Dual Energy-Aware based Trajectory Optimization for UAV Emergency Wireless Communication Network: A Multi-armed Bandit Approach

Amr Amrallah*1, Ehab Mahmoud Mohamed†‡2, Gia Khanh Tran*3, and Kei Sakaguchi*§4
*Department of Electrical and Electronic Engineering, Tokyo Institute of Technology, Tokyo, Japan
†Electrical Engineering Department, Prince Sattam Bin Abdulaziz University, Wadi Addwasir, Saudi Arabia
‡Electrical Engineering Department, Aswan University, Aswan, Egypt
§Academy for Super Smart Society, Tokyo Institute of Technology, Tokyo, Japan

Emails: {1amrallah, 3khanhtg, 4sakaguchi}@mobile.ee.titech.ac.jp, 2ehab_mahmoud@aswu.edu.eg

***Abstract*—In a post-disaster area, Unmanned Aerial Vehicles (UAVs) are considered one of the most effective ways to provide emergency wireless communication services, especially when the wireless infrastructure becomes malfunctioned due to the extensive damage. In this paper, we investigate the UAV-based emergency wireless communication network for a post-disaster area, where the UAV acts as a flying Base Station (BS) to provide wireless connectivity from the sky. UAV should collect as much data as possible from ground users in the affected area. Considering the malfunction of power supplies in the post-disaster area, the available energy for ground users is very limited. Moreover, UAV operates with an onboard battery with a limited capacity. Aiming to maximize the uplink throughput by maximizing the number of visited ground users during the flight round, the UAV trajectory optimization problem is formulated under the concern of dual limited available energy (i.e., limited ground user and UAV energy capacities). Considering that both energy terms are dynamic and cumulative over time, this optimization problem becomes hard to be solved using conventional optimization methods. Therefore, a multi-armed bandit (MAB)-based algorithm controlled with dual limited energy capacities is proposed to tackle this problem. The simulation results show that the proposed algorithm could solve the optimization problem and maximize the achievable throughput under these energy constraints.**

***Index Terms*—unmanned aerial vehicles, trajectory optimization, emergency wireless communication, multi-armed bandit**

I. Introduction

Large-scale natural disasters always wreak unpredictable casualties on life and severe havoc on property. During the last decades, various types of natural disasters, such as floods, hurricanes, earthquakes, tsunamis, wild fires, etc., caused thousands of injuries, deaths, and about 100% –150% additional increase in material losses over the globe [1]. The first few hours after the disaster occurrence are considered the golden relief time for victims in the post-disaster area. Hence, an emergency wireless communication network becomes crucial in this situation, especially when the communication infrastructure is devastated due to the damage caused by this disaster. Moreover, the paralysis of the power system caused by a natural disaster makes the situation more complicated in the post-disaster area. In 2011, a massive earthquake with a magnitude of 9.0 caused a tsunami on the eastern coast of Japan which destroyed more than 6000 base stations (BSs) in that area. Therefore, the still operating BSs are heavily overloaded with a large volume of voice and data traffic that caused a high call block rate, resulting in the loss of communication services for four days after the tsunami [2].

The ultimate solution would be to deploy a specific wireless network that is independent of the existing broadband network. An unmanned Aerial Vehicle (UAV) wireless network is considered one of these feasible and efficient wireless networks for emergency wireless communications. UAVs are known for their flexible deployment and immediate response that can be used as ubiquitous temporary mobile BS to establish such an emergency wireless communication network [3]. Therefore, during the last few years, UAVs have been utilized to support different applications in emergency wireless communication networks including disaster management, surveillance and early warnings, post-disaster fusion centers, damage assessment, supply-aid drop, etc.

