Time-adapted Early Arrival Path for Drone Parcel Delivery through Public Transportation Vehicles: Using Q-learning.

Mohammed Rahmani[1], Florian Delavernhe[1], Sidi Mohammed Senouci[1], Marion Berbineau[2]

[1] DRIVE EA1859, Univ. Bourgogne Franche Comté, F58000, Nevers, France
{Mohammed.Rahmani,Florian.Delavernhe, sidi-mohammed.senouci}@u-bourgogne.fr
[2] Univ Gustave Eiffel, COSYS, F-59650 Villeneuve d'Ascq, France
marion.berbineau@univ-eiffel.fr

Abstract : *Drone-based delivery has gained popularity in recent years, and many different delivery systems and schemes have been proposed. One of the most promising scheme concepts is based on the collaboration between a drone and a public transportation network to expand the delivery range while conserving drone battery energy and reducing delivery costs. Path planning is the main problem with this design, as the public transportation network is stochastic and time-dependent. In this paper, an inspection time-adapted early arrival path problem is formulated, which seeks the path for a drone that ensures: (i) reaching the customer as soon as possible, (ii) adapting to random fluctuations in public transportation schedules, and (iii) taking into acount the battery consumption of the drone. To achieve these requirements, a Q-learning-based planning method is proposed. The simulation results validate the effectiveness and feasibility of Q-learning on the planning path for parcel deliveries: at any departure instant, the arrival of the drones at the customer's location was guaranteed, i.e., the resulting path is 100% reliable. In addition, the convergence of the Q-Learning algorithm was reached after only 1000 learning epochs. Furthermore, the experimental results show that the Q-Learning solution can achieve a lower early arrival time and lower power consumption compared to another algorithm.*

Keywords : *Drone delivery, Early Arrival Path, Energy, Q-learning, Public transportation.*

1 Introduction

With the increasing involvement of drones in civilian domains such as surveillance [5], wireless communications [2], and package delivery [4], several new research directions have been opened. Recently, many researchers and logistics companies have been interested in researching and testing unmanned aerial vehicles (UAVs) in last-mile delivery systems [4]. As a result, several schemes based on drones have been proposed with various benefits and drawbacks. A first scheme focuses on using drones only to deliver parcels from the warehouse to the customer [1], thus resulting with a quick delivery process with little human labor. However, the delivery range is limited due to the restricted battery of the drone. To increase delivery range, a new scheme (see [3]) used trucks to transport the drones to specific points near the customers, and then the drones took off from the trucks to deliver parcels to customers. Obviously, the primary drawbacks of this scheme are the high operating costs induced by truck driver expenses and fuel costs. A third method proposed to avoid these impacting costs as well as increase the delivery area of drones, is to utilizing public transportation vehicles as drone transporters during daily journeys [6]. The present work focuses on this scheme.

The main challenge facing this scheme is path planning. Indeed, public transportation vehicles have pre-determined routes and timetables that logistics companies cannot control. Additionally, the travel time of a vehicle between two stations depends on the instant of departure and is stochastic as it is affected by time uncertainties such as traffic congestion, vehicle breakdowns, etc. Therefore, computing a path offline (*i.e.*, before the drone leaves the warehouse) is far from optimal due to the stochastic nature of the transportation network.

Online path planning is a more efficient option and it is based on the drone communicating with public transportation and collecting real-time information to adapt its path. Logistically, many suppliers prefer to deliver their parcels as quickly as possible to satisfy their customers, which leads to the problem of planning an early arrival path rather than the shortest path. Alternatively, they also want to serve all consumers, including those far from the warehouse; therefore, the delivery path for them will be longer, resulting in more energy consumption. Therefore, the capacity of the drone battery must be taken into account during the path planning.

