

# UAV-Assisted Emergency Communications: An Extended Multi-Armed Bandit Perspective

Yu Lin, Tianyu Wang, *Member, IEEE*, and Shaowei Wang<sup>✉</sup>, *Senior Member, IEEE*

**Abstract**—In this letter, we investigate how a rotary-wing unmanned-aerial vehicle (UAV) acts as a wireless base station to provide emergency communication service for a post-disaster area with unknown user distribution. The formulated optimization task is to find out the optimal path starting and ending at the same point to serve as many as possible users under limited battery capacity. We show that this problem can be transformed to an extended multi-armed bandit (MAB) problem, for which we propose two efficient path planning algorithms. Simulation results show that, in terms of the total number of served users, our proposed algorithms outperform the straightforward helical path that scans the entire area by circles with increasing radius.

**Index Terms**—Emergency communications, multi-armed bandit, online learning, unmanned-aerial vehicle.

## I. INTRODUCTION

UNMANNED-AERIAL vehicle (UAV) has called people's attention due to its advantages over traditional aircrafts. Compared to the manned aircraft, UAVs can perform high-risk and long-time missions which are too dangerous or too heavy for human pilots, and thus, they originated mostly in military applications such as reconnaissance, strikes, and communications. Also, UAV is much cheaper to operate and maintain than the manned aircraft, which helps the UAV to develop many novel applications in the commercial world, the scientific field and the recreational domain, e.g., public security monitoring, crop monitoring and irrigation, data collection, aerial photography and drone racing.

Recently, UAV has been introduced in wireless communication systems for dynamic relaying, large-area environmental sensing and emergency communications, which are referred to as the UAV-assisted networks. In the environmental sensing systems, the UAVs are utilized as data collectors that fly by a large area of depopulated zone to collect data of the environmental sensors deployed in this zone [1]. In mobile relaying systems, UAVs are utilized as dynamic relays to replace fixed relay nodes such that the relay locations can be

dynamically adjusted to match communication scenarios as well as possible [2]–[4]. In the emergency communications systems, UAVs can be utilized as the platforms that are equipped with small base stations (BSs) to deliver broadband wireless services for hotspots with temporary events such as assemblies and sports games, as well as for post-disaster areas when the existing communications infrastructure is damaged by flood, hurricane or earthquake [5]–[7].

In the post-disaster emergency communications, a rotary-wing UAV can hover above the target area to serve mobile users within a certain horizontal distance. In order to cover a large post-disaster area, the UAV needs to set a path that covers the target area and sequentially serves mobile users along its path. In [8], the total power consumption of the UAV is minimized by jointly designing the scheduling scheme of the sensor node, the power allocation strategy and the flight trajectory. In [9], a UAV-mounted BSs deployment scheme is proposed to minimize the number of BSs needed to provide wireless coverage for a group of distributed ground terminals. According to the investigations of the existing work, the major challenge of UAV path planning in the emergency communications is that the user distribution in the target area is unknown by the UAV. Therefore, the UAV can not pre-plan an optimal path that maximizes the number of served users. Also, due to the limited battery capacity of UAV, an energy-efficient trajectory is critical for the UAV-assisted emergency communication [10].

In this letter, we develop an online learning framework to characterize the UAV path planning problem for the post-disaster emergency communication where the UAV acts as a BS to serve mobile users along its path. The optimization task is formulated as an extended multi-armed bandit (MAB) problem [11] and we propose two path planning schemes to gradually learn an energy-efficient path while serving the mobile users within the coverage of the UAV. Simulation results show that our proposed algorithms outperform the baseline helical path one by 85% and 100%, respectively, in terms of the number of served users.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

As shown in Fig. 1, a single rotary-wing UAV serves as a mobile BS platform to serve a post-disaster area, where the entire ground infrastructure is shut down due to earthquake or flood. The UAV takes off at the center point of the area and provides the area radio access service while flying. It comes back to the center point to recharge before the battery is used up. We assume that the UAV flies to a point and then hovers above the point to serve mobile users within horizontal distance  $d$ . Thus, the UAV needs to fly and hover alternatively to cover the post-disaster area. The flying speed of the UAV

Manuscript received January 21, 2019; revised February 28, 2019; accepted March 10, 2019. Date of publication March 19, 2019; date of current version May 8, 2019. This work was partially supported by the National Natural Science Foundation of China (61801208, 61671233), the Jiangsu Science Foundation (BK20170650), the Postdoctoral Science Foundation of China (BX201700118, 2017M621712), the Jiangsu Postdoctoral Science Foundation (1701118B), and the open research fund of National Mobile Communications Research Laboratory (2019D02). The associate editor coordinating the review of this letter and approving it for publication was K. Rajawat. (Corresponding author: Shaowei Wang.)

