sUAS Separation Assurance with Mixed Aircraft Equipage and Heterogeneous Tactical Deconfliction Policies

Iman Sharifi, Hyeong T. Kim, Maheed H. Ahmed, Mahsa Ghasemi, Peng Wei

Elmore Family School of Electrical and Computer Engineering

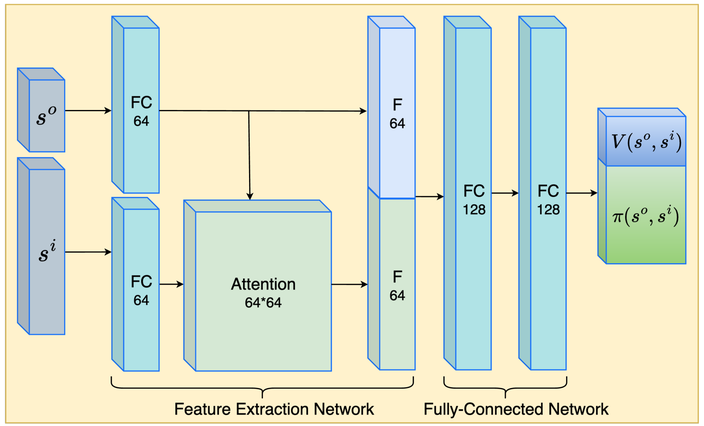
MOTIVATION

- In the envisioned dense urban airspace, multiple companies operate heterogeneous drones with distinct configurations, e.g., velocity/acceleration, sensing/communication ranges, and proprietary policies, making tactical deconfliction of these small unmanned aerial systems (sUASs) highly challenging.
- Existing multi-agent reinforcement learning (MARL) frameworks are often unrealistic as they consider either fully heterogeneous or fully homogeneous policies for all drones!

This research aims to answer three key questions:

1. Can heterogeneous and homogeneous drones with MARL policies reach an equilibrium in dense traffic?
2. If they reach, is the training fair to all drones, or discriminatory?
3. How does MARL policies interact with other policies? Are they still fair?

MARL: PPOA2C



Attention-based proximal-policy-optimization-driven advantage actor-critic (PPOA2C)

Configurations

Parameter	Notation	Configuration X (strong)	Configuration Y (weak)
Speed Range (m/s)	$[v_{\min}, v_{\max}]$	[0, 44.88]	[0, 30.12]
Acceleration (m/s²)	Δv / Δt	{−1.71, 0, 1.71}	{−1.02, 0, 1.02}
Sensing Range (m)	$\mathcal{R}$	1000	750

MARL Components

States

Ownship: $s_t^o = \{d^o, v^o, \theta^o, v'^o\}$, Intruder: $s_t^i = \{d^i, v^i, \theta^i, v'^i\}$

Actions

$\mathbb{A} = [-\Delta v,\ 0,\ +\Delta v\ ]$

Rewards

$$R(s,a,t) = R_1(s) + R_2(s) + R_3(a) + R_4(s) + R_5(t)$$

LoS Reward

$$R_1(s) = \begin{cases} -1, & \text{if } d_o^i < d_{NMAC} \\ \alpha(-1 + \dfrac{d_o^i - d_{NMAC}}{d_{LoWC} - d_{NMAC}}), & \text{if } d_{NMAC} \le d_o^i \le d_{LoWC} \\ 0, & \text{otherwise} \end{cases}$$

Velocity Reward

$$R_2(s) = \begin{cases} -\psi_1^v, & \text{if } v^o < v_{min}^o + \eta_1^v \\ \psi_2^v, & \text{if } v^o > v_{max}^o - \eta_2^v \\ 0, & \text{otherwise} \end{cases}$$

Action Reward

$$R_3(a) = \begin{cases} -\psi_1^a, & \text{if } a_t^o \neq a_{t-1}^o \\ -\psi_2^a, & \text{if } a_t^o \neq hold \\ 0, & \text{otherwise} \end{cases}$$

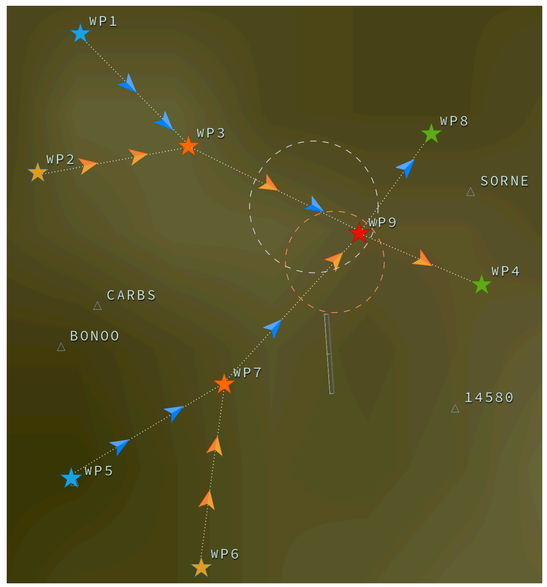
Mission Completion Reward

$$R_4(s) = \begin{cases} \psi_m, & \text{if } d_f^o < \eta_m \\ 0, & \text{otherwise} \end{cases}$$

