues in the field of wireless communications networks [8]. In this context, modern UAVs are equipped with wireless communications, ML, and image processing techniques. These techniques can support a UAV's trajectory optimization while avoiding obstacles and dealing with a limited battery capacity, which leads to serving more spots and enhancing the whole mission's energy efficiency. Recently, "follow me" drones have boomed in market value [9]. These drones are capable of filming a moving person with intelligent target-tracking and obstacle-avoidance algorithms, resulting in fabulous camera footage. Furthermore, novel UAV-related applications such as area surveillance, disaster relief, and traffic control are just a few applications that can be intelligently developed for future cities [10].

Multi-armed bandit (MAB) algorithms are considered one of the RL algorithms which are preferred in dealing with online optimization problems [11]. MAB algorithms can be defined as a set of arms, i.e., actions, of a bandit machine. At any given moment, pulling an arm leads to an instantaneous reward that is sampled from a certain distribution. A player wants to maximize his accumulated reward over the playing period by choosing an arm to pull during each moment of playing. Nevertheless, this player has no idea about the instantaneous reward behind each arm, since it will be revealed when the player decides to choose it. Therefore, some amount of the reward could be missed out due to this hidden setting. This loss is denoted by the term regret [12,13]. Thus, a player should develop a strategy to choose the arm that leads to the highest reward. On the other hand, this strategy should keep an eye on balancing between playing with the previously discovered arms that have high rewards or playing with the still-undiscovered ones that might have higher rewards. This is a common MAB dilemma, and it is called the exploration–exploitation trade-off [14,15]. Aiming to bolster disaster resilience, this paper describes a method of leveraging the latest advances in MAB algorithms and UAV wireless communications networks to improve the functionality of emergency wireless communication services for post-disaster response and assessment.

1.1. Prior Works and Motivations

One of the main benefits of deploying UAVs in emergency wireless communication networks is their capability of gathering extensive data from scattered ground devices, such as ground BSs, ground users, and even ground sensors [16]. The paper just cited gives

a broad overview of different techniques but does not dive deeply in a specific direction. Furthermore, a UAV can operate as a flying edge server or a BS to support various traffic offloading scenarios [17], but it has a limited size of state action space. Due to its mobility, the planning and optimization of the UAV's trajectory and radio resource management of its wireless network are crucial issues. Researchers conducted many investigations on this topic during the past few years [18]. The UAV's speed and the location of its waypoints were used in [19] to design an optimal trajectory. However, the discussion was limited to cases where UAVs are used as relay stations in ad hoc networks. Minimizing the total energy consumption was studied in [20] using UAV speed control and a UAV data-scheduling-based heuristic algorithm, but it can be considered a theoretical approach only due to its large approximation factors. The authors of [21] considered UAVs with small cell capabilities to work as UAV-BSs. Particularly, the UAV movement, charging, and coverage action are considered in terms of jointly optimizing the energy and throughput through revenue and cost components. The UAV task scheduling was investigated in [22], where a mathematical framework for the optimization of UAV-aided video monitoring of a set of points of interest (PoI) distributed in a large urban area was proposed. Using this framework, which is based on mixed integer linear programming (MILP) techniques and real experimental data, particular energy-constrained UAVs are selected for recharging using public transportation buses, which also transfer the UAVs to desired PoIs in order to increase reliability and coverage.

UAV trajectory optimization may be carried out using traditional optimization approaches when realistic models of UAV wireless networks, including their flight dynamics, are available. Still, building these realistic network models is quite challenging; thus, model-free machine-learning methods can be used to manage the operation of UAVs that utilize wireless communication networks. By utilizing data gathered from prior experiences, machine learning algorithms are able to create autonomous control policies [23]. The authors of [24] studied the optimal deployment of UAVs equipped with directional antennas, using circle packing theory, where the 3D locations of the UAVs are determined such in a way that the total coverage area is maximized. The policy gradient approach for trajectory optimization used by the authors of [25] was able to maximize the overall distance covered by the UAV. However, this method took a lot of time and effort to find the best answer due to the large number of possible trajectories that the UAV must fly. The authors of [26] used the deep Q-learning method to optimize the UAV's flight path to maximize data rate during the flight period in an unknown environment. One major limitation of this proposed Q-learning approach for trajectory optimization is the long learning time, which makes it unfeasible even for moderate state spaces. By planning the UAV's flight trajectory, the authors of [27] were able to maximize the uplink transmission rate in a UAV cellular network. The deterministic policy gradient (DPG) approach was used to solve the optimization problem after it was converted into a Markov decision process (MDP). However, the characteristics of mmWave channels and beamforming were not taken into consideration during the optimization process.

Despite the existence of numerous excellent studies on UAV wireless communication networks, there are only a few works that focus on UAV-assisted emergency wireless communication networks. In our earlier studies [28,29], we investigated the radio resource allocation for a UAV emergency wireless communication network using a dynamic spectrum access system. The purpose of the deployment of UAVs as a cognitive radio network (CRN) was to maximize the downlink data rate in a post-disaster environment. Moreover, the limited transmission power of each UAV was used to control the constructed two multi-player MAB-based optimization problems called the power budget aware upper confidence bound (PBA-UCB) algorithm and the power budget aware Thompson sampling (PBA-TS) algorithm. The problem of gateway selection in a post-disaster area was addressed in [30], where a decentralized MAB algorithm was adapted to each UAV to let it maximize its data throughput by optimally choosing a suitable gateway. However, the optimization algorithm encountered some data loss due to not choosing the optimal strategy at the be-

ginning of the optimization process. The authors of [31] built a system of a re-configurable intelligent surface (RIS) attached to a UAV. With the aid of a modified version of the MAB algorithm, the optimization problem aimed to find the optimum trajectory of the UAV that maximizes the total throughput while reducing the consumed flying power of the UAV. For a UAV with a limited battery capacity, the maximization problem for the number of served users was studied in [32] using two MAB algorithms called the ϵ-greedy algorithm and the D-UCB algorithm. The UAV trajectory optimization problem was studied in [33] to maximize the accumulated data volume from ground sensors under unknown network information. The optimization problem was transformed into a finite MDP and solved using two Q-learning-based UAV trajectory optimization frameworks called SUTOA and QUTOA. A Lyapunov-based deep Q-learning framed work called Safe-DQN was proposed in [34] to study the UAV trajectory optimization problem in a UAV-based emergency wireless communication network. The joint optimization problem aimed to maximize the total system rate under the constraints of the limited flight time of the UAV, the power capacity of the ground user, and the need to avoid obstacles in the disaster area. All the previous research was controlled by the limited capacity of the attached onboard battery for each UAV.

All of these studies on UAV emergency wireless communication networks focused on the optimization issue under a single power restriction, either a restricted UAV battery capacity or a limited amount of energy accessible to ground users (i.e., ground UE or ground sensors). We argue that these two elements together should be taken into account while constructing a UAV emergency wireless communication network. This is because the natural disaster destroys or at least renders the power supply network inoperable. Therefore, the goal of our suggested framework is to solve the UAV trajectory optimization problem under these two limited power constraints. In order to do this, our goal was to investigate a dual constraint optimization problem that might increase the UAV emergency wireless network's reliability in comparison to earlier studies. It should be noted that, to the best of our knowledge, our earlier work in [35] was the first study to investigate this sort of optimization issue with dual constrained energy capacity for both UAV and UEs at the same time. Furthermore, in the research, we extend our problem formulation by deeply evaluating the performance of our proposed framework against different benchmark methods. This evaluation was conducted in terms of the accumulated long-term uplink throughput of all UEs, the energy consumed by all UEs during the data-offloading process, and the energy efficiency of the UEs.