Despite the advantages of using UAVs to operate emergency wireless communication networks in the post-disaster area, there are some technical issues that should be considered to realize these types of network, such as: 1) the available energy for victims is fugacious due to the destruction of the power supply infrastructure as a resulted of the natural disaster [4]; 2) the UAV working time is limited to the capacity of the attached on-board battery. Once it becomes nearly depleted, the UAV should return to its base to recharge its battery [5]; 3) the UAV should plan and optimize its flight trajectory in this harsh environment caused by the natural disaster. This required a fast online optimization algorithm to deal with the abrupt change in the geographical field [3]. Therefore, all these factors should be deemed while designing an emergency wireless communication network. Furthermore, since it is considered a critical operation, the UAV must serve as many victims as possible in the disaster area before its batteries run out of energy.

A. Prior Works and Motivations

One of the key features of using UAVs in emergency wireless communication networks is their ability to collect wide-ranging data from geographically scattered ground devices, such as ground BSs, ground users, and even ground sensors [6]. Moreover, UAV can act as a flying BS or edge server to aid different traffic offloading scenarios [7]. Therefore, not only the radio resource management of UAV wireless network but also the planning and optimization of the UAV's trajectory becomes a critical issue due to its mobility. In recent years, many investigations have been conducted on this topic. The authors in [8] used the speed of the UAV with the location of the waypoints to design the UAV trajectory. In this way, they could minimize the mission completion time in a UAV-based multi-cast system. A heuristic algorithm based on UAV speed control and UAV data scheduling was proposed in [9] to minimize the total energy consumption.

When accurate models of UAV wireless networks including their flight dynamics are available, UAV trajectory optimization can be performed using conventional optimization techniques. However, it is difficult to construct these accurate network models, leading to the use of model-free machine learning algorithms to control the operation of UAVs that support wireless communication networks. Machine learning algorithms are capable of formulating autonomous control policy by exploiting collected information from past experiences [10]. The authors of [11] maximized the total distance traveled by the UAV using a policy gradient method for trajectory optimization. The deep Q-learning algorithm is utilized in [12] to optimize the flight trajectory of the UAV to maximize the data rate during the flight time in an unknown environment. In [13], the authors maximized the uplink transmission rate in the UAV cellular network by designing the flying trajectory of the UAV. They could transform the optimization problem into a Markov Decision Process (MDP) and solved it using the Deterministic Policy Gradient (DPG) algorithm.

Despite the existence of many excellent research on UAV wireless communications, there are a few works focused on UAV-assisted emergency wireless communication networks. In our previous works [14], [15] , we studied the radio resource allocation for UAV emergency wireless communications using dynamic spectrum access system. UAVs were deployed as a Cognitive Radio Network (CRN), aiming to maximize the downlink data rate in a post-disaster area. This optimization problem was implemented as a multi-player multi-armed bandit problem controlled by the limited transmission power of each UAV. Mohamed *et al.* in [16] addressed the gateway selection for gateways in a post-disaster area. Where, UAVs try to find the best gateway to offload its data. The whole process was done in a decentralized manner using a constrained MAB-based algorithm. Also, a constrained MAB-based algorithm is adopted in [17] to support an attached Reconfigurable Intelligent Surface (RIS) to the UAV. The optimization problem amid to find the optimum trajectory of the UAV that maximises the total throughput while reducing the consumed flying power of the UAV. The UAV trajectory optimization problem that maximized the accumulated data volume from ground sensors was studies in [18] under unknown network information. The optimization problem is transformed into a finite MDP and solved using two reinforcement learning frameworks, called state-action-reward-state-action (Sarsa) and Q-learning-based, UAV trajectory optimization algorithms (i.e., SUTOA and QUTOA). Authors in [19] studied the UAV trajectory optimization problem in a UAV emergency wireless communication network aimed to maximize the total system rate and constrained by the limited flight time of the UAV, the power capacity of the ground user, and the need to avoid obstacles in the disaster area. The whole process was done using a Lyapunov-based deep Q-learning framework called Safe-DQN.

