Covertness-Aware Trajectory Design for UAV: A Multi-Step TD3-PER Solution

Presenter: Yuanjian Li & Prof. Hamid Aghvami *Fellow, IEEE*

Department of Enginnering
Faculty of Natural, Mathematical & Engineering Sciences
King's College London

20th March 2022

Outline

Brief Overview

System Model

- Pathloss Model

The Proposed Algorithm

Simulation Results

Brief Overview

Although covert transmissions have been intensively studied in the field of terrestrial communications, covert transmissions in UAV-aided networks have not drawn much attention so far, especially on the topic of how to help the UAV achieve low probability of being detected via deep reinforcement learning (DRL)-aided trajectory design. Motivated by the above observations, this paper investigates transmission throughput maximization problem for UAV-mounted network via path planning, subject to covert, velocity and mobility constraints. The main contributions are concluded as follows.

Brief Overview

- With the building-distribution-based A2G pathloss model and assuming that the Warden has no exact knowledge of its received noise power, the optimal detection threshold adopted by the Warden is derived. Considering that the UAV cannot gain perfect Warden's location information in practice, the estimated Warden's overall detection error rate on the perspective of the UAV is formulated, which then plays the role as the covert constraint.

Brief Overview

- The considered maximization problem is difficult to be solved via standard optimization methods, which is alternatively mapped into a Markov decision process (MDP) and tackled via DRL-aided approach. Specifically, a twin delayed deep deterministic policy gradient (TD3) agent is invoked to help the UAV find proper velocity from continuous action space, alongside the UAV's flight from the initial location to the destination. Furthermore, multi-step learning and prioritized experience replay (PER) techniques are integrated to help the TD3 agent hit a neater training performance.
- To highlight the advantages offered by the proposed multi-step TD3-PER solution, performance comparisons against DRL-based and non-learning baselines, i.e., double duelling deep Q network (D3QN) and straight-line solutions, are provided in numerical results.

System Model

Covertness-aware transmissions within UAV-aided network are considered, where a UAV u tries to deliver messages to legitimate nodes $l \in \mathbb{L}$ with low probability of being detected by a Warden w. Note that both the legitimate nodes and the Warden are located on the earth, while all the involved transceivers are equipped with single antenna. The UAV is supposed to reach a predefined destination from its initial location, with static flying altitude A. The UAV's horizontal location at time instant $t \in [0, T]$ remains in the range of $\vec{q}_{\mathsf{lo}} \preceq \vec{q}_u(t) \preceq \vec{q}_{\mathsf{up}}$, where $\vec{q}_{\mathsf{lo}} = (x_{\mathsf{lo}}, y_{\mathsf{lo}})$, $\vec{q}_{\mathsf{up}} = (x_{\mathsf{up}}, y_{\mathsf{up}})$, T represents the overall flight duration and $\preceq$ denotes the element-wise inequality. The flight duration T is uniformly divided into N time slots. The time slot length is delicately defined as a relatively small value $\Delta_t = T/N$, and thus the velocity and the distances from the UAV to the ground nodes can be treated as unchanged within each time slot. Moreover, the horizontal coordinates of legitimate nodes and Warden are indicated as $\vec{q}_l = (x_l, y_l)$ and $\vec{q}_w = (x_w, y_w)$, respectively.

Pathloss Model

In accordance with 3GPP urban-micro (UMi) pathloss model, the A2G pathloss in dB from $i \in \{l, w\}$ to the UAV within n-th time slot is given by

$$\aleph_i\left[\vec{q}_u(n)\right]=\begin{cases} \max\{\aleph', 30.9+[22.25-0.5\lg(A)] \\ \qquad\qquad \lg(d_{iu})+20\lg(f_c)\}, & \text{LoS} \\ \max\{\aleph_i^{\text{LoS}}\left[\vec{q}_u(n)\right], 32.4+ \\ [43.2-7.6\lg(A)]\lg(d_{iu})+20\lg(f_c)\}, & \text{NLoS} \end{cases}, \qquad (1)$$

where $\aleph' = 20\lg(d_{iu})+20\lg(f_c)+32.45$ represents the free space pathloss, f_c in GHz indicates the carrier frequency and $d_{iu}=\sqrt{||\vec{q}_u(n)-\vec{q}_i||^2+A^2}$ outputs the Euclidean distance between the UAV and i. From (1), it is straightforward to conclude that the availability of $\vec{q}_i$ is of essence for the UAV to estimate the corresponding A2G pathlosses.

