Cooperative Dynamic Weapon-Target Assignment in a Multiagent Engagement

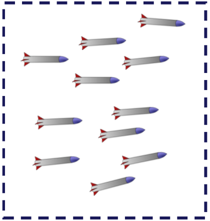
Gleb Merkulov*, Eran Iceland°, Shay Michaeli*,
Oren Gal†, Ariel Barel*, and Tal Shima*

* Technion Israel Institute of Technology
° Hebrew University of Jerusalem
† University of Haifa

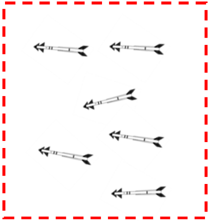
2025 IAAC3, 28 Apr. 2025

Scenario and Objective I

Swarm Attack Scenario



Interceptor
single wave

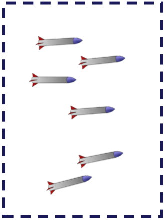


Swarm attack

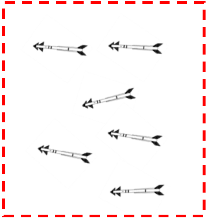
Solution Approaches

- **Single Shot**
  - Improved miss probability
  - Bad resource management

Swarm Attack Scenario



Interceptor
first wave



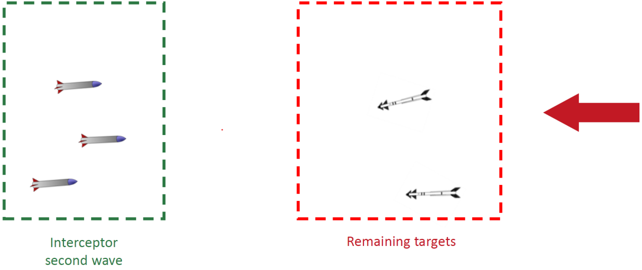
Swarm attack

Solution Approaches

- **Single Shot**
  - Improved miss probability
  - Bad resource management
- **Shoot-Look-Shoot**
  - Better resource management
  - Time constraints, reallocation

Scenario and Objective I

Swarm Attack Scenario

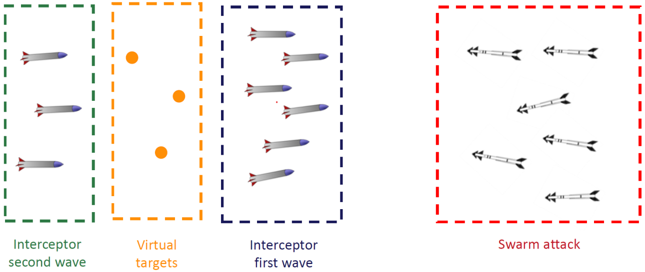


Solution Approaches

- **Single Shot**
  - Improved miss probability
  - Bad resource management
- **Shoot-Look-Shoot**
  - Better resource management
  - Time constraints, reallocation

Scenario and Objective I

Swarm Attack Scenario



Solution Approaches

- **Single Shot**
  - Improved miss probability
  - Bad resource management
- **Shoot-Look-Shoot**
  - Better resource management
  - Time constraints, reallocation
- **Shoot-Shoot-Look**
  - Dynamic allocation
  - Highest complexity