All these related works to UAV emergency wireless communication networks addressed the optimization problem under a single power constraint, either a limited UAV battery capacity or a limited available energy for ground users (i.e., ground UE or ground sensors). We believe that these two factors are considered the most critical factors in designing a UAV emergency wireless communication network. This is due to the devastation or, at least, the paralysis of the power supply network as a result of the natural disaster. Therefore, our proposed framework aims to study the optimization problem of the UAV trajectory under these two power conditions. To this end, our objective is to study a dual constraint optimization problem that could improve the reliability of the UAV emergency wireless network compared to previous works.

B. Contributions and Organization

As discussed in the previous section, most of the current studies related to emergency wireless communication network focused on the limited energy capacity of UAV and just a few of them considered the limited energy capacity of ground users (i.e., ground users equipment (UEs)). Here comes our idea to fill this gap by studying an optimization scenario under both of limited UAU and UEs energy capacity. UAV is considered as flying BS that provide wireless connectivity from the sky to ground UEs in the post-disaster area. The data collected from the UEs is considered extremely important to estimate the situation of the victims and evaluate the damage in the post-disaster area. To that end, these valuable data can be analyzed to guide rescue teams in saving these precious lives. Our main target is to maximize the collected data from ground UEs under the limited power capacity for both UAV and ground UEs. However, since the coverage of UAV is relatively small compared to ground BSs, our target is to optimize the UAV flying trajectory to maximize the number of visited ground UEs before its battery being used up. The main contribution of this paper can be summarized as follows:

- We proposed a framework for an emergency wireless communication network where the UAV collects user data in a post-disaster area. Ground BSs malfunctioned as a result of the damage caused by a natural disaster, however ground UEs that are located in the UAV coverage area

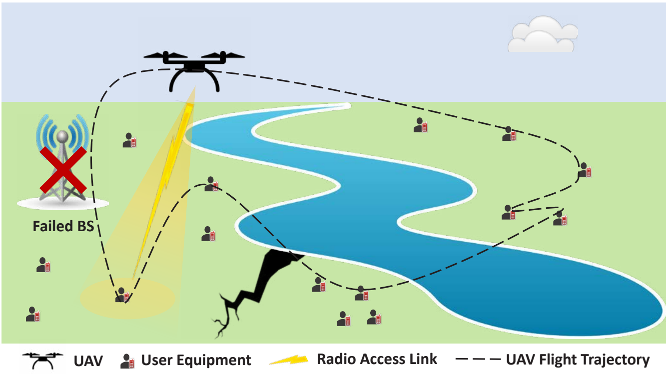

Fig. 1. UAV emergency wireless communication network

can upload their data using an alternative way of connectivity from the sky by the support of UAV emergency wireless communication network. Taking into account the limited available energy for both the UAV and ground UEs in the post-disaster area, we formulate a dynamic optimization problem to maximize the uplink throughput for the UAV emergency wireless communication network by optimizing the flying trajectory of the UAV under these assumptions.

- The optimization problem is transformed into a constrained multi-armed bandit (MAB) problem where action, reward, and cost are defined as the flight direction, the throughput of the uploaded data, and the dissipated power for both the UAV and UEs, respectively.
- It should be mentioned that, to the best of our knowledge, this is the first research work to study this type of optimization problem under dual-constrained energy capacity for both UAV and UEs, simultaneously.

The rest of our paper is organized as follows. Section II shows the system model used in our study and formulates the optimization problem to maximize the long-term throughput. A constrained multi-armed bandit algorithm is adapted in Section III to solve this optimization problem. The results and analysis of the simulation are provided in Section IV. Finally, the paper is concluded in Section V.

II. System Model and Problem Formulation

This section discusses the network architecture for the UAV-assisted emergency wireless network, the flying model used for the UAV, the channel model used in uploading data, and formulates the optimization problem.