Pathloss Model

However, it is difficult (or, even impossible) for the UAV to gain perfect location estimations of malicious equipments on the ground. Therefore, this paper adopts a practical assumption on location availability, i.e., the UAV can only obtain the Warden's location information with uncertainty, while the exact locations of the legitimate nodes are known by the UAV *a prior*. Specifically, uncertain location estimation model is invoked to characterize the noised location information (e.g., Gaussian estimation noises) of the Warden, expressed as

$$\vec{q}_w = \hat{\vec{q}}_w + \vec{\varepsilon}, \tag{2}$$

where $\hat{\vec{q}}_w = (\hat{x}_w, \hat{y}_w)$ and $\vec{\varepsilon} = (\ddot{x}_w, \ddot{y}_w) \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I})$ represent the estimated Warden's location and the corresponding estimation error, respectively.

Pathloss Model

To practically trace the type of experienced A2G pathlosses, building distribution within $\mathbb{A}$ should be taken into consideration. Then, the type of large-scale pathloss of A2G channels for UAV at arbitrary location $\vec{q}_u(n)$, i.e., LoS or NLoS in (1), can be accurately determined via checking the potential blockages between the UAV and ground receiver i.

Transmission Rate

For the ϖ-th channel use over the n-th time slot, the received signals at the legitimate node can be given by

$$y_l(\varpi,n)=\sqrt{P_u 10^{\frac{-\aleph_l[\vec{q}_u(n)]}{10}}}x_u(\varpi)+\varkappa_l(\varpi), \tag{3}$$

where $x_u(\varpi)\sim\mathcal{CN}(0,1)$ is the transmitted signal from the UAV to the legitimate node, P_u means the UAV's transmit power and $\varkappa_l(\varpi)\sim\mathcal{CN}(0,\sigma_l^2)$ denotes the AWGN. Note that $\varpi=\{1,2,\dots,c\}$ indicates the symbol index within a time slot and c measures the slot length. Then, the transmission rate in bps/Hz from the UAV to the legitimate node over the n-th time slot can be derived as

$$R_l(n)=\log_2(1+\Gamma_l(n)), \tag{4}$$

where $\Gamma_l(n)=P_u 10^{\frac{-\aleph_l[\vec{q}_u(n)]}{10}}/\sigma_l^2$ represents the corresponding signal-to-noise-ratio (SNR).

Warden's Uncertain AWGN Model

In practice, it is impractical for transceivers to gain perfect information regarding their received AWGN. Therefore, uncertain AWGN model is adopted to characterize the dynamics of Warden's AWGN power. Specifically, the Warden only knows the distribution of its received AWGN's variance, given by

$$f_{\sigma_w^2}(x) = \begin{cases} \frac{1}{2\ln\left(10^{\frac{\iota}{10}}\right)x}, & x \in [10^{-\frac{\iota}{10}}\hat{\iota}, 10^{\frac{\iota}{10}}\hat{\iota}] \\ 0, & \text{otherwise} \end{cases}, \tag{5}$$

where ι in dB measures the degree of noise uncertainty and $\hat{\iota}$ indicates nominal noise power.

Warden's Optimal Detection Threshold

Proposition 1

The optimal detection threshold $\tau^*(n)$ adopted by the Warden within the n-th time slot is given by

$$\tau^*(n) = \begin{cases} 10^{\frac{\iota}{10}}\hat{\iota}, & P_u 10^{\frac{-\aleph_w[\vec{q}_u(n)]}{10}} \geq 10^{\frac{\iota}{10}}\hat{\iota} - 10^{-\frac{\iota}{10}}\hat{\iota} \\ P_u 10^{\frac{-\aleph_w[\vec{q}_u(n)]}{10}} + 10^{-\frac{\iota}{10}}\hat{\iota}, & \textit{otherwise} \end{cases}. \qquad (6)$$

Warden's Optimal Detection Threshold

Then, the corresponding minimum overall detection error rate can be calculated as

$$\mathbb{P}_{de}^{*}(n)= \begin{cases} 0, & P_u 10^{\frac{-\aleph_w[\vec{q}_u(n)]}{10}} \geq 10^{\frac{\iota}{10}}\hat{\iota} - 10^{-\frac{\iota}{10}}\hat{\iota} \\ \frac{1}{2\ln\left(10^{\frac{\iota}{10}}\right)} \ln\left(\frac{10^{\frac{\iota}{10}}\hat{\iota}}{P_u 10^{\frac{-\aleph_w[\vec{q}_u(n)]}{10}} + 10^{-\frac{\iota}{10}}\hat{\iota}}\right), & \text{otherwise} \end{cases} . \qquad (7)$$

Warden's Optimal Detection Threshold

Proposition *1* enables the Warden minimize its overall detection error rate within arbitrary time slot, via providing the optimal detection threshold. Then, the UAV needs to estimate the Warden's overall detection error rate within each time slot, based on which it can make adaptive decisions to counter detections from the Warden, e.g., trajectory design. With the considered uncertain location estimation model (2), the UAV can estimate the expected version of overall detection error rate suffered by the Warden.