Challenges

- Computationally hard
- Dynamic adaptation
- VT allocation

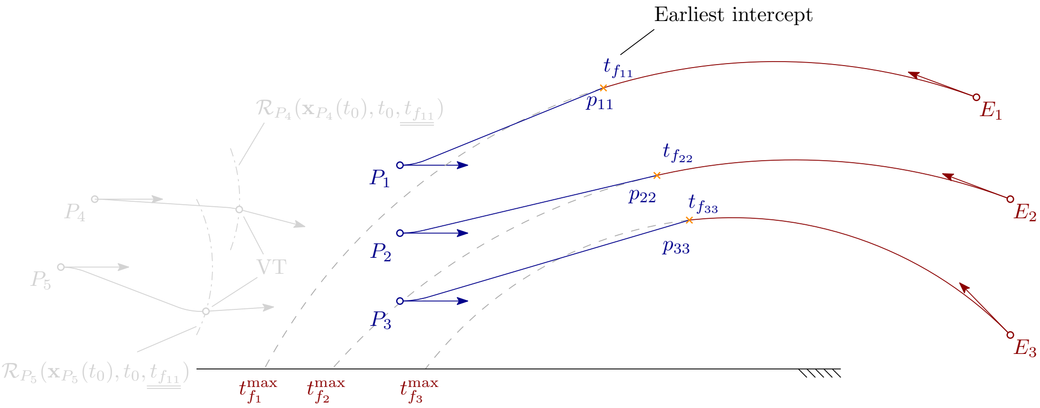
Objective

Dynamic WTA strategy

Global Assumptions

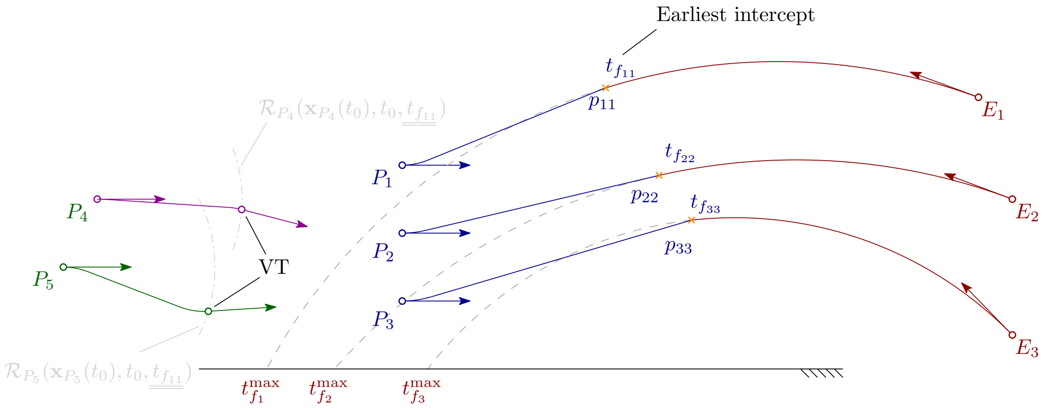
- Planar nonlinear engagement
- Predictable Evader motion
- Unicycle models for Pursuers and Evaders
- Constant speed
- Perfect information

Shoot-Shoot-Look Scenario



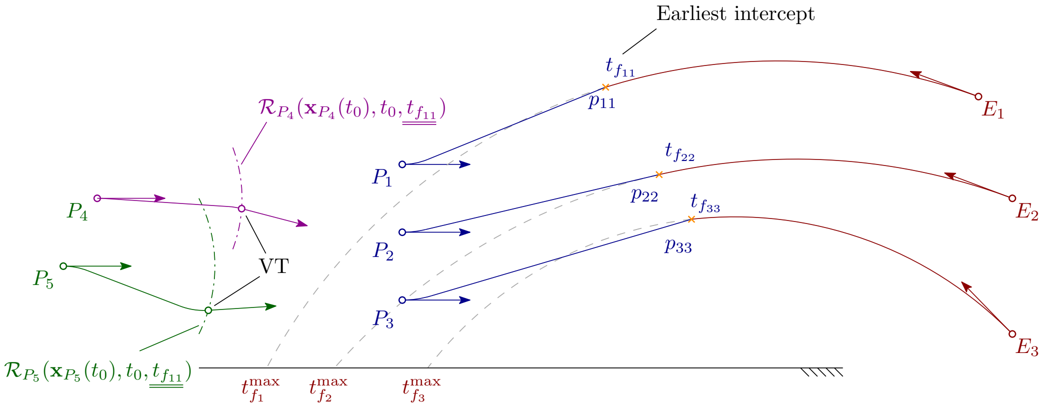
- First wave is allocated a-priori
- Intercept times define allocation decision instances