1.2. Contributions and Organization

According to the discussion in the preceding subsection, the majority of recent research on UAV emergency wireless communication networks concentrated primarily on the limited battery energy capacity of UAVs; just a small number of studies took into account the restricted energy capacity of ground users, i.e., ground users' equipment (UEs). We created a suggestion to fill this gap by examining an optimization scenario with constrained energy capacity for both UEs and UAVs. UAVs are seen as flying BSs that provide a wireless connection to ground UEs in the disaster-affected region from the sky. The information gathered from the UEs is deemed critical for estimating the status of the victims and assessing the damage in the post-disaster area. As a result, this critical data may be processed to help rescue crews save these precious lives. Our major goal is to acquire as much data from ground UEs as possible given the restricted power capacity of both the UAV and the ground UEs. However, since UAV coverage is somewhat limited in comparison to terrestrial BSs, our goal is to optimize the UAV flight trajectory to maximize the number of ground UEs visited before the battery runs out. Considering this limited battery capacity, another interesting idea is to have the UAVs maximize the scanned area while capturing photos to aid the rescue teams or to estimate the damage caused by the natural disaster. This goal was kept for our future work. The primary contributions of this work can be summarized as follows:

- In our situation, a UAV would gather user data in a disaster-affected region as part of a wireless emergency communication network. Ground BSs fail as a consequence of natural catastrophe damage, but ground UEs in the UAV coverage area may upload data using an alternate mode of connection from the sky thanks to the assistance of the UAV emergency wireless communication network. We propose an online optimization problem to optimize the uplink throughput for the UAV emergency wireless communication network by optimizing the flight trajectory of the UAV under these assumptions, taking into consideration the limited available energy for both the UAV and ground UEs in the post-disaster region.
- The optimization problem is adapted into a constrained MAB problem, with action, reward, and cost defined as the flight direction, uploaded data throughput, and dissipated energy for both the UAV and UEs, respectively.
- The numerical analysis of our proposed framework shows a considerable increase in long-term throughput and a slight increase in the energy consumption of the UEs in the post-disaster area, resulting in better energy efficiency for our proposed framework compared to other benchmark UAV trajectory optimization methods.

The rest of this paper is organized as follows. Section 2 presents the network architecture and formulates the online optimization problem for the long-term uplink throughput maximization problem. In Section 3, the general MAB framework is illustrated, followed by our proposed MAB-based framework for UAV trajectory optimization under dual energy constraints. Simulation results and numerical analysis are given in Section 4, and finally, the paper is concluded in Section 5.

2. Network Architecture and Problem Formulation

In this section, we discuss the architecture for the UAV-assisted emergency wireless communication network, including the flying model used for the UAV, the channel model for data uploading, and the optimization problem formulation.

2.1. UAV Flying Model

The system architecture for the UAV-assisted emergency wireless communication network is shown in Figure 1. In this scenario, a natural disaster, such as an earthquake or flood, strikes a specific location and causes the power grid and wireless network to fail. Our plan is to use the UAV to enable wireless access from the sky in this post-disaster area. In this approach, wireless connectivity may be enabled for victims, i.e., ground UEs, in this devastated region, allowing them to offload data that will be useful in guiding rescue crews and evaluating the damage. We assumed that there are M UEs trapped in this post-disaster area, denoted by $\mathcal{M} = \{1, \ldots, M\}$. Each of them has a fixed position designated by the following in Cartesian coordinates $l_m = (x_m, y_m)$. The UEs locations are supposed to be known to the UAV through self-reported global positioning system (GPS) coordinates. The discussion on how these data are transferred to the UAV is beyond the scope of this paper. It is assumed that the UAV will begin flying from the center of the post-disaster area, i.e., the center of the simulation area, which is denoted by $l_0 = (x_0, y_0)$. Additionally, it flies according to a constant speed of v and an altitude of H. We assume that this altitude is relatively high and that the data transmission duration is reasonably short and denoted by τ. As a result, the UAV is regarded immobile when uploading the UE data.

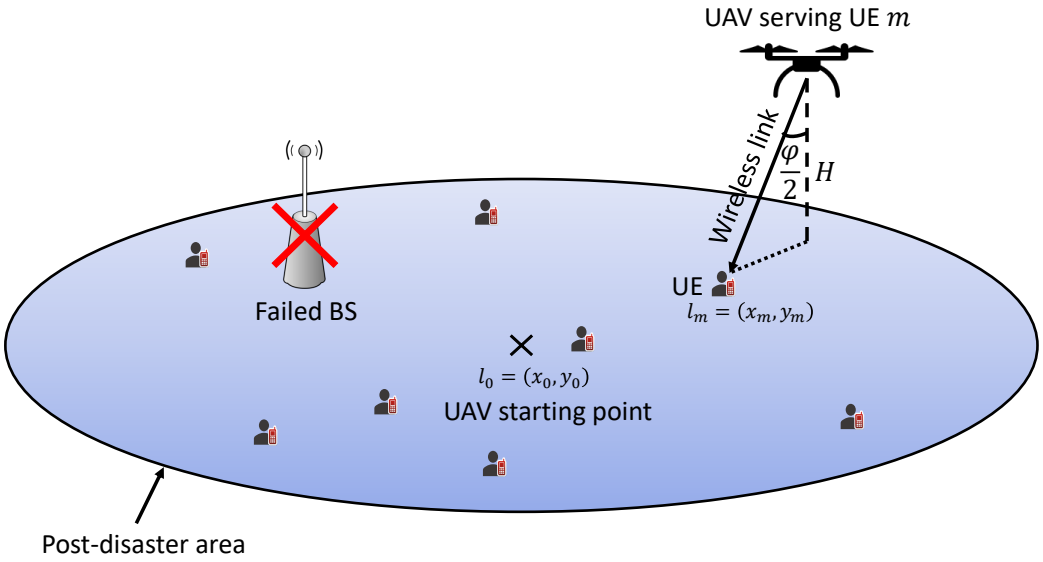


Figure 1. UAV emergency wireless communication network.

2.2. Wireless Communication Channel Model

For the convenience of designing an emergency wireless communication network, our designed system should utilize a channel in the unlicensed band, i.e., 2.4 GHz. In such a way, this system can be easily integrated with the hardware of modern UEs. Hence, the utilized channel model is expounded at [34], in accordance with the 3rd Generation Partnership Project (3GPP) specification in the technical report presented in [36]. This channel model represents the wireless communication link between the UAV and each of the served UEs into two components, i.e., the line-of-sight (LOS) component and the non-line-of-sight (NLOS) component, according to their corresponding probabilities, and can be calculated by (1).

$$L_m = \begin{cases} 30.9 + (22.25 - 0.5\log_{10} H)\log_{10} d_m + 20\log_{10} f, & \text{if LOS link} \\ \max\left(L_m^{\text{LOS}}, 32.4 + (43.2 - 7.6\log_{10} H)\log_{10} d_m + 20\log_{10} f\right), & \text{if NLOS link} \end{cases} \tag{1}$$

where H denotes the UAV flight altitude, f is the carrier frequency, and d_m is the distance between the UAV and any corresponding UE m, which can be calculated as follows:

$$d_m = \sqrt{H^2 + \|l_m - l_0\|^2}, \ \forall m \in \mathcal{M} \tag{2}$$

Since the calculation of path loss due to the NLOS component is a function of the path loss due to the LOS component L_m^{LOS}, the term L_m^{LOS} should be calculated prior to estimate the path loss of the NLOS component. The probability of the LOS link is denoted by $\mathcal{P}_m^{\text{LOS}}$ and given in (3).

