




Article

# Enhanced Dynamic Spectrum Access in UAV Wireless Networks for Post-Disaster Area Surveillance System: A Multi-Player Multi-Armed Bandit Approach

Amr Amrallah <sup>1,\*</sup>, Ehab Mahmoud Mohamed <sup>2,3</sup>, Gia Khanh Tran <sup>1</sup> and Kei Sakaguchi <sup>1</sup>

<sup>1</sup> Department of Electrical and Electronic Engineering, School of Engineering, Tokyo Institute of Technology, Meguro, Tokyo 152-8550, Japan; khanhtg@mobile.ee.titech.ac.jp (G.K.T.); sakaguchi@mobile.ee.titech.ac.jp (K.S.)

<sup>2</sup> Electrical Engineering Department, College of Engineering, Prince Sattam Bin Abdulaziz University, Wadi Addwasir 11991, Saudi Arabia; ehab\_mahmoud@aswu.edu.eg

<sup>3</sup> Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan 81542, Egypt

\* Correspondence: amrallah@mobile.ee.titech.ac.jp

**Abstract:** Modern wireless networks are notorious for being very dense, uncoordinated, and selfish, especially with greedy user needs. This leads to a critical scarcity problem in spectrum resources. The Dynamic Spectrum Access system (DSA) is considered a promising solution for this scarcity problem. With the aid of Unmanned Aerial Vehicles (UAVs), a post-disaster surveillance system is implemented using Cognitive Radio Network (CRN). UAVs are distributed in the disaster area to capture live images of the damaged area and send them to the disaster management center. CRN enables UAVs to utilize a portion of the spectrum of the Electronic Toll Collection (ETC) gates operating in the same area. In this paper, a joint transmission power selection, data-rate maximization, and interference mitigation problem is addressed. Considering all these conflicting parameters, this problem is investigated as a budget-constrained multi-player multi-armed bandit (MAB) problem. The whole process is done in a decentralized manner, where no information is exchanged between UAVs. To achieve this, two power-budget-aware PBA-MAB) algorithms, namely upper confidence bound (PBA-UCB (MAB) algorithm and Thompson sampling (PBA-TS) algorithm, were proposed to realize the selection of the transmission power value efficiently. The proposed PBA-MAB algorithms show outstanding performance over random power value selection in terms of achievable data rate.

**Keywords:** unmanned aerial vehicles; dynamic spectrum access; quality of service; reinforcement learning; multi-armed bandit



**Citation:** Amrallah, A.; Mohamed, E.M.; Tran, G.K.; Sakaguchi, K. Enhanced Dynamic Spectrum Access in UAV Wireless Networks for Post-Disaster Area Surveillance System. *Sensors* **2021**, *21*, 7855. <https://doi.org/10.3390/s21237855>

Academic Editor: Margot Deruyck

Received: 5 November 2021

Accepted: 23 November 2021

Published: 25 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The fast development of UAVs, which are commonly known as drones, has received much attention in various domains [1,2]. Recently, UAVs have been leveraged for future civil applications although their usage was restricted to military applications only during the last few years. This is considered a promising direction since UAVs have unique properties that can support this goal. UAVs are capable of various functions as they are able to fly, are maneuverable, and are easy to deploy. Hence, UAVs can handle different tasks as delivery services, traffic monitoring, aerial photography, disaster management, rescue operations, and wireless communications [1,2]. In recent years, major disasters have occurred around the world such as the great Tohoku earthquake and tsunami, which hit Japan in 2011; Hurricane Sandy on the northeastern coast of the USA in 2012; the Nepal earthquake in 2015, the massive explosion in the port of Beirut, Lebanon, in 2020; and the global wildfires in North America and Europe in 2021. All these natural disasters caused terrible damage to infrastructure and loss of human lives. The first few hours after the disaster are considered the golden relief time to provide support and emergency aid to save these precious lives. Therefore, this paper focuses on wireless communications applications

for UAVs to support a post-disaster area surveillance system. Specifically, UAVs can fly over the post-disaster areas to collect live photos of the current situation and send this collected information to a disaster management center to be analyzed. This will enable rescue teams to get information promptly about the actual situation in the affected area, which will enhance their response time [3].

On the other hand, and due to the persistent increase in demand for mobile services, spectrum resources are becoming more and more scarce [4]. Therefore, it is expected that future mobile networks will host a modern communications technology that supports unsurpassed networking architecture and energy-efficient devices. To realize these novel concepts, new fundamental challenges have appeared on the surface. Unlike wired communications systems, due to the national spectrum regulations and the hardware limitation, the wireless world has limited links to distribute. Consequently, it will be mandatory for the traditional regulation of the spectrum to have a fundamental reform so that it can allow more efficient use of spectrum resources. Spectrum inefficiency has become a major concern; hence it is imperative to search for an effective solution to deal with the resource allocation problems from the spectrum and power-efficiency points of view. This solution should achieve three main goals. Firstly, it should be amenable to the distributed implementation. Secondly, it should be capable of dealing with the uncertainty caused by the lack of information. Thirdly, it should deal with users' selfishness. One of the most promising solutions is the DSA system [5], which can be implemented as a CRN [6]. A DSA system has the ability to enhance the spectrum utilization efficiency [7]. Hence, CRN allows unlicensed Secondary Users (SUs) to coexist with the licensed Primary Users (PUs) in the licensed band without causing any harmful impact on PUs in terms of different Quality of Services (QoS) aspects. In other words, SUs can utilize a portion of the licensed PUs spectrum under certain QoS constraints [6]. Therefore, to enhance the network efficiency, SUs' spectrum utilization should be maximized while keeping an eye on the QoS level of the high-priority traffic, i.e., the PUs traffic, to avoid any services interruption to the highly prioritized data transmission.

The concept of this resource allocation issue is considered a challenging problem for two reasons. First, the resource allocation process can be made with a large number of orthogonal communication dimensions such as time, frequency, code, space, and antenna direction [8]. Second, in order to enhance the spectrum utilization, QoS for both PUs and SUs should be maximized. To achieve this, there are different conflicting parameters that need to be jointly optimized as transmitted power, channel occupation, total throughput, and mutual interference level between simultaneous users. Therefore, for a certain number of PUs and SUs, there are indispensable targets for the optimization algorithm such as the interference threshold for each PU, the channel state information, and the geographical location for both of PUs and SUs. Moreover, this optimization scenario can be decentralized;; in other words, there is no need to deploy a fusion center to collect enough information from the environment and complete the optimization process to the end. Since energy levels are not observed in general, and both PUs and SUs form a distributed network, it can benefit from that distribution to sense the available energy at each node. From this point of view, the design of an efficient future wireless network needs to deal with the uncertainty of information besides different users' competition and selfishness. Hence, it becomes mandatory to search for a powerful mathematical tool that can deal with such unprecedented network problems.

Machine learning (ML) algorithms, more precisely reinforcement learning (RL) algorithms, are leveraged to deal with these kinds of optimization problems [9]. The reason behind selecting RL algorithms is their capability to achieve tremendous results in generalization and efficiency, leading to their capability to tackle real-life problems, and especially in field of wireless communications [9]. Furthermore, RL algorithms are able to deal with conflicting optimization parameters of the resource allocation problem for the DSA system [10]. Without prior information about the environment, an agent can learn to enhance its future actions based on its past experience. MAB algorithms are considered one of such

RL algorithms. MAB algorithms can be described as a set of actions (arms) of a bandit machine that each arm leads to a certain reward [11]. A player needs to maximize their accumulated reward over the playing epoch by choosing one arm to pull in each playing round. Moreover, this player has no idea about the reward behind each arm. So, this instantaneous reward behind each arm is revealed once the player decides to select this arm. Therefore, for this hidden setting, the player may lose some reward in each trial due to not selecting the arm that leads to the highest reward value instead of the chosen arm. This loss is denoted by regret [12]. Thus, each player should select a sequence of arms to pull to maximize their total reward over horizon, in other words, to minimize their total regret over horizon. This is a common dilemma faces MAB algorithms and it is called the exploration–exploitation trade off [13–15].

Over the last decade, with the rapid increase in the number of natural disasters occurring throughout the world, there has become an urgent need to develop a smart post-disaster surveillance system. This smart system should operate in a fully decentralized manner, i.e., without having a controlling center, to speed up collection and analysis of data for a post-disaster area to enhance the performance and reduce the response time of the rescue operations. DSA systems are considered a rich topic that was deeply investigated in the early 2000s for some quite old applications such as analog TV white spaces, especially in the Very High Frequency (VHF) and the Ultra High Frequency (UHF) bands [16]. Hence, we aimed to refurbish the well-known DSA system by exploiting the benefit of using ML algorithms as a modern optimization tool. Furthermore, UAVs, which are capable of flying and capturing high-resolution videos using attached cameras, were leveraged recently to support various applications in the civilian life. All these ideas motivated us to develop a smart and cheap post-disaster surveillance system by combining the advantages of DSA system, UAVs, and ML algorithms. In addition, this system is presented as unconventional method to solve the spectrum scarcity problem. In this way, DSA-system-aided ML algorithms can open the gate to unprecedented applications in the field of UAVs wireless communication networks.