A. System Model

Fig. 1 shows the system architecture for the UAV-assisted emergency wireless communication network, where a natural disaster, such as an earth quake, floods, etc., hits a certain area and results in a malfunction of power supplies and the wireless network. Our idea is to deploy UAV in this post-disaster area to provide the wireless connectivity from the sky. In such a way, wireless connectivity can be provided for victims, i.e., ground UEs, in this damaged area, so that they can offload their data, that should be valuable in guiding rescue teams and estimating damage. It is assumed that there are M UEs trapped in this post-disaster area, denoted by $\mathcal{M} = \{1, ..., M\}$. Each of them has a stationary location denoted in Cartesian coordinates by $l_m = (x_m, y_m)$. It is supposed that the UAV starts to fly from the middle of the disaster are, i.e., the middle of the simulation area, which is denoted by $l_0 = (x_0, y_0)$. Also, it flies according to a constant speed ν m/s and altitude H m. We suppose that this altitude is relatively high and the data transmission period is quite small. Therefore, the UAV is considered stationary during offloading the UE data. We make use of the channel model that is illustrated in [19] according to the 3GPP specifications in the technical report [20]. The wireless communication link between the UAV and each of the served UEs is determined by two components, i.e., line-of-sight (LoS) component and nonline-of-sight (NLoS) component according to the probability of each of them. Equations (1), (2), at the bottom of this page, show the path-loss equation and the probability of LoS and NLoS equation, respectively, where d_m denotes the direct link between the UAV and any connected UE. The following equations (3), (4), and (5) show how to calculated parameters d_m, d_0, and p_0, respectively.

$$d_m = \sqrt{H^2 + \|l_m - l_0\|^2}, \forall m \in \mathcal{M} \tag{3}$$

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

$$p_0 = 233.98 \log_{10} H - 0.95 \tag{5}$$

The probability of NLoS can be obtained as:

$$P_m^{NLoS} = 1 - P_m^{LoS} \tag{6}$$

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

$$g_m = P_m^{LoS}\left(10^{L_m^{LoS}/10}\right)^{-1} + P_m^{NLoS}\left(10^{L_m^{NLoS}/10}\right)^{-1} \tag{7}$$

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

$$P_m^{LoS} = \begin{cases} 1, & \text{if} \sqrt{d_m^2 - H^2} \le d_0 \\ \frac{d_0}{\sqrt{d_m^2 - H^2}} + \exp\left\{\left(\frac{-\sqrt{d_m^2 - H^2}}{p_0}\right)\left(1 - \frac{d_0}{\sqrt{d_m^2 - H^2}}\right)\right\}, & \text{if} \sqrt{d_m^2 - H^2} > d_0 \end{cases} \tag{2}$$

For the sake of simplicity, we assume that the effective radiation angle of the attached UAV antenna is denoted by ϕ, then the maximum direct transmission link is $H/\cos(\phi)$. In addition, it is assumed that only a single UE can establish a wireless link to the UAV at a time. Hence, there is no simultaneous wireless connections from UEs towards UAV. A connection indicator α_m is set to 1 when a UE m successfully established a wireless link to the UAV, and otherwise it set to 0. In this way, we ensure that there is no interference from the simultaneous transmission of multiple UEs. Then, the transmission data rate of any UE m towards the UAV can be calculated as:

$$R_m = \alpha_m B \log_2 \left(1 + \frac{g_m P_m^{tx}}{\sigma_0}\right) \tag{8}$$

where B is the available bandwidth of the wireless channel, P_m^{tx} is the transmission power for UE m, σ_0 denotes the power of additive white Gaussian noise (AWGN) on the UAV side.

From the limited energy capacity point of view, the power consumption consists of two terms: 1) the power consumed by each UE during data offloading and idle mode; 2) the power consumed by the UAV during flying over the post-disaster area to provide wireless network connectivity to UEs. The power consumed in the UAV by the receiver circuit and signal processing is relatively small compared to the flying power and can be neglected. These two power terms can be denoted as follows:

$$e_m(t) = \begin{cases} \alpha_m P_m^{tx}, & \text{if UE at Tx mode} \\ (1-\alpha_m) e_{idle}, & \text{if UE at idle mode} \end{cases} \tag{9}$$