Y. Lin and S. Wang are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: wangsw@nju.edu.cn).

T. Wang is with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China, and also with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: tianyu.alex.wang@nju.edu.cn).

Digital Object Identifier 10.1109/LCOMM.2019.2906194

1558-2558 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.  
See [http://www.ieee.org/publications\\_standards/publications/rights/index.html](http://www.ieee.org/publications_standards/publications/rights/index.html) for more information.

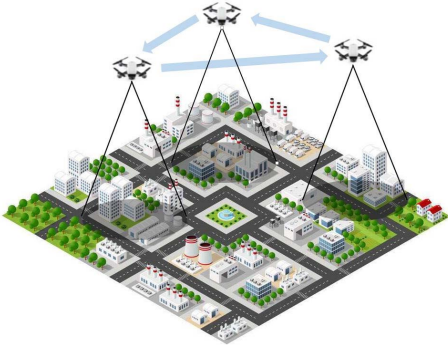


Fig. 1. UAV-assisted emergency communication, where the hovering UAV is utilized as a wireless BS covering a certain area.

is  $V_f$  and the hovering interval is  $T_h$ . The battery capacity is  $E$ , and the average engine power for hovering and flying is given by  $e_h$  and  $e_f$ , respectively.

To simplify analysis, we divide the entire area into  $K \times K$  equal grids and assume that the UAV can serve all users within the grid when it is hovering above the center of the grid. The set of all grids is denoted by  $\mathcal{N} = \{1, 2, \dots, K^2\}$ . We denote by  $s_p = \{n_1, n_2, \dots, n_{t_p-1}, n_{t_p}\}$  as a UAV path, where each element  $n_i \in \mathcal{N}$  represents a grid in this area. The path starts at grid  $n_1$ , serves  $t_p$  grids, and returns to grid  $n_1$  to recharge. Also, we denote  $\mathcal{S}$  as the set of all possible paths that start and end at the center grid. A fixed total number of users are located randomly in the entire area and the users need radio access service with the same probability. The traffic demand of mobile users within grid  $n_i$  is denoted by  $r_{n_i}$ , which is a random variable unknown to the UAV. A user usually needs to be served only once in the real post-earthquake scenario so serving a few grids repeatedly would cause power waste. Then we can assign a weight factor to the traffic of the visited grids before to reduce the probability of visiting the grids again. To simplify notations, we define  $n_{t_p+1} \doteq n_1$ . Therefore, the key issue of UAV-assisted emergency communication is to find the most efficient path that provides communication service to as many as possible users while satisfying the battery constraint. The optimization task can be formulated as follows:

$$\begin{aligned} \max_{s_p \in \mathcal{S}} \quad & \sum_{t=1}^{t_p} r_{n_t} \\ \text{s.t.} \quad & t_p e_h T_h + \sum_{t=1}^{t_p} e_f \frac{d_{n_t, n_{t+1}}}{V_f} \leq E, \end{aligned} \quad (1)$$

where  $d_{n_t, n_{t+1}}$  represents the distance between the centers of grid  $n_t$  and grid  $n_{t+1}$ .

### III. PATH PLANNING ALGORITHMS

#### A. Analysis of Our Problem

We first give a brief description of the MAB problem. There exists a bandit with multiple arms denoted by  $\mathcal{J} = \{1, 2, \dots, J\}$ . Each arm provides a reward drawn from a distribution associated to it. At round  $t$ , a gambler pulls one of the arms, e.g. arm  $i$ , and gets a reward,  $x_i(t)$ . Initially, the gambler has no knowledge about the rewards of the arms.

The goal of the gambler is to maximize the total reward of all trials. Therefore, the gambler needs to explore different arms to find the most valuable one, while at the same time, exploit the most valuable arm as many times as possible so as to maximize the total reward. Thus, an efficient algorithm for the MAB problem should be able to balance between exploration and exploitation.

Then we show that Eq. (1) defines an extended MAB problem: The arms are the grids, the gambler is the UAV, and the reward is the number of users served by the UAV hovering over the grids. The difference between our problem and the MAB is as follows: The gambler in the MAB problem can pull each arm arbitrarily while the UAV needs to fly to different grids to serve mobile users, introducing a battery consumption that depends on the distance between its current location and the target grid. We extend two classic algorithms for the MAB problem to address our formulated UAV path planning task.