$$\mathcal{P}_m^{\text{LOS}} = \begin{cases} 1, & \text{if } \sqrt{d_m^2 - H^2} \le d_0 \\ \frac{d_0}{\sqrt{d_m^2 - H^2}} + \exp\left\{\left(\frac{-\sqrt{d_m^2 - H^2}}{p_1}\right)\left(1 - \frac{d_0}{\sqrt{d_m^2 - H^2}}\right)\right\}, & \text{if } \sqrt{d_m^2 - H^2} > d_0 \end{cases} \tag{3}$$

$$d_0 = \max\left(294.05\log_{10} H - 432.94, 18\right) \tag{4}$$

$$p_1 = 233.98\log_{10} H - 0.95 \tag{5}$$

Furthermore, the probability of NLOS can be obtained naturally for the probability of LOS as follows:

$$\mathcal{P}_m^{\mathrm{NLOS}} = 1 - \mathcal{P}_m^{\mathrm{LOS}} \tag{6}$$

The channel gain between the UAV and any connected UE can be calculated as follows:

$$g_m = \mathcal{P}_m^{\mathrm{LOS}} \left(10^{L_m^{\mathrm{LOS}}/10}\right)^{-1} + \mathcal{P}_m^{\mathrm{NLOS}} \left(10^{L_m^{\mathrm{NLOS}}/10}\right)^{-1} \tag{7}$$

where L_m^{LOS} and L_m^{NLOS} are the path loss for the LOS and NLOS, respectively, and can be calculated from (1).

2.3. Data Transmission Model

For the sake of simplicity, we assumed that the UAV emergency wireless communication network can be established between the UAV and only one UE at any certain time. Hence, there are no simultaneous wireless connections from different UEs to the UAV. The effective radiation angle of the UAV antenna is denoted by φ, so the maximum distance between the UAV and any UE that permits the establishment of a wireless communication link is $H/\cos(\varphi)$. Additionally, it can be observed that the relationship between the channel gain g_m in (7) and the distance d_m in (2) is an inverse relationship. Therefore, our definition of the effective radiation angle φ is used as a parameter to make sure that this distance is suitable for establishing a wireless communication link. This can be done by evaluating the signal-to-noise ratio (SNR) value for a covered UE. When it reaches a certain threshold that permits the establishment of a wireless communication link, this covered UE can access the UAV to offload its data. Additionally, the value of φ can be chosen to be very narrow to shrink the UAV coverage. In such a way, the simultaneous transmission from different UEs can be easily eliminated. Hence, a UE can be within the UAV coverage if and only if it belongs to the following set:

$$\mathcal{M}_{\mathrm{cov}} = \{m \in \mathcal{M} \ : \ d_m \leq H/\cos(\varphi)\} \tag{8}$$

It is assumed that each UE in the post-disaster area has an amount of data equal to Ψ bits. Then, a UE access indicator, denoted by α_m, is used to show whether the m-UE is connected to the UAV or not. This access indicator depends on two factors, i.e., the distance from the UAV, d_m, and the total uploaded bits from the m-UE to the UAV, Ω_m. Thus, α_m can be expressed as follows:

$$\alpha_m(t) = \begin{cases} 1, & \text{if } m \in \mathcal{M}_{\mathrm{cov}},\ \Omega_m(t) < \Psi \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where $t \in \mathcal{T}$, $\mathcal{T} = \{1, \ldots, T\}$ is the time elapsed while the UAV flies over the post-disaster area. The total uploaded bits from the m-UE to the UAV can be calculated as:

$$\Omega_m(t) = \sum_{i=1}^{t} \omega_m(i) \tag{10}$$

where ω_m is the instantaneous uploaded data size at time t and can be calculated as follows:

$$\omega_m(t) = R_m(t)\ \tau \tag{11}$$

where R_m is the transmission data rate from the m-UE to the UAV and can be calculated according to Shannon's theorem as follows:

$$R_m(t) = \alpha_m(t)\ B\ \log_2\left(1 + \frac{g_m\ P_m^{\mathrm{Tx}}}{\sigma_0}\right) \tag{12}$$

where B is the available wireless channel bandwidth, P_m^{Tx} is the transmission power from m-UE, and σ_0 denotes the power of the additive white Gaussian noise (AWGN) at the UAV receiver.

2.4. Energy Model

From the perspective of the limited energy capacity, the consumed energy can be classified as follows: (1) the energy consumed by each m-UE while it is idle and during the data offloading period; (2) the energy consumed by the UAV while it is flying over the post-disaster area to provide the wireless connectivity for the trapped UEs. Thus, at any time t, these two consumed terms of energy can be denoted as follows:

$$e_m(t) = \begin{cases} \alpha_m(t)\ P_m^{\mathrm{Tx}}\ \tau, & \text{if } m\text{-UE at Tx mode} \\ (1-\alpha_m(t))\ e^{\mathrm{idle}}, & \text{if } m\text{-UE at idle mode} \end{cases} \tag{13}$$

$$E(t) = \Xi\, t \tag{14}$$

where e^{idle} is the energy consumed by each of m-UE during the idle mode, and Ξ is the UAV's flying power. Of course, there are many factors that control the UAV's energy consumption, such as the flying speed, acceleration, and mass of the UAV. However, we tried to simplify the notation of the energy consumption to be averaged per unit of time. In such a way, we can study the ability of our proposed solution to handle this dynamic energy consumption over time. Furthermore, the energy consumed by the UAV's receiver circuit and signal processing are relatively low compared to the energy consumed during flying, so it can be neglected. To expand this research to more detailed power consumption, the work presented in [3] is a straightforward extension, and it will be considered for our future work.

2.5. Problem Formulation

The ultimate goal of the post-disaster surveillance system is to improve the rescue success rate of victims and also to reduce casualties. This goal can be achieved by maximizing the data uploaded from the trapped victims in the post-disaster area over the UAV trajectory. At the same time, we must take into account the valuable limited energy of both UEs and the UAV. Mathematically speaking, our optimization problem can be expressed as follows:

$$\max_{m \in \mathcal{M}} \quad \frac{1}{T} \sum_{t=1}^{T} \sum_{m=1}^{M} \omega_m(t) \tag{15}$$

$$\text{s.t.} \sum_{t=1}^{T} e_m(t) \leq e_0,\ \forall m \in \mathcal{M} \tag{16}$$

$$\sum_{t=1}^{T} E(t) \leq E_0 \tag{17}$$

The optimization problem shown in (15) is considered an online optimization problem that aims to maximize the long-term throughput of the whole network by optimizing the UAV's flight trajectory. Since there is an unlimited number of routes that can be existed by changing the order of how the UAV serves the UEs, our optimization problem is an NP-hard problem. However, by considering energy constraints introduced in equations (16) and (17), the optimization problem can be viewed as an NP-complete problem. The whole optimization process is done not only in an online manner but also in a decentralized way where there is no information exchange between different network elements. Furthermore, for any conventional programming solvers, all information should be gathered at one centralized entity to solve the optimization problem, which cannot be satisfied while designing an emergency wireless communications network for a post-disaster surveillance system. In such a case, we suggest using a reinforcement-learning-based algorithm to deal with this kind of online optimization problem.

The decision variables can be defined as the accumulated instantaneous throughput $\omega_m(t)$ for all the M UEs throughout the UAV's flight time T. The constraint (16) shows that

the maximum energy available for each UE is limited by e_0, and the other constraint (17) limits the energy available for the UAV by E_0; both are considered the feasibility constraints of the optimization problem. Furthermore, the right-hand sides of constraints (16) and (17) are also long-term cumulative variables related to the UAV flight trajectory. Hence, the whole flight-trajectory process should be taken into account when solving the position of the UAV at any time t. Therefore, this optimization problem becomes difficult to figure out using conventional optimization methods. Additionally, sharing information on the remaining battery capacity for every UE in the post-disaster area is quite a changeling, especially when the commercial mobile network has malfunctioned. Therefore, for the sake of simplicity and without loss of generalization, our optimization problem was designed for the worst-case scenario for the available battery capacity for each UE. This value was chosen to be around 10% of modern UE's average total battery capacity [37]. In the next section, we introduce an MAB-based framework to tackle this issue.