$$E_{UAV}(t) = P_{fly} \tag{10}$$

Where $t = \{1, ..., T\}$ is the elapsed time while the UAV scans the post-disaster area. Our goal is to maximize the data collected from the ground UEs so that it can improve the efficiency of the rescue operation. Also, it should keep an eye on the valuable and limited energy on both UEs and the UAV sides. Therefore, the optimization problem is formulated to maximize the long term UAV uplink throughput via optimizing the flight trajectory of the UAV. To this end, the optimization problem can be defined as follows:

$$\begin{aligned} &\max_{m \in \mathcal{M}} \frac{1}{T} \sum_{t=1}^{T} \sum_{m=1}^{M} R_m(t) \\ \text{s.t.} \; &\sum_{t=1}^{T} e_m(t) \leq e_0, \forall m \in \mathcal{M} \\ &\sum_{t=1}^{T} E_{UAV}(t) \leq E_0 \end{aligned} \tag{11}$$

Where e_0 is the total available energy for each UE m, and E_0 is the total available energy for the UAV.

III. DEA-MAB Based UAV Trajectory Optimization Framework

In this section, we will introduce the general MAB algorithm and the proposed dual-energy-aware MAB (DEA-MAB) framework aiming to solve the pre-described trajectory optimization problem.

A. The General MAB Algorithm

MAB algorithm can be described as a set of arms of a bandit machine. Each arm leads to a certain reward. A player who wants to maximize his total reward during a certain playing period; however, the amount of reward behind each arm is not revealed to the player in advance. In this context, this instantaneous reward behind each arm is just revealed to the player once he plays with a certain arm. In addition, the player may lose some reward in each trial due to not selecting the arm that leads to the highest known reward. This loss is denoted by *regret* [21]. Therefore, each player should develop a strategy to maximize the total reward over the horizon, in other words, minimize the regret over this horizon. Therefore, he should find the best strategy to continue exploiting the discovered arm that leads to the highest reward, and at the same time, he should continue exploring other arms that may lead to a higher reward. This is considered a common dilemma facing the MAB algorithm, and is called the exploration-exploitation trade-off [22].

B. The Proposed DEA-MAB Framework

Our proposed DEA-MAB framework is inspired by the Cost-Subsidized Explore-Then-Commit algorithm illustrated in [23]. In any real world application, a certain action that leads to a higher reward always has a higher cost. One of the ideas to balance this reward/cost trade-off is to deduct the paid cost from the achieved reward. However, it is not always meaningful, especially when the reward and the cost are defined in different quantities [23]. Thus, it becomes necessary to find a new MAB algorithm that can optimize both cost and reward. In other words, it can avoid incur excessive cost for just a marginal increase in the reward. In the proposed DEA-MAB framework, both the upper confidence bound (UCB) and the lower confidence bound (LCB) values are examined for each candidate arm. Then, a feasibility region of acceptable reward is constructed, where arms that achieve a reward larger than the maximum LCB value are counted. For each of these arms, we check for the corresponding remaining UE energy and construct a list of the most critical UEs whose energy is about to be depleted. Of this list, we choose the UE that has the least UAV flying power consumption to be selected for data offload operation.

To illustrate the proposed DEA-MAB framework, in each time period t, the UAV should fly towards a certain UE that the DEA-MAB selects to provide the emergency wireless connectivity from the sky. The input of the DEA-MAB is the space for all available M UEs in the post-disaster area and the tuning parameters ω and ζ , and the output is the next selected UE to be served. In this way, the UAV can optimize the flight trajectory by selecting the most suitable UE to be served at each time t. The process is divided into two phases. For the initialization phase, the UAV flies toward the UEs in a random way, estimates the achievable data rate R_m at each time t, and counts the number of selected UEs as follows:

