End-to-end Performance Optimization for Mixed FSO/RF-aided Non-Terrestrial Networks: A DRL Approach

NGUYEN Van Tinh - m5272016
Supervisor: Prof. LE Doan Hoang

The University of Aizu, Japan

Feb. 14th, 2025

Outline

I. Introduction

II. System Description

III. MARL for UAV Placement

IV. Simulation Results

V. Conclusion

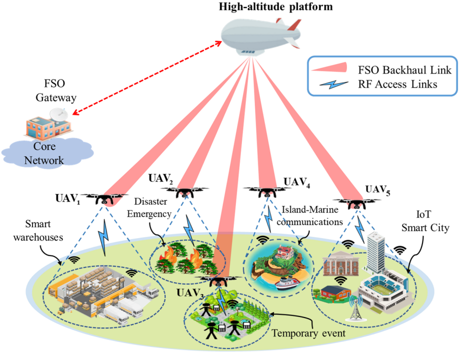
FSO-Based Air-Ground Integrated Networks

- FSO is a line-of-sight technology using infrared frequency bands (187 - 400 THz) for data transmission in space

  ⟹ *Large bandwidth, high-speed connections (∼ hundreds of Gbps or even Tbps)*

  ⟹ *Potential solution for backhaul networks*

- Space networks, employing Unmanned Aerial Vehicle (UAV), and High-Altitude Platform (HAP) as relay/aerial base station (ABS)



⟹ *Low cost, wide coverage, flexible deployment, and strong light-of-sight connectivity*

E.g., SoftBank's HAPS, Airbus's Zephyr S, and Huawei Digital Sky projects

The integration of HAP for FSO backhaul and UAV for last-mile access, forming the FSO-based air-ground integrated network (AGIN), is the promising and pragmatic NTN architecture for the future 6G era.

Critical Issues: End-to-end Dynamic Network

FSO Backhaul links:

- **Atmospheric turbulence:** air pockets with different refractive indexes cause the scintillation effect
- **Pointing error:** misalignment between the center of satellite beam footprint and that of the UAV detector
- **Cloud attenuation:** The liquid water particles in clouds cause the scattering phenomenon
  - Clouds are moving/changing over time

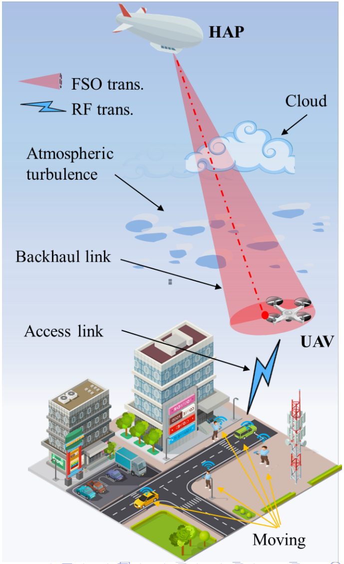
$\Longrightarrow$ *One of the most limiting factors for laser beams*

➡ *Unstable and time-varying channel → Limited capacity of the backhaul link*

Access link:

- Different distributions and the movement of end-users

➡ *Difficulty in the deployment of UAV-mounted BS*



Research Question

How to optimize the end-to-end network performance under the dynamic conditions from both backhaul (e.g., movement of clouds) and last-mile access (movement of end-users) networks?

$\Longrightarrow$ *Find the optimal position of UAV-mounted BS to maximize the end-to-end network performance. An efficient algorithm, indeed, is needed to tackle this critical issue.*

Literature Review

Extensive research endeavors have been conducted to find effective UAV placement algorithms for end-to-end networks comprising FSO backhaul and RF access links[1][2][3][4][5].

- However, the current studies mainly focus on terrestrial networks and ignore the effects of the FSO channel conditions[1][2][3][4][5]
- The current works mainly address the UAV placement problem with static networks[1][2][3][4]

➡ *An efficient algorithm for UAV placement in dynamic air-ground integrated networks is needed*

---

[1] S. Liu, H. Dahrouj, and M.-S. Alouini, "Joint user association and beamforming in integrated satellite-haps-ground networks," *IEEE Trans. Veh. Technol.*, vol. 73, no. 4, pp. 5162–5178, 2024.

[2] L. Yu, X. Sun, S. Shao, *et al.*, "Backhaul-aware drone base station placement and resource management for fso-based drone-assisted mobile networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 3, pp. 1659–1668, 2023.

[3] S. Zhang and N. Ansari, "3d drone base station placement and resource allocation with fso-based backhaul in hotspots," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3322–3329, 2020.

[4] D. Wu, X. Sun, and N. Ansari, "An fso-based drone assisted mobile access network for emergency communications," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 1597–1606, 2020.