3. Dual-Energy-Aware MAB-Based UAV Trajectory Optimization Approach

In this section, we explain the general MAB framework and then illustrate how the proposed dual-energy-aware MAB approach could address our previously described optimization problem.

3.1. General MAB Framework

Generally speaking, in any MAB-based framework, a player aims to maximize his long-term reward while playing with a set of arms of the bandit machine, $j \in \{1, \ldots, J\}$. This can be performed in a sequential way by selecting an arm at time t, i.e., $j(t)$, and observing their corresponding reward, i.e., $r_j(t)$. In the first few moments, the player tries to explore candidate arms as much as possible and observes their corresponding rewards. After that, the player exploits the arm with the highest reward, based on the gathered information from the already explored arms, to maximize the cumulative reward over the episode. This dilemma is quite well-known in the world of the MAB framework and is known as the exploration–exploitation trade-off [15]. The MAB framework can be classified as stochastic or adversarial based on the distribution of the rewards [14,15]. For the stochastic MAB framework, the rewards behind each arm are drawn from independent and identical distribution (i.i.d); however, for the adversarial MAB framework, rewards are selected arbitrarily with no prior distribution. For these two types of MAB frameworks, extensive research has been done to deal with these kinds of problems, resulting in the introduction of different algorithms, such as the ϵ-greedy algorithm [38], the upper confidence bound (UCB) algorithm [39], the Thompson sampling (TS) algorithm [40], and the exponential-weight algorithm for exploration and exploitation (EXP3) [41]. Furthermore, in real-world optimization problems, choosing an arm with a higher reward will have a high cost as well. Thus, cost-effective and budget-constrained MAB algorithms are introduced to deal with these kinds of scenarios [42,43].

3.2. The DEA-MAB Approach

To address the online optimization problem with the dynamic energy consumption over time that is given in (15), and which constrained by conditions (16) and (17), an MAB-based framework that is dual-energy-aware called DEA-MAB is proposed. Our DEA-MAB approach is inspired by the cost-subsidized explore-then-commit algorithm proposed in [43], where the chosen arm is accompanied by a certain cost. One of the traditional ways to optimize this reward/cost is to directly deduct the cost from the reward in the control formula. However, this is not usually meaningful in real-world problems, especially when the reward and the cost are defined in different quantities [43], such as the achievable throughput and the energy consumed, as illustrated in our problem formulation. Hence, it is necessary to find a better way to optimize for both the reward and the cost. In other words, the algorithm should avoid incurring an excessive cost for just a marginal increase in the reward. This may be done by building a feasible set of arms which is an estimate of

all arms with a mean reward greater than the least tolerable value in each round, based on the upper confidence bound (UCB) and lower confidence bound (LCB) of the reward of each arm. Then, the arm with the lowest cost in this feasible set is selected to be played by it.

Though the cost-subsidized explore-then-commit algorithm is considered a good solution for separating the reward and the cost functions, it still needs some adaptation to tackle our optimization problem that is given in (15). Precisely speaking, our optimization problem considers two different energy costs, so the DEA-MAB algorithm adds a further step for checking the second cost. Thus, some controlling functions were added in the proposed algorithm to precisely address this issue.

Algorithm 1 summarizes how the DEA-MAB algorithm works. The DEA-MAB algorithm's input attributes are the state spaces of all available $\mathcal{M}$ UEs, including their corresponding locations $l_m \; \forall \; m \in \mathcal{M}$; the total flight time T; tuning parameters δ and ϵ; the available energy for each piece of UE e_0; and the total flight time of the UAV till its battery is completely depleted, T. At each time period t of the total flight time T, the UAV should select one of M UEs distributed in the post-disaster area via the DEA-MAB algorithm; then it will fly towards it to offload its data. In the beginning, the algorithm is initialized at $t = 0$ by setting the number of times each m-UE is selected, $Q_m(t)$, and their average achievable throughput, $\overline{\omega}_m(t)$, to 0. The DEA-MAB algorithm is divided into two phases, i.e., the pure exploration phase and the selection phase. During the exploration phase, the UAV randomly selects a UE to visit as follows:

$$m^*(t) = t \mod M \tag{18}$$

Then, the corresponding throughput $\omega_{m^*}(t)$ is observed, and the selection number, $Q_m(t)$, and the average throughput, $\overline{\omega}_m(t)$, are updated as in the following equations:

$$Q_{m^*}(t) = Q_{m^*}(t-1) + 1 \tag{19}$$

$$\overline{\omega}_{m^*}(t) = \frac{1}{Q_{m^*}(t)} \sum_{i=1}^{Q_{m^*}(t)} \omega_m(i) \tag{20}$$

The exploration phase is performed for a time period equal to $M\,\pi$, where $\pi = (T/M)^{2/3}$ is as given in [43]. After that, the DEA-MAB algorithm goes for the selection phase during each time $t \in [M\pi + 1, T]$, where both the UCB and LCB are calculated as follows:

$$\gamma_m^{\mathrm{UCB}}(t) = \overline{\omega}_m(t) + \sqrt{2\ln(t)/Q_m(t)}, \; \forall \; m \in \mathcal{M} \tag{21}$$

$$\gamma_m^{\mathrm{LCB}}(t) = \overline{\omega}_m(t) - \sqrt{2\ln(t)/Q_m(t)}, \; \forall \; m \in \mathcal{M} \tag{22}$$

Then, the UE index corresponding to the maximum value of the $\gamma_m^{\mathrm{LCB}}(t)$ is calculated as follows:

$$\eta_t = \arg\max_m \gamma_m^{\mathrm{LCB}}(t) \tag{23}$$

Afterwards, the feasibility region of all UEs having $\gamma_m^{\mathrm{UCB}}(t) \geq (1-\delta)\gamma_{\eta_t}^{\mathrm{LCB}}(t)$ is enumerated as follows:

$$F(t) = \left\{ m : \gamma_m^{\mathrm{UCB}}(t) \geq (1-\delta)\gamma_{\eta_t}^{\mathrm{LCB}}(t) \right\} \tag{24}$$

Algorithm 1: The proposed algorithm: DEA-MAB.

Output: $m^*(t)$
Input: $\mathcal{M}$, $l_m \; \forall \; m \in \mathcal{M}$, T, δ, ε, e_0, ν, Ξ, E_0
Initialization: at $t = 0$, Set $Q_m(0) = 0$, $\overline{\omega}_m(0) = 0$, $\forall m \in \mathcal{M}$
Exploration Phase:
Explore available UEs and calculated the corresponding throughput
for $t = 1$ **to** $M\pi$ **do**
  $m^*(t) = t \mod M$
  Fly towards a UE $m^*(t)$ and obtain $\omega_{m^*}(t)$
  $Q_{m^*}(t) = Q_{m^*}(t-1) + 1$
  $\overline{\omega}_{m^*}(t) = \frac{1}{Q_{m^*}(t)} \sum_{i=1}^{Q_{m^*}(t)} \omega_m(i)$
end for
Selection Phase:
for $t = M\pi + 1$ **to** T **do**
  $\gamma_m^{\mathrm{UCB}}(t) \leftarrow \overline{\omega}_m(t) + \sqrt{2\ln(t)/Q_m(t)}$, $\forall \; m \in \mathcal{M}$
  $\gamma_m^{\mathrm{LCB}}(t) \leftarrow \overline{\omega}_m(t) - \sqrt{2\ln(t)/Q_m(t)}$, $\forall \; m \in \mathcal{M}$
  $\eta_t = \arg\max_m \gamma_m^{\mathrm{LCB}}(t)$
  $F(t) = \left\{ m : \gamma_m^{\mathrm{UCB}}(t) \geq (1-\delta)\gamma_{\eta_t}^{\mathrm{LCB}}(t) \right\}$
  Obtain $e_m \; \forall \; m \in F(t)$
  if $\sum_{i=1}^{t} e_m(i) \geq (1-\varepsilon)e_0$ **then**
    $C(t) = \left\{ m : \sum_{i=1}^{t} e_m(i) \geq (1-\varepsilon)e_0 \right\}$
  else
    $C(t) = F(t)$
  end if
  $m^*(t) = \arg\min_{m \in C(t)} E(t)$
  The UAV fly towards UE $m^*(t)$ and obtain $\omega_{m^*}(t)$
  $Q_{m^*}(t) = Q_{m^*}(t-1) + 1$
  $\overline{\omega}_{m^*}(t) = \frac{1}{Q_{m^*}(t)} \sum_{i=1}^{Q_{m^*}(t)} \omega_m(i)$
  if $E_0 - \sum_{i=0}^{t} E(i) < 2\,\Xi\,\sqrt{\|l_{m^*} - l_0\|^2\,\nu^{-1}}$ **then**
    Break the data offloading loop and the UAV returns to its base
  end if
end for