$$Q_m(t+1) = Q_m(t) + 1 \tag{12}$$

$$R_m(t+1) = \frac{1}{Q_m(t+1)} \sum_{i=1}^{Q_m(t+1)} R_m(i) \tag{13}$$

where $Q_m(t)$ is the count of times that the UAV selects the UE m until time t. This initialization phase is considered a pure exploration phase and is carried out over a period of time equal to $(T/M)^{2/3}$ as illustrated in [23]. In the second phase, i.e., the selection phase, at each time t, the UCB and LCB are evaluate as follows:

$$\gamma_m^{UCB}(t) = R_m(t) + \sqrt{2\ln(t)/Q_m(t)}, \forall m \in \mathcal{M} \tag{14}$$

$$\gamma_m^{LCB}(t) = R_m(t) - \sqrt{2\ln(t)/Q_m(t)}, \forall m \in \mathcal{M} \tag{15}$$

The UE index that achieved the highest value of $\gamma_m^{LCB}(t)$ at time t is calculated as follows:

$$\eta(t) = \arg\max_m \Big(\gamma_m^{LCB}(t)\Big) \tag{16}$$

Then, the feasibility region of UEs can be calculated as follows:

$$F_{eas}(t) = \Big\{m : \gamma_m^{UCB}(t) \geq (1-\omega)\gamma_m^{LCB}(t)\Big\} \tag{17}$$

Out of this feasibility region, we construct a list of the most critical UEs that need to be served before their power being depleted as follows:

$$C_{rtcl}(t) = \Big\{m : \sum_{t=1}^{\tau} e_m(t) \geq (1-\zeta)e_0\Big\} \tag{18}$$

Finally, out of this list, the UE corresponding the lowest energy dissipated in UAV flying towards this UE is selected for the UAV trajectory as follows:

$$m^*(t) = \underset{m\in C_{rtcl}(t)}{\arg\min} \Big(E_{UAV}(t)\Big) \tag{19}$$

In this way, the DEA-MAB framework can optimize the UAV flying trajectory under dual-energy constraints.

IV. Simulation Results

In this section, we evaluate the proposed DEA-MAB framework for the trajectory optimization of the UAV emergency wireless communication network. It is assumed that the UAV should find the best trajectory to maximize the data throughput from ground UEs in the post-disaster area. Table I shown the simulation parameters.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Simulation area	10 km^2
Flight speed (ν)	20 km/h
Flight altitude (H)	100 m
UAV radiation angle (ϕ)	$\pi/8$ rad
Carrier frequency (f)	5.8 GHz
Channel bandwidth (B)	10 MHz
UE transmission power (P_m^{tx})	23 dBm
UAV battery capacity (E_0)	100-200 Wh
Total number of UEs (M)	20-50
AWGN spectral density (σ_0)	-130 dBm/Hz
UE power dissipation in idle mode (e_{idle})	0.05 mW
Data rate feasibility region factor (ω)	0.6
Critical power feasibility region factor (ζ)	0.5