[5] Y. Guan, S. Zou, H. Peng, *et al.*, "Cooperative uav trajectory design for disaster area emergency communications: A multiagent ppo method," *IEEE Internet Things J.*, vol. PP, pp. 1–1, Jan. 2023. DOI: `10.1109/JIOT.2023.3320796`.

Goals of the Study

1. We deploy an optimization framework leveraging multi-agent reinforcement learning (MARL) to address the UAV placement problem in the end-to-end AGIN network
   - The framework accounts for the dual constraints of both backhaul and access links
   - The dynamic conditions include the mobility of end-users and the movement of clouds
   - The algorithm seeks to determine the optimal UAV positions to maximize end-to-end throughput while adapting to the continuously changing network conditions
2. We conduct extensive simulations to validate the effectiveness of the framework
   - The results reveal that the trained agents can adapt to the dynamic nature of network environment and maintain high throughput performance

Outline

I. Introduction

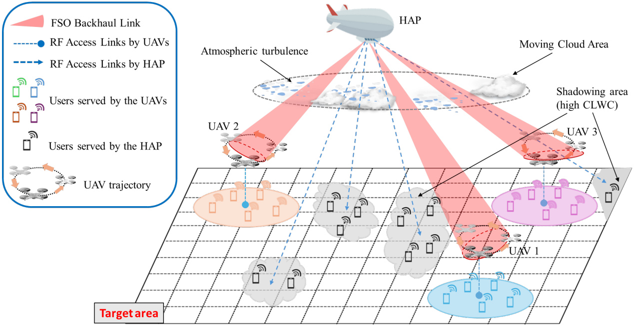
II. System Description

III. MARL for UAV Placement

IV. Simulation Results

V. Conclusion

System Model (1)



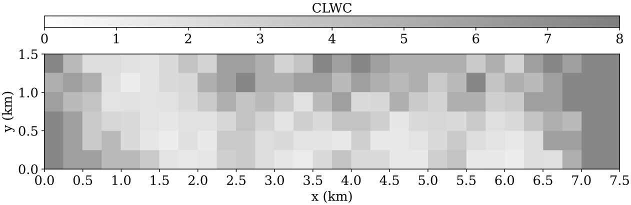
One HAP and multiple UAVs are deployed to provide internet connection to rural/remote areas or temporary events. The service duration is divided into T equal time slots. The end-to-end network scenario includes 2 main transmission links:

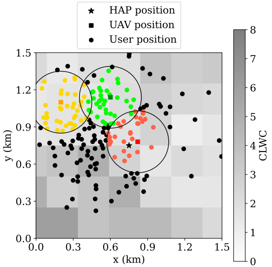
1. **Backhaul links:** from HAP to M UAVs → FSO transmission
2. **Access links:** from HAP and M UAVs to N users → RF transmission

System Model (2)

We consider a target area of 1.5 km x 1.5 km

- **Users**:
  - N mobile users (MU) are normally distributed with a random mean and a standard deviation of 300m
  - A part of users follow the Gauss-Markov mobility model
- **Cloud model:**
  - We consider a 7.5 km × 1.5 km cloud with heterogeneous CLWC[1]
  - The cloud moves to the west with a velocity of 8 m/s





---

[1] CLWC (cloud liquid water content) - a measure of the total liquid water contained in a cloud in a vertical column of the atmosphere (the less, the better)

Problem Formulation (1)

- Link association:
  - We consider a greedy approach that assigns the user to either the HAP or one of the UAVs that provides the strongest signal.
  - Let $\alpha_{n,i,t} \in \{0,1\}$ and $\beta_{n,t} \in \{0,1\}$ be binary indicate whether user n is associated with the i-th UAV or the HAP. The constraint is formulated as

$$\sum_{i=1}^{M} \alpha_{n,i,t} + \beta_{n,i,t} = 1$$

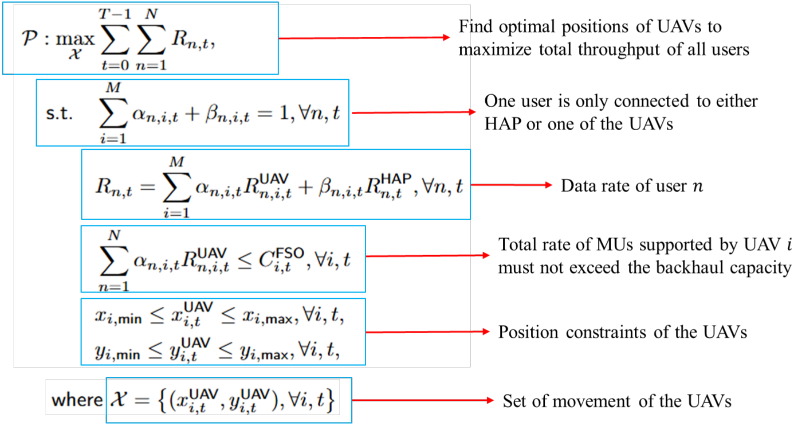
  - The data rate of user n at time slot t is given as