For this set of UEs, $F(t)$, the dissipated energy for each of the m-UE contained in this $F(t)$ list is obtained. Then, a control set, $C(t)$, is constructed out of all UEs in $F(t)$. A check is performed for the UEs' energy consumption; then priority is given to all UEs in the $F(t)$ list in case they exceed their energy consumption with a value of $1-\varepsilon$ of the total available energy e_0. Otherwise, $C(t)$ is set to be equal to $F(t)$. This can be illustrated as follows:

$$C(t) = \begin{cases} m : \sum_{i=1}^{t} e_m(i) \geq (1-\varepsilon)e_0, & \sum_{i=1}^{t} e_m(i) \geq (1-\varepsilon)e_0 \\ F(t), & \text{otherwise} \end{cases} \tag{25}$$

Out of this list, $C(t)$, the UE corresponding to the minimum UAV energy cost, $E(T)$, is selected as a next-served UE for data offloading in the UAV flight trajectory as follows:

$$m^*(t) = \arg\min_{m \in C(t)} E(t) \tag{26}$$

Afterwards, values of the selection number, $Q_m(t)$, and the average throughput, $\overline{\omega}_{m^*}(t)$, are updated for the selected UE, $m^*(t)$, as given in Algorithm 1. Since the UAV should accomplish the whole data offloading task and ensure flying back to its base before the

battery is used up, the UAV should confirm that there is enough remaining battery energy for returning. Otherwise, the UAV could be lost or damaged if it cannot arrive at its base before the battery becomes empty. Therefore, a checking step is provided to confirm this critical condition at each time before deciding to choose the next UE to be served. In this way, the DEA-MAB algorithm can optimize the UAV's flight trajectory considering limited energy of both the UAV and the UEs.

3.3. Complexity Analysis of The Proposed Approach

In the previous section, the task of the UAV finding the best trajectory in the post-disaster area is spotlighted. This is accomplished by finding the optimal policy to choose the next UE to be served through the learning process in Algorithm 1. In the beginning, the uplink throughput that can be achieved while the UAV connects to this UE is examined. Then, a higher priority is given to UEs whose batteries are nearly depleted. The consumed energy during UAV flying is also minimized. Moreover, it is assumed that the action space is deterministic; i.e., all actions are well-known to the UAV. Hence, the fundamental source of the computational complexity of the DEA-MAB algorithm comes from calculating both the UCB and the LCB. Then, other parameters are updated according to this selection. It should be mentioned that these parameters have the same computational complexity order as UCB or LCB. Hence, the overall computational complexity order of our proposed algorithm is a polynomial of $M + 1$, and can be expressed as $\mathcal{O}(M + 1)$ [43].

4. Simulation Results

In this section, the performance of the DEA-MAB algorithm is evaluated. In the simulation, it was assumed that the UAV will provide wireless connectivity for a previously allocated area where there are M trapped UEs which are randomly distributed. However, for a large post-disaster area, more than one UAV can be deployed to support the data offloading while considering the coordination between UAVs to facilitate rescue operations. This larger system is left for future work.

Table 1 shows the simulation parameters used in verifying our proposed algorithm. In order to investigate the effectiveness of our proposed framework, two trajectory optimization methods were used as benchmarks for the sake of comparison. These two methods can be described as follows:

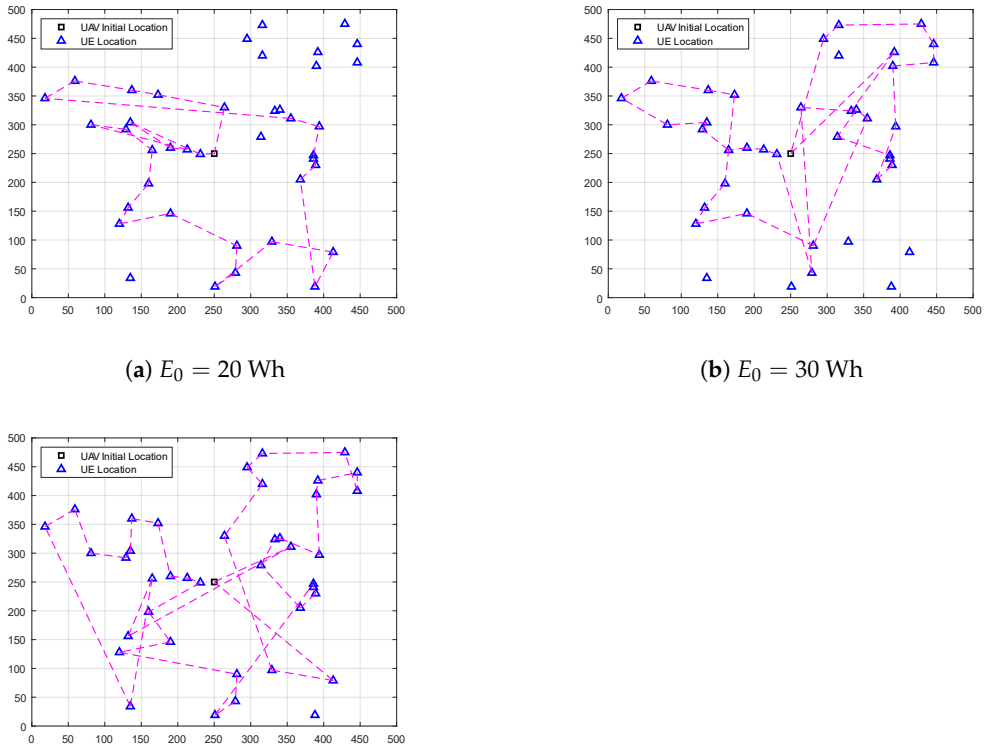
1. The post-disaster area spiral scanning (PASS) method: This method is designed to scan the whole area using the spiral path where the UAV starts to fly from the center of the post-disaster area. With respect to the UAV antenna's radiation angle, a projected circle is created on the ground. This circle scans the whole post-disaster area from the center to the borders.
2. Shortest flight path (SFP) method: In this method, the UAV starts to fly from the center of the post-disaster area and then selects the UE with the shortest path. Then, the UAV flies toward this UE and hovers above it to offload its data. After that, the UAV searches for the next close UE and flies toward it. This operation is performed till the last UE.

In the following, the performance of the proposed framework is evaluated by comparing it with benchmark algorithms during the varying of both the number of trapped UEs in the post-disaster area and the UAV's battery capacity. For the sake of accuracy, and due to the randomness in UEs' distributions, all simulations were performed for a long enough time, i.e, 10^4 iterations. The average value of each case is provided for a better estimation of the result.