#### B. Distance-Aware Upper Confidence Bound Algorithm

For the upper confidence bound (UCB) algorithm, the gambler first pulls each arm once. Then at any round  $t > |J|$ , the gambler pulls arm  $j^* \in \mathcal{J}$  that satisfies

$$j_{UCB}^* = \arg \max_{j \in \mathcal{J}} \{\bar{x}_j(t) + \sqrt{2 \ln t / T_{j,t}}\}, \quad (2)$$

where  $\bar{x}_j(t)$  is the average reward obtained from arm  $j$  at round  $t$ , and  $T_{j,t}$  is the number of times that arm  $j$  has been pulled. If an arm has been pulled too many times, that is,  $T_{j,t}$  is large, the confidence interval  $\sqrt{2 \ln t / T_{j,t}}$  decreases and the gambler intends to explore other less pulled arms. When an arm achieves a high reward in the past, that is,  $\bar{x}_j(t)$  is large, the gambler intends to exploit this high-profit arm to yield the possible maximal reward. An efficient trade-off between exploration and exploitation can be achieved in this way.

We propose a distance-aware UCB (D-UCB) algorithm to increase the battery usage efficiency of the UAV so that it can serve more users in its entire trajectory. Introducing two new terms  $\beta d_{n_{t-1}, n}$  and  $\gamma d_{n_1, n} / B_r$  into Eq. (2), to represent the flight cost and the remaining battery power for the UAV to visit the next grid, we drop the first  $K^2$  trials to set an initial reward for each grid as the UCB does. Then we check the remaining battery power at each round  $t$ . If the remaining battery power is sufficient for a trip, the UAV chooses the next grid  $n_{D-UCB}^*$  from  $\mathcal{N} = \{1, 2, \dots, K^2\}$  satisfying

$$\begin{aligned} n_{D-UCB}^* \\ = \arg \max_{n \in \mathcal{N}} \{ \bar{r}_{n,t}^* + \alpha \sqrt{\frac{2 \ln t}{T_{n,t}}} - \beta d_{n_{t-1}, n} - \gamma \frac{d_{n_1, n}}{B_r} \}, \end{aligned} \quad (3)$$

where  $\bar{r}_{n,t}^*$  represents the average reward of grid  $n$ ,  $T_{n,t}$  represents the times that the UAV visited grid  $n$ ,  $d_{n_{t-1}, n}$  is the distance between the next grid and the current one,  $d_{n_1, n}$  is the distance between the next grid and the final recharging grid, and  $B_r$  is remaining battery power.

The introduced flight cost term prompts the D-UCB to choose the nearby grids so that the flight energy consumption can be reduced. Then more energy can be used for hovering and serving users. The remaining power term prompts the

TABLE I  
SIMULATION PARAMETERS

$\alpha$	1
$\beta$	2
$\gamma$	1
$\epsilon$	0.2
$\tau$	1
$\theta$	0.5
Area of grid	100m × 100m
Average engine power for hovering	$e_h = 4$
Average engine power for flying	$e_f = 2$
Cover range of the UAV	100m
Flying speed	$V_f = 20\text{km/h}$
Hover interval	$V_h = 120\text{s}$
User density	48 users per grid

D-UCB algorithm to choose the grids near the recharging point so as to return to the point in time before the battery is used up.

C.  $\epsilon$ -Exploration Algorithm

For the  $\epsilon$ -exploration algorithm, the gambler pulls an arm with probability  $\epsilon$ , or pulls the arm that yields the highest average reward with probability  $1 - \epsilon$ . After checking the remaining battery power at each round  $t$ , with probability  $\epsilon$ , the UAV chooses the grid  $n_{\epsilon-e}^*$  from  $\mathcal{N} = \{1, 2, \dots, K^2\}$  according to a softmax function which converts the rewards into probabilities

$$p(n) = \frac{e^{\frac{\bar{r}'_{n,t}}{\tau}}}{\sum_{i=1}^{K^2} e^{\frac{\bar{r}'_{i,t}}{\tau}}}, \tag{4}$$

and with probability  $1 - \epsilon$ , the UAV chooses the grid  $n_{\epsilon-e}^*$  from  $\mathcal{N} = \{1, 2, \dots, K^2\}$  satisfying

$$n_{\epsilon-e}^* = \arg \max_{n \in \mathcal{N}} \{\bar{r}'_{n,t} - \theta d_{n_{t-1},n}\}, \tag{5}$$

where  $\tau (>0)$  is the temperature parameter of the softmax function. For high temperatures, e.g.,  $\tau \rightarrow \infty$ , all grids have nearly the same probability to be selected; and the lower the temperature, the more average rewards affect the probability. For low temperature, e.g.,  $\tau \rightarrow 0$ , the probability of selecting the grid with the highest average reward tends to 1. We can see that the  $\epsilon$ -exploration intends to choose the nearby grids by introducing the flight cost term to Eq. (5).