$$R_{n,t} = \sum_{i=1}^{M} \alpha_{n,i,t} R_{n,i,t}^{\text{UAV}} + \beta_{n,i,t} R_{n,t}^{\text{HAP}},$$

where $R_{n,i,t}^{\text{UAV}}$ and $R_{n,t}^{\text{HAP}}$ denote the data rates provided by the UAV i and the HAP, respectively.

Problem Formulation (2)

- Problem formulation
  - We aim to maximize the cumulative data rate of users by optimizing the UAV positions. The optimization problem can be formulated as



$$\mathcal{P}: \max_{\mathcal{X}} \sum_{t=0}^{T-1} \sum_{n=1}^{N} R_{n,t},$$

Find optimal positions of UAVs to maximize total throughput of all users

$$\text{s.t.} \quad \sum_{i=1}^{M} \alpha_{n,i,t} + \beta_{n,i,t} = 1, \forall n, t$$

One user is only connected to either HAP or one of the UAVs

$$R_{n,t} = \sum_{i=1}^{M} \alpha_{n,i,t} R_{n,i,t}^{\text{UAV}} + \beta_{n,i,t} R_{n,t}^{\text{HAP}}, \forall n, t$$

Data rate of user n

$$\sum_{n=1}^{N} \alpha_{n,i,t} R_{n,i,t}^{\text{UAV}} \leq C_{i,t}^{\text{FSO}}, \forall i, t$$

Total rate of MUs supported by UAV i must not exceed the backhaul capacity

$$x_{i,\text{min}} \leq x_{i,t}^{\text{UAV}} \leq x_{i,\text{max}}, \forall i, t,$$
$$y_{i,\text{min}} \leq y_{i,t}^{\text{UAV}} \leq y_{i,\text{max}}, \forall i, t,$$

Position constraints of the UAVs

$$\text{where } \mathcal{X} = \{(x_{i,t}^{\text{UAV}}, y_{i,t}^{\text{UAV}}), \forall i, t\}$$

Set of movement of the UAVs

Outline

I. Introduction

II. System Description

III. MARL for UAV Placement

IV. Simulation Results

V. Conclusion

Multi-agent Reinforcement Learning (MARL)

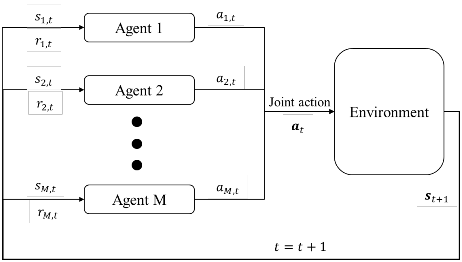


Figure: Diagram of the agent-environment interaction loop in MARL

- **RL:** a process in which an agent learns to make decisions *through trial and error*
- The problem is often modeled mathematically as *a Markov decision process (MDP)*, where the agents interact with the environment based on a particular policy
- At time step t, each agent observes its current observation $s_{i,t}$ and chooses an action $a_{i,t}$ from the action space. The combination of all actions from the agents is called a joint action. The environment then returns the new observation $s_{i,t+1}$ and reward/punishment $r_{i,t}$ for each agent.
- The algorithm aims to *maximize the cumulative received rewards*

Markov Decision Process (MDP)

Considered state, action, and reward function of UAV i:

- **Observation** $s_{i,t}$: the current location of the UAV, the locations of other UAVs, the heatmap of users (currently supported by that UAV or HAP), and the cloud heatmap of the area
- **Action** $a_{i,t}$: move (1)-north, (2)-west, (3)-south, (4)-east, (5)-remain stationary
- **Reward function**: we consider individual and teamwork (global) rewards to encourage collaboration among agents
  - *Global reward:*

$$r_t^{\text{glob}} = \begin{cases} c_1 \sum_{i=1}^{M} \sqrt{R_{i,t}^{\text{UAV}} \times C_{i,t}^{\text{FSO}}}, & R_{i,t}^{\text{UAV}} \geq R_i^{\text{thres}}, \forall i, \\ c_1 \sum_{i=1}^{M} \sqrt{R_{i,t}^{\text{UAV}} \times C_{i,t}^{\text{FSO}}} - 1, & \text{otherwise}, \end{cases}$$

where

c_1: normalization factor

$R_{i,t}^{\text{UAV}}$: Data rate provided to users by UAV i

$C_{i,t}^{\text{FSO}}$: FSO backhaul capacity of UAV i

R_i^{thres}: Data rate threshold of UAV i

Markov Decision Process (MDP)