To the best of our knowledge, a path planning problem that considers the stochastic time-dependent public transportation network and the optimization of both the delivery time and the drone battery life capacity together has not been examined. In this paper, we address the time-adapted early arrival path planning for drone parcel delivery via public transportation. Q-Learning is a popular method for solving this type of path planning problem. The reason of this is its ability to solve optimization problems in an unmodeled environment, as well as its adaptability to change the environmental conditions. The idea is to obtain the optimal policy by selecting the best actions to reach a defined location based on the observed state of the environment as well as on the accumulated historical experience. Therefore, we are motivated to use Q-learning to find the earliest arrival path for the drone from the warehouse to the customer, with the goal of minimizing arrival time to the customer and energy consumption. The main contributions of this paper are summarized as follows: **(i)** We introduce a new problem statement regarding the determination of the time-adapted earliest arrival path in a stochastic time-dependent parcel delivery system that combines a public transportation network and a drone, **(ii)** We formulate a path planning problem as a multiobjective problem, in which one must search for the optimal path that minimizes both the arrival time to reach the customers and the amount of energy consumed, **(iii)** The Q-learning algorithm is proposed and tested to validate its efficiency and feasibility on the planning path compared to the random approach

The remainder of this paper is organized as follows. The system model and problem statement are described in Section 2. Section 3 presents the proposed Q-learning algorithm for finding the earliest-arrival path. In Section 4, simulation results are analyzed and conclusions of the paper are drawn in 5.

2 System Model and Problem Statement

Let $G(V, E, \Omega)$ denotes the system model, where V is the set of nodes that comprise transportation stops, warehouses, and customers. E is the set of links between pairs of nodes, which represent the links between two stops through the lines of the transportation network, or direct links by drone fly between the various nodes. $E = \{\prec v, v', l_i, t_j \succ | v, v' \in V, l_i \in L, t_j \in T^{l_i}\}$ where l_i denote the public transport line number i, l_0 the direct unmanned flight and L is the set of these lines. Whereby, t_j denotes trip number j on one of the lines (there are several trips per day on each line). We denote T^{l_i} as the set of trips on line l_i. Let Ω denotes a set of weights between pairs of nodes. It represents the traversal time between nodes: $\Omega = \{\omega_{vv'}^{l_i t_j} | v, v' \in V, l_i \in L, t_j \in T^{l_j}\}$. Additionally, for the traversal time, we define two functions that represent the arrival and departure time of trip t_j of line l_i at node v, $\alpha(l_i, t_j, v)$ and $\tau(l_i, t_j, v)$, respectively. (when next trip is l_i=l_0 then $\tau - \alpha = 0$ i.e. no waiting time)

Définition 1 *We define path p_{uv} between pair of nodes $u = n_0, v = n_m \in V$ in G as a sequence of links $(n_z, n_{z+1}, l_i, t_j) \in E$ as: $p_{uv} = \Big\{(u, n_1, l_i, t_j), (n_1, n_2, l_{i'}, t_{j'})..(n_{m-1}, v, l_m, t_k)\Big\}$ where each consecutive couple of links in this path should satisfies (1).*

$$\forall (n_z, n_{z+1}, l_i, t_j), (n_{z+1}, n_{z+2}, l_{i'}, t_{j'}) \in p_{uv} : \alpha(l_i, t_j, v') \leq \tau(l_{i'}, t_{j'}, v') \tag{1}$$

In the rest of the paper, we define P_{uv} as set of all possible paths from u to v.

2.1 The earliest-arrival path

When dealing with a drone-based delivery system, the goal is to get parcels to customers as early as possible, which raises the issue of calculating the shortest path. However, in a system that combines public transportation and drones, our problem is not only dependent on traversal time, but also on the departure and arrival times of transportation vehicles. As a result, in this work, we addressed the earliest-arrival path problem.

Définition 2 *The earliest-arrival path p^*_{uv} is the path that provides the earliest arrival time from a source u to a target v. i.e., the following condition is satisfied:*

$$\forall p \in P_{uv} : \alpha(l^*, t^*, v) \leqslant \alpha(l, t, v) | t^* \in p^*_{uv}, t \in p_{uv} \tag{2}$$

2.2 Drone energy model

For the energy consumption model, we consider hovering and flight as the main phases in which the drone's energy is consumed, while energy consumption is assumed to be zero when the drone travels on top of a public vehicle. Note that the controller's power usage is ignored. Let e_f and e_h the energy consumed by the drone while flying and hovering, respectively. Then, we can associate the drone's energy consumption with traveling time. For that, let e_p denote the energy consumption along a path, which can be calculated as follows:

$$e_p = \sum_{(v,v',l_i,t_j) \in p_{uv}} (\tau(l_{i'}, t_{j'}, v) - \alpha(l_i, t_j, v))e_h + \omega^{lt}_{vv'} e_f \tag{3}$$

Where the first term represent the energy spent on the waiting time for the trip t_j of line l_i at station v, equal to zero if $l = l_0$. Whilst the last term represents the energy consumed during the traveling between v and v' through a trip t_j of line l_i , equal to zero if $l \neq l^0$.

