A Triple Learner Based Energy Efficient Scheduling for Multi-UAV Assisted Mobile Edge Computing

Jiayuan Chen*, Changyan Yi*, Jialiuyuan Li*, Kun Zhu* and Jun Cai†

*Computer Science & Technology, Nanjing University of Aeronautics and Astronautics, China
†Electrical and Computer Engineering, Concordia University, Montréal, QC, Canada
Email: jiayuan.chen@nuaa.edu.cn

May 30, 2023

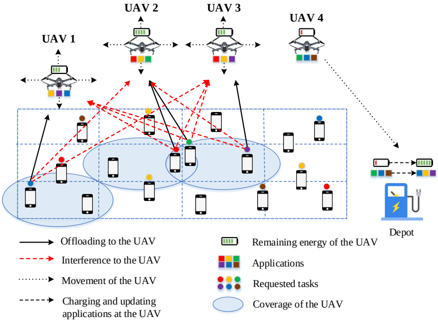
Outline

- Introduction
- System Model and Problem Description
- Problem Reformulation and Solution
- Simulation Results
- Conclusion

Background and Motivation

- Multi-UAV assisted MEC:
  - To provide prompt and flexible computing services to support mobile end users (e.g., IoT devices);
  - Unmanned aerial vehicles (UAVs) with attributions of high-flexibility in providing computing services are recruited.
- Open problems:
  - Computing tasks offloaded from different end-users are required to be processed by specific service applications;
    - the limited storage capacities of UAVs impede their abilities to store all applications.
  - The limited energy capacities of UAVs also hinders the implementation of multi-UAV assisted MEC in providing the long-term MEC services.
- Motivation:
  - Despite many researches have worked for them, there are still some critical issues:
    - how UAVs' installed applications should be updated (with severely restricted wireless backhauls)?
    - how UAVs energy replenishment should be jointly scheduled?

System Model



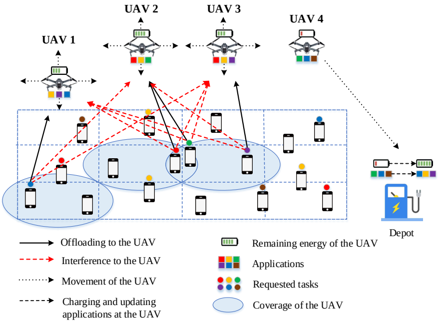
- M UAVs acting as mobile edge servers and N IoT devices.
- The distance of IoT device $n \in \mathcal{N}$ and UAV $m \in \mathcal{M}$ at time slot t:

$$d_{m,n}(t) = \sqrt{(x_m^U(t) - x_n^I(t))^2 + (y_m^U(t) - y_n^I(t))^2 + H^2}.$$

- The path loss between IoT device $n \in \mathcal{G}_m$ and UAV $m \in \mathcal{M}$:

$$\begin{aligned}\lambda_{m,n}(t) = \quad & 20log(\sqrt{H^2 + d_{m,n}(t)^2}) + \delta_{m,n}(t)(\eta_{LoS} - \eta_{NLoS}) \\ & +20log[(4\pi f)/c] + \eta_{NLoS}.\end{aligned}$$

System Model (Cont'd)



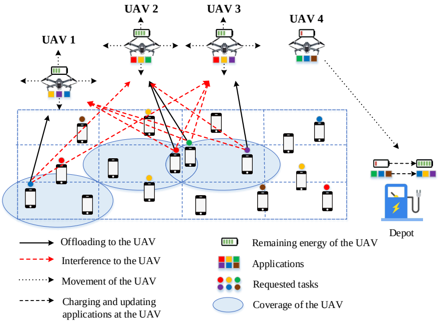
- The transmission time of IoT devices $n \in \mathcal{G}_m$ in offloading a task to UAV $m \in \mathcal{M}$:

$$t_{m,n}^{off}(t) = \frac{\boldsymbol{v}_n(t)(\boldsymbol{w}_m(t)^\top)D_n}{B\log_2(1+\gamma_{m,n}(t))}.$$

