

Reinforcement Learning-Based Deadline and Battery-Aware Offloading in Smart Farm IoT-UAV Networks

Anne Catherine Nguyen[†], Turgay Pamuklu[†], *Member, IEEE*, Aisha Syed[‡],
W. Sean Kennedy[‡], Melike Erol-Kantarci[†], *Senior Member, IEEE*

[†]*School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada*

[‡]*Nokia Bell Labs*

Emails: {anguy087, turgay.pamuklu, melike.erolkantarci}@uottawa.ca, {aisha.syed, william.kennedy}@nokia-bell-labs.com

Abstract—Unmanned aerial vehicles (UAVs) with mounted base stations are a promising technology for monitoring smart farms. They can provide communication and computation services to extensive agricultural regions. With the assistance of a Multi-Access Edge Computing infrastructure, an aerial base station (ABS) network can provide an energy-efficient solution for smart farms that need to process deadline critical tasks fed by IoT devices deployed on the field. In this paper, we introduce a multi-objective maximization problem and a Q-Learning based method which aim to process these tasks before their deadline while considering the UAVs' hover time. We also present three heuristic baselines to evaluate the performance of our approaches. In addition, we introduce an integer linear programming (ILP) model to define the upper bound of our objective function. The results show that Q-Learning outperforms the baselines in terms of remaining energy levels and percentage of delay violations.

Index Terms—Aerial base station, Smart farm, Unmanned aerial vehicle, Reinforcement Learning.

I. INTRODUCTION

One of the challenges of agriculture is battling against nature, this includes: pest control, fire control, and monitoring crop growth. A farm consists of hundreds of acres of land and it is difficult to monitor all the crops for pests. Fires can be less detrimental to the crops if they are detected and pinpointed early. With the introduction of smart agriculture, farmers can use internet of things (IoT) devices with cameras to monitor their crops and use image recognition to detect pests, fires, and the growth stages of crops.

Image classification has emerged as a useful tool for agriculture. Aldabbagh et al. introduced a Deep Learning algorithm to predict the growth stage of chili plants from images in [1]. In [2], Yu et al. proposed a pest monitoring system that comprised of unmanned aerial vehicles (UAV) with a mounted camera and ground sensors to monitor fields of crops. The UAVs flew across the fields and captured images and performed image classification to detect pest infestations on crops.

IoT devices are limited in computing capacity and cannot perform complex image recognition tasks. They need to offload computational tasks to nearby devices that have the processing capacity to execute the tasks. These image processing tasks must be done in a time-critical manner since fire and pest detection are time sensitive tasks.

In [3], Zhao et al. proposed using a network that consists of multiple hovering UAVs and Multi-Access Edge Computing (MEC) devices to aid farm monitoring. The UAVs provided: connectivity between the IoT devices, and a central processing unit (CPU) that can perform the image classification task. The UAVs have mounted CPUs and can compute the image processing tasks themselves or they can forward the task to a nearby MEC device to compute the task. Although our architecture is similar, [3] focuses on optimal placement of UAVs.

UAVs are battery operated devices and are therefore limited to a finite amount of energy and CPU capacity. The MEC device can assist the UAVs by taking over some of the computationally heavy tasks. MEC devices have a more powerful CPU, and can share the load of the UAVs so that the UAVs do not need to drain their battery quickly. Offloading some tasks to MEC will also prevent a bottleneck situation where all of the tasks are waiting to be executed in the UAV's processing queues. In this paper, we aim to extend the UAV's energy level because we want the UAVs to hover over the smart farm for as long as possible. We also aim for the tasks to be completed before their deadline.

In this study, we introduce a problem that has two objectives: *i*) extend the battery life of the UAVs, and *ii*) minimize the number of delay violations. The longer the aerial base stations (ABS) can hover, the more they can assist the IoT devices in computing the intensive tasks. The UAVs also need to ensure that these time-sensitive tasks will be computed before their deadlines.

The remainder of this paper is organized as follows. The related works are explained in Section II. Then, we detail the system model and the problem definition in Section III. Section IV introduces the proposed method and the baselines. Section V provides the computational experiment. Finally, in Section VI, we conclude the paper.

II. RELATED WORKS

There have been many studies emerging recently that suggest using UAVs in a smart farm. Zhou et al. provided an extensive survey for the use of UAVs with MEC in [4]. Lottes

TABLE I
SUMMARY OF THE NOTATIONS.