UAV's estimated detection error rate

Proposition 2

On the perspective of UAV, the expected overall detection error rate under uncertain Warden's location estimation model within the n-th time slot can be derived as

$$\bar{\mathbb{P}}_{de}(n) = \begin{cases} 0, & P_u 10^{\frac{-\aleph_w[\vec{q}_u(n)]}{10}} \geq 10^{\frac{\iota}{10}}\hat{\iota} - 10^{-\frac{\iota}{10}}\hat{\iota} \\ \hat{\mathbb{P}}_{de}(n), & \textit{otherwise} \end{cases}, \tag{8}$$

where

$$\bar{\mathbb{P}}_{de}(n) \simeq \hat{\mathbb{P}}_{de}(n) = \frac{\sum_{\check{c}=1}^{\hat{c}} \ln\left(\frac{10^{\frac{\iota}{10}}\hat{\iota}}{P_u 10^{\frac{-\aleph_w[\vec{q}_u(n),\check{c}]}{10}} + 10^{-\frac{\iota}{10}}\hat{\iota}}\right)}{2\hat{c}\ln\left(10^{\frac{\iota}{10}}\right)}. \tag{9}$$

Problem Formulation

This paper aims to maximize transmission throughput from the UAV to legitimate nodes alongside the UAV's trajectory from the initial UAV location to the destination, via designing UAV's marching direction within each time slot, subject to covert, velocity and mobility constraints. Then, the corresponding optimization problem can be formulated as

$$\text{(P1)}:\max_{\vec{v}_u(n)} \sum_{n=1}^{N} \sum_{l\in\mathbb{L}} R_l(n), \tag{10a}$$

$$\text{s.t.} \bar{\mathbb{P}}_{de}(n) \geq 1-\varsigma, \tag{10b}$$

$$\vec{q}(n+1) = \vec{q}(n) + \Delta_t \vec{v}_u(n), \|\vec{v}_u(n)\| = V, \tag{10c}$$

$$\vec{q}_{\text{lo}} \preceq \vec{q}_u(n) \preceq \vec{q}_{\text{up}}, \vec{q}_u(0) = \vec{q}_u(I), \vec{q}_u(N) = \vec{q}_u(D). \tag{10d}$$

Problem Formulation

The constraint (10b) makes sure that a certain level of covert transmission can be achieved, while the constraints (10c) and (10d) indicate the velocity and mobility regulations. For simplicity, the factor Δ_t in (10a) is omitted. Intuitively, the UAV needs to adopt proper flying direction within arbitrary time slot during its flight, for not only avoiding "covert holes" where the covert requirement cannot be satisfied but also directing itself to visit possible locations where greater transmission rate is achievable. Hence, it is non-trivial for the UAV to carefully design its velocity for arbitrary time slot.

Because of the building-distribution-based pathloss and uncertain location estimation models, it is challenging to tackle (P1) via standard optimization techniques (e.g., convex optimization), if not impossible. Alternatively, this paper aims to design a DRL-aided approach to efficiently solve the formulated optimization goal.

MDP Formulation

To design the DRL-aided solution, the first step is to map the considered problem into a MDP, stated as follows.

- $\mathcal{S}$: The state space is continuous, which contains possible UAV locations within $\mathbb{A}$, subject to $\vec{q}_{\mathsf{lo}} \preceq \vec{q}_u \preceq \vec{q}_{\mathsf{up}}$.
- $\mathcal{A}$: The action space is continuous, which involves possible velocity options $\vec{v}_u \in \mathbb{R}^{1*2}$, subject to $\|\vec{v}_u\| = V$.
- $\mathcal{T}$: State transition is deterministic, governed by (10c).
- r: According to the optimization objective (10a), it is direct to design the reward as $r(\vec{q}_u) = \sum_{l\in\mathbb{L}} R_l(\vec{q}_u) - 1$, where the penality -1 is used to encourage the UAV to reach the destination with fewer steps.