- The size of tasks computed by UAV $m \in \mathcal{M}$:

$$\begin{aligned} Task_m^{comp}(t) = \quad & \min\{\textstyle\sum_{n\in\mathcal{G}_m} \boldsymbol{v}_n(t)(\boldsymbol{w}_m(t)^\top)D_n f_m^U, \\ & (t^{hover} - \min\{t_{m,n}^{off}(t)_{n\in\mathcal{G}_m}\}) f_m^U\}. \end{aligned}$$

System Model (Cont'd)



- The energy consumption of UAV $m \in \mathcal{M}$ for computing tasks at time slot t:

$$E_m^{comp}(t) = \xi (f_m^U)^2 Task_m^{comp}(t).$$

- The propulsion energy consumption (consisting of the energy consumption of horizontal moving and hovering) of the UAV m:

$$E_m^{pro} = P_m^{pro}(V)\frac{q}{V} + P_m^{pro}(0)t^{hover}.$$

Problem Description

- We aim to maximize the energy efficiency of all UAVs, i.e., the total amount of offloaded tasks computed by all UAVs over their total energy consumption:

$$E^{effi}(t) = \frac{\sum_{m=1}^{M} \kappa_m(t) Task_m^{comp}(t)}{\sum_{m=1}^{M} (\kappa_m(t)(E_m^{comp}(t) + E_m^{pro}) + (1 - \kappa_m(t))E_m^{dep})}.$$

Problem Description (Cont'd)

- The joint optimization of multiple UAVs' trajectory planning, energy renewal and application placement for long-term energy efficiency can be formulated as:

$$[\mathcal{P}1]: \max_{\mathcal{U}_m(t), \boldsymbol{w}_m(t), \kappa_m(t)} \lim_{T \to +\infty} \frac{1}{T} \sum_{t=1}^{T} E^{effi}(t) \tag{1}$$

$$s.t., \ \kappa_m(t) \in \{0, 1\}, \forall m \in \mathcal{M}, t \geq 1, \tag{2}$$

$$w_{m,c}(t) \in \{0, 1\}, \forall m \in \mathcal{M}, \forall c \in C, t \geq 1, \tag{3}$$

$$\sum_{c=1}^{C} \mu_c w_{m,c}(t) \leq S_m, \forall m \in \mathcal{M}, \tag{4}$$

$$\sum_{m=1}^{M} w_{m,c}(t) \kappa_m(t) \geq 1, \forall c \in C, t \geq 1, \tag{5}$$

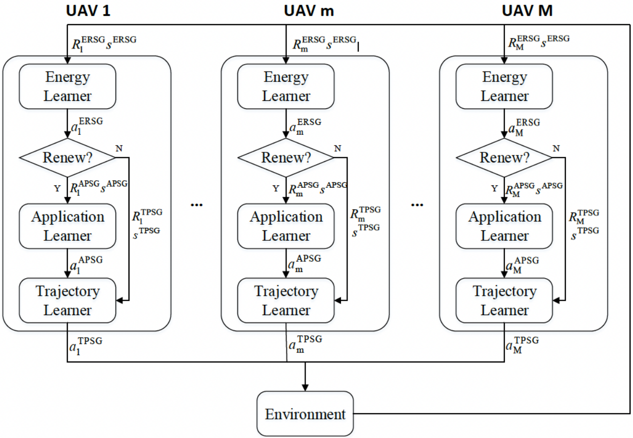
$$|\mathcal{U}_m(t) - \mathcal{U}_m(t-1)|^2 \kappa_m(t) = q^2, \forall m \in \mathcal{M}, t \geq 1, \tag{6}$$

$$(x_m^U(t) - x_m^U(t-1))(y_m^U(t) - y_m^U(t-1)) \kappa_m(t) = 0, \forall m \in \mathcal{M}, t \geq 1, \tag{7}$$

$$|\mathcal{U}_m(t) - \mathcal{U}_{m'}(t)| \kappa_m(t) \geq q, \forall m \in \mathcal{M}, \forall m' \in \mathcal{M} \backslash \{m\}, t \geq 1. \tag{8}$$