In this paper, we aim to design and evaluate a spectrum allocation for a DSA system using MAB algorithms to support a post-disaster surveillance system. From a MAB perspective, UAVs, which are considered SU transmitters, will act as the player who aims to maximize their long-term reward, i.e., data rate. Furthermore, this player is constrained by a limited power budget. On the other hand, different transmitting power levels will act as arms of the bandit machine. The MAB algorithm is considered the most suitable algorithm for our optimization problem as it can deal with online optimization problems without any prior information about the environment except the player's observations of the achieved reward while playing. Our paper adapts two different MAB algorithms, the Upper Confidence Bound (UCB) [15] and Thompson Sampling (TS) [17], to address such an optimization problem. In this paper, a modified version of MAB algorithms is proposed to treat our optimization problem. This is called the Power-Budget-Aware PBA-MAB (MAB) algorithm. The key idea behind the PBA-MAB algorithm is to include the available power budget for each UAV in the decision-making process when choosing the most appropriate transmitting power value.

From the point of view of the DSA system, the SU network, which consists of UAVs and temporary base stations, shared the spectrum resources as a CRN with the PU network, which consists of highway Electronic Toll Collection (ETC) gates and cars passing these ETC gates, under certain QoS constrains. Hence, SU transmitters are allowed to send their data without causing a harmful interference to the most precious data of the PU network. It should be mentioned that our design allows both the PU network and the SU network to coexist at the same time under a certain signal-to-interference-plus-noise ratio (SINR) threshold. Furthermore, we need to utilize the multi-objective formulation. Given the location of each PU and the power budget of each SU, we seek to design for a joint optimization problem considering different conflicting objects such as interference coordination, sum-rate maximization, and total number of active SUs in the network,

subject to QoS constraints for both PUs and SUs. Despite the adversarial problem definition and the selfish behavior of each UAV toward achieving its maximum data rate, modified MAB algorithms learn how to select the most suitable action over time to enhance the overall system performance as discussed in [18–20] and illustrated in our paper. The main contributions of this paper can be summarized in the following points:

- The selection of the transmitted power value for UAVs aiding a post-disaster area surveillance system is formulated as an optimization problem aiming to maximize the achievable data rate while considering the limited available power budget for each UAV. This is done in a decentralized manner as there is no exchange of information among UAVs.
- Integrating the post-disaster surveillance system as a CRN is considered an unconventional solution for the spectrum scarcity problem. Furthermore, it can reduce the overhead cost of renting dedicated frequency channels for post-disaster surveillance operations, while they are rarely used just when a disaster occurs.
- Despite the nature of original MAB algorithms to maximize the long-term reward, i.e., the achieved data rate, MAB algorithms are modified to take into account the limited power budget for transmission. Therefore, the selection of the transmitted power not only aims to maximize the data rate for the current channel but also considers the remaining power budget to maximize the data rate for the next available channel.

The rest of the paper is organized as follows. Section 2 overviews the related work. Section 3 introduces the system model and the power value selection optimization problem. Section 4 introduces proposed PBA-MAB algorithms and how these algorithms can deal with this kind of optimization problem. Section 5 gives simulation and analysis of the proposed optimization scenario. Finally, we summarize the result and point out the future research in Section 6.

## 2. Related Works

Since the early 21st century, the idea of DSA gained increasing attention, especially in the US and Europe, due to the spectrum congestion [21]. An overview of the major technical and regularity issues of DSA systems was presented in [21]. The authors of [22] introduced the concept of multi-dimensional spectrum sensing and discussed the challenges associated with it. They developed prediction algorithms based on the past multi-dimensional spectrum utilization information to predict the future usage of the spectrum. With the aid of the DSA system, CRN can be established to support different applications as public safety, smart grid, broadband cellular, and medical applications. Ref. [23] discussed some challenges that faced the practical application toward this idea. An overview of CRN design layers, such as the physical layer (PHY), the medium-access control layer (MAC), and the network layer, is presented in [24]. Furthermore, the authors showed how these layers can interact with each other. The authors of [25] investigated the throughput improvements in a CRN using different channel selection techniques such as frequency hopping, frequency tracking, and frequency coding. Ref. [26] investigated the CRN formed by the incorporating radio capabilities of a Wireless Sensor Network (WSN). It addressed both advantages and limitations of CRN for WSN in conjunction with the existing applications and techniques. A continuous-time Markov chain model is implemented in [27] for a DSA system in an open spectrum wireless network. The authors of [28] examined how CRN devices can find an available spectrum channel under different system capabilities, spectrum policies, and environmental conditions. They defined this problem as a “rendezvous” problem. With the aid of RL algorithms, the authors of [29] proposed a framework for Internet of Things (IoT) devices to capture and model the traffic behavior of short-time spectrum occupancy in order to determine the existing interference in the shared bands. In [30], a novel information and energy cooperation method were introduced for cognitive Heterogeneous Networks (HetNets). This method aimed to enhance energy efficiency by solving an energy efficiency maximization problem with respect to joint time allocation and power control. The authors of [31] proposed an enhanced fusion center rule

for soft decision cooperative spectrum sensing using energy detection to mitigate the noise uncertainty effect and to enhance the sensing performance of CRNs.

In recent years, there have been research efforts for using UAVs to support post-disaster area applications. In [32], the authors used UAVs with conjunction with cellular network and WSN to aid disaster management applications. A genetic algorithm was used in [33] for UAV location optimization to enhance the overall coverage and data rate of the wireless network. The authors of [34] proposed an effective method to support rescue operations in locating victims of a natural disaster. This was done with the aid of lidar and infrared depth cameras attached to UAVs to build a detecting system independent of the illumination intensity. A video recorder and a geolocation module attached to UAV were used in [35] to search for survivors in a post-disaster area. In [36], the authors examined flying communication services using Wi-Fi, video camera, and web servers attached to UAVs. They aimed to enable affected users after a disaster to use their smartphones for texting and video communication in real-time. The authors of [37] proposed a mobility model based on self-deployment of an aerial ad hoc network based on the Jaccard dissimilarity metric for a post-disaster area. The software simulation integrates the mobility of victims and generate a corresponding UAVs mobility model to trace those victims. In [38], authors proposed an energy efficient task scheduling for the collected data by UAVs from ground IoT network to support a disaster management system.

In [39], UAVs were used as on-demand airborne relays to connect remote users with a cellular BS when they were separated by vast obstacles. Furthermore, UAVs can be used in WSNs to distribute and collect information in both of Control Plane (CP) and Data Plane (DP) from wireless sensors deployed on the ground level [40,41]. UAVs are being used to assist the management and control of Vehicle Ad hoc NETWORKS (VANETs) and extend its coverage [42]. All the above existing research works assume a full awareness of the network parameters, which is not the case of our paper, where there is no information change among UAVs while trying to maximize the achievable data rate, as the network is fully decentralized.

On the other hand, RL algorithms have become a promising optimization technique for solving chronic UAV problems that have occurred as a result of integrating UAVs in wireless communication applications. RL algorithms are well known for their capability to achieve near optimal results in generalization and efficiency. Therefore, they are used to tackle real-time problems in the field of wireless communications. Detailed discussion about different MAB algorithms can be found in [43,44]. It has been shown in several works that MAB algorithms can be adapted to tackle such problems related to DSA systems. The authors of [45] proposed MAB learning algorithms for CRN, and particularly for spectrum sensing in a DSA system in licensed bands [7]. Different MAB algorithms, such as UCB and TS, have been used to improve the spectrum access in unlicensed Wi-Fi networks [45,46]. The authors of [47] considered a set of policies for multiple-user-independent and identical distributed (iid) and rested MAB problems with the assumption that each SU declares its action to others, e.g., the selected channel, which is considered a strong constraint. A disputed learning and spectrum access policy for iid rewards is discussed in [48], and it was proven that this policy has a logarithmic order regret. In [49], the decentralized learning for DSA system with multiple SUs spectrum access has been studied. The authors of [50] proposed a modified MAB algorithm to solve the gateway selection in UAV wireless network for post-disaster area applications. These algorithms are considering the battery life while searching for the most suitable gateway UAV to maximize the total system throughput. A dynamic wireless channel selection based on the MAB algorithm with laser chaos time sequence is proposed in [51]. The adaptive channel selection achieved a higher throughput using four channels Wireless Local Area Network (WLAN) based on IEEE802.11a system. The authors of [52] proposed a simple and powerful tug-of-war MAB algorithm. Since this algorithm is very simple, it can be applied in wireless network selection for devices with small processing capabilities as IoT devices and smartphones. Ref. [53] studied the millimeter-wave (mmWave) two-hop relaying as a single-player MAB

problem in order to enable one relay probing while maximizing the achievable spectral efficiency. This was done by using modified versions of MAB algorithms. The authors of [54] studied the problem of joint neighbor discovery and selection in mmWave device to device (D2D) networks using a stochastic budget-constraint MAB algorithm.