2.3 Problem statement

Now, for a given network $G(V, E, \Omega)$, t_0 (instant when the customer makes order), and E_0 (initial energy), the problem can be expressed as finding a path from the warehouse to the customer such that the earliest arrival time and total consumed energy are minimized. To do so, we use a weighted sum of these objectives and define the optimal path in terms of the equations 2 and 3.

In his work, the optimization objective is:

$$p^* = \arg\min_{P_{uv}} (\gamma \alpha(., ., v) + (1 - \gamma)e_p) \tag{4}$$

where P_{uv} is all possible paths from u to v. γ is a user-defined parameter, $\gamma \in [0, 1]$, such that γ is used to determine a trade-off between energy and early arrival time. The larger the γ is, the more the minimum arrival time is favored.

3 Q-learning for Time-adapted Early Arrival Path

In the literature, dynamic programming methods are generally used to exhaustively search for the optimal path. In such works, the optimal path is calculated based on a priori information (previously provided). It is known as offline path planning. However, when the uncertainty and time dependency of public transportation, as well as the lack of information, are considered, these methods become inefficient in such scenarios. In the present work, we used the Q-learning algorithm to find the earliest arrival path for the drone. The main idea is to give the drone the ability to determine its path and update it according to the information available in real time, adapted to random changes in public transportation, which is known as online path planning.

3.1 The idea of Q-learning

Due to the capability of Q-learning to solve optimization problems without relying on the environment model, we are motivated to use it for online path planning without the need to model the uncertainty of public transportation times. In this work, we use a drone as an agent to find the possible paths between the warehouse and the customer. It seeks to minimize arrival time and the energy consumption. The Markov Decision Process (MDP) suitable for drone path planning in our system can be defined by the following tuple $< S, A, T, R >$. Among them, S is a finite set of states that represents stops, warehouses, and consumers. A is a finite set of actions described by tuple $< line, trip >$, that represents transportation trips and flying trips. T is the transition probability function, $T : S \times A \times S \rightarrow [0, 1]$, is the probability of a drone to take $a =< l, t >$ to move from state s_k to state s_{k+1}. R is the reward function will be explained in 3.2.

3.2 Reward and Q-value functions

The reward function is a real-time reward. After the drone performs an action, *i.e.*, choosing a trip, the environment generates feedback on the chosen trip that is used to evaluate the performance of the action. The reward function is the combination between energy consumption and the related expected waiting and traversal times. It is designed as follows:

$$r = \begin{cases} \gamma(W + T) + (1 - \gamma)e_h \times W & if \quad l_i \neq l_0 \\ \gamma T + (1 - \gamma)e_f \times T & if \quad l_i = l_0 \end{cases} \tag{5}$$

where W and T are the waiting time and traversal time of trip t_j of line l_i, respectively.These can be calculated as follows:

$$W = \tau(\prec l_i, t_j \succ, s_k) - \alpha(\prec l_{i''}, t_{j''} \succ, s_k), \qquad T = \alpha(\prec l_i, t_j \succ, s_{k+1}) - \tau(\prec l_i, t_j \succ, s_k),$$

Equation 5, when $l_i \neq l_0$, means that the drone uses vehicles to travel as action, and the related reward ignores the part corresponding to energy consumption during the fly, whereas when $l_i = l_0$, the drone performs flying travel trip as action, and the related reward ignores the part corresponding to energy consumption during waiting time and the waiting time.

Similarly, the Q-value function is important in the Q-Learning method for problem solving. In this paper, we use random values to initialize the Q-value at the beginning of learning. Hence, during the learning progress in each action-selection the Bellman equation 7 is used to update the estimated Q-value:

$$Q_{new}\left(s_k, a\right\rangle = Q_{old}\left(s_k, a\right\rangle + \lambda \left[r + \sigma \min_{a'} Q_{old}\left(s_{k+1}, a'\right\rangle - Q_{old}\left(s_k, a\right\rangle)\right] \tag{7}$$

where a and a' the current action and future action, respectively. $Q_{old}\left(s_{k+1}, a'\right\rangle$ is the old value, $\min_{a'} Q_{old}\left(s_{k+1}, a'\right\rangle$ is the estimate of optimal future value, λ and σ are the learning rate and discount factor of the Q-learning algorithm, respectively. At the end of training, the optimal Q-value function $Q^*(s, \prec l_i, t_j \succ)$ (as given by eq. 8) is the maximum action-value function over all policies.