Sets	Size	Description
$t \in \mathcal{T}$	T	Set of time intervals
$j \in \mathcal{J}$	J	Set of UAVs
$l \in \mathcal{L}$	L	Set of MEC devices
$j' \in \mathcal{J}^+$	$J+L$	Set of CPUs
$k \in \mathcal{K}$	K	Set of task types
Variables	Domain	Description
$x_{jtj'}$	$\{0, 1\}$	Task offload decision
$p_{jtj't'}$	$\{0, 1\}$	CPU allocation
$p_{jtj't'}^+$	$\{0, 1\}$	CPU allocation first time interval
$p_{jtj't'}^-$	$\{0, 1\}$	CPU allocation last time interval
v_{jt}	$\{0, 1\}$	Deadline violation
Task Parameters	Range	Description
α_{jt}^B	$\{0, 1\}$	Task demand indicator
α_{jt}^P	\mathbb{R}	Processing time
α_{jt}^D	\mathbb{R}	Deadline
\mathcal{V}_j^L	\mathbb{N}	Violation reward level
Δ_{jt}	\mathbb{R}	Scheduling + processing delay
Energy Parameters	Range	Description
Υ_j^B	\mathbb{N}	Battery capacity
Υ_j^H	\mathbb{R}	Hovering energy cons.
Υ_j^A	\mathbb{R}	Antenna energy cons.
Υ_j^I	\mathbb{R}	CPU idle energy cons.
Υ_j^C	\mathbb{R}	CPU active energy cons.
Υ_j^R	\mathbb{R}	Remaining energy in a battery
Υ_j^L	\mathbb{N}^+	Battery reward level
W	$[0, 1]$	Energy consumption weight
Θ	\mathbb{R}	Scaling factor

et al. detailed the usage of UAVs in a smart farm in [5]. They presented a real-world smart farm case in which a UAV device is used for image classification. In another study, Islam et al. introduced pesticide usage for a similar scenario and discuss the tradeoff between latency and battery usage [6]. Lastly, Zhao et al. aimed to increase the throughput by considering the delay critical tasks in their smart farm environment [3].

Energy efficiency and time critical task approaches for UAVs are not limited to smart farm scenarios. Yang et al. dealt with resource allocation and task offloading problems to reduce the overall power consumption in their networks [7]. In another study, Zhou et al. included satellite computation/communication as an alternative for UAVs [8]. They modelled their problem as a constrained Markov decision process (MDP) and provided a deep reinforcement learning solution. Meanwhile, Ghdiri et al. provided a cluster-based approach for deadline aware task computation in UAVs [9]. Lastly, Yao et al. preferred a generalized Nash equilibrium approach for the offloading problem in a UAV swarm [10]. Unlike the previous works, we focus on tasks deadlines and maximizing hover time by providing an integer linear programming (ILP) model and Q-Learning solution.

III. SYSTEM MODEL

We consider a set of UAVs, $j \in \mathcal{J}$, is hovering above a rural area, communicating with IoT devices and providing them with guaranteed service. In addition, a set of MEC servers $l \in \mathcal{L}$, is located at the edge of the smart farm and is available

for task processing. One of the primary advantages of sharing the tasks with a MEC server is the extension of the hovering time of the UAVs, which have limited battery capacities (Υ_j^B). Overall, CPUs ($j' \in \mathcal{J}^+$), located either in a UAV or MEC, can be selected to process the IoT tasks.

In our time interval based ($t \in \mathcal{T}$) model, the IoT devices may demand to process K types of tasks from their associated UAV, $j \in \mathcal{J}$, in any of these time intervals (α_{jt}^B). Each task type has a unique CPU processing time (α_{jt}^P), and a deadline (α_{jt}^D) that should not be violated to provide a reasonable quality of service to these IoTs. We propose a task offloading problem that will complete the tasks before their deadline, and improve the hovering time of the UAVs. The problem will be explained in the following subsection.

A. Task Offloading Optimization for Increasing UAV Hover Time and Reducing the Deadline Violations

We combine our two significant key performance indicators (KPIs): increasing the hover time and reducing the task deadline violations, as a multi-objective maximization problem in Eq. 1. Here, W is the weight of the increasing hover time goal. Θ is a scale value used to normalize energy consumption (Watts), and the total number of deadline violations (v_{jt}). In order to improve overall hover time, we maximize the minimum remaining energy ($\Upsilon_{j'}^R$) in the batteries of the set of UAVs ($j' \in \mathcal{J}$). Therefore we can extend the operation time of all UAVs without the need to recharge their batteries. Eq. 2 calculates the remaining energy in a battery by subtracting the hovering ($\Upsilon_{j'}^H$), antenna ($\Upsilon_{j'}^A$), and CPU idle ($\Upsilon_{j'}^I$) energy consumptions¹ from the full battery capacity ($\Upsilon_{j'}^B$), respectively. Lastly, we calculate the total CPU active energy consumption by multiplying the total number of time intervals that CPU j' spent to process a task with the difference of CPU active ($\Upsilon_{j'}^C$) and idle energy consumptions.