- Since UAVs are intelligent, to solve problem [$\mathcal{P}_1$], we can allow each UAV to make its own decisions while regulate the underlying cooperation and competition among them.
  - UAVs are expected to cooperatively conduct the trajectory planning, energy renewal and application placement to maximize the energy efficiency of all UAVs while guaranteeing QoS of IoT devices;
  - Allowing UAVs to make decisions themselves may also lead to competitions in trajectory planning, energy renewal and application placement;
  - The uncertainty that the future environment information (e.g., task requirements of IoT devices) is not available to UAVs.
- Reformulate [$\mathcal{P}1$] to three coupled multi-agent stochastic games:
  - Trajectory planning stochastic game (TPSG): $\langle \mathcal{M}, \mathcal{S}^{TPSG}, \mathcal{A}^{TPSG}, \mathcal{P}^{TPSG}, \mathcal{R}^{TPSG} \rangle$;
  - Energy renewal stochastic game (ERSG): $\langle \mathcal{M}, \mathcal{S}^{ERSG}, \mathcal{A}^{ERSG}, \mathcal{P}^{ERSG}, \mathcal{R}^{ERSG} \rangle$;
  - Application placement stochastic game (APSG): $\langle \mathcal{M}, \mathcal{S}^{APSG}, \mathcal{A}^{APSG}, \mathcal{P}^{APSG}, \mathcal{R}^{APSG} \rangle$.

Triple-learner based reinforcement learning (TLRL) approach for multiple UAVs trajectory planning,energy renewal and application placement:



- Markov decision processes (MDP) construction for each UAV in TPSG: $(\mathcal{S}^{TPSG}, \mathcal{A}_m^{TPSG}, \mathcal{R}_m^{TPSG}, \mathcal{P}^{TPSG})$.
  - 1) Environment State $s^{TPSG}(t) = (\mathcal{U}_m(t), \boldsymbol{w}_m(t))_{m\in\mathcal{M}}$:
    - All UAVs' positions $\mathcal{U}_m(t)$, $m \in \mathcal{M}$;
    - Application placement $\boldsymbol{w}_m(t)$, $m \in M$.
  - 2) Action $a_m^{TPSG}(t)_{m\in\mathcal{M}}$:
    - $a_m^{TPSG}(t) \in \mathcal{A}_m^{TPSG}$, where $\mathcal{A}_m^{TPSG} = \{forward, backward, left, right\}$.
  - Reward $\mathcal{R}_m^{TPSG}(t)_{m\in\mathcal{M}}$:
    - $$\mathcal{R}_m^{TPSG}(t) = \frac{\kappa_m(t)\, Task_m^{comp}(t)}{E_m^{comp}(t) + E_m^{pro}},$$

      where the numerator indicates the size of tasks computed by UAV m, and the denominator represents the energy consumption of UAV m at time slot t.

- Q-Learning for TPSG (Trajectory Learner):
  - Policy:

$$\pi_m^{TPSG}(s^{TPSG}, a_m^{TPSG}) = \begin{cases} 0, \text{if UAV } m \text{ decides to return to the depot,} \\ 1-\epsilon, \text{if } Q_m^{TPSG}(s^{TPSG}, \cdot, \cdot) \text{ of } a_m^{TPSG} \text{ is the highest,} \\ \epsilon, \text{otherwise.} \end{cases}$$

  - Q Function:

$$Q_m^{TPSG}(s^{TPSG}, \boldsymbol{a}^{TPSG}, \pi_m^{TPSG}) = \mathbb{E}(\sum_{\tau=0}^{\infty} \gamma^{\tau} \mathcal{R}_m^{TPSG}(t+\tau+1) | s^{TPSG}(t) = s^{TPSG}, \boldsymbol{a}(t)^{TPSG} = \boldsymbol{a}^{TPSG}, \pi_m^{TPSG})$$

  - Update Rule:

$$Q_m^{TPSG}(s^{TPSG}, \boldsymbol{a}^{TPSG}, t+1) = Q_m^{TPSG}(s^{TPSG}, \boldsymbol{a}^{TPSG}, t) + \beta^{TPSG}(\mathcal{R}_m^{TPSG}(t) + \gamma \max_{\boldsymbol{a}^{TPSG'} \in \mathcal{A}^{TPSG}} Q_m^{TPSG}(s^{TPSG'}, \boldsymbol{a}^{TPSG'}, t) - Q_m^{TPSG}(s^{TPSG}, \boldsymbol{a}^{TPSG}, t))$$