### 3. System Model

This section discusses the network architecture of the post-disaster area surveillance system using UAVs and the used channel model for transmitting the collected data.

#### 3.1. Post-Disaster Area Surveillance System Architecture

Figure 1 shows a simplified version of the system architecture of the UAV wireless network in a metropolitan post-disaster area. Since the first few hours after the occurrence of the natural disaster (such as flood or earthquake) are considered the golden relief time to save human lives, as discussed in the introduction section, UAVs should collect pivotal information about victims in the damaged area using an attached high-definition camera. The collected data can be further analyzed by the disaster management center to identify victim's exact location, number, age, gender, and injury status. On the other hand, temporary base stations are deployed in the disaster area to collect this information from surveillance UAVs and send them to the disaster management center to aid rescue teams. These temporary base stations are used as charging stations for UAVs. Furthermore, they are considered the starting flying points. UAVs fly over the disaster area to capture live photos of certain points at the damaged area. The way in which these temporary base stations transmit the collected data to the disaster management center, and the method for selecting surveillance points, are outside the scope of this paper. Moreover, we assumed in this paper that the different locations in the affected area have the same weight of importance, so these points were chosen on random bases.



**Figure 1.** UAV surveillance-system-assisted DSA for a metropolitan post-disaster area.

On the other hand, our system aims to build this surveillance system using CRN. Therefore, the SU network, which is represented by UAVs and temporary base stations, will utilize the same frequency band of the PU network. The PU network is represented by ETC gates and bypassing vehicles in a nearby highway. In this way, we aimed to reduce the cost of reserving dedicated channels for surveillance system while it is being used during the time of natural disasters only. Each UAV collects and sends data to its

corresponding temporary base station. Furthermore, each UAV should not deal harmful interference to the transmitted data between ETC gates and vehicles on the nearby highway. It should be mentioned that our optimization problem design is considered a soft-spectrum allocation. The difference between conventional spectrum allocation that have been studied in [55,56] and our optimization problem is that the conventional optimization problem treats the spectrum allocation as a hard allocation problem; i.e., no two users (PU and SU) can share same channels at the same time. However, our design introduces other orthogonal dimensions of the threshold to enable more than one user to coexist at the same frequency band if their QoS constraints are not violated. Furthermore, for the sake of generalization, we supposed that all PU channels, which connect every ETC gate and nearby vehicles which are passing this ETC gate, are always active and occupied with the PU network traffic. In this way, we considered the worst-case design scenario in which the QoS constraints should be carefully verified during the optimization process.

### 3.2. Problem Formulation

In the following, our design employs the physical model proposed in [57], which provides a path-loss model to realize the communication environment. It is assumed UAVs can communicate to temporary base station via air-to-air wireless communication link. Basically, this type of link can be called a Line of Sight (LoS) wireless communication link. Since the design is built using CRN, which shares the spectrum between PU network and SU network, this shared frequency band is split into  $Q$  independent sub-bands, and each sub-band has a bandwidth  $W$  in Hertz. Each primary and secondary transmitter receiver pair, referred to as primary and secondary users, is numbered by indices  $\psi \in \Psi = \{PU_1, \dots, PU_\Psi\}$  and  $\omega \in \Omega = \{SU_1, \dots, SU_\Omega\}$ . Hence, at any time  $r$ , the general path-loss formula between any transmitter  $\alpha$  and any receiver  $\beta$  can be expressed by:

$$L_{\alpha\beta,q}(r) = \frac{G_{Tx,\alpha}G_{Rx,\beta}}{d_{\alpha\beta}^\zeta} \left( \frac{c}{4\pi f_q(r)} \right)^2 \quad (1)$$

where  $G_{Tx,\alpha}$  and  $G_{Rx,\beta}$  are the transmit and receive antenna gains, respectively,  $d_{\alpha\beta}$  is the distance between  $\alpha$  transmitter and  $\beta$  receiver,  $c$  is the speed of light,  $f_q(r)$  is the carrier frequency of sub-band  $q$ , and  $\zeta$  is the attenuation constant for the LoS wireless communication link. For the current design, it is assumed that the pass loss is the dominant loss factor for the received power. Hence, the effect of multi-path fading and shadowing is ignored. Furthermore, we assumed the transmitted signal is affected by an Additive White Gaussian Noise (AWGN) channel with zero mean and  $N_0$  variance. Therefore, the SINR of SU  $\omega$  in carrier  $q$  at time  $r$  can be given by:

$$\gamma_{\omega,q}(r) = \frac{p_{\omega,q}(r)L_{\omega\omega,q}(r)}{N_0 + \sum_{\lambda \in \Psi \cup \Phi, \lambda \neq \omega} p_{\lambda,q}(r)L_{\lambda\omega,q}(r)} \quad (2)$$

where  $p_{\omega,q}(r)$  and  $p_{\lambda,q}(r)$  denote the transmitted power of the  $\omega$ -th SU and the  $\lambda$ -th PU or SU, respectively. For a successful established communication link, the SINR should satisfy a condition that the achievable SINR must be greater than the threshold SINR, which is given by  $\gamma_{\omega,q}(r) > \gamma_{\omega TH,q}(r)$ . Under these assumptions, the achievable data rate can be calculated by:

$$R_{\omega,q}(r) = \begin{cases} W \sum_{q=1}^Q \log_2(1 + \gamma_{\omega,q}(r)), & \text{if } \gamma_{\omega,q}(r) > \gamma_{\omega TH,q}(r) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where  $W$  is the bandwidth of the communication channel.

Since the data rate is measured from the receiver side, we assumed this value is reported to the SU transmitter through a feedback channel. The concept behind this

assumption comes from how modern communication systems are supposed to offer high flexibility in different ways. One of these ways is to split user and control planes to support software defined networking applications to allow flexible placement of processing function between different network nodes [58]. For PUs, it is assumed that they operate in a narrowband network, which means a pre-determined power value is assigned to each PU. This design criterion is suitable when licensed users have to operate on narrowband channels. On the other hand, for a wideband PU network, straightforward extension can be done without affecting this methodology. Since SUs need to utilize these multiband channels, where each sub-band is previously assigned to a certain PU, each SU has a power budget denoted by  $P_{\max}$ . Whereas it is assumed that our PUs and SUs networks use omnidirectional antennas, the communication channel can be established according to (1) with considering antennas gain  $G_{Tx,\alpha} = G_{Rx,\beta} = 1, \forall \alpha, \beta$ . Furthermore, it is assumed that each PU transmits using only single sub-band, and PUs operate in disjoint sub-bands. As a result, we have the number of PUs equal to number of channels and hence  $\Psi = Q$ . The main target of the optimization algorithm is to maximize the sum-rate, the total throughput, for the SUs network. This can be achieved by optimizing the power levels allocated for each SU within each shared traffic channel. The power allocation vector can be defined as  $\mathbf{p}_\omega = [p_{\omega,1}, \dots, p_{\omega,Q}]^T$ , where each element represents the power value for SU  $\omega$  for each sub band  $q$ . In case that a SU has a power vector equal to zero, it means that this SU is inactive. On the other hand, for PUs, it is allowed for a single PU to transmit only on a single sub-band so that they are operating in disjoint sub-bands. Moreover, during data transmission of SUs, they should avoid causing any harmful interference to the high priority traffic that belongs to PUs network. It is mandatory for each SU to satisfy this condition and not exceed its allowed power budget during transmission as well. Considering all these power budget limitations and interference constraints, the sum-rate maximization problem can be formulated as:

$$\begin{aligned} \max \quad & \frac{1}{\mathcal{R}} \sum_r \sum_\omega \sum_q R_{\omega,q}(r) \\ \text{s.t.} \quad & \gamma_{\psi,q}(r) > \gamma_{\psi\text{TH},q}(r) \\ & \gamma_{\omega,q}(r) > \gamma_{\omega\text{TH},q}(r) \end{aligned} \quad (4)$$

where  $\mathcal{R}$  is the total time spent for data transmission,  $r = 1, \dots, \mathcal{R}$ , and  $\gamma_{\psi,q}(r) > \gamma_{\psi\text{TH},q}(r)$ ,  $\gamma_{\omega,q}(r) > \gamma_{\omega\text{TH},q}(r)$  are the SINR constraint conditions for all PUs and all SUs, respectively. Thus, for SUs, it is mandatory to satisfy both SINR conditions to utilize a sub-band channel from PUs channels.

Since our network is designed in a decentralized way with no information exchange between different network elements, the only information available to UAVs are the location, the channel frequency and the transmission power of each ETC gate system. Therefore, we have developed a method to let UAVs estimate the interference caused by self-transmission and calculate the corresponding SINR value for each PU's receiver. With the aid of Equations (1) and (2), each UAV will calculate the expected SINR value at each ETC gate under the interference effect of its own data transmission. Then, each UAV can check individually for the satisfaction of SINR conditions for both the PU network and the SU network. In such a way, there is no need to deploy a fusion center to share the SINR information between different SU network nodes, and therefore the network can be implemented in a decentralized way.