Figure 2 shows a sample of the UAV's flight trajectory in the post-disaster area. To visualize how our DEA-MAB algorithm could optimize the UAV's flight trajectory considering its available battery power, three different values were used, i.e., $E_0 = 20, 30, 40$ Wh, while keeping the number of UEs equal to 40. Obviously, increasing available UAV battery power increases the chance of serving more UEs in the post-disaster area.

Table 1. Simulation parameters.

Parameter	Value
Simulation area	500 m × 500 m
Number of UEs in the simulation area (M)	$20, 30, 40, 50$
UAV flight speed (v)	20 km/h
UAV flight altitude (H)	100 m
UAV antenna radiation angle (φ)	$\pi/8$ rad
Carrier frequency (f)	2.4 GHz
Channel bandwidth (B)	10 MHz
Data transmission duration (τ)	1 s
UE Transmission power (P_m^{Tx})	23 dBm
AWGN spectral density (σ_0)	-130 dBm/Hz
UAV battery capacity (E_0)	$20, 30, 40$ Wh
UAV flying power (Ξ)	120 W
UE battery capacity (e_0)	1 Wh
UE energy dissipation in idle mode (e^{idle})	0.01 J
Data rate feasibility region factor (δ)	0.6
Critical power feasibility region factor (ε)	0.5



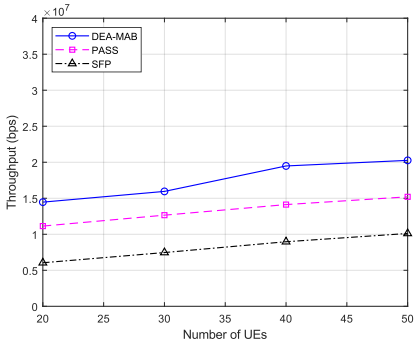
(**a**) $E_0 = 20$ Wh (**b**) $E_0 = 30$ Wh

(**c**) $E_0 = 40$ Wh

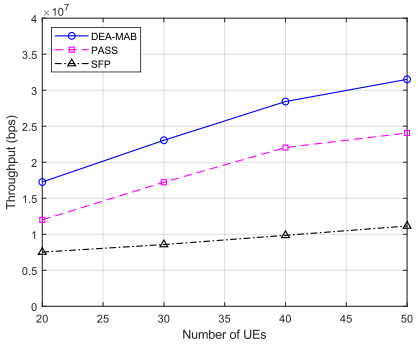
Figure 2. A sample of the UAV flight trajectory using the DEA-MAB algorithm.

Figure 3 gives the long-term throughput for the data uploaded from UEs in the emergency wireless communication network. It is clearly visible that regardless of the

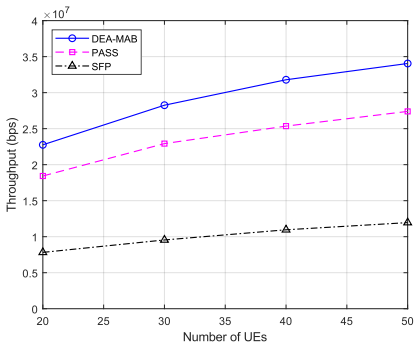
value of the UAV's battery capacity or the algorithm used, as the number of UEs trapped in the post-disaster area increases, the uplink data throughput increases as well. Nevertheless, this upward trend gradually decreases, and all curves would saturate at a certain number of UEs. This is because the maximum capacity of a communication system with a fixed bandwidth is fixed. Hence, while the number of UEs increases, the accumulated uplink throughput of the emergency wireless network continues to approach this maximum capacity. When comparing the throughput performance of the DEA-MAB algorithm with other benchmark methods at various values of UAV battery capacity, it is clear that our proposed algorithm can achieve more uplink throughput than the PASS method, and much higher than the SFP method. For example, when ($E_0 = 20$ Wh, $M = 30$), ($E_0 = 30$ Wh, $M = 40$), and ($E_0 = 40$ Wh, $M = 50$), the DEA-MAB algorithm achieved higher throughput performance by 26%, 28%, and 24% compared to the PASS method, and high performance by 113%, 188%, and 184% than the SFP method, respectively.



(**a**) $E_0 = 20$ Wh



(**b**) $E_0 = 30$ Wh

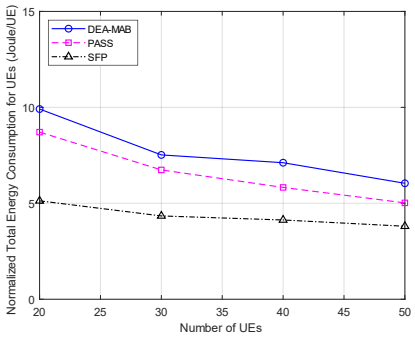


(**c**) $E_0 = 40$ Wh

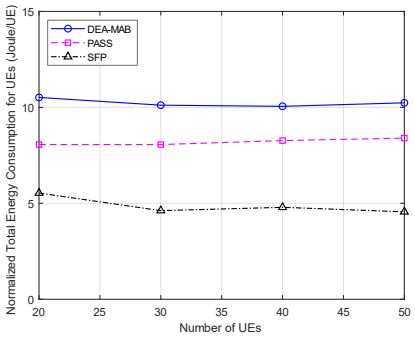
Figure 3. The DEA-MAB algorithm's throughput versus the number of users.

In Figure 4, the normalized total energy consumption of all UEs trapped in the post-disaster area is compared among the three methods. It can be seen clearly that regardless of the used method, as the number of UEs increases, the total normalized energy consumption of UEs increases as well. Furthermore, for the same method with a certain number of UEs, the higher the UAV's battery capacity, the more energy consumed per UE. This can be justified, as when the UAV has a higher battery capacity, it can have a higher chance to offload data from a larger number of UEs before its battery becomes depleted. Additionally, since $P_m^{\mathrm{Tx}}\tau >> e^{\mathrm{idle}}$, more UEs tend to consume energy in the data-offloading process rather than just staying in idle mode. When comparing the normalized energy consumption performance of the DEA-MAB algorithm with other benchmark methods at the same values of UAV battery capacity, it can be shown that the DEA-MAB algorithm always has higher energy consumption than the PASS method, and much higher than the SFP method. This can be explained by the overall system throughput being increased at the

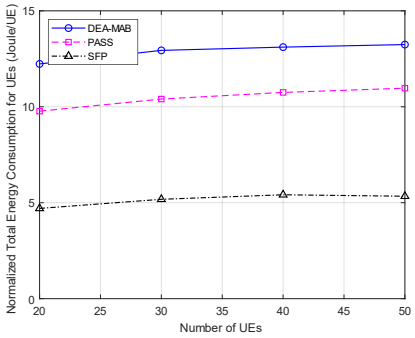
cost of more energy consumption by the UEs. For the sake of comparison, let us observe the same points at ($E_0 = 20$ Wh, $M = 30$), ($E_0 = 30$ Wh, $M = 40$), and ($E_0 = 40$ Wh, $M = 50$): the total energy consumption of all UEs using the DEA-MAB algorithm was increased by 11%, 24%, and 23% compared to the PASS method, and by 73%, 109%, and 169% compared to the SFP method.



(**a**) $E_0 = 20$ Wh



(**b**) $E_0 = 30$ Wh



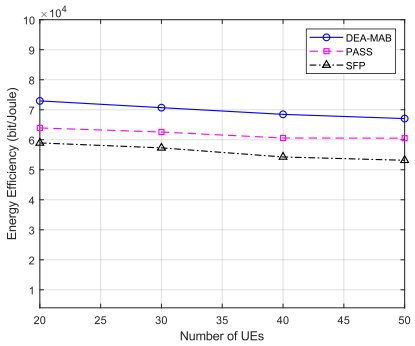
(**c**) $E_0 = 40$ Wh

Figure 4. Normalized total energy consumption versus the number of users.