Simulation Settings

Param.	Value	Param.	Value	Param.	Value
M	3	B	10 MHz	C	10
N	300	D_n	$[2, 5]$ MB	V	20 m/s
t^{hover}	5 s	ξ	10^{-18}	f	3 GHz
q	100 m	p_n^{tran}	$[0.2, 0.5]$ W	H	120 m
S_m	6 GB	μ_c	$[1, 3]$ GB	φ	-174 dBm/Hz
a, b	9.6117, 0.1581	f_m^U	2 $Mbps$	Target region	1000 $m \times$ 1000 m

Simulation Settings (Cont'd)

Benchmark:

- Energy efficient oriented trajectory planning (EOTP): determines the trajectories of all UAVs with the aim of maximizing the energy efficiency but asks UAVs to return to the depot only when the batteries are approaching exhausted, and ignoring the update of application placement.
- Decentralized multiple UAVs cooperative reinforcement learning (DMUCRL) [1]: designed to maximize the energy efficiency of UAVs by controlling all UAVs to work collaboratively based on a double Q-learning (where each UAV contains a trajectory learner and an energy learner), and thus it also ignores the management of application placement in UAVs.

[1] Zhao, Chenxi, et al."Multi-UAV trajectory planning for energy-efficient content coverage: A decentralized learning-based approach." *IEEE Journal on Selected Areas in Communications* 2021.

Simulation Results

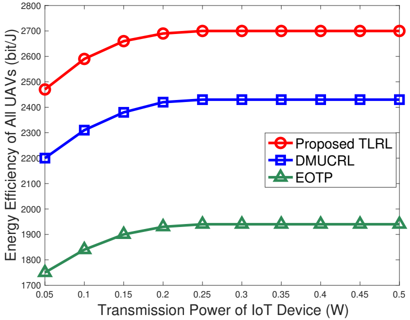


Figure: Energy efficiency w.r.t. transmission power of IoT devices.

- Energy efficiency first increases and then becomes stable due to limited computing capacity despite the increasing offloading tasks;
- TLRL outperforms both DMUCRL and EOTP because TLRL jointly optimizes all UAVs' trajectory planning, energy renewal and application placement.

Simulation Results (Cont'd)

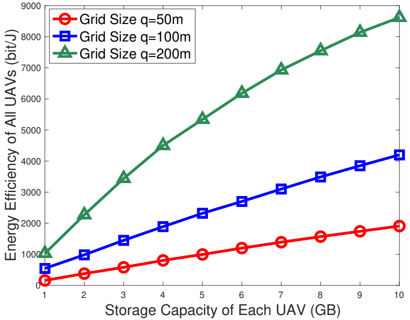


Figure: Energy efficiency w.r.t. storage capacity of each UAV.

- The larger the grid size is, the higher energy efficiency is obtained, because more IoT devices are included in a grid and, therefore, more offloading tasks;
- Energy efficiency increases monotonically with the storage capacity, because more types of applications can be placed in each UAV.

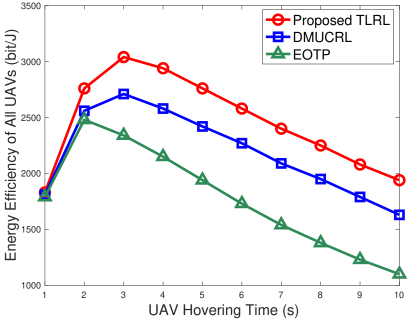


Figure: Energy efficiency w.r.t. UAV hovering time.

- The energy efficiency first increases with the UAV hovering time, and then decreases. Because with the growth of hovering time, more offloaded tasks can be computed when hovering. But when all tasks have been completely processed, they will become idle and consume hovering energy over the target region until hovering time expires.

Conclusion

- An energy efficient scheduling problem for multi-UAV assisted MEC has been studied.
- With the aim of maximizing the long-term energy-efficiency of all UAVs, a joint optimization of UAVs' trajectory planning, energy renewal and application placement is formulated.
- By taking the inherent cooperation and competition among UAVs, we reformulate such optimization problem as three coupled multi-agent stochastic games, and then propose a novel TLRL approach for reaching the equilibrium.
- Simulation results show that, compared to counterparts, the proposed TLRL approach can significantly increase the energy efficiency of all UAVs.

Thank you

Q&A

(Please feel free to contact Jiayuan Chen (Email: jiayuan.chen@nuaa.edu.cn).)