Eq. 3 ensures that each task generated in this smart farm (α_{jt}^B) is allocated to one of the CPUs in this network. For that purpose, this equation includes two critical decision variables. First, the task offloading binary indicator $x_{jtj'}$, is equal to one if the CPU j' processes the task received by UAV j in time interval t . Second, the binary indicator $p_{jtj't'}$, is equal to one if that task is processed in the time interval t' . Note that if the required processing time (α_{jt}^P) is longer than one time interval, multiple time intervals need to be allocated for that task. Lastly, the time intervals earlier than the task arrival time ($t' < t$) can not be allocated in a CPU to process this task (Eq. 4). A CPU's core can serve at most one task in a single time interval (Eq. 5). However, it could be possible to extend this model to multitasking CPUs without loss of generality. Moreover, a task can be processed at most by one CPU in the same time interval (Eq 6). Lastly, we have to limit the number

¹Note that UAVs also need to consume energy for their other operations, such as communicating with other devices. However, the energy consumption associated with these are not impacted by our decisions in this system model. Thus, we do not include them to improve the readability of the problem definition.

of offloading decisions to process the entire task in the same CPU (Eq 7).

Maximize:

$$W * \min_{j' \in \mathcal{J}} \Upsilon_{j'}^R - \frac{1-W}{\Theta} \sum_{\substack{j \in \mathcal{J} \\ t \in \mathcal{T}}} v_{jt} \quad (1)$$

$$\Upsilon_{j'}^R = \Upsilon_{j'}^B - (\Upsilon_{j'}^H + \Upsilon_{j'}^A + \Upsilon_{j'}^I) * T - \sum_{\substack{j \in \mathcal{J} \\ t \in \mathcal{T} \\ t' \in \mathcal{T}}} (\Upsilon_{j'}^C - \Upsilon_{j'}^I) * p_{jtj't'} \quad (2)$$

Subject to:

$$\sum_{t'=t}^T \sum_{j' \in \mathcal{J}^+} p_{jtj't'} * x_{jtj'} = \alpha_{jt}^B * \alpha_{jt}^P, \quad \forall j \in \mathcal{J}, \forall t \in \mathcal{T} \quad (3)$$

$$\sum_{t'=0}^{t-1} \sum_{j' \in \mathcal{J}^+} p_{jtj't'} * x_{jtj'} = 0, \quad \forall j \in \mathcal{J}, \forall t \in \mathcal{T} \quad (4)$$

$$\sum_{j \in \mathcal{J}} \sum_{t \in \mathcal{T}} p_{jtj't'} \leq 1, \quad \sum_{j \in \mathcal{J}} \sum_{t \in \mathcal{T}} p_{jtj't'}^+ \leq 1, \quad \sum_{j \in \mathcal{J}} \sum_{t \in \mathcal{T}} p_{jtj't'}^- \leq 1, \quad \forall j' \in \mathcal{J}^+, \forall t' \in \mathcal{T} \quad (5)$$

$$\sum_{j' \in \mathcal{J}^+} p_{jtj't'} \leq 1, \quad \sum_{j' \in \mathcal{J}^+} p_{jtj't'}^+ \leq 1, \quad \sum_{j' \in \mathcal{J}^+} p_{jtj't'}^- \leq 1, \quad \forall j \in \mathcal{J}, \forall t' \in \mathcal{T}, \forall t \in \mathcal{T} \quad (6)$$

$$\sum_{j' \in \mathcal{J}^+} x_{jtj'} \leq 1, \quad \forall j \in \mathcal{J}, \forall t \in \mathcal{T} \quad (7)$$

$$\Delta_{jt} = \sum_{\substack{j' \in \mathcal{J}^+ \\ t' \in \mathcal{T}}} [p_{jtj't'}^+ * (t') - t + \alpha_{jt}^P] \quad (8)$$

$$\alpha_{jt}^D \geq \Delta_{jt} - M * v_{jt}, \quad \forall j \in \mathcal{J}, \forall t \in \mathcal{T} \quad (9)$$

$$\alpha_{jt}^D \leq \Delta_{jt} + M * (1 - v_{jt}), \quad \forall j \in \mathcal{J}, \quad \forall t \in \mathcal{T} \quad (10)$$

$$p_{jtj'(t+1)} = p_{jtj't'} + p_{jtj'(t+1)}^+ - p_{jtj'(t+1)}^- \quad (11)$$

$$p_{jtj'(t+1)}^+ + p_{jtj'(t+1)}^- \leq 1, \quad \forall j \in \mathcal{J}, \forall t \in \mathcal{T}, \forall j' \in \mathcal{J}^+, \forall t' \in \mathcal{T} \quad (12)$$