Shoot-Shoot-Look Scenario



- Backup pursuers are assigned to virtual targets (VT)
- Virtual target = position + heading = future pursuer state

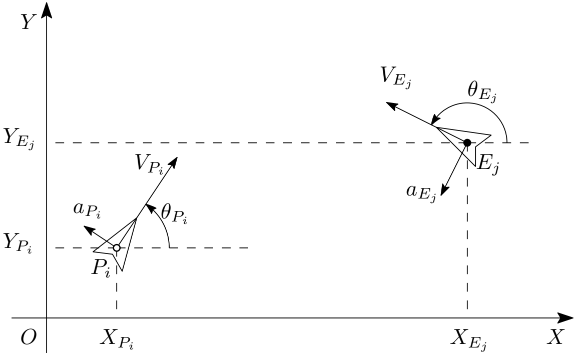
Shoot-Shoot-Look Scenario



- Virtual targets are samples from reachable sets
- Reachable set – all states that can be attained at time $t_{f_{11}}$ from the initial state $\mathbf{x}_P(t_0)$

Engagement Kinematics

Engagement Geometry



Motion Models

$$\begin{aligned} \dot{x}_k &= V_k \cos\theta_k \\ \dot{y}_k &= V_k \sin\theta_k \qquad k \in \{P_i, E_j\} \\ \dot{\theta}_k &= a_k / V_k \end{aligned}$$

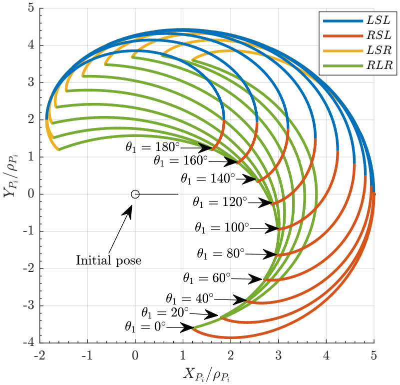
- $a_{E_j} = \text{const}$ – circular motion
- $a_{P_i} \in \{0, \pm a_{P_i}^{\max}\}$ – Dubins vehicle

- Pursuer employs min-time trajectories against VT [1] and Evaders [2]

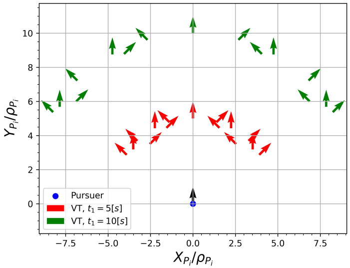
[1] Dubins, L. E. (1957). On curves of minimal length with a constraint on average curvature, and with prescribed initial and terminal positions and tangents. American Journal of mathematics, 79(3), 497-516.
[2] Zheng, Y., Chen, Z., Shao, X., & Zhao, W. (2021). Time-optimal guidance for intercepting moving targets by Dubins vehicles. Automatica, 128, 109557.

Reachable Set & Virtual Target Selection

Pursuer Reachable Set



VT Sampling



- Analytical RS description from [3]
- *LSL*, *RSL*, *LSR* to ensure min-time paths

[3] Patsko, V. S., & Fedotov, A. A. (2022). Three-dimensional reachable set for the Dubins car: Foundation of analytical description. Commun. Optim. Theory, 2022, 1-42.