Considered state, action, and reward function:

- **Reward function**:
  - *Local reward:*

$$r_{i,t}^{\text{local}} = \begin{cases} c_2\sqrt{R_{i,t}^{\text{UAV}} \times C_{i,t}^{\text{FSO}}} - n_{\text{c},i}p, & R_{i,t}^{\text{UAV}} \geq R_i^{\text{thres}}, \\ c_2\sqrt{R_{i,t}^{\text{UAV}} \times C_{i,t}^{\text{FSO}}} - n_{\text{c},i}p - 1, & \text{otherwise}, \end{cases}$$

where

c_2: normalization factor

$n_{c,i}p$: penalty term for overlapping coverage among UAVs

$\Longrightarrow$ *The combined reward for the i-th UAV:*

$$r_{i,t} = wr_t^{\text{glob}} + (1-w)r_{i,t}^{\text{local}},$$

where w is the weight ratio between global and local reward

UAV Placement Algorithm

Algorithm MARL-based UAV Placement Algorithm

```
for each UAV $i$ do
    Initialize the replay buffer $D_i$ with capacity $C_i$
    Initialize the $Q_i$ network with random weights $\boldsymbol{\theta_i}$
    Initialize the target $\hat{Q}_i$ network with weights $\hat{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i$
end for
for each episode do
    Initialize/Reset the network environment
    for each time slot $t$ do
        for each UAV $i$ do
            Obtain the observation $\boldsymbol{s}_{i,t}$ from the environment
            Obtain action $a_{i,t}$ according to the $\epsilon$-greedy policy
            Execute $a_{i,t}$ and perform link association
        end for
        Update the environment
        for each UAV $i$ do
            Observe reward $r_{i,t}$ and next state $\boldsymbol{s}_{i,t+1}$
            Store transition $(\boldsymbol{s}_{i,t}, a_{i,t}, r_{i,t}, \boldsymbol{s}_{i,t+1})$ in replay buffer $D_i$
            Sample a random minibatch of transitions from $D_i$
            Obtain the target Q-value $y_{i,t}$
            Perform a gradient descent step with respect to $\boldsymbol{\theta}_i$
            Reset $\hat{Q}_i = Q_i$ every fixed number of steps
        end for
    end for
end for
```

Outline

I. Introduction

II. System Description

III. MARL for UAV Placement

IV. Simulation Results

V. Conclusion

System Parameters

Name	Value
FSO Backhaul Links	
FSO transmit power	6 dBm
FSO bandwidth	3 GHz
HAP altitude	20 km
CLWC range	0.5 - 7.5 mg/m^3
RF Access Links	
HAP transmit power	35 dBm
UAV transmit power	25 dBm
Total RF bandwidth	600 MHz
UAV altitude	250 m
UAV coverage	250 m
RL Model	
DRL framework	DQN
Learning rate	0.001
Discount factor	0.99
Time slot duration	1 s
Total time slot	300 (training), 900 (testing)
Scale factor	$c_1 = 5.56 \times 10^{-10}$, $c_2 = 1.67 \times 10^{-9}$
Data rate threshold	0.6 Gbps
Weight ratio	0.35

Deployed Models

For comparison, we employ two benchmark algorithms:

1. **Multi-agent reinforcement learning without consideration of clouds (MARLwC):**
    - The agents do not have any information related to the status of clouds
      → *observation spaces do not include the current CLWC heatmap*
2. **Single-agent reinforcement learning (SARL):**
    - Each UAV is controlled by a DQN agent and makes the movement decision independently without collaboration with other UAVs
      → *positions of other UAVs are not included in the observation space, and the reward function only takes into account local rewards*

Episode Return

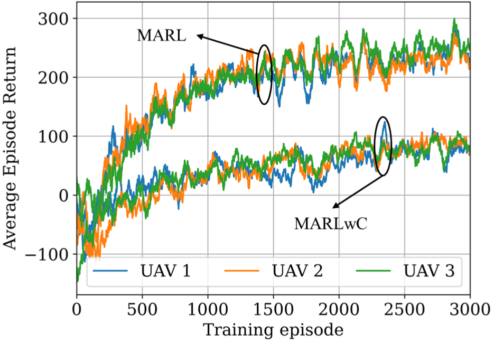


Figure: Average episode reward versus training episodes.