$$p_{jtj'(0)} = p_{jtj'(0)}^+ \quad (13)$$

$$\sum_{t' \in \mathcal{T}} p_{jtj't'}^+ \leq 1, \quad \forall j \in \mathcal{J}, \forall t \in \mathcal{T}, \forall j' \in \mathcal{J}^+ \quad (14)$$

$$\sum_{t' \in \mathcal{T}} p_{jtj't'}^- \leq 1, \quad \forall j \in \mathcal{J}, \forall t \in \mathcal{T}, \forall j' \in \mathcal{J}^+ \quad (15)$$

In order to calculate the summation of the scheduling and processing delay (Δ_{jt}) of a task that arrived to UAV j in time interval t in Eq 8, we introduce a new decision variable, $p_{jtj't'}^+$, to render the first time interval, t' , that we processed this task in CPU j' . If this binary decision variable equals to one, CPU j' starts to process that task. After we multiply that decision variable with t' and subtract the task arrival time (t), we can then find the scheduling delay for this task. After adding the

processing delay of this task (α_{jt}^P) in to that equation, we can find the summation of the scheduling and processing delay.

We use the delay calculation to identify the deadline violation variable (v_{jt}). The association between that binary decision variable and the deadline of the task accomplished with Eqs. 9 and 10 are done by using the Big-M method which is a common linear programming solving method that uses a very large constant in a constraint. In the case of a higher deadline value, v_{jt} should be zero to satisfy the constraint defined by Eq. 9. Otherwise, v_{jt} should be one to satisfy Eq. 10.

As mentioned before, $p_{jtj't'}^+$ equals to one if the task arrives at the UAV j in time interval t and starts to be processed in j' in time interval t' . Meanwhile, $p_{jtj't'}^-$ equals to one if the same task is completed in time interval t' . Therefore we can identify the exact start time and completion time of a task in a CPU. In addition, we want to process a task in a CPU without an interruption which is called contiguity. Eqs. 11-15 ensure that CPU time interval allocation contiguity. If we detail these equations, Eq. 11 ensures that $p_{jtj't'}^+$ or $p_{jtj't'}^-$ should be one in a case when the values of consecutive allocation variables ($p_{jtj't'}$ and $p_{jtj'(t'+1)}$) are different, which actually means that we start to process the related task or finish to process it, respectively. Eq. 12 eliminates the ping-pong effect. Eq. 13 ensures that $p_{jtj't'}^+$ equals to one in case that we start to process the task in the first time interval. Lastly, Eq. 14 and Eq. 15 limit the start and completion time of a task, respectively.

IV. PROPOSED METHOD

A. Q-Learning Approach

We propose a finite-horizon multi-agent MDP framework to solve the problem explained in the previous section. Each UAV independently make decisions with a tuple $\mathcal{F} = \{\mathbb{P}, \mathbb{A}, \mathbb{R}, \mathbb{S}, \Pi\}$ in which:

- **State Transitions:** $\mathbb{P} : \mathbb{S}_1 \times \mathbb{A} \times \mathbb{S}_2 \implies \mathbb{R}$

The framework has a task-based state transition model. After each task arrives at the UAV, we update the environment and calculate the delays and battery levels to find the first state \mathbb{S}_1 . After the action, we update the environment to find \mathbb{S}_2 and then reward \mathbb{R} . Due to unpredictable task arrivals, state transitions are stochastic.

- **Action:** $\mathbb{A} = \{x_{j' \in \mathcal{J}^+}\}$

After a UAV receives a task, it has three options: processing that task locally, offloading it to another UAV, or offloading the task to the MEC. That decision-making process is accomplished by choosing a CPU in the set of CPUs (\mathcal{J}^+) that includes all of these three alternatives.

- **State:** $\mathbb{S} = \{k, \Delta_{j' \in \mathcal{J}^+}, \Upsilon_{j' \in \mathcal{J}}^L\}$

Here k is the type of the task received by the UAV, $\Delta_{j' \in \mathcal{J}^+}$ are the delays in all CPUs, and $\Upsilon_{j' \in \mathcal{J}}^L$ is the battery levels of all UAVs calculated by Eq. 17.

- **Policy(II):** We use an epsilon-greedy policy in this framework.