Intercept Model

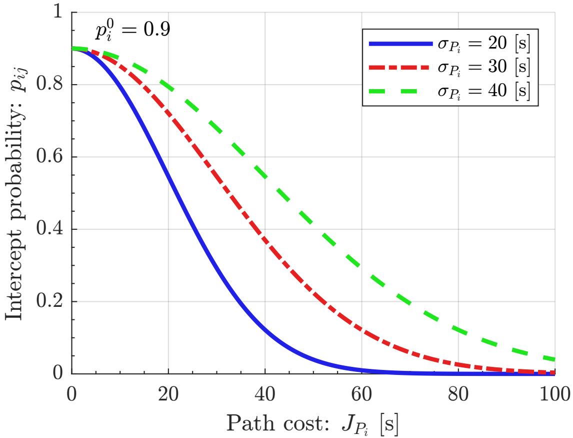
- Pursuer motion related to intercept probability

$$\boxed{p_{ij}(t_{d_k}) = p_i^0 \exp\left(-\frac{J_{P_i}^2(t_{f_{ij}})}{2\sigma_{P_i}^2}\right)}$$

- J_{P_i} – path cost = time + control effort

$$J_{P_i}(t) = t + \alpha \int_0^t a_{P_i}^2(\xi)\,\mathrm{d}\xi$$

- Allows small corrections
- Penalizes large corrections
- Can extend to better model



Objective and Reward Functions

Objective – intercept maximal number of evaders s.t. time constraints

Status dynamics

$$E_j(t_{d_{k+1}}) = E_j(t_{d_k}) - A_{ij}(t_{d_k}) w_j(t_{d_k})$$

- t_{d_k} – decision time (engagement outcome instance)
- $E_J(t_{d_k}) \in \{0, 1\}$ – evader status
- $A_{ij}(t_{d_k}) \in \{0, 1\}$ – allocation variable
- $w_j(t_{d_k}) \in \{0, 1\}$ – random engagement outcome indicator (1 with prob. $p_{ij}(t_{d_k})$)

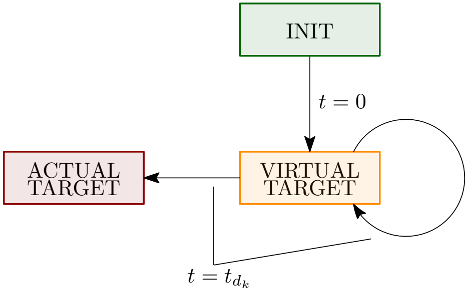
Equivalent exact reward function

$$R = \max_{\text{allocation}} \left\{ \sum_{k=1}^{K-1} p_{ij}(t_{d_k}) \right\}$$

- Allocation – VT & Evaders
- Exact reward is sparse

Backup Pursuer Decision Making

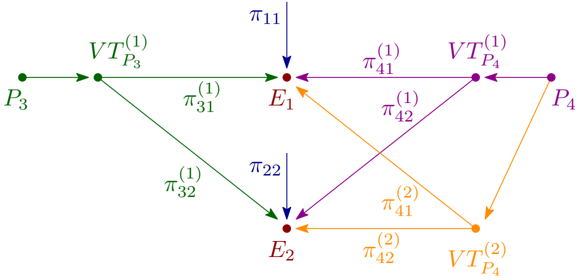
Decision Flowchart



Allocation Policy

- Greedy centralized allocation to free evader
- Sequential decentralized VT allocation – Greedy vs RL

Information Available to Pursuer



- Kinematics $\rightarrow$ intercept probabilities
  - π_{ij} – first-wave intercept probs.
  - $\pi_{ij}^{(l)}$ – predicted intercept prob. through $VT_{P_i}^{(l)}$

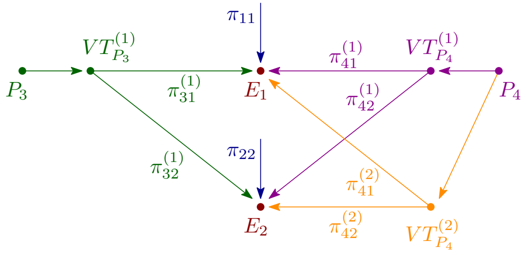
Greedy Heuristic Algorithm

Heuristic Idea

Greedy altruism

VT: maximize intercept prob. addition

Example VT Evaluation



Greedy Algorithm

1. If there is an unengaged evader – greedy allocation (max. intercept probability)
2. If all evaders engaged
   2.1 Initialize cumulative evader intercept probabilities