IV. SIMULATIONS

We compare the performance of our proposed algorithms with the helical path one which scans the entire area by circles with increasing radius and backtracks to the starting point. The users are randomly dropped and the traffic in a grid follows Binomial distribution  $B(u_n, p)$ , where  $u_n$  is the number of users in grid  $n$  and  $p$  is the possibility that the user needs radio access service and is set to 0.2 in all simulations. The weight factor of the visited grids is set to 0.2. The key parameters of the proposed algorithms are selected via a series of numerical simulations, all of which are given in Table I.

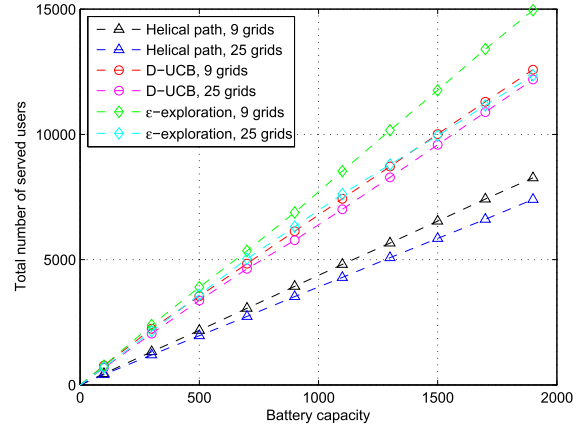


Fig. 2. Total number of served users as a function of battery capacity.

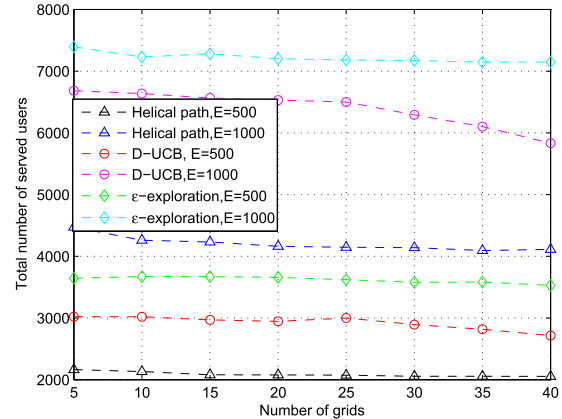


Fig. 3. Total number of served users as a function of grids.

Figure 2 shows the total number of served users as a function of battery capacity. It can be seen that the total number of served users is roughly in a linear relationship with the battery capacity. For all three algorithms, the total number of served users for the case of 9 grids is larger than that of 25 grids. The reason is that the UAV is more likely to use more energy to hover for communication when the number of grids is small. The D-UCB and the  $\epsilon$ -exploration outperform the helical path algorithm by 85% and 100%, respectively, in terms of the total number of served users when the service area is divided into 25 grids and the battery capacity is 1000.

In Fig. 3, we show the total number of served users as a function of the number of grids. The total number of served users decreases slightly as the number of grids becomes larger, especially for the D-UCB. Though visiting more grids could potentially serve more users, it inevitably leads to consuming more power on flying, not on hovering for communication. From Fig. 2 and Fig. 3 we can observe that the  $\epsilon$ -exploration can serve more users than that of the D-UCB.

Figure 4 shows the percentage of hover battery consumption and the visited grids as the function of grids, where the battery capacity is 1000. As can be seen from Fig. 4(a), more than 90% power of the UAV is used to hover and communication for the proposed D-UCB and  $\epsilon$ -exploration algorithms. In contrast, less than 70% power of the UAV is used to hover for the helical path scheme when the number of grids is

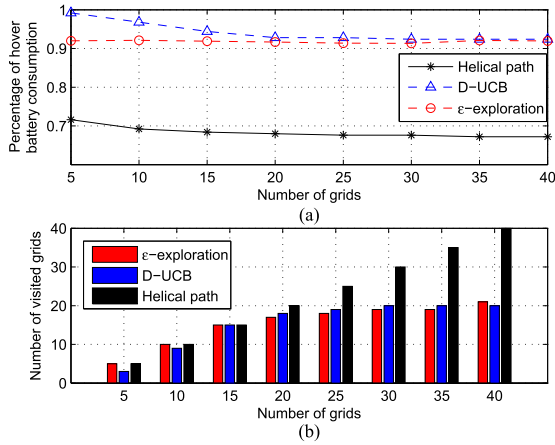


Fig. 4. (a) Hover battery consumption; (b) Number of visited grids.