$$Q^*(s, \prec l_i, t_j \succ) = \min_{a'} Q(s_{k+1}, a') \tag{8}$$

Finally, the ϵ-greedy approach was used during the deployment of the Q-learnig learned model to maintain a balance between exploration and exploitation. In other words, we use a random generated value $\in [0, 1[$ and compare it to pre-defined ϵ ($\epsilon = 0.1$ in our work). When a generated value is greater than ϵ, the action that corresponds to maximum Q-value be chosen, rather than a random action. In this manner, the discrepancy between local optima is avoided.

4 Simulation Results

To demonstrate the effectiveness of the proposed Q-Learning for the time-adapted early arrival path, we use the simple network illustrated in Fig. 1. This network is made up of 5 public transportation stations, 1 warehouse, 1 customer, and 5 public transportation line (the label 0 indicates the drone fly). Table 1 provide the mean and standard deviations of the link traversal and departure instants times. The smaller deviations are related to l_0 as the drone flight links are more reliable than public vehicles.

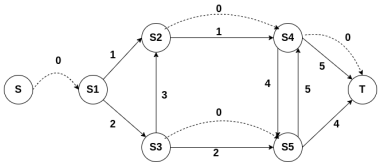
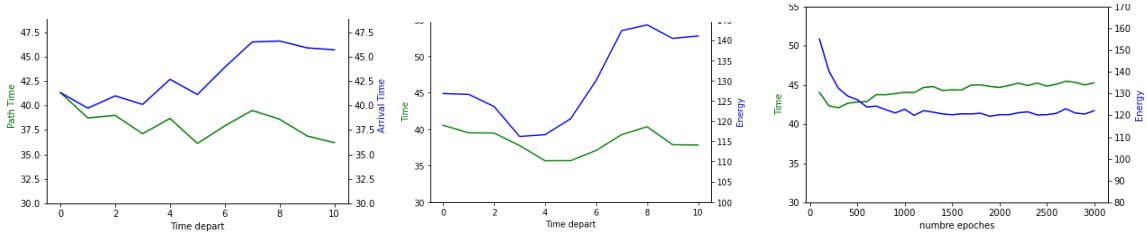


FIG. 1: Adopted network used to illustrate how the proposed Q-learning works

TAB. 1: The distributions of links traversal times and departure instants

Line	Link	Traversal times	Departure instants
L0	SS_1	3(0.1)	
	S_2S_4	6(0.2)	
	S_4T	2(0.1)	
	S_3S_5	6(0.2)	
L1	S_1S_2	13(2), 13(2),12(2), 11(1.5),11(1.5)	-1(1), 9(1),19(1), 29(1), 39(1)
	S_2S_4	10(2.5), 10(2.5), 11(2), 11(2), 10(2.5)	12(3), 22(3), 31(3), 40(2.5), 50(2.5)
L2	S_1S_3	15(1.4), 15(1.4), 16(1),16(1)	-5(2), 2(2), 9(2), 16(1)
	S_3S_5	10(1.3), 10(1.2), 10(1), 11(1.1)	10(3.4), 17(3.4), 25(3.4), 32(2)
L3	S_3S_2	5(1),7(1),7(1),6(1.5)	5(1),17(1.5),29(1),41(1)
L4	S_4S_5	7(1), 8(1.2), 8(1.2), 7(1.2), 7(1)	8(3), 19(3.2),29(2.7), 38(2.7), 48(2.5)
	S_5T	3(1), 3(1),3(1.2), 4(1), 4(1)	8(3), 19(3.2),29(2.7), 38(2.7), 48(2.5)
L5	S_5S_4	10(2),10(2),10(2),10(2)	5(1), 20(1), 35(1.5), 50(1.5)
	S_4T	5(1), 5(1), 5(1.2), 5(0.5)	15(3), 30(3), 45(3.5), 60(3.5)



(a) Path arrival and traversal times (b) Path traversal times and consumed energy (c) The convergence of Q-learning.