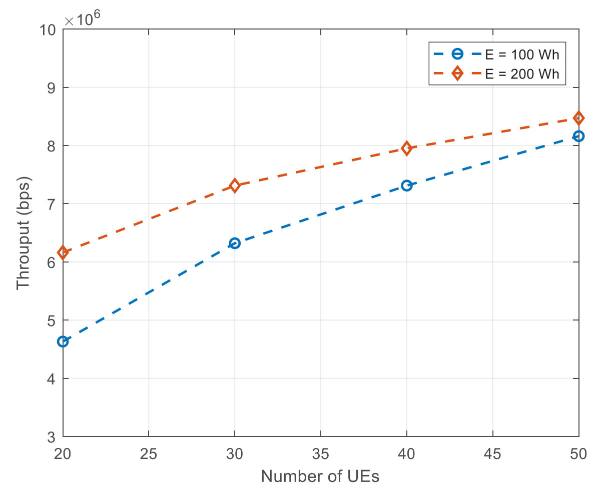

Fig. 2. Cumulative uplink throughput with varying number of UEs

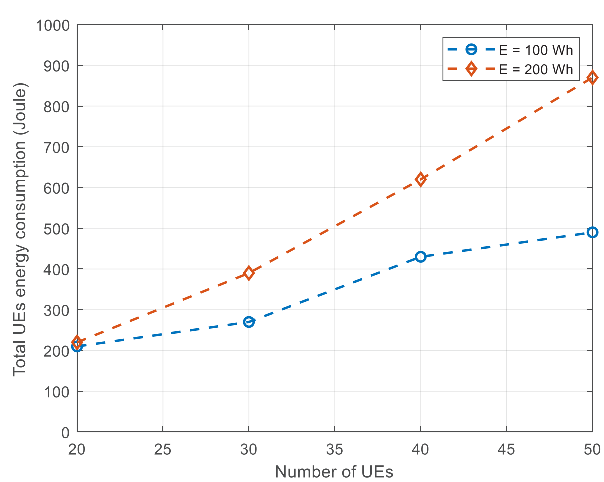

Fig. 3. Total UEs energy consumption with varying number of UEs

Fig. 2 shows the cumulative uplink throughput of the UAV emergency wireless communication network. These two curves show an upward trend, which means that, regardless of the battery capacity of the UAV, the total system throughput increases with the increase in the number of UEs. This upward performance gradually tends to approach saturation with increasing UEs. This can be justified that for any communication system with a certain bandwidth, there is a maximum channel capacity regardless the number of users in that area. Furthermore, when the UAV battery capacity is increased to 200 Wh, the system achieves a higher throughput with a lower number of users compared to UAV with a battery capacity of 100 Wh. However, this curve tends to saturation at larger number of users due to the previous fact.

In Fig. 3, the total energy consumption of UEs is compared to the number of available UEs in the post-disaster area. It can be observed that with the increase in UEs number, the total energy consumption is increased as well. However, for a lower UAV battery capacity, i.e., 100 Wh, there is no considerable increase in the total UEs power consumption when the number of UEs are increased from 40 to 50. This can be described as the UAV battery cannot support a longer flight distance to cover the increased number of UEs that are spread in the post-disaster area. So, for 50 UEs, it seems some of them do not have a chance to offload their data and stay in idle mode which is reflected in decreasing in the total power consumption of UEs. However, for a UAV battery capacity equal to 200 Wh, the total power consumption of UEs is still increasing when the number of UEs reaches 50. In this situation, the larger UAV battery capacity could support a longer flight distance which is reflected in the increase of the total UEs power dissipation as well.

V. Conclusions

In this paper, we have studied trajectory optimization for a UAV-assisted emergency wireless communication network. The UAV is deployed to provide emergency wireless connectivity from the sky to ground UEs when ground BSs malfunction as a result of the damage caused by a natural disaster. The UAV tried to optimize its flying trajectory to maximize the cumulative uplink throughput. However, due to the energy limitation for both the UAV and UEs, the optimization problem is constrained by this dynamic energy consumption over time. We proposed a dual-energy-aware based multi-armed bandit framework to tackle this constrained problem. The proposed framework could optimize the UAV flight trajectory with respect to the limited available energy for both the UAV and UEs. In addition, it could maximize cumulative throughput and accomplish the task of offloading UEs in the post-disaster area.

Acknowledgement

This work was supported in part by a research study grant from the Telecommunications Advancement Foundation.

References

[1] R. Munich, "Natcatservice loss events worldwide 1980–2014," *Munich Reinsurance: Munich, Germany*, vol. 10, 2015.

[2] G. Deepak, A. Ladas, Y. A. Sambo, H. Pervaiz, C. Politis, and M. A. Imran, "An overview of post-disaster emergency communication systems in the future networks," *IEEE Wireless Communications*, vol. 26, no. 6, pp. 132–139, 2019.

[3] M. Erdelj, E. Natalizio, K. R. Chowdhury, and I. F. Akyildiz, "Help from the sky: Leveraging uavs for disaster management," *IEEE Pervasive Computing*, vol. 16, no. 1, pp. 24–32, 2017.