#### 4. Proposed Power Budget Aware MAB Algorithm

This section discusses two proposed algorithms to tackle this rate maximization problem. These algorithms are called Power Budget Aware Upper Confidence Bound (PBA-UCB) and the Power Budget Aware Thomson Sampling (PBA-TS).



#### 4.1. Proposed PBA-UCB Algorithm

UCB is considered one of the efficient MAB algorithms that can achieve balancing for the exploration-exploitation dilemma of the MAB algorithm. UCB enhances the confidence of the arm selection by decreasing the uncertainty behind the reward that will be revealed. Algorithm 1 illustrates a modified version of the UCB algorithm, which is called the PBA-UCB algorithm. This algorithm is applied to each UAV to select the most suitable transmission power in a selfish way to maximize the system rate. It is assumed that each UAV has information about the location of surrounding ETC gates operating in the surveillance area. Furthermore, they know the transmitting frequency for each ETC gate. The method of how UAVs can detect the location and the operating frequency of each ETC gate is behind the scope of this paper. Hence, each UAV tries to maximize its own data rate while competing with other UAVs to increase its transmission power while keeping an eye on the SINR threshold. At the beginning, i.e., the first  $\mathcal{N}$  rounds, PBA-UCB algorithm, which is enabled on each UAV, tests the data rate that can be achieved by transmitting on all available channels with random transmission power and observes the achievable data rate. Afterwards, for the remaining rounds,  $\mathcal{N} + 1 \leq r \leq \mathcal{R}$ , the PBA-UCB algorithm picks a power value in a way that satisfies:

$$p_{\omega,q}^*(r) = \arg \max_{p_{\omega,q} \in \mathbf{p}_{\omega}} \left( \hat{\mu}_{\omega,q}(r-1) + \sqrt{\frac{\eta \ln(r)}{T_{\omega,q}^{(p)}(r-1)}} - \frac{p_{\omega,q}}{p_{\omega,M}} \right) \quad (5)$$

where  $p_{\omega,q} \in \mathbf{p}_{\omega}$  is the average reward obtained for transmission power value  $p$  in channel  $q$  up to the last previous round ( $r-1$ ),  $\hat{\mu}_{\omega,q}(r-1)$  is the average achievable data rate to the last previous round ( $r-1$ ) using transmission power value  $p$  in channel  $q$ , and it can be calculated as:

$$\hat{\mu}_{\omega,q}(r-1) = \frac{1}{T_{\omega,q}^{(p)}(r-1)} \sum_{m=1}^{T_{\omega,q}^{(p)}(r-1)} R_{\omega,q}(m) \quad (6)$$

where  $R_{\omega,q}(m)$  is the achievable data rate, which can be obtained from Equation (3).  $T_{\omega,q}^{(p)}(r-1)$  is a count of the number of selections of this transmitting power value until the last previous round ( $r-1$ ).  $p_{\omega,q}$  is the selected power value for transmission and  $p_{\omega,M}$  is the total available power budget for UAV that can be used. This equation illustrates how PBA-UCB works. If a transmission power value is selected many times, which makes  $T_{\omega,q}^{(p)}(r-1)$  become large, the confidence bond term  $\sqrt{\frac{\eta \ln(r)}{T_{\omega,q}^{(p)}(r-1)}}$  decreases, and that causes the UAV to seek to explore other power values that are less selected in the previous rounds. On the other hand, when a transmission power value achieved a high reward, i.e., high data rate, during the past rounds, which means  $\hat{\mu}_{\omega,q}(r-1)$  becomes large, the UAV seeks to exploit this high-gain arm in order to achieve the maximum achievable reward during this round. Originally, the PBA-UCB algorithm sets parameter  $\eta$  to a positive value of 2 in most cases [13], but empirically, when it is set to  $\eta = 0.5$ , the performance is improved [12]. In that way, the PBA-UCB algorithm can solve the exploration–exploitation trade-off in an efficient way. Furthermore, the term  $\frac{p_{\omega,q}}{p_{\omega,M}}$  shows how a UAV can balance between selecting a power value to achieve a high data rate and consider for the remaining power budget to be used in transmission on next available channels. It should be mentioned that this last term defines the contribution behind our proposed PBA-UCB algorithm. Since the original UCB algorithm could achieve only balancing between exploration and exploitation, our proposed PBA-UCB algorithm enables a novel way to keep an eye on the remaining power budget while balancing between exploration and exploitation. Furthermore, when selecting a transmission power, the PBA-UCB algorithm checks for the satisfaction of both PU and SU SINR conditions. Once it is satisfied, the algorithm confirms the use of this transmission power value, starts to transmit data, and calculates the corresponding rate. Otherwise, it sets the transmission power to zero and also sets the corresponding data rate to zero. In this

way, the PBA-UCB algorithm can make sure there is no harmful interference that affects the PU data transmission. On the other hand, it also counts for the interference threshold on other SUs data transmission. Since the SINR condition is considered a critical design issue, this operation is done in both of the initialization phase and the rate maximization phase to ensure the feasibility of the proposed PBA-UCB algorithm. Algorithm 1 illustrates the proposed PBA-UCB algorithm.

---

**Algorithm 1** PBA-UCB transmission power selection
 

---

```

1: for  $\omega \leftarrow 1$  to  $\Omega$  do
2:   for  $1 \leq r \leq \mathcal{N}$  do ▷ initialization phase
3:     for  $q \leftarrow 1$  to  $Q$  do
4:       Select a random value for  $p_{\omega,q}(r)$ 
5:       if  $\gamma_{\psi,q}(r) > \gamma_{\psi,TH,q}(r)$  then
6:         if  $\gamma_{\omega,q}(r) > \gamma_{\omega,TH,q}(r)$  then
7:           Obtain  $R_{\omega,q}(r)$ 
8:            $T_{\omega,q}^{(p)}(r) \leftarrow 1$ 
9:         else
10:           $p_{\omega,q}(r) \leftarrow 0$ 
11:        end if
12:      else
13:         $p_{\omega,q}(r) \leftarrow 0$ 
14:      end if
15:    end for
16:  end for
17:  for  $r \leftarrow \mathcal{N} + 1$  to  $\mathcal{R}$  do ▷ rate maximization phase
18:    Set  $p_{\omega,M}$  max SU Tx power
19:    for  $q \leftarrow 1$  to  $Q$  do
20:       $p_{\omega,q}^*(r) = \arg \max_{p_{\omega,q} \in \mathbf{p}_{\omega}} \left( \hat{\mu}_{\omega,q}(r-1) + \sqrt{\frac{\eta \ln(r)}{T_{\omega,q}^{(p)}(r-1)}} - \frac{p_{\omega,q}}{p_{\omega,M}} \right)$ 
21:      if  $\gamma_{\psi,q}(r) > \gamma_{\psi,TH,q}(r)$  then
22:        if  $\gamma_{\omega,q}(r) > \gamma_{\omega,TH,q}(r)$  then
23:          Obtain  $R_{\omega,q}(r)$  using  $p_{\omega,q}^*(r)$ 
24:           $T_{\omega,q}^{(p^*)}(r) \leftarrow T_{\omega,q}^{(p^*)}(r-1) + 1$ 
25:           $\hat{\mu}_{\omega,q}(r) \leftarrow \frac{1}{T_{\omega,q}^{(p^*)}(r)} \sum_{m=1}^{T_{\omega,q}^{(p^*)}(r)} R_{\omega,q}(m)$ 
26:           $p_{\omega,M} \leftarrow p_{\omega,M} - p_{\omega,q}^*$ 
27:        else
28:           $p_{\omega,q}^*(r) \leftarrow 0, R_{\omega,q}(r) \leftarrow 0$ 
29:        end if
30:      else
31:         $p_{\omega,q}^*(r) \leftarrow 0, R_{\omega,q}(r) \leftarrow 0$ 
32:      end if
33:    end for
34:  end for
35: end for

```

---

#### 4.2. Proposed PBA-TS Algorithm

TS algorithm copes with the exploration–exploitation dilemma using a different method than the previously discussed UCB algorithm. Basically, the reward gained by laying with different arms using the TS algorithm is drawn from a pure Bayesian probabilistic model [59]. In the beginning, TS uses a prior distribution for the reward based on the initialization of parameters of the probabilistic model. Afterward, it tries to keep tracking of the reward posterior distribution using the observation from the environment during the learning process. Thus, it can randomly choose a suitable arm that is matched

to be optimal according to the probability model. Thus, at each round, random samples are drawn from the constructed reward's posterior distribution. TS selects an arm to play that can maximize the selected sampled value. Then, the arm's posterior distribution is updated by modifying its model parameters. This updated distribution will be used for the arm selection of the upcoming rounds. It is known that TS has a superb empirical performance and even better than the achieved performance of the UCB algorithm.