We can see that

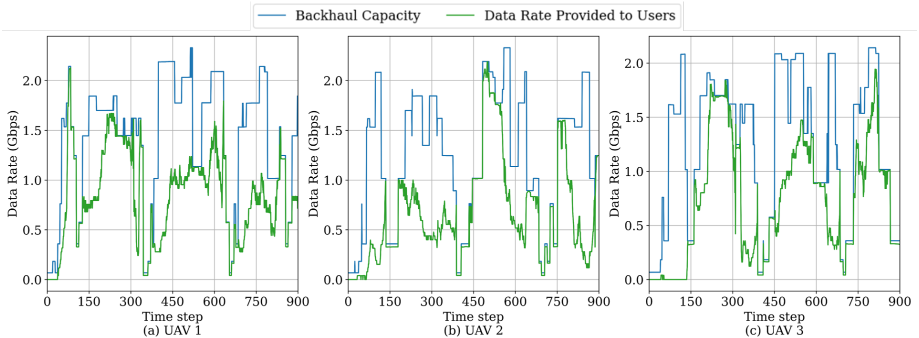
- The episode reward for each UAV gradually increases and converges after about 1500 episodes

⟹ *All agents are learning from the environment and making more efficient decisions to achieve higher returns.*

- MARL model gains significantly higher returns than MARLwC model.

⟹ *The MARL model can adapt well to the changing cloud conditions*

Backhaul Capacity vs. Data Rate



(a) UAV 1
(b) UAV 2
(c) UAV 3

Figure: Backhaul capacity and data rate provided to users using MARL algorithm.

- The trained MARL can maintain a high and stable backhaul capacity
- Also, the data rate delivered to users closely follows the upper bound dictated by the available FSO backhaul capacity

⟹ *The MARL algorithm successfully optimizes UAV positioning to mitigate moving cloud conditions and adapt to user distribution/mobility*

Total data rate of users (1)

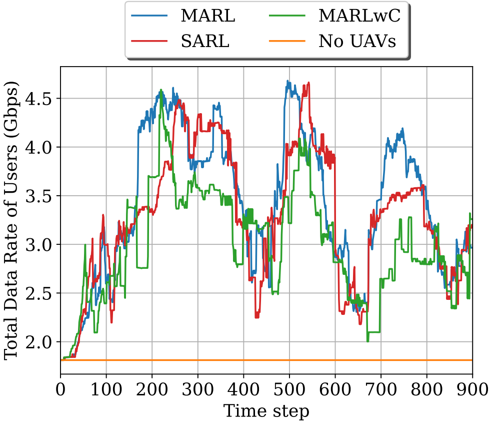


Figure: Total data rate of users over one test episode.

$\Longrightarrow$ The proposed MARL approach consistently achieves the highest overall performance while SARL follows closely behind. In contrast, the MARLwC algorithm exhibits the lowest overall throughput

$\Longrightarrow$ The deployment of UAVs as aerial base stations leads to a substantial increase in the total data rate delivered to users

Total data rate of users (2)

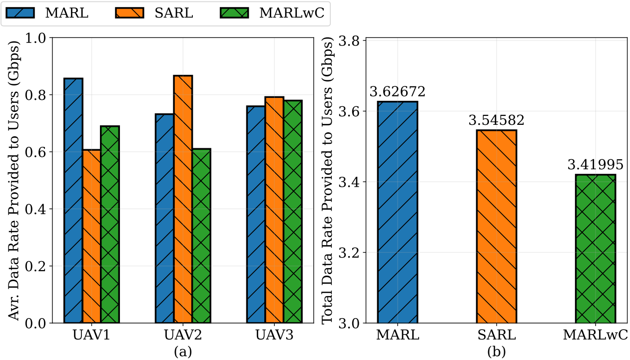


Figure: (a) Average data rate provided to users by each UAV; (b) Total data rate of users. (The results are taken across 100 test episodes).

- The data rate trends vary across each UAV. However, the MARL provides the most stable performance
- In total, the MARL algorithm can greatly enhance network performance, resulting in significant gaps with other algorithms

Outline

I. Introduction

II. System Description

III. MARL for UAV Placement

IV. Simulation Results

V. Conclusion

Conclusion

1. This thesis introduced an optimization framework leveraging MARL to address the UAV placement problem to optimize the end-to-end throughput of users in dynamic air-gorund integrated networks

2. Remarkable observations from the results:
    - The trained agents can adapt to the dynamic nature of network environment and maintain high throughput performance
    - The proposed algorithm performs significant gaps with other baseline algorithms, which highlights the collaboration and environmental awareness in optimizing network performance