- **Reward:**

$$\mathbb{R} = (\Upsilon_{j_a}^L - 1) + (1 - \mathbb{E}(v_{j_a})) + \mathcal{V}_{j_a}^L * \mathbb{E}(v_{j_a}) \quad (16)$$

$$\Upsilon_{j_a}^L = \begin{cases} 2, & \text{if } \mathbb{E}(\Upsilon_{j_a}^R) - \max_{j' \in \mathcal{J}}(\mathbb{E}(\Upsilon_{j'}^R)) \geq -\epsilon \\ 0, & \text{if } \mathbb{E}(\Upsilon_{j_a}^R) - \max_{j' \in \mathcal{J}}(\mathbb{E}(\Upsilon_{j'}^R)) \leq -2 * \epsilon \\ 1, & \text{otherwise,} \end{cases} \quad (17)$$

$$\mathcal{V}_{j_a}^L = \begin{cases} -40, & \text{if } \mathbb{E}(v_{j_m}) = 0 \\ -20, & \text{if } \mathbb{E}(v_{j_r}) = 0 \\ -10, & \text{if } \exists j' \in (\mathcal{J} / (j_r \cup j_a))(\mathbb{E}(v_{j'})) = 0 \\ -1, & \text{otherwise,} \end{cases} \quad (18)$$

We introduce the battery reward level concept ($\Upsilon_{j_a}^L$) to map the action (j_a) into the reward function (Eq. 16). A battery reward may have three levels [0,1,2]; thus, we return a negative (-1) reward for the battery level 0, and we return a positive (1) reward² for the battery level 2. That battery level is calculated by Eq. 17, in which $\mathbb{E}(\Upsilon_{j'}^R)$ is a UAV battery's expected remaining energy when that UAV starts to process the delegated task. If the difference between maximum energy level and the energy level of the selected UAV is lower than a certain level (ϵ), we promote this action with a positive reward. Therefore we can balance the remaining energy levels of the UAVs and increase their hovering time. Lastly, we also introduce a hysteresis approach to that calculation by adding an extra battery level for the energy differences between $[-\epsilon, -2\epsilon]$.

In addition, the reward function (Eq. 16) includes the expected deadline violation $\mathbb{E}(v_{j_a})$, which equals zero if the task does not yield to a deadline violation at the delegated CPU (UAV or MEC j_a). In that case, the reward function returns 1. If a deadline violation has occurred, the reward function returns the violation reward level ($\mathcal{V}_{j_a}^L$), calculated by Eq. 18. In that equation, we focus on the expected delay violations if we delegated the task to a different CPU instead of the original action (j_a). There can be several possible scenarios. The first case is when we could have prevented a delay violation if the task had been sent and processed in the MEC. We want to encourage the tasks to be processed in the MEC if it has an idle CPU, therefore we return a significant penalty in that condition. The second case occurs if we could have prevented a delay violation if the task was never offloaded from the received UAV in the first place. The third case represents the condition if we do not expect a violation from a different offloading decision. Otherwise, if a deadline violation would have occurred in any action, we return a small penalty. The numeric values selected as rewards or penalties are empirically set.

In the following subsections, we explain the baseline schemes.

²We assume that MEC's energy is supplied by the main electricity grid; therefore, we always return a positive reward in the case of MEC selection

B. Baseline Approaches

1) *Round Robin (RR)*: Our first baseline is the simple Round Robin method in which, when a UAV needs to offload a task, it selects the other UAVs and the MEC in a round robin fashion.

2) *Highest Energy First (HEF)*: In this algorithm, the offloading decision is based on the UAV's remaining battery level. The UAVs in the network regularly update one another with their current remaining battery levels and they store this information in a table. When a UAV receives a task from an IoT device, it scans the table and finds the UAV with the highest remaining energy level. Once it has found the UAV with the highest energy level, it checks the difference between the highest remaining energy level and its energy level. If the difference is greater than the 1% threshold, then the UAV will offload the task to the UAV with the higher energy level. Otherwise, the UAV will compute the task locally. Since the MEC device has unlimited power capacity, we limit the UAV's ability to offload a task to MEC to 20% of the time.

3) *Lowest Queue Time and Highest Energy First (QHEF)*: We improve the HEF scheme by adding the queue time to the offloading decision. Queueing delay is defined as the sum of the processing delays of all the tasks that are currently in the UAV's CPU queue. After a UAV receives a task from the IoT device, it finds the UAV with the lowest queuing delay. It checks to see if the difference between the current UAV's queuing delay and the lowest queuing delay is greater than the threshold of 0.5 seconds. If it is greater than the queueing delay threshold, then that UAV will be a contender for an offloading destination. Then it will try and find a UAV that has the lowest queuing delay but higher remaining battery level. The difference between the possible offloading destination and current UAV's battery levels must be above an energy threshold of 1%. If such a UAV exist, then it will offload the task to that UAV, else, it will add the task to the list of tasks that the current UAV will process locally. Hence, this algorithm considers both the UAVs' battery levels and queuing delay in the decision-making process.