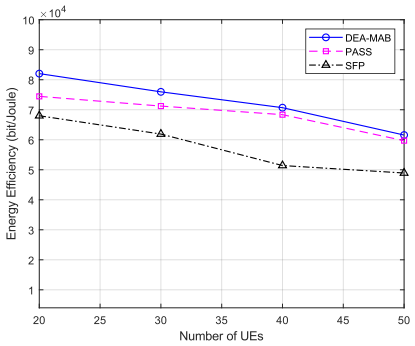
As observed from the analysis of results in Figures 3 and 4, it can be concluded that the DEA-MAB algorithm can achieve a considerable increase in the uplink throughput of UEs with a reasonable increase in the UEs energy consumption. Hence, for a better understanding of the advantages of using the DEA-MAB algorithm, the UEs' energy efficiency (μ) is compared using our proposed algorithm against benchmark methods. μ can be defined as the ratio of the long-term UEs uplink throughput over the total UEs energy consumption in bit/Joule as follows:

$$\mu = \frac{\sum_{t=1}^{T} \sum_{m=1}^{M} \omega_m(t)}{\sum_{t=1}^{T} \sum_{m=1}^{M} e_m(t)} \tag{27}$$

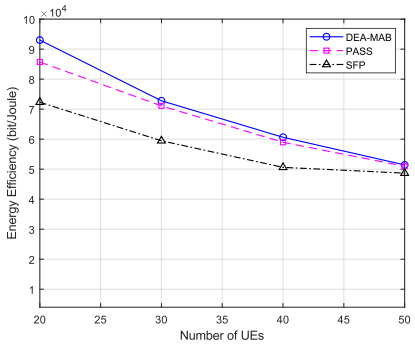
In the energy efficiency performance shown in Figure 5, it is observed clearly that whatever the UAV's battery capacity or the number of UEs trapped in the post-disaster area, the DEA-MAB algorithm can surpass benchmark methods in terms of energy efficiency, which, of course, means enhancing the overall performance of the emergency wireless communication network. It should be mentioned that, when increasing the UAV's battery capacity to 40 Wh, as in Figure 5c, the PASS method achieved a performance that is very close to that of the DEA-MAB algorithm. This can be justified, as the UAV's battery at this point becomes quite enough to accomplish the spiral scanning for a major part of the post-disaster area.



(**a**) $E_0 = 20$ Wh



(**b**) $E_0 = 30$ Wh



(**c**) $E_0 = 40$ Wh

Figure 5. UEs' energy efficiency versus the number of users.

5. Conclusions

In this paper, the trajectory optimization for a UAV-assisted emergency wireless communication network was investigated. The UAV is deployed as a temporary BS to provide wireless connectivity from the sky for trapped UEs in a post-disaster area where all BSs are damaged or have malfunctioned due to a natural disaster. The UAV's target is to optimize its flying trajectory to maximize the long-term uplink throughput from UEs. However, due to the malfunctioning of the power supplies in the disaster area as well, this optimization problem is performed with limited battery capacity of not only the UAV but also UEs in the post-disaster area. We proposed an MAB-based algorithm constrained with these two energy limitations to address this optimization problem. The proposed algorithm can solve the trajectory optimization problem with respect to this dynamic energy consumption over time. Simulation results showed that our algorithm outperforms benchmark methods in terms of long-term uplink throughput and energy efficiency. Furthermore, it could increase the energy consumption of the UEs during the data offloading process, which reflects success in maximizing the UEs served in a post-disaster area and accomplishing the task of information collection in the post-disaster area. A straightforward extension could be to expand the simulation area to be served with more than one UAV. In such a case, each UAV would have to develop a strategy to not only maximize the objective function but also to avoid collisions with other UAVs. One of these strategies would be to keep a certain operating distance between each pair of UAVs. This distance could be designed using optical sensors attached to the UAV to recognize the surrounding UAVs, or by detecting a low-power beacon signal transmitted from each operating UAV. A detailed system design was kept for our future work. Additionally, for a more realistic scenario, UEs might be considered as moving objects, and the UAV should consider an accurate methodology for estimating the location of each UE that should be served.

Author Contributions: Conceptualization, A.A., E.M.M., and G.K.T.; methodology, A.A. and G.K.T.; software, A.A. and G.K.T.; validation, A.A. and G.K.T.; formal analysis, A.A., E.M.M., and G.K.T.; investigation, A.A. and G.K.T.; resources, A.A., E.M.M., and G.K.T.; data curation, A.A. and G.K.T.; writing—original draft preparation, A.A.; writing—review and editing, A.A., E.M.M., G.K.T., and K.S.; visualization, A.A.; supervision, E.M.M., G.K.T., and K.S.; project administration, G.K.T., and K.S.; funding acquisition, G.K.T., and K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work is partially supported by the Telecommunications Advancement Foundation, Japan.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BS	Base station
UAV	Unmanned aerial vehicle
ML	Machine learning
RL	Reinforcement learning
MAB	Multi-armed bandit
DPG	Deterministic policy gradient
MDP	Markov decision process
CRN	Cognitive radio network
UCB	Upper confidence bound
TS	Thompson sampling
RIS	Re-configurable intelligent surface
SUTOA	state-action-reward-state-action based UAV-trajectory optimization algorithm
QUTOA	Q-learning based UAV-trajectory optimization algorithm
UE	User equipment
GPS	Global positioning system
3GPP	3rd generation partnership project
LOS	Line-of-sight
NLOS	Non-line-of-sight
SNR	Signal-to-noise ratio
AWGN	Additive white Gaussian noise
EXP3	The exponential-weight algorithm for exploration and exploitation
LCB	Lower confidence bound
DEA	Dual-energy aware
PASS	Post-disaster area spiral scanning
SFP	Shortest flight path
PoI	Points of interest
MILP	Mixed Integer Linear Programming

References

1. Munich, R. Natcatservice loss events worldwide 1980–2014. *Munich Reinsur. Munich Ger.* **2015**, *10*.
2. Deepak, G.; Ladas, A.; Sambo, Y.A.; Pervaiz, H.; Politis, C.; Imran, M.A. An overview of post-disaster emergency communication systems in the future networks. *IEEE Wirel. Commun.* **2019**, *26*, 132–139.
3. Tran, G.K.; Ozasa, M.; Nakazato, J. NFV/SDN as an Enabler for Dynamic Placement Method of mmWave Embedded UAV Access Base Stations. *Network* **2022**, *2*, 479–499.
4. Saad, W.; Bennis, M.; Mozaffari, M.; Lin, X. *Wireless Communications and Networking for Unmanned Aerial Vehicles*; Cambridge University Press: Cambridge, UK, 2020.