$$\pi_j = p_{ij}, \quad i \in \text{first wave}$$

   2.2 For each backup pursuer P_i
   - select the VT as

$$l_i^* = \arg\max_{l=1\ldots L} \left\{ \sum_{j=0}^{M} \pi_{ij}^{(l)} (1 - \pi_j) \right\}, \quad VT^*_{P_i} = VT^{l_i^*}_{P_i}$$

   - Update cumulative intercept probabilities

$$\pi_j \leftarrow \pi_j + \frac{(1 - \pi_j)\pi_{ij}^{l_i^*}}{\text{num. live evaders}}$$

RL Algorithm

Algorithm steps:

1. If there is an unengaged evader – greedy allocation (max. intercept probability)
2. Otherwise select VT as RL action

Rewards:

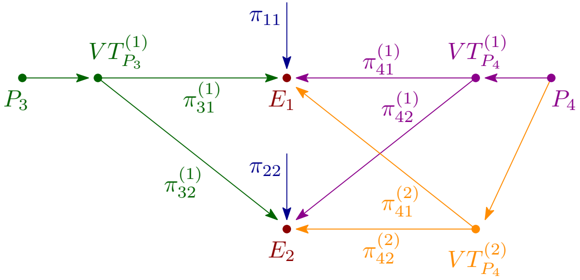
1. *Exact (sparse)* – next intercept probability $p_{i_k,j_k}(t_{d_k})$
2. *Non-sparse VT* – for current backup pursuer P_i:

   1. assign a score for each potential VT:

   $$S_i^{(l)} = \sum_{j=1}^{M} (1-\pi_j) \cdot \left[\max_{1 \le j \le M} \pi_{ij}^{(l)} \right]$$

   2. assign reward as added score

   $$R = \frac{S_i^{(l)} - S_{i-1}^{(l)}}{\text{num. live evaders}}$$



Scenario Example

- # Pursuers $= 15$, 10 – first wave, 5 – backup
- # Evaders $= 10$
- # VT $= 9$

Example scenario video

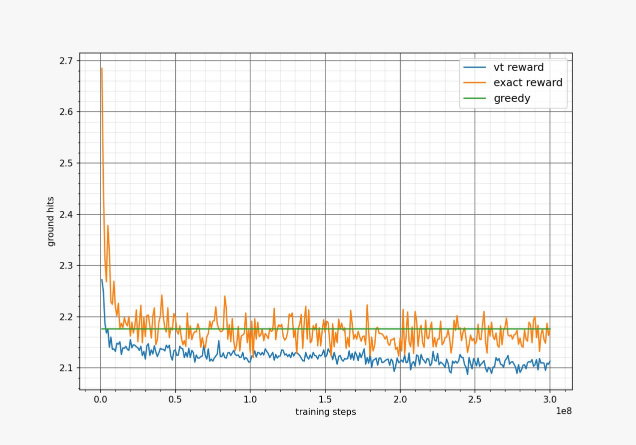
Statistical Analysis

RL vs Greedy

		Mean ground hits	
		13 vs 9	**15 vs 10**
Greedy		2.1310	2.2178
RL	Exact	2.0928	2.1917
	Non-sparse	2.0798	2.1581

- Greedy close to RL
- Non-sparse reward better than exact
- Non-sparse reward continues to improve ground hits

Learning curves



Conclusions

- **Generalized formulation of dynamic WTA problem in Shoot-Shoot-Look scenario**
- Exact reward function
- Greedy and RL algorithms
  - Both used as mutual optimality measures
  - Greedy close to RL
  - Non-sparse reward is better than exact for RL
  - Kinematic features did not improve performance $\Rightarrow$ **probabilistic features are sufficient**

Thank you for your attention!