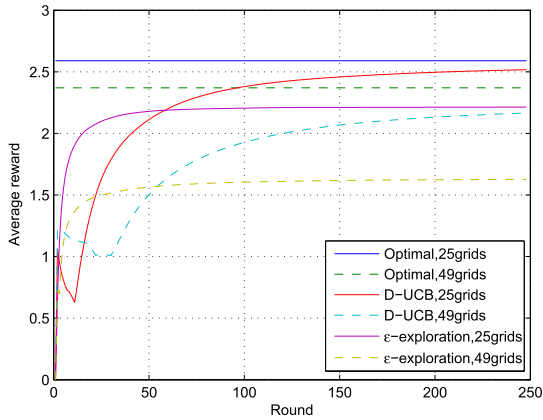


Fig. 5. The average reward as a function of round.

more than 10. We can see from Fig. 4(b) that the number of visited grids of the D-UCB and the  $\epsilon$ -exploration only changes slightly when the number of grids is more than 20, which means that the two algorithms are always weighted in visiting the promising grids even though the service area is large.

Figure 5 shows the convergence of the D-UCB and the  $\epsilon$ -exploration, where the optimal is determined with priori knowledge. We can see that the  $\epsilon$ -exploration algorithm converges much faster than the D-UCB algorithm. As the increasing of the number of grids, both of the algorithms need more rounds for convergence. As for the reward, it can be seen from Fig. 5 that the D-UCB yields more reward than the  $\epsilon$ -exploration. However, the complexity of the D-UCB is higher than that of the  $\epsilon$ -exploration algorithm. Choosing which one

for application should be in accordance with the practical requirement so as to achieve the tradeoff between reward and complexity.

## V. CONCLUSION

In this letter, we have investigated the UAV-assisted emergency communication for post-disaster areas. The considered optimization task is transformed into an extended MAB problem, for which we have proposed the D-UCB algorithm and the  $\epsilon$ -exploration algorithm to solve efficiently, respectively. Our proposed algorithms can explore unknown user distribution and serve as many as possible users with limited battery capacity. Simulation results show that the proposed algorithms outperform the baseline helical path in terms of the total number of served users.

## REFERENCES

- [1] N. Goddemeier, K. Daniel, and C. Wietfeld, "Role-based connectivity management with realistic air-to-ground channels for cooperative UAVs," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 5, pp. 951–963, Jun. 2012.
- [2] Y. Zeng, R. Zhang, and T. J. Lim, "Throughput maximization for UAV-Enabled mobile relaying systems," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 4983–4996, Dec. 2016.
- [3] F. Jiang and A. L. Swindlehurst, "Optimization of UAV heading for the ground-to-air uplink," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 5, pp. 993–1005, Jun. 2012.
- [4] Y. Sun, T. Wang, and S. Wang, "Location optimization for unmanned aerial vehicles assisted mobile networks," in *Proc. IEEE Int. Conf. Commun.*, May 2018.
- [5] J. Wang, C. Jiang, Z. Han, Y. Ren, R. G. Maunder, and L. Hanzo, "Taking drones to the next level: Cooperative distributed unmanned-aerial-vehicular networks for small and mini drones," *IEEE Veh. Technol. Mag.*, vol. 12, no. 3, pp. 73–82, Sep. 2017.
- [6] N. Rupasinghe, A. S. Ibrahim, and I. Guvenc, "Optimum hovering locations with angular domain user separation for cooperative UAV networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016.
- [7] A. Merwaday, A. Tuncer, A. Kumbhar, and I. Guvenc, "Improved throughput coverage in natural disasters: Unmanned aerial base stations for public-safety communications," *IEEE Veh. Technol. Mag.*, vol. 11, no. 4, pp. 53–60, Dec. 2016.
- [8] M. Hua, Y. Wang, Z. Zhang, C. Li, Y. Huang, and L. Yang, "Power-efficient communication in UAV-Aided wireless sensor networks," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1264–1267, Jun. 2018.
- [9] J. Lyu, Y. Zeng, R. Zhang, and T. J. Lim, "Placement optimization of UAV-Mounted mobile base stations," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 604–607, Mar. 2017.
- [10] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.
- [11] J. Vermorel and M. Mohri, "Multi-armed bandit algorithms and empirical evaluation," in *Proc. Eur. Conf. Mach. Learn.*, 2005, pp. 437–448.