In our proposed PBA-TS algorithm, it is assumed that the reward, i.e., the achieved data rate, is affected by AWGN noise and mutual interference from other PUs and SUs occupying the same channel. Hence, the assumption of the Gaussian distribution is compatible with our problem formulation. The selection of the most suitable power value for transmission, which can maximize the achieved data rate, can be expressed as:

$$p_{\omega,q}^*(r) = \arg \max_{p_{\omega,q} \in \mathbf{p}_{\omega}} \left( \varphi_{\omega,q}(r-1) - \frac{p_{\omega,q}}{p_{\omega,M}} \right) \quad (7)$$

where  $\varphi_{\omega,q}(r-1)$  is a sample for the previously constructed posterior distribution from the achieved data rate by a UAV  $\omega$  at channel  $q$  with transmission power  $p_{\omega,q}$ . The posterior distribution is constructed from the Gaussian distribution  $\mathcal{N}(\hat{\mu}_{\omega,q}(r), \sigma^2(r))$ , where  $\hat{\mu}_{\omega,q}(r)$  and  $\sigma^2(r)$  are the mean and the variance of the distribution according to the model in [20], and they can be calculated as:

$$\hat{\mu}_{\omega,q}(r) = \frac{1}{T_{\omega,q}^{(p)}(r)} \sum_{m=1}^{T_{\omega,q}^{(p)}(r)} R_{\omega,q}(m) \quad (8)$$

$$\sigma^2(r) = \frac{1}{T_{\omega,q}^{(p)}(r) + 1} \quad (9)$$

where  $R_{\omega,q}(m)$  is the achievable data rate and can be obtained from Equation (3),  $T_{\omega,q}^{(p)}(r)$  is the counted number of selections of this transmitting power value until the last previous round ( $r-1$ ), and  $R_{\omega,q}(m)$  is the achieved data rate. The term  $\frac{p_{\omega,q}}{p_{\omega,M}}$  is deduced from the distribution to balance between the rate maximization process and the remaining power budget that should be used to transmit data over the next channels. At each round  $r$ , a sample  $\varphi_{\omega,q}(r-1)$  is taken from the previously constructed Gaussian distribution. Then, the optimum power value  $p_{\omega,q}^*$  that maximizes Equation (7) will be selected for transmission. After that, UAV  $\omega$  starts to transmit over a channel  $q$  using  $p_{\omega,q}^*$ , its corresponding number of selections  $T_{\omega,q}^{(p^*)}(r)$  is updated, and the achievable data rate  $R_{\omega,q}(r)$  is observed to construct the Gaussian distribution for the next round  $r+1$ . This process is conducted till the last round  $\mathcal{R}$ . Furthermore, along with the PBA-UCB algorithm, the SINR conditions of both of PU and SU networks are examined at each time when choosing a certain power value for data transmission. If both SINR conditions are satisfied, the PBA-TS algorithm starts to use this transmission power value and counts the corresponding data rate. Otherwise, the PBA-TS algorithm sets the transmission power to zero, which leads to zero achievable data rate. The whole process of the proposed PBA-TS algorithm is summarized in Algorithm 2.

#### 4.3. Complexity Analysis of the Proposed Algorithms

In this paper, we spotlight the task of UAVs to build a post-disaster surveillance system as a CRN by finding the optimal policy for each UAV. In Algorithms 1 and 2, learning processes can find the optimal transmission power value for both PBA-UCB and PBA-TS by examining various transmission power values over every channel for all UAVs using different policies. On the other hand, it tries to keep the interference level under certain thresholds. Let  $\Xi$  represent the total number of available arms, i.e., total elements of the power vector  $\mathbf{p}$ . It is assumed that the action space is deterministic; i.e., all actions are

well known to each UAV. Therefore, the number of iterations of PBA-UCB is at most of the order of  $\mathcal{O}(\Omega \cdot Q \cdot \Xi)$  steps. In particular, the complexity of PBA-UCB can be expressed as  $\mathcal{O}(\Omega \cdot Q^2)$ , if the total number of the available power levels  $\Xi$  in the power vector  $\mathbf{p}$  is equal to the total number of channels  $Q$ . This means the complexity of the PBA-UCB algorithm is a polynomial in  $\Omega$  and  $Q$ . Moreover, the PBA-TS has the same computational complexity  $\mathcal{O}(\Omega \cdot Q^2)$  as the PBA-UCB algorithm. However, the update strategy in the PBA-TS algorithm is based on sampling from the Gaussian distribution  $\mathcal{N}(\hat{\mu}_{\omega,q}(r), \sigma^2(r))$ ; hence it may impose a slightly higher complexity depending on the sampling process.

---

**Algorithm 2** PBA-TS transmission power selection
 

---

```

1: for  $\omega \leftarrow 1$  to  $\Omega$  do
2:   Set  $\hat{\mu}_{\omega,q} \leftarrow 0, \sigma^2 \leftarrow 1$ 
3:   for  $r \leftarrow 1$  to  $\mathcal{R}$  do
4:     Set  $p_{\omega,M} = \max$  SU Tx power
5:     for  $q \leftarrow 1$  to  $Q$  do
6:       Draw a sample  $\varphi_{\omega,q}(r-1)$  from the distribution
        $\mathcal{N}(\hat{\mu}_{\omega,q}(r), \sigma^2(r))$ 
7:        $p_{\omega,q}^*(r) = \arg \max_{p_{\omega,q} \in \mathbf{p}_{\omega}} \left( \varphi_{\omega,q}(r-1) - \frac{p_{\omega,q}}{p_{\omega,M}} \right)$ 
8:       if  $\gamma_{\psi,q}(r) > \gamma_{\psi_{\text{TH}},q}(r)$  then
9:         if  $\gamma_{\omega,q}(r) > \gamma_{\omega_{\text{TH}},q}(r)$  then
10:          Obtain  $R_{\omega,q}(r)$  using  $p_{\omega,q}^*(r)$ 
11:           $T_{\omega,q}^{(p^*)}(r) \leftarrow T_{\omega,q}^{(p^*)}(r-1) + 1$ 
12:           $\hat{\mu}_{\omega,q}(r) \leftarrow \frac{1}{T_{\omega,q}^{(p^*)}(r)} \sum_{m=1}^{T_{\omega,q}^{(p^*)}(r)} R_{\omega,q}(m)$ 
13:           $\sigma^2(r) \leftarrow \frac{1}{T_{\omega,q}^{(p^*)}(r)+1}$ 
14:           $p_{\omega,M} \leftarrow p_{\omega,M} - p_{\omega,q}^*$ 
15:        else
16:           $p_{\omega,q}^*(r) \leftarrow 0, R_{\omega,q}(r) \leftarrow 0$ 
17:        end if
18:      else
19:         $p_{\omega,q}^*(r) \leftarrow 0, R_{\omega,q}(r) \leftarrow 0$ 
20:      end if
21:    end for
22:  end for
23: end for

```

---

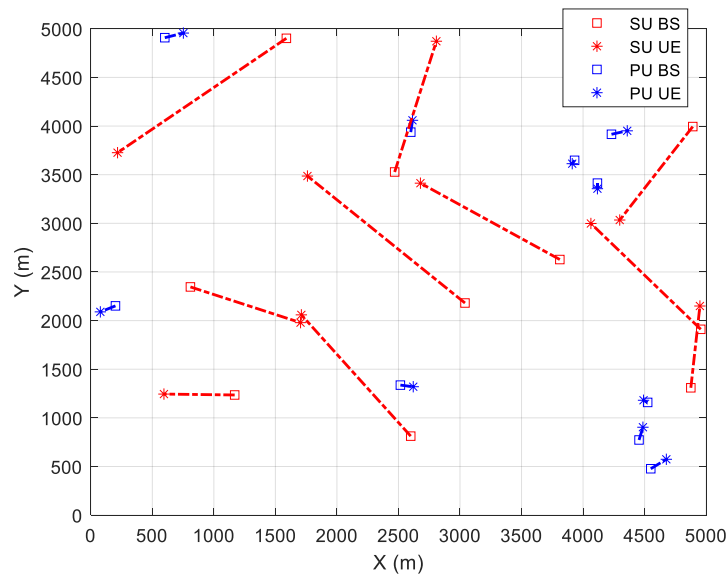
## 5. Simulation Results

In this section, the simulation results of our proposed algorithms are evaluated in terms of solution performance. We distributed each PU and SU transmitter randomly in a  $5 \text{ km} \times 5 \text{ km}$  area, while PUs and SUs receivers are deployed in a certain area from PUs and SUs transmitters to comply with the SINR constraint. The SINR threshold is chosen to be 30 dB for the PUs network, which is relatively high to ensure that the accumulated data transmission from SUs will not cause any harmful interference to the most valuable traffic. On the other hand, the SINR value for SUs network is set to 5 dB to ensure a successful data transmission. The transmission powers for PUs and SUs networks are set to 24 dBm and 30 dBm, respectively. We deployed 10 armed bandits to represent 10 different levels of UAVs' transmission power. These power levels are uniformly distributed with separation equal to the maximum transmission power divided by number of armed bandits. Both PU and SU networks operate at 5.8 GHz band with a bandwidth equal to 10 MHz. Since both PUs and SUs networks operate in an open area, the attenuation constant parameter is set to 3 for a free-space communication in a metropolitan area. Table 1 summarizes the system's parameters which are used for simulation.