The Proposed Algorithm

Algorithm 1: The Proposed Multi-Step TD3-PER Solution

```
Initialization: Initialize the twin critic networks $Q(s, a|\boldsymbol{\theta}_{C_1})$, $Q(s, a|\boldsymbol{\theta}_{C_2})$ and the actor network $\mu(s|\boldsymbol{\theta}_{\mu})$, then update their target networks via $\boldsymbol{\theta}^{-}_{C_1} \leftarrow \boldsymbol{\theta}_{C_1}$, $\boldsymbol{\theta}^{-}_{C_2} \leftarrow \boldsymbol{\theta}_{C_2}$ and $\boldsymbol{\theta}^{-}_{\mu} \leftarrow \boldsymbol{\theta}_{\mu}$. Initialize the PER buffer R with capacity $C$. Set the size of mini-batch as $N_{mb}$. Set the step length of multi-step learning as $N_{ms}$. Set the policy update delay as $N_{pud}$. Set target network update factor as $\tau$;
for $te = [1, te_{\max}]$ do
    Set time step $n = 0$. Rest the UAV to its initial location as $\vec{q}u(n) \in \mathcal{S}$. Initialize a sliding buffer $\hat{\text{R}}$ with capacity $N_{ms}$;
    repeat
        Select and execute action $a_n = \mu(\vec{q}u(n)|\boldsymbol{\theta}_{\mu}) + \vartheta$, then observe the next state $\vec{q}_u(n+1)$ and the immediate reward $r_n = r[\vec{q}_u(n+1)]$;
        Get and record 1-step transition $\{\vec{q}_u(n), a_n, r_n, \vec{q}_u(n+1)\}$ into $\hat{\text{R}}$;
        if $n \geq N_{ms}$ then
            Generate the $N_{ms}$-step reward $r_{n-N_{ms}:n} = \sum_{n_{ms}=0}^{N_{ms}-1} \gamma^{n_{ms}} r_{n-N_{ms}+n_{ms}}$ from $\hat{\text{R}}$ and record $N_{ms}$-step experience $\{\vec{q}_u(n - N_{ms}), a_{n-N_{ms}}, r_{n-N_{ms}:n}, \vec{q}_u(n)\}$ into R;
        end
        Sample a mini-batch of $N_{mb}$ $N_{ms}$-step transitions from R with priorities $p_{n_{mb}}$;
        $y_{n_{mb}} = r_{n_{mb}:n_{mb}+N_{ms}} + \gamma^{N_{ms}} \min_{j=1,2} Q(\vec{q}_u(n_{mb}+N_{ms}), \mu(\vec{q}_u(n_{mb}+N_{ms})|\boldsymbol{\theta}^{-}_{\mu}) + \vartheta^{-}|\boldsymbol{\theta}^{-}_{C_j})$;
        Compute the mean squared losses of the twin critics as
          $\mathcal{L}^{sq}(\boldsymbol{\theta}_{C_j}) = \frac{1}{N_{mb}} \sum n_{mb} \frac{1}{Cp_{n_{mb}}} (y_{n_{mb}} - Q(\vec{q}_u(n_{mb}), a_{n_{mb}}|\boldsymbol{\theta}_{C_j}))^2$;
        Update the twin critic networks via gradient decent aiming to minimize $\mathcal{L}^{sq}(\boldsymbol{\theta}_{C_j})$;
        Every $N_{pud}$ times the twin critic netowrks are trained, update the actor network via gradient ascent to maximize $J(\boldsymbol{\theta}) = \frac{1}{N_{mb}} \sum n_{mb} Q(s_{n_{mb}}, \mu(s_{n_{mb}}|\boldsymbol{\theta}_{\mu})|\boldsymbol{\theta}_{C_1})$, then update the target actor and target twin critics in a soft copy fashion as $\boldsymbol{\theta}^{-} \leftarrow \tau\boldsymbol{\theta} + (1-\tau)\boldsymbol{\theta}^{-}$;
        Let $n+ = 1$;
    until $\vec{q}u(n) = \vec{q}u(T) \ || \ \vec{q}u(n) \notin \mathcal{S} \ || \ n = N_{\max}$;
end
```