4) *ILP Solver*: In addition to the heuristics we explained above, we use GUROBI Solver [11] to find the optimum solution for our multi-objective maximization problem (Eqs. 1-15). Despite the NP-hard property of our problem, that method could be used for only small solution space problems. We detail the findings in the following section.

V. PERFORMANCE EVALUATION

A. Simulation Platform

To simulate the UAV network in the smart farm, we used Simu5G which is a 5G network simulator library developed over Omnet++ [12]. Simu5G contains modules that model the different nodes found in a 5G and LTE network following the 3GPP standards [13].

TABLE II
ENERGY CONSUMPTION PARAMETERS.

$\Upsilon_{j'}^B$	$\Upsilon_{j'}^H$	$\Upsilon_{j'}^A$	$\Upsilon_{j'}^I$	$\Upsilon_{j'}^C$
570	211	17	4320	12960

TABLE III
SIMULATION PARAMETERS.

Parameter	Value			
J	4			
L	1			
\mathcal{K}	Fire detection (FD) Pesticide detection (PD) Growth monitoring (GM)			
Task Type	$(1/\lambda)^4$	α_{jt}^D	$\alpha_{jt}^P(UAV)$	$\alpha_{jt}^P(MEC)$
FD	0.25s	0.3s	0.1s	0.05s
PD	0.25s	0.8s	0.5s	0.25s
GM	0.5s	5s	0.1s	0.05s

B. Energy Consumption Parameters

We use the parameters³ in Table II and Equation 2 to model the remaining UAV energy in our simulations. We assume that we are using a battery that is similar to the one found in [14]. To calculate the power consumption of hovering, we used eq. 2 from [15], we also used their assumptions such that that the UAV will have: 4 rotors, fluid density of $1.204kg/m^3$, rotor disc area of $0.2m^2$, frame's mass $M = 1.5kg$, and battery and payload $m = 3$ kg. The equation from [15] is given in eq. 19, where M is the mass of UAV in kg, m is the mass of the battery and other payloads in kg, g is gravity in Newtons, ρ is the fluid density in kg/m^3 , ς is the rotor disc area in m^2 , and n is the number of rotors.

$$\Upsilon_{j'}^H = (M + m)^{\frac{3}{2}} \sqrt{\frac{g^3}{2\rho\varsigma n}} \quad (19)$$

C. Simulation Results

The results shown in this section are the average of ten runs with different seeds. Interarrival times between the tasks are exponentially distributed and their processing times are deterministic. Table III presents the parameters used in all simulations. Fig. 1 shows the Q-Learning convergence in four UAVs. We used 0.05 as a learning rate, 0.85 for discount value and ran the scenario one million times (episodes). The solid lines show the average of 10k episodes, and the shaded area shows the variation of the cumulative reward. The average of the rewards in UAVs are close to each other in an episode, which means UAVs can learn independently with the same performance. Also, when we zoom in on the last 60k episodes, as seen in the subfigure, the difference in the cumulative reward is very low.

The UAV's remaining energy percentage is the amount of energy left in the battery after the simulation has completed.

³Due to limited simulation time, we set the energy consumption level of idle and busy CPU periods to be at the level they would be if they ran for ten hours, to highlight the performance of methods in terms of energy consumption.

⁴Mean interarrival rate.

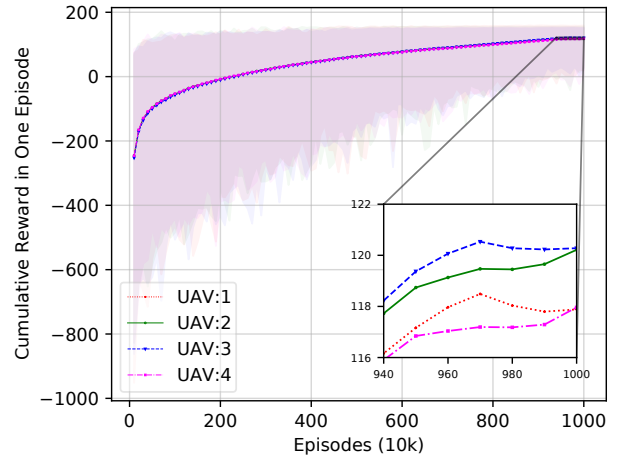


Fig. 1. Convergence of multi-UAV Q-Learning method.