Time Reward

$$R_5(t) = \begin{cases} -\psi_t & \text{if } t < T \\ 0, & \text{otherwise} \end{cases}$$

BlueSky Scenario



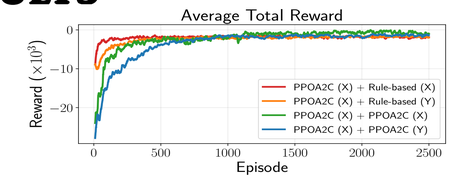
Frisco City, Dallas, Texas

Acknowledgement

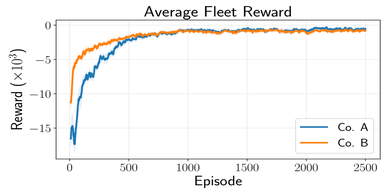
This work was supported in part by the National Aeronautics and Space Administration under Grant 80NSSC24M0070.

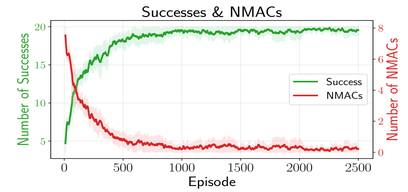
TRAINING RESULTS

Policy-Configuration Results:

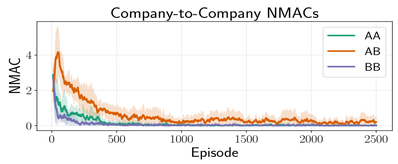


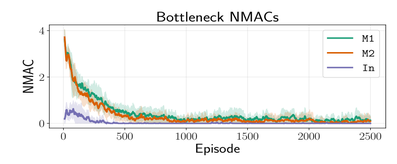
PPOA2C (X) + PPOA2C (Y):





Key takeaway: All policy-configuration models reach an equilibrium to maintain a conflict-free airspace. NMACs decrease over episodes and converges to the minimum.





Key takeaway: All the NMAC categories converge to zero, but Co. A and Co. B still have NMACs at highly dense scenarios due to the complexity of the scenario.

EVALUATIONS

$F_t = \left(1 - \frac{|T_A - T_B|}{\max(T_A, T_B)}\right) * 100$

		XY Configurations				XX Configurations		
Policy A (Configuration A): **Policy B (Configuration B):**		**Random (X) Random (Y)**	**Rule-based (X) Rule-based (Y)**	**PPOA2C (X) PPOA2C (Y)**	**PPOA2C (X) Rule-based (Y)**	**Rule-based (X) Rule-based (X)**	**PPOA2C (X) PPOA2C (X)**	**PPOA2C (X) Rule-based (X)**
Average NMAC (↓)	**AA**	3.30	0.24	0.03	0.000	0.19	0.02	0.01
	AB	1.23	0.25	0.15	0.005	0.30	0.26	0.13
	BB	3.44	0.26	0.02	0.001	0.04	0.02	0.00
	M1	3.92	0.00	0.00	0.005	0.16	0.13	0.06
	M2	3.96	0.00	0.00	0.000	0.13	0.17	0.08
	IN	0.10	0.77	0.22	0.001	0.24	0.00	0.00
	Total	7.99	0.77	**0.22**	**0.006**	0.53	**0.30**	**0.14**
Success (↑)	N_s / 20	3.96 ± 1.77	18.53 ± 1.65	**19.55 ± 0.98**	**19.98 ± 0.10**	19.42 ± 0.7	**19.79 ± 0.60**	**19.90 ± 0.34**
Reward (↑)	R (×10³)	—	−3.15	**−1.41**	−1.76	−3.27	**−0.80**	−1.87
Mission Time	T_A (min)	—	5.88	7.23	6.58	5.05	7.42	5.46
	T_B (min)	—	7.13	8.94	7.60	5.05	7.54	5.02
Fairness (↑)	F_t (%)	—	82.4	80.8	86.5	**100**	**98.4**	91.9

Takeaway I: Two heterogeneous PPOA2C policies outperformed two strong rule-based policies in both XY and XX configuration settings, finishing the mission with an average 98.35% success rate.

Takeaway II: PPOA2C has shown great adaptation to the rule-based method as well, successfully finishing the missions with an average 99.2% success rate. However, the rule-based method receives lower evaluation rewards since it only cares about NMACs.

Takeaway III: From the mission time standpoint, the interaction can be discriminative against one of the fleets of agents. Especially when the configurations are different, the interaction is in favor of the fleet with stronger configurations.

Takeaway IV: Even with similar configurations, the interaction can be in favor of one of the heterogeneous policies, e.g., with similar configurations, we observed that the interaction was in favor of the rule-based policy rather than the PPOA2C policy.

Iman Sharifi, Hyeong T. Kim, Maheed H. Ahmed, Mahsa Ghasemi, Peng Wei, "sUAS Separation Assurance with Mixed Aircraft Equipage and Heterogeneous Tactical Deconfliction Policies", International Conference on Intelligent Transportation Systems (ITSC), 2026 (Under preparation).