**Table 1.** Simulation parameters.

Notation	Value
No. of armed bandits	10
Simulation area	5 km × 5 km
PU Tx power	24 dBm
$P_{\max}$	30 dBm
$W$	10 MHz
$f_q$	5.8 GHz
$c$	$3 \times 10^8$ m/s
$\zeta$	3
$\gamma_{\psi\text{TH},q}$	30 dB
$\gamma_{\omega\text{TH},q}$	5 dB
$N_0$	−100 dBm
$\eta$	0.5

Figure 2 shows an example of PUs and SUs transmitter/receiver pairs deployment. The deployment of PU receivers, i.e., cars, in the simulation area was done in a random way within  $\delta$  distance from their corresponding transmitters, while  $\delta$  is chosen to achieve 30 dB at the boundary of their deployment region. The number of sub-bands is set to be equal to the number of PUs, and hence  $\Psi = Q$ , as described previously in Section 3.

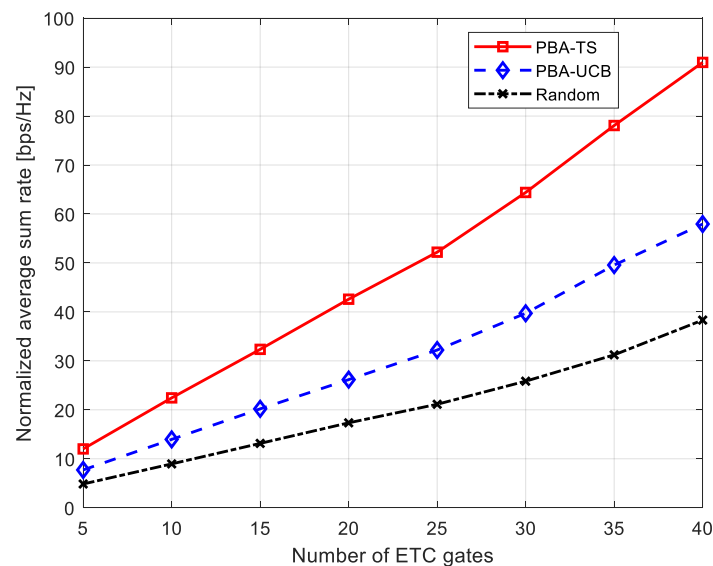
**Figure 2.** Distribution of PUs and SUs Tx/Rx pairs.

### 5.1. Average Total System Rate

This section shows the performance of the total average system rate in bps/Hz against different values of UAVs and ETC gates.

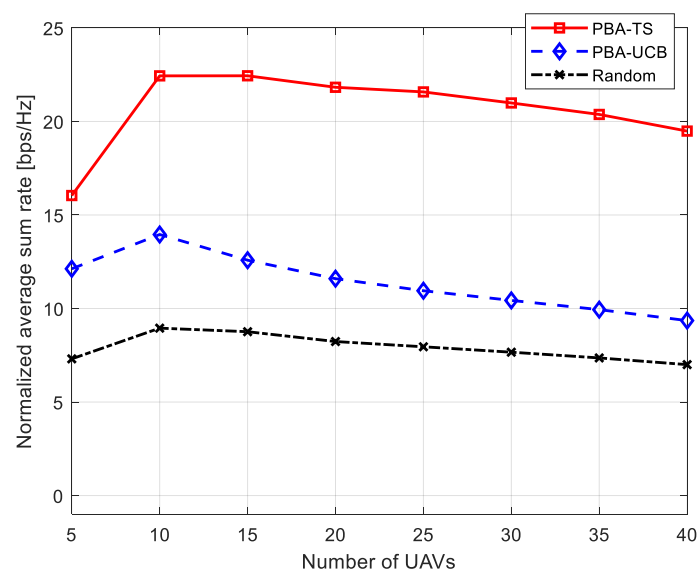
Figure 3 shows the total average system rate using 10 UAVs while increasing the number of ETC gates. It is shown in this figure that the PBA-TS algorithm achieved the highest data rate performance compared to both the PBA-UCB algorithm and transmission using a random power value. The reason behind this is that PBA-TS algorithm is constructed using posterior distributions for the obtained data rates through the integrated Bayesian strategy. On the other hand, transmission using a random power value has the worst performance due to the randomness in the selection of this power value for transmission in each round. Thus, each UAV experiences random interference from not only ETC gates but also other UAVs that share these channels. Furthermore, when the number of ETC gates increases and each ETC gate has its own separate channel, the number of available spectrum resources increases as well. This leads to each UAV becoming able to transmit data over a wider

band of channels and causes the total achievable average system rate to increase for both the PBA-TS algorithm and the PBA-UCB algorithm. On the other hand, and due to the randomness illustrated in this section, the increase in the achievable total average system rate using a random power value data transmission is not as high as the achievable data rate using either the PBA-TS algorithm or the PBA-UCB algorithm.



**Figure 3.** Normalized average sum rate against number of ETC gates using 10 UAVs.

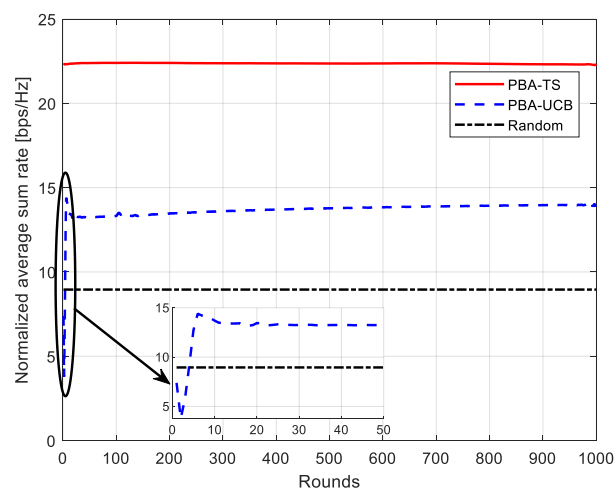
Figure 4 shows the performance of the achievable total average system rate against an increasing number of UAVs while keeping the number of ETC gates equal to 10. It is interesting that at the beginning with a few increments of the number of UAVs, the achievable data rate, using our proposed PBA-MAB algorithms, is increased till a certain point. Then, the achievable data rate begins to decrease with any increment in the number of deployed UAVs. The reason behind that is that while increasing the number of UAVs, the mutual interference between UAVs increases as well. Our proposed PBA-MAB algorithms succeeded in mitigating the interference effect, which is reflected in the achievable data rate reduction. Furthermore, the proposed PBA-TS algorithm can still achieve the highest data rate performance compared to the proposed PBA-UCB algorithm and the transmission using a random power value.



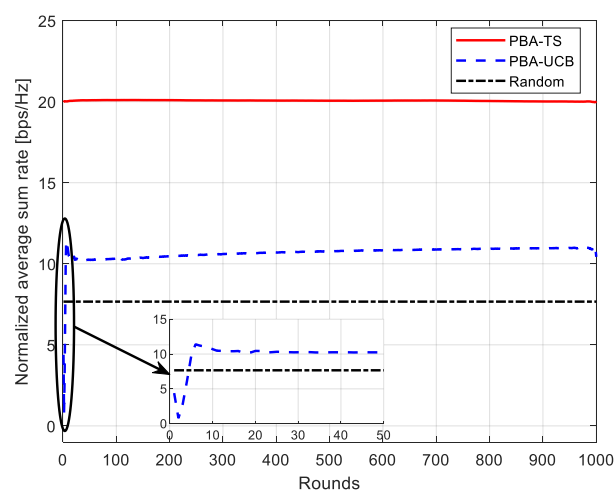
**Figure 4.** Normalized average sum rate against number of UAVs using 10 ETC gates.