5. Erdelj, M.; Natalizio, E.; Chowdhury, K.R.; Akyildiz, I.F. Help from the sky: Leveraging UAVs for disaster management. *IEEE Pervasive Comput.* **2017**, *16*, 24–32.
6. Kwasinski, A.; Weaver, W.W.; Chapman, P.L.; Krein, P.T. Telecommunications power plant damage assessment for hurricane katrina–site survey and follow-up results. *IEEE Syst. J.* **2009**, *3*, 277–287.
7. Fotouhi, A.; Qiang, H.; Ding, M.; Hassan, M.; Giordano, L.G.; Garcia-Rodriguez, A.; Yuan, J. Survey on UAV cellular communications: Practical aspects, standardization advancements, regulation, and security challenges. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3417–3442.
8. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
9. Authority, F.A. FAA Aerospace Forecast: Fiscal Years 2019–2039; U.S. Department of Transportation: Washington, DC, USA, 2019. Available online: https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/fy2019-39_faa_aerospace_forecast.pdf (accessed on 26 January 2023)
10. Hashesh, A.O.; Hashima, S.; Zaki, R.M.; Fouda, M.M.; Hatano, K.; Eldien, A.S.T. AI-Enabled UAV Communications: Challenges and Future Directions. *IEEE Access* **2022**, *10*, 92048–92066.
11. Chen, X.; Nie, Y.; Li, N. Online Residential Demand Response via Contextual Multi-Armed Bandits. *IEEE Control. Syst. Lett.* **2021**, *5*, 433–438. https://doi.org/10.1109/LCSYS.2020.3003190.
12. Katehakis, M.N.; Veinott Jr, A.F. The multi-armed bandit problem: Decomposition and computation. *Math. Oper. Res.* **1987**, *12*, 262–268.
13. Bubeck, S.; Cesa-Bianchi, N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv* **2012**, arXiv:1204.5721. https://doi.org/arxiv.org/abs/1204.5721v2.
14. Auer, P.; Cesa-Bianchi, N.; Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* **2002**, *47*, 235–256. https://doi.org/10.1023/A:1013689704352.
15. Audibert, J.Y.; Munos, R.; Szepesvári, C. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theor. Comput. Sci.* **2009**, *410*, 1876–1902. https://doi.org/10.1016/j.tcs.2009.01.016.
16. Lahmeri, M.A.; Kishk, M.A.; Alouini, M.S. Artificial intelligence for UAV-enabled wireless networks: A survey. *IEEE Open J. Commun. Soc.* **2021**, *2*, 1015–1040.
17. Mamaghani, M.T.; Hong, Y. Intelligent trajectory design for secure full-duplex MIMO-UAV relaying against active eavesdroppers: A model-free reinforcement learning approach. *IEEE Access* **2020**, *9*, 4447–4465.
18. Han, S.I. Survey on UAV Deployment and Trajectory in Wireless Communication Networks: Applications and Challenges. *Information* **2022**, *13*, 389.
19. Zeng, Y.; Xu, X.; Zhang, R. Trajectory design for completion time minimization in UAV-enabled multicasting. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 2233–2246.
20. Sugihara, R.; Gupta, R.K. Speed control and scheduling of data mules in sensor networks. *ACM Trans. Sens. Netw. (TOSN)* **2010**, *7*, 1–29.
21. Chiaraviglio, L.; D'Andreagiovanni, F.; Liu, W.; Gutierrez, J.A.; Blefari-Melazzi, N.; Choo, K.K.R.; Alouini, M.S. Multi-area throughput and energy optimization of UAV-aided cellular networks powered by solar panels and grid. *IEEE Trans. Mob. Comput.* **2020**, *20*, 2427–2444.
22. Trotta, A.; Andreagiovanni, F.D.; Di Felice, M.; Natalizio, E.; Chowdhury, K.R. When UAVs ride a bus: Towards energy-efficient city-scale video surveillance. In Proceedings of the IEEE Infocom 2018-IEEE Conference on Computer Communications, Honolulu, HI, USA, 16–19 April 2018; pp. 1043–1051.
23. Chen, M.; Challita, U.; Saad, W.; Yin, C.; Debbah, M. Artificial neural networks-based machine learning for wireless networks: A tutorial. *IEEE Commun. Surv. Tutors* **2019**, *21*, 3039–3071.
24. Mozaffari, M.; Saad, W.; Bennis, M.; Debbah, M. Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage. *IEEE Commun. Lett.* **2016**, *20*, 1647–1650.
25. Pearre, B.; Brown, T.X. Model-free trajectory optimisation for unmanned aircraft serving as data ferries for widespread sensors. *Remote Sens.* **2012**, *4*, 2971–3005.
26. Bayerlein, H.; De Kerret, P.; Gesbert, D. Trajectory optimization for autonomous flying base station via reinforcement learning. In Proceedings of the 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Kalamata, Greece, 25–28 June 2018; pp. 1–5.
27. Yin, S.; Zhao, S.; Zhao, Y.; Yu, F.R. Intelligent trajectory design in UAV-aided communications with reinforcement learning. *IEEE Trans. Veh. Technol.* **2019**, *68*, 8227–8231.
28. Amrallah, A.; Mohamed, E.M.; Tran, G.K.; Sakaguchi, K. Enhanced dynamic spectrum access in UAV wireless networks for post-disaster area surveillance system: A multi-player multi-armed bandit approach. *Sensors* **2021**, *21*, 7855.
29. Amrallah, A.; Mohamed, E.M.; Tran, G.K.; Sakaguchi, K. Radio Resource Management Aided Multi-Armed Bandits for Disaster Surveillance System. In Proceedings of the Proc. 2020 International Conference on Emerging Technologies for Communications (ICETC2020), Virtual, 2–4 December 2020.
30. Mohamed, E.M.; Hashima, S.; Aldosary, A.; Hatano, K.; Abdelghany, M.A. Gateway selection in millimeter wave UAV wireless networks using multi-player multi-armed bandit. *Sensors* **2020**, *20*, 3947.
31. Mohamed, E.M.; Hashima, S.; Hatano, K. Energy Aware Multi-Armed Bandit for Millimeter Wave Based UAV Mounted RIS Networks. *IEEE Wirel. Commun. Lett.* **2022**, *11*, 1293–1297.

32. Lin, Y.; Wang, T.; Wang, S. UAV-assisted emergency communications: An extended multi-armed bandit perspective. *IEEE Commun. Lett.* **2019**, *23*, 938–941.
33. Cui, J.; Ding, Z.; Deng, Y.; Nallanathan, A.; Hanzo, L. Adaptive UAV-trajectory optimization under quality of service constraints: A model-free solution. *IEEE Access* **2020**, *8*, 112253–112265.
34. Zhang, T.; Lei, J.; Liu, Y.; Feng, C.; Nallanathan, A. Trajectory optimization for UAV emergency communication with limited user equipment energy: A safe-DQN approach. *IEEE Trans. Green Commun. Netw.* **2021**, *5*, 1236–1247.
35. Amrallah, A.; Mohamed, E.M.; Tran, G.K.; Sakaguchi, K. Dual Energy-Aware based Trajectory Optimization for UAV Emergency Wireless Communication Network: A Multi-armed Bandit Approach. In Proceedings of the 2022 Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN), Barcelona, Spain, 5–8 July 2022; pp. 43–48.
36. 3GPP. Study on Enhanced LTE Support for Aerial Vehicles (Release 15), 2017. Available online: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3231 (accessed on 26 January 2023)
37. González-Cañete, F.J.; Casilari, E. Consumption analysis of smartphone based fall detection systems with multiple external wireless sensors. *Sensors* **2020**, *20*, 622.
38. Vermorel, J.; Mohri, M. Multi-armed bandit algorithms and empirical evaluation. In *European Conference on Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 437–448.
39. Agrawal, R. Sample mean based index policies by o (log n) regret for the multi-armed bandit problem. *Adv. Appl. Probab.* **1995**, *27*, 1054–1078.
40. Scott, S.L. A modern Bayesian look at the multi-armed bandit. *Appl. Stoch. Model. Bus. Ind.* **2010**, *26*, 639–658.
41. Auer, P.; Cesa-Bianchi, N.; Freund, Y.; Schapire, R.E. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.* **2002**, *32*, 48–77.
42. Ding, W.; Qin, T.; Zhang, X.D.; Liu, T.Y. Multi-armed bandit with budget constraint and variable costs. In Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, Bellevue, WA, USA, 14–18 July 2013.
43. Sinha, D.; Sankararaman, K.A.; Kazerouni, A.; Avadhanula, V. Multi-armed bandits with cost subsidy. In Proceedings of the International Conference on Artificial Intelligence and Statistics, PMLR, Virtual Event, 13–15 April 2021, pp. 3016–3024.