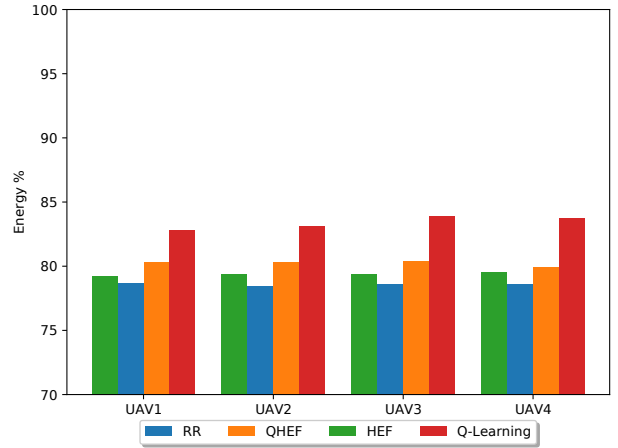


Fig. 2. Comparison of remaining energy levels of UAVs.

The remaining energy percentage is directly related to the UAV's hover time. A higher remaining energy percentage means that the UAV will be able to hover over the network for a longer period of time. Fig. 2 compares the remaining energy levels of the UAVs. The Q-Learning algorithm has the highest remaining energy level for all four of the UAVs.

In Q-Learning and QHEF, the UAVs offload their tasks to the MEC device more than the other UAVs. This allows the UAVs to not only preserve their own batteries, but also preserve the batteries of the other UAVs in the network. In HEF and RR, the majority of the offloaded tasks went to other neighbouring UAVs. Because the UAVs are not allowed to further offload an offloaded task, they must compute the offloaded task locally to prevent loops. Therefore, the additional offloaded tasks caused the UAVs to drain their batteries faster than the UAVs that used QHEF and Q-Learning.

A delay violation occurs when the task has reached its predetermined deadline. Fig. 3 compares the percentage of delay violations out of the total number of tasks in the network. Q-Learning has the lowest percentages of delay violations,

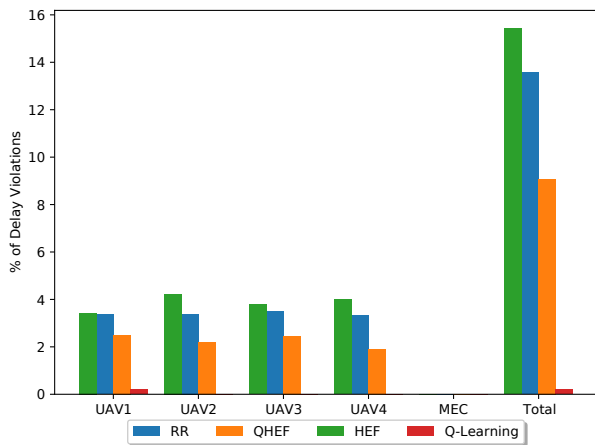


Fig. 3. Distribution of delay violations amongst smart farm nodes.

TABLE IV
Q-LEARNING AND ILP SOLUTION COMPARISON

Methods	Υ_0^R	Υ_1^R	Υ_2^R	Υ_3^R	$\sum_{\substack{j \in \mathcal{J} \\ t \in \mathcal{T}}} v_{jt}$
ILP ($W = 0$)	97.7%	97.9%	97.9%	98.2%	7
ILP ($W = 1$)	99.1%	99.1%	99.1%	99.1%	24
ILP ($W = 0.5$)	97.9%	97.9%	97.9%	97.9%	9
Q-Learning	98.6%	98.3%	98.6%	98.2%	13

because it considers the task processing and scheduling delays in their decision-making process. QHEF selects the destination with the lowest queueing delay. Both of these algorithms consider the delay of a task, which is why they outperform RR and HEF in terms of delay violations.

In addition to the heuristic baselines, we used an ILP solver to find the upper bound for hover time and lower bound for delay violation KPIs. However, due to the problem's NP-hard property, we had to limit the simulation time to four seconds ($T = 4s$) to find a solution with a 48 hours solver running time. In addition, we reduced the interarrival time to 0.125s for all task types and used more strict deadlines for fire detection (0.2s) and pesticide detection tasks (0.6s).

Table IV shows that the ILP solver can reduce the number of delay violations to seven if the solver only takes into account delay violations as the objective ($W = 0$). On the other hand, when the solver focuses only on increasing the hover time ($W = 1$), it can reduce the energy consumption in UAVs more than the other methods by offloading all tasks to MEC. However, that approach caused a significant rise in delay violations. It provided an adequate deadline violation if the solver is run as a multi-objective solver with a balanced value between these two KPIs ($W = 0.5$). Lastly, if we compare the proposed Q-Learning method with this balanced solver method, Q-Learning provides a slight improvement in increasing the hover time, while it could not outperform the solver with regard to reducing the number of deadline violations. Nonetheless, the ILP solver needed 48 hours run-time which is significantly high.