## 5.2. Convergence Rate

The convergence rate is considered one of the most important parameters to judge the efficiency of online learning algorithms such as MAB algorithms; the faster the algorithm can converge, the better the reward that can be gained in just a few attempts. Hence, this section studies the convergence rate of the achievable total average system rate for our proposed PBA-MAB algorithms with different settings. Figures 5 and 6 show the convergence rate of the achievable total average system rate using 10 ETC gates while changing the number of UAVs to be 10 and 30. This can show the convergence rate for each algorithm under different network setup and different interference values. As shown in these figures, the horizontal axis indicates the count for rounds. Each algorithm runs its iterative process over counts till the algorithm converges toward a higher data rate. The proposed PBA-TS algorithm can converge faster than the PBA-UCB algorithm due to the fact that it uses Bayesian strategy over the posterior distributions of the reward. On the other hand, the PBA-UCB fluctuates during the few beginning rounds, and it takes more time to converge than the PBA-TS algorithm. Furthermore, it has a less convergence rate that the PBA-TS algorithm when both of the algorithms saturate by the end of the simulation rounds. These results can be concluded that both proposed PBA-MAB algorithms can deal with the adversarial network setup and selfish behavior of the UAVs. Hence, it means that every UAV learns how to select the most suitable transmission power value to enhance the overall system performance at every round. Furthermore, without loss of generality, it keeps an eye on the interference level while choosing this most suitable action.



**Figure 5.** Convergence of normalized average sum rate using 10 ETC gates and 10 UAVs.



**Figure 6.** Convergence of normalized average sum rate using 10 ETC gates and 30 UAVs.

## 6. Conclusions

In this paper, we have investigated the radio resource allocation for a CRN through DSA system to support a disaster surveillance system using UAVs wireless networks. To tackle this problem, we proposed two MAB algorithms, i.e., the PBA-UCB algorithm and the PBA-TS algorithm. The idea behind deploying MAB algorithms, as a class of RL algorithms, is the ability of MAB algorithms to solve online optimization problems with conflicting parameters that need to be jointly optimized. Since there is no information exchange between all UAVs, multi-player PBA-MAB algorithms were introduced to deal with this selfish configuration. Proposed PBA-MAB algorithms show outstanding performance over transmission using a random power value selection. Furthermore, the proposed algorithms showed a moderate convergence rate. The obtained results showed the capability of different MAB algorithms to deal with such problems with a high degree of randomness. Therefore, it can open the way for applying ML algorithms and more precise MAB algorithms to handle various wireless communication problems.

**Author Contributions:** Conceptualization, A.A. and E.M.M.; methodology, A.A. and G.K.T.; software, A.A.; validation, A.A. and G.K.T.; formal analysis, A.A.; investigation, A.A.; resources, A.A.; data curation, A.A.; writing—original draft preparation, A.A.; writing—review and editing, A.A. and G.K.T.; 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:** We would like to acknowledge the KDDI Foundation International Students Scholarship and the Telecommunications Advancement Foundation for the financial support to complete this research.

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

## Abbreviations

The following abbreviations are used in this manuscript:

DSA	Dynamic Spectrum Access
UAV	Unmanned Aerial Vehicle
CRN	Cognitive Radio Network
ETC	Electronic Toll Gate
MAB	Multi-armed Bandit
PBA-MAB	Power-Budget-Aware Multi-armed Bandit
UCB	Upper Confidence Bound
TS	Thompson Sampling
PBA-UCB	Power-Budget-Aware Upper Confidence Bound
PBA-TS	Power-Budget-Aware Thompson Sampling
PU	Primary User
SU	Secondary User
QoS	Quality of Service
ML	Machine Learning
RL	Reinforcement Learning
SINR	Signal-to-Interference-Plus-Noise Ratio
WSN	Wireless Sensor Network
CP	Control Plane
DP	Data Plane



VANET	Vehicle Ad hoc NETwork
iid	independent and identical distribution
WLAN	Wireless Local Area Network
LoS	Line of Sight
AWGN	Additive White Gaussian Noise
HetNets	Heterogeneous Networks
PHY	Physical layer
MAC	Medium Access Control layer
IoT	Internet of Things
VHF	Very High Frequency
UHF	Ultra High Frequency
mmWave	millimeter-wave
D2D	Device to Device

## References

- Mkiramweni, M.E.; Yang, C.; Li, J.; Zhang, W. A Survey of Game Theory in Unmanned Aerial Vehicles Communications. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3386–3416. [[CrossRef](#)]
- Mozaffari, M.; Saad, W.; Bennis, M.; Nam, Y.H.; Debbah, M. A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2334–2360. [[CrossRef](#)]
- Panda, K.G.; Das, S.; Sen, D.; Arif, W. Design and Deployment of UAV-Aided Post-Disaster Emergency Network. *IEEE Access* **2019**, *7*, 102985–102999. [[CrossRef](#)]
- Hu, F.; Chen, B.; Zhu, K. Full Spectrum Sharing in Cognitive Radio Networks Toward 5G: A Survey. *IEEE Access* **2018**, *6*, 15754–15776. [[CrossRef](#)]
- Zhao, Q.; Swami, A. A Survey of Dynamic Spectrum Access: Signal Processing and Networking Perspectives. In Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing—ICASSP '07, Honolulu, HI, USA, 15–20 April 2007; Volume 4, pp. IV-1349–IV-1352. [[CrossRef](#)]
- Akyildiz, I.F.; Lee, W.y.; Vuran, M.C.; Mohanty, S. A survey on spectrum management in cognitive radio networks. *IEEE Commun. Mag.* **2008**, *46*, 40–48. [[CrossRef](#)]
- Haykin, S. Cognitive radio: Brain-empowered wireless communications. *IEEE J. Sel. Areas Commun.* **2005**, *23*, 201–220. [[CrossRef](#)]
- Drozd, A.L.; Mohan, C.K.; Varshney, P.K.; Werner, D.D. Multiobjective joint optimization and frequency diversity for efficient utilization of the RF transmission hyperspace. In Proceedings of the 2004 International Waveform Diversity Design Conference, Edinburgh, UK, 8–10 November 2004; pp. 1–7. [[CrossRef](#)]
- Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press, Cambridge, MA, USA, 2018.
- Kwasinski, A.; Wang, W.; Mohammadi, F.S. Reinforcement learning for resource allocation in cognitive radio networks. *Mach. Learn. Future Wirel. Commun.* **2020**, *2020*, 27–44. doi:10.1002/9781119562306.ch2. [[CrossRef](#)]
- Katehakis, M.N.; Veinott, A.F., Jr. The multi-armed bandit problem: Decomposition and computation. *Math. Oper. Res.* **1987**, *12*, 262–268. [[CrossRef](#)]
- Bubeck, S.; Cesa-Bianchi, N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv* **2012**, arxiv:1204.5721v2.
- Auer, P.; Cesa-Bianchi, N.; Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* **2002**, *47*, 235–256.:1013689704352. [[CrossRef](#)]
- Audibert, J.Y.; Munos, R.; Szepesvári, C. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theor. Comput. Sci.* **2009**, *410*, 1876–1902. [[CrossRef](#)]
- Francisco-Valencia, I.; Marcial-Romero, J.R.; Valdovinos-Rosas, R.M. A comparison between UCB and UCB-Tuned as selection policies in GGP. *J. Intell. Fuzzy Syst.* **2019**, *36*, 5073–5079. [[CrossRef](#)]
- Olivieri, M.; Barnett, G.; Lackpour, A.; Davis, A.; Ngo, P. A scalable dynamic spectrum allocation system with interference mitigation for teams of spectrally agile software defined radios. In Proceedings of the First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005), Baltimore, MD, USA, 8–11 November 2005; pp. 170–179. [[CrossRef](#)]
- Agrawal, S.; Goyal, N. Further optimal regret bounds for thompson sampling. *Artif. Intell. Stat.* **2013**, *31*, 99–107.
- Sun, Y.; Peng, M.; Zhou, Y.; Huang, Y.; Mao, S. Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3072–3108. [[CrossRef](#)]
- Huang, Y.; Xu, C.; Zhang, C.; Hua, M.; Zhang, Z. An Overview of Intelligent Wireless Communications using Deep Reinforcement Learning. *J. Commun. Inf. Netw.* **2019**, *4*, 15–29. [[CrossRef](#)]
- Wilhelmi, F.; Cano, C.; Neu, G.; Bellalta, B.; Jonsson, A.; Barrachina-Muñoz, S. Collaborative spatial reuse in wireless networks via selfish multi-armed bandits. *Ad Hoc Netw.* **2019**, *88*, 129–141. [[CrossRef](#)]
- Zhao, Q.; Sadler, B.M. A Survey of Dynamic Spectrum Access. *IEEE Signal Process. Mag.* **2007**, *24*, 79–89. 361604. [[CrossRef](#)]