FIG. 2: The simulation results of the Q-Learning. $\gamma = 0.5$

Figure 2(a) shows both the total time and the arrival times of a path from the warehouse to the customer by using Q-learning algorithm, for different departure moments. From this figure, we can remark that each departure time results in different traversal and arrival times, which indicates that the routes in our system are time dependent. Fig 2(b) shows the travel time and energy consumed on the path according to the different instants of departure. We can remark that by using a static γ in the reward function that prioritizes traversal time over power consumption, the Q-learning gives a small and approximately constant traversal time as opposed to large and fluctuating power consumption. Additionally, Fig. 2(c) investigated the convergence of the learning algorithm. As we see, both the energy consumption and the traversal time curves converge after 1000 iterations. The results presented in Fig. 2 prove that our inspection on time-adapted early arrival path problem is well formulated, as well as prove the effectiveness of the proposed Q-learning-based algorithm to solve it.

The results of Fig. 3 demonstrate that the Q Learning solution is able to achieve a lower early arrival time and lower power consumption compared to a semi-random algorithm at any departure instant. This is due to the fact that the random algorithm selects the next trip, from among the available trips that do not backtrack, randomly without looking at previous experiences, as opposed to the suggested Q-learning, where next trip choices are based on available flights and previous Q-values.

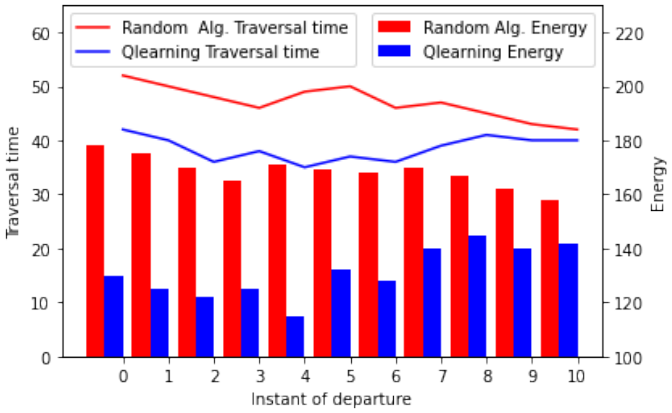


FIG. 3: Comparison of the random algorithm and the proposed Q-Learning

5 Conclusion

An inspection on time-adapted early arrival path problem in drone-vehicles based delivery schema, using public transportation, has been proposed in this paper. This path seeks to guarantee an early arrival time for a drone delivery while minimizing energy consumption and being resilient to stochastic fluctuations in public transportation schedules. The problem has been formulated as a Markov decision process. To address this problem, a Q-learning-based planning method is proposed. Simulation results show that our model of the time-adaptive early arrival path problem is well developed, and validate the effectiveness and feasibility of the proposed Q-learning for the delivery package planning path. The arrival of the drones to customers is 100% assured at any departure time. Moreover, the Q-Learning method achieved convergence after only 1000 learning epochs. In addition, experimental results reveal that the Q Learning approach has a lower early arrival time and power consumption than a random algorithm. This work allows the computation of the expected arrival path for a single drone.

References

[1] Taha Benarbia and Kyandoghere Kyamakya. A literature review of drone-based package delivery logistics systems and their implementation feasibility. *Sustainability*, 14(1):360, 2021.

[2] Elhadja Chaalal, Sidi-Mohammed Senouci, and Laurent Reynaud. A new framework for multi-hop abs-assisted 5g-networks with users' mobility prediction. *IEEE Transactions on Vehicular Technology*, 71(4):4412–4427, 2022.

[3] Sung Hoon Chung, Bhawesh Sah, and Jinkun Lee. Optimization for drone and drone-truck combined operations: A review of the state of the art and future directions. *Computers & Operations Research*, 123:105004, 2020.

[4] https://www.amazon.com/Amazon Prime-Air/b?ie=UTF8node=8037720011, 2022.

[5] Hailong Huang and Andrey V Savkin. An algorithm of reactive collision free 3-d deployment of networked unmanned aerial vehicles for surveillance and monitoring. *IEEE Transactions on Industrial Informatics*, 16(1):132–140, 2019.

[6] Hailong Huang, Andrey V Savkin, and Chao Huang. Scheduling of a parcel delivery system consisting of an aerial drone interacting with public transportation vehicles. *Sensors*, 20(7):2045, 2020.