[4] A. Kwasinski, W. W. Weaver, P. L. Chapman, and P. T. Krein, "Telecommunications power plant damage assessment for hurricane katrina–site survey and follow-up results," *IEEE Systems Journal*, vol. 3, no. 3, pp. 277–287, 2009.

[5] A. Fotouhi, H. Qiang, M. Ding, M. Hassan, L. G. Giordano, A. Garcia-Rodriguez, and J. Yuan, "Survey on uav cellular communications: Practical aspects, standardization advancements, regulation, and security challenges," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3417–3442, 2019.

[6] M.-A. Lahmeri, M. A. Kishk, and M.-S. Alouini, "Artificial intelligence for uav-enabled wireless networks: A survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1015–1040, 2021.

[7] M. T. Mamaghani and Y. Hong, "Intelligent trajectory design for secure full-duplex mimo-uav relaying against active eavesdroppers: a model-free reinforcement learning approach," *IEEE Access*, vol. 9, pp. 4447–4465, 2020.

[8] Y. Zeng, X. Xu, and R. Zhang, "Trajectory design for completion time minimization in uav-enabled multicasting," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2233–2246, 2018.

[9] R. Sugihara and R. K. Gupta, "Speed control and scheduling of data mules in sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 7, no. 1, pp. 1–29, 2010.

[10] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3039–3071, 2019.

[11] B. Pearre and T. X. Brown, "Model-free trajectory optimisation for unmanned aircraft serving as data ferries for widespread sensors," *Remote Sensing*, vol. 4, no. 10, pp. 2971–3005, 2012.

[12] H. Bayerlein, P. De Kerret, and D. Gesbert, "Trajectory optimization for autonomous flying base station via reinforcement learning," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2018, pp. 1–5.

[13] S. Yin, S. Zhao, Y. Zhao, and F. R. Yu, "Intelligent trajectory design in uav-aided communications with reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 8227–8231, 2019.

[14] A. Amrallah, E. M. Mohamed, G. K. Tran, and K. Sakaguchi, "Enhanced dynamic spectrum access in uav wireless networks for post-disaster area surveillance system: a multi-player multi-armed bandit approach," *Sensors*, vol. 21, no. 23, p. 7855, 2021.

[15] ——, "Radio resource management aided multi-armed bandits for disaster surveillance system," in *Proc. 2020 International Conference on Emerging Technologies for Communications (ICETC2020), Virtual, K1-4*, 2020.

[16] E. M. Mohamed, S. Hashima, A. Aldosary, K. Hatano, and M. A. Abdelghany, "Gateway selection in millimeter wave uav wireless networks using multi-player multi-armed bandit," *Sensors*, vol. 20, no. 14, p. 3947, 2020.

[17] E. M. Mohamed, S. Hashima, and K. Hatano, "Energy aware multi-armed bandit for millimeter wave based uav mounted ris networks," *IEEE Wireless Communications Letters*, 2022.

[18] J. Cui, Z. Ding, Y. Deng, A. Nallanathan, and L. Hanzo, "Adaptive uav-trajectory optimization under quality of service constraints: A model-free solution," *IEEE Access*, vol. 8, pp. 112 253–112 265, 2020.

[19] T. Zhang, J. Lei, Y. Liu, C. Feng, and A. Nallanathan, "Trajectory optimization for uav emergency communication with limited user equipment energy: a safe-dqn approach," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 3, pp. 1236–1247, 2021.

[20] 3GPP, "Study on enhanced lte support for aerial vehicles (release 15)," 2017.

[21] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2, pp. 235–256, 2002.

[22] J.-Y. Audibert, R. Munos, and C. Szepesvári, "Exploration–exploitation tradeoff using variance estimates in multi-armed bandits," *Theoretical Computer Science*, vol. 410, no. 19, pp. 1876–1902, 2009.

[23] D. Sinha, K. A. Sankararaman, A. Kazerouni, and V. Avadhanula, "Multi-armed bandits with cost subsidy," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 3016–3024.