High-Level Workflow

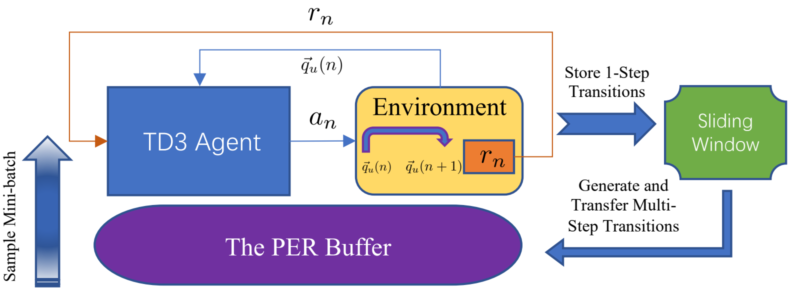


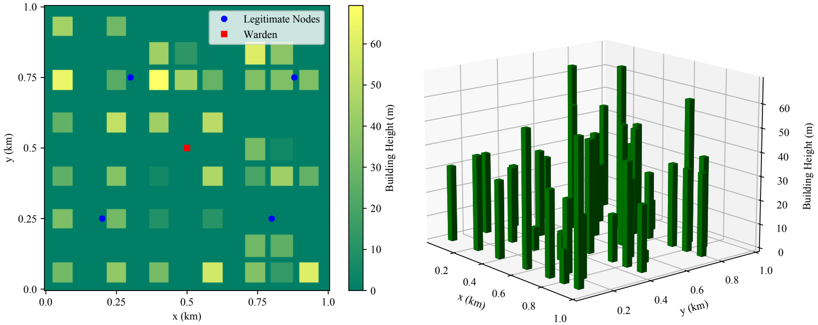
Figure 1: High-level workflow of the proposed multi-step TD3-PER solution

Simulation Parameter Settings

Parameters	Values	Parameters	Values	Parameters	Values
Side length of $\mathbb{A}$ D	1 km	Duration of time slot Δ_t	0.5 s	Speed of the UAV V	50 m/s
Carrier frequency f_C	2 GHz	Flying altitude of UAV A	100 m	Location estimation error variance σ_e^2	0.025
AWGN variance σ_l^2	-90 dBm	UAV's transmit power P_u	30 dBm	Covert requirement ς	0.001
Noise uncertainty degree ι	3dB	Nominal noise power $\hat{\iota}$	10^{-6}	Amount of numerical evaluation $\hat{c}$	1000
ITU building distribution parameter $\hat{\alpha}$	0.2	ITU building distribution parameter $\hat{\beta}$	40	ITU building distribution parameter $\hat{\gamma}$	25
Amount of buildings $\hat{\beta}D^2$	40	Expected size of each building $\hat{\alpha}/\hat{\beta}$	0.005 km^2	Maximum height of buildings	70 m
Replay buffer capacity C	10^6	Mini-batch size N_{mb}	32	Multi-step learning length N_{ms}	6
Policy update delay N_{pud}	10	Target network update factor τ	10^{-5}	Actor noise power variance σ_μ^2	1
Target Actor noise power variance $\sigma_{\mu^-}^2$	1	UAV exploration step threshold $N_{\max}$	150	Reaching destination bonus	4000
Hitting boundary penality	-10000	Visiting covert hole penality	-1000	Learning rates for actor/critic	$10^{-4}/10^{-3}$

Table 1: Simulation Parameter Settings

Building Distribution

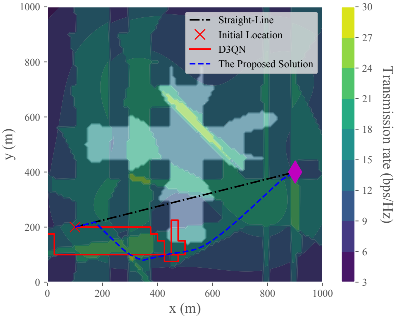


(a) Local building distribution

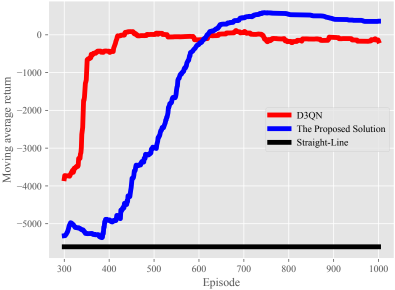
(b) 3D view of local building distribution

Figure 2: The building distribution under consideration

Performance Comparison



(a) Comparison on designed trajectories



(b) Comparison on moving average returns

Figure 3: Comparison on designed trajectory and moving average returns

Q & A

Thanks for your attentions

This is the end of this demonstration
Any Question?