VI. CONCLUSION

In this paper, we used a MEC assisted ABS network to provide a deadline aware service to the IoT tasks of a smart farm. In addition, we balanced the energy usage of the UAVs in this network to increase their hover time. We provided a Q-Learning based approach and several baselines to analyze our method. The results showed that our Q-Learning method performed better than the baselines in terms of remaining energy level and percentage of delay violations. Furthermore, ILP results demonstrated that the Q-Learning method is close to the optimum solution and exceeded the ILP approach by providing a solution for more extensive problems and being adaptive to changes in the environment.

ACKNOWLEDGEMENT

This work is supported by MITACS Canada Accelerate program under collaboration with Nokia Bell Labs.

REFERENCES

- [1] A. D. A. Aldabbagh, C. Hairu, and M. Hanafi, "Classification of Chili Plant Growth using Deep Learning," in *10th International Conference on System Engineering and Technology (ICSET)*, IEEE, Nov 2020.
- [2] Y. Lina and Y. Xiuming, "Design of Intelligent Pest Monitoring System Based on Image Classification Algorithm," in *3rd International Conference on Control and Robots (ICCR)*, pp. 21–24, IEEE, Dec 2020.
- [3] J. Zhao, Y. Wang, Z. Fei, and X. Wang, "UAV Deployment Design for Maximizing Effective Data with Delay Constraint in a Smart Farm," in *International Conference on Communications in China (ICCC)*, IEEE, Aug 2020.
- [4] F. Zhou, R. Q. Hu, Z. Li, and Y. Wang, "Mobile Edge Computing in Unmanned Aerial Vehicle Networks," *IEEE Wireless Communications*, vol. 27, pp. 140–146, Feb 2020.
- [5] P. Lottes, R. Khanna, J. Pfeifer, R. Siegwart, and C. Stachniss, "UAV-based crop and weed classification for smart farming," in *International Conference on Robotics and Automation (ICRA)*, IEEE, May 2017.
- [6] N. Islam, M. M. Rashid, F. Pasandideh, B. Ray, S. Moore, and R. Kadel, "A Review of Applications and Communication Technologies for Internet of Things (IoT) and Unmanned Aerial Vehicle (UAV) Based Sustainable Smart Farming," *Sustainability*, vol. 13, p. 1821, Feb 2021.
- [7] Z. Yang, C. Pan, K. Wang, and M. Shikh-Bahaei, "Energy Efficient Resource Allocation in UAV-Enabled Mobile Edge Computing Networks," *IEEE Transactions on Wireless Communications*, vol. 18, pp. 4576–4589, Sep 2019.
- [8] C. Zhou, W. Wu, H. He, P. Yang, F. Lyu, N. Cheng, and X. Shen, "Deep Reinforcement Learning for Delay-Oriented IoT Task Scheduling in SAGIN," *IEEE Transactions on Wireless Communications*, vol. 20, pp. 911–925, Feb 2021.
- [9] O. Ghdiri, W. Jaafar, S. Alfattani, J. B. Abderrazak, and H. Yanikomeroğlu, "Energy-Efficient Multi-UAV Data Collection for IoT Networks with Time Deadlines," in *IEEE Global Communications Conference*, 2020.
- [10] K. Yao, J. Chen, Y. Zhang, L. Cui, Y. Yang, and Y. Xu, "Joint Computation Offloading and Variable-width Channel Access Optimization in UAV Swarms," in *IEEE Global Communications Conference*, 2020.
- [11] Gurobi, "Optimizer reference manual," 2021. Available online at: <https://www.gurobi.com/documentation/9.1/refman/index.html>, last accessed on 2021-05-15.
- [12] OpenSim Ltd., "What is OMNeT++?," 2019. Available online at: <https://omnetpp.org/intro/>, last accessed on 2021-10-19.
- [13] G. Nardini, D. Sabella, G. Stea, P. Thakkar, and A. Virdis, "Simu5G—An OMNeT++ Library for End-to-End Performance Evaluation of 5G Networks," *IEEE Access*, vol. 8, pp. 181176–181191, 2020.
- [14] HSE - UAV, "High Power Drone Battery 6S-HV (LiHV)," 2021. Available online at: <https://hse-uav.com/product/high-power-drone-battery-6s-hv-lihv-25000mah-22-8v/>, last accessed on 2021-10-19.
- [15] K. Dorling, J. Heinrichs, G. G. Messier, and S. Magierowski, "Vehicle Routing Problems for Drone Delivery," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, pp. 70–85, Jan 2017.