22. Yucek, T.; Arslan, H. A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE Commun. Surv. Tutor.* **2009**, *11*, 116–130. [[CrossRef](#)]
23. Wang, J.; Ghosh, M.; Challapali, K. Emerging cognitive radio applications: A survey. *IEEE Commun. Mag.* **2011**, *49*, 74–81. [[CrossRef](#)]
24. Liang, Y.C.; Chen, K.C.; Li, G.Y.; Mahonen, P. Cognitive radio networking and communications: An overview. *IEEE Trans. Veh. Technol.* **2011**, *60*, 3386–3407. [[CrossRef](#)]
25. Srinivasa, S.; Jafar, S.A. Cognitive Radios for Dynamic Spectrum Access—The Throughput Potential of Cognitive Radio: A Theoretical Perspective. *IEEE Commun. Mag.* **2007**, *45*, 73–79. [[CrossRef](#)]
26. Akan, O.B.; Karli, O.B.; Ergul, O. Cognitive radio sensor networks. *IEEE Netw.* **2009**, *23*, 34–40. [[CrossRef](#)]
27. Xing, Y.; Chandramouli, R.; Mangold, S.; N, S. Dynamic spectrum access in open spectrum wireless networks. *IEEE J. Sel. Areas Commun.* **2006**, *24*, 626–637. [[CrossRef](#)]
28. Theis, N.C.; Thomas, R.W.; DaSilva, L.A. Rendezvous for Cognitive Radios. *IEEE Trans. Mob. Comput.* **2011**, *10*, 216–227. [[CrossRef](#)]
29. Homssi, B.A.; Al-Hourani, A.; Krusevac, Z.; Rowe, W.S.T. Machine Learning Framework for Sensing and Modeling Interference in IoT Frequency Bands. *IEEE Internet Things J.* **2021**, *8*, 4461–4471. [[CrossRef](#)]
30. Xiao, Z.; Li, F.; Jiang, H.; Bai, J.; Xu, J.; Zeng, F.; Liu, M. A Joint Information and Energy Cooperation Framework for CR-Enabled Macro-Femto Heterogeneous Networks. *IEEE Internet Things J.* **2020**, *7*, 2828–2839. [[CrossRef](#)]
31. Farag, H.M.; Mohamed, E.M. Soft decision cooperative spectrum sensing with noise uncertainty reduction. *Pervasive Mob. Comput.* **2017**, *35*, 146–164. [[CrossRef](#)]
32. Erdelj, M.; Natalizio, E. UAV-assisted disaster management: Applications and open issues. In Proceedings of the 2016 International Conference on Computing, Networking and Communications (ICNC), Kauai, HI, USA, 15–18 February 2016; pp. 1–5. [[CrossRef](#)]
33. Merwaday, A.; Tuncer, A.; Kumbhar, A.; Guvenc, I. Improved Throughput Coverage in Natural Disasters: Unmanned Aerial Base Stations for Public-Safety Communications. *IEEE Veh. Technol. Mag.* **2016**, *11*, 53–60. [[CrossRef](#)]
34. Lee, S.; Har, D.; Kum, D. Drone-Assisted Disaster Management: Finding Victims via Infrared Camera and Lidar Sensor Fusion. In Proceedings of the 2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), Nadi, Fiji, 4–6 December 2016; pp. 84–89. [[CrossRef](#)]
35. Rivera, A.; Villalobos, A.; Monje, J.; Mariñas, J.; Oppus, C. Post-disaster rescue facility: Human detection and geolocation using aerial drones. In Proceedings of the 2016 IEEE Region 10 Conference (TENCON), Singapore, 22–26 November 2016; pp. 384–386. [[CrossRef](#)]
36. Kobayashi, T.; Matsuoka, H.; Betsumiya, S. Flying Communication Server in case of a Largescale Disaster. In Proceedings of the 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), Atlanta, GA, USA, 10–14 June 2016; Volume 2, pp. 571–576. [[CrossRef](#)]
37. Sánchez-García, J.; García-Campos, J.; Toral, S.L.; Reina, D.G.; Barrero, F. A Self Organising Aerial Ad Hoc Network Mobility Model for Disaster Scenarios. In Proceedings of the 2015 International Conference on Developments of E-Systems Engineering (DeSE), Dubai, United Arab Emirates, 13–14 December 2015; pp. 35–40. [[CrossRef](#)]
38. Ejaz, W.; Ahmed, A.; Mushtaq, A.; Ibnkahla, M. Energy-efficient task scheduling and physiological assessment in disaster management using UAV-assisted networks. *Comput. Commun.* **2020**, *155*, 150–157. [[CrossRef](#)]
39. Zhan, P.; Yu, K.; Swindlehurst, A.L. Wireless Relay Communications with Unmanned Aerial Vehicles: Performance and Optimization. *IEEE Trans. Aerosp. Electron. Syst.* **2011**, *47*, 2068–2085. [[CrossRef](#)]
40. Zhang, S.; Shi, S.; Gu, S.; Gu, X. Power Control and Trajectory Planning Based Interference Management for UAV-Assisted Wireless Sensor Networks. *IEEE Access* **2020**, *8*, 3453–3464. [[CrossRef](#)]
41. Baek, J.; Han, S.I.; Han, Y. Energy-Efficient UAV Routing for Wireless Sensor Networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 1741–1750. [[CrossRef](#)]
42. Seliem, H.; Shahidi, R.; Ahmed, M.H.; Shehata, M.S. Drone-Based Highway-VANET and DAS Service. *IEEE Access* **2018**, *6*, 20125–20137. [[CrossRef](#)]
43. Lai, T.L.; Robbins, H. Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math.* **1985**, *6*, 4–22. [[CrossRef](#)]
44. Thompson, W.R. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* **1933**, *25*, 285–294. [[CrossRef](#)]
45. Toldov, V.; Clavier, L.; Loscrí, V.; Mitton, N. A Thompson sampling approach to channel exploration-exploitation problem in multihop cognitive radio networks. In Proceedings of the 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain, 4–7 September 2016; pp. 1–6. [[CrossRef](#)]
46. Bonnefoi, R.; Moy, C.; Palicot, J. Advanced metering infrastructure backhaul reliability improvement with cognitive radio. In Proceedings of the 2016 IEEE International Conference on Smart Grid Communications (SmartGridComm), Sydney, Australia, 6–9 November 2016; pp. 230–236. [[CrossRef](#)]
47. Kalathil, D.; Nayyar, N.; Jain, R. Decentralized Learning for Multiplayer Multiarmed Bandits. *IEEE Trans. Inf. Theory* **2014**, *60*, 2331–2345. [[CrossRef](#)]

48. Liu, K.; Zhao, Q.; Krishnamachari, B. Decentralized multi-armed bandit with imperfect observations. In Proceedings of the 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 29 September–1 October 2010; pp. 1669–1674. [[CrossRef](#)]
49. Anandkumar, A.; Michael, N.; Tang, A.K.; Swami, A. Distributed Algorithms for Learning and Cognitive Medium Access with Logarithmic Regret. *IEEE J. Sel. Areas Commun.* **2011**, *29*, 731–745. [[CrossRef](#)]
50. 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. [[CrossRef](#)]
51. Takeuchi, S.; Hasegawa, M.; Kanno, K.; Uchida, A.; Chauvet, N.; Naruse, M. Dynamic channel selection in wireless communications via a multi-armed bandit algorithm using laser chaos time series. *Sci. Rep.* **2020**, *10*, 1574. [[CrossRef](#)]
52. Oshima, K.; Onishi, T.; Kim, S.J.; Ma, J.; Hasegawa, M. Efficient wireless network selection by using multi-armed bandit algorithm for mobile terminals. *Nonlinear Theory Its Appl. IEICE* **2020**, *11*, 68–77. [[CrossRef](#)]
53. Mohamed, E.M.; Hashima, S.; Hatano, K.; Fouda, M.M.; Fadlullah, Z.M. Sleeping Contextual/Non-Contextual Thompson Sampling MAB for mmWave D2D Two-Hop Relay Probing. *IEEE Trans. Veh. Technol.* **2021**, *70*, 12101–12112. [[CrossRef](#)]
54. Hashima, S.; Hatano, K.; Takimoto, E.; Mahmoud Mohamed, E. Neighbor Discovery and Selection in Millimeter Wave D2D Networks Using Stochastic MAB. *IEEE Commun. Lett.* **2020**, *24*, 1840–1844. [[CrossRef](#)]
55. Bhardwaj, P.; Panwar, A.; Ozdemir, O.; Masazade, E.; Kasperovich, I.; Drozd, A.L.; Mohan, C.K.; Varshney, P.K. Enhanced Dynamic Spectrum Access in Multiband Cognitive Radio Networks via Optimized Resource Allocation. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 8093–8106. [[CrossRef](#)]
56. Miao, G.; Himayat, N.; Li, G.Y.; Talwar, S. Low-Complexity Energy-Efficient Scheduling for Uplink OFDMA. *IEEE Trans. Commun.* **2012**, *60*, 112–120. [[CrossRef](#)]
57. Gupta, P.; Kumar, P. The capacity of wireless networks. *IEEE Trans. Inf. Theory* **2000**, *46*, 388–404. [[CrossRef](#)]
58. Arnold, P.; Bayer, N.; Belschner, J.; Zimmermann, G. 5G radio access network architecture based on flexible functional control/user plane splits. In Proceedings of the 2017 European Conference on Networks and Communications (EuCNC), Oulu, Finland, 12–15 June 2017; pp. 1–5. [[CrossRef](#)]
59. Chapelle, O.; Li, L. An empirical evaluation of thompson sampling. *Adv. Neural Inf. Process. Syst.* **2011**, *24*, 2249–2257.