

Transformer-based Multi-agent Reinforcement Learning for Separation Assurance in Structured and Unstructured Airspaces

Arsyi Aziz

Graduate Research Assistant, Intelligent Aerospace Systems Lab

PhD Student, Department of Computer Science

The George Washington University

SESAR Innovation Days

Bled, Slovenia

December 4, 2025

Why separation assurance is hard for future operations

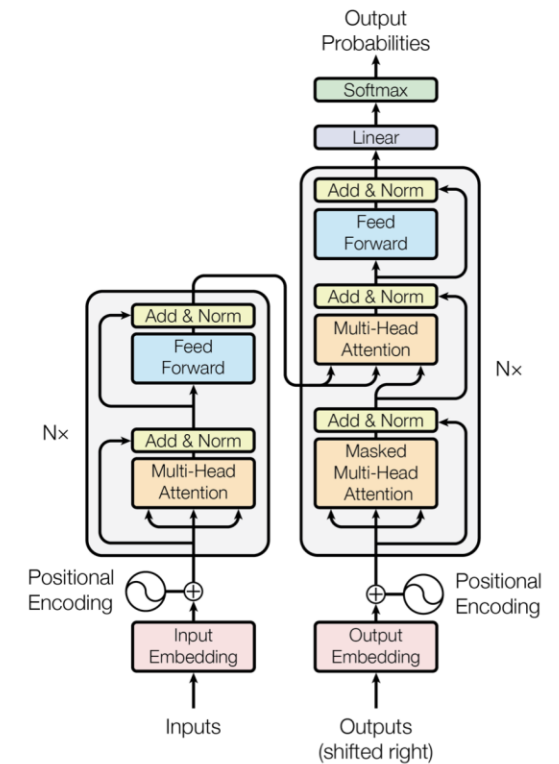
- Current ATM relies on optimization based metering and precomputed schedules
- AAM operations adds:
 - Point-to-point travel
 - High-density traffic
 - Unpredictable operations
- Can break fixed schedules



Image source: NASA

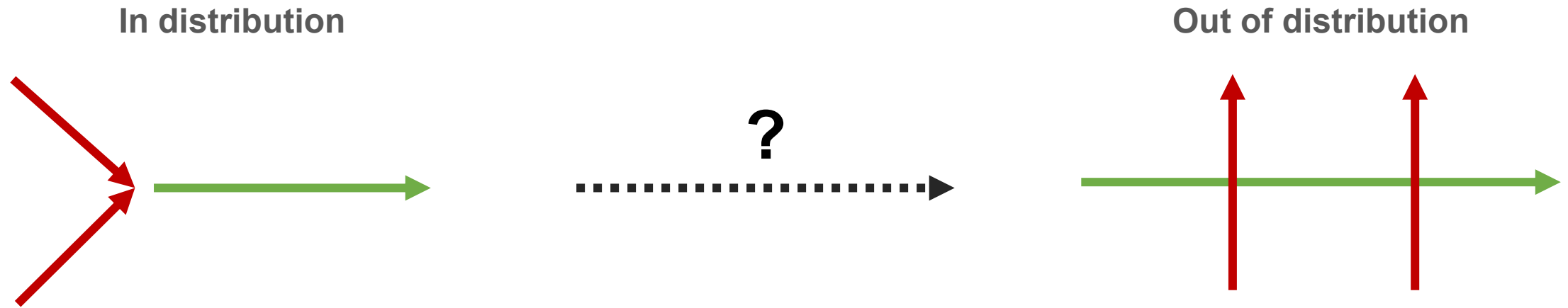
Why MARL and transformers

- Multi-agent reinforcement learning (MARL) learns policies through simulation
- Enables adaptive policies that react in real time
- Transformers naturally handles multiple aircraft
 - Originally proposed to handle variable length sentences
 - Now widely used to solve wide range of machine learning tasks
 - Basis of ChatGPT: **T** stands for transformers



Transformer Architecture
“Attention Is All You Need”,
Vaswani et. al. 2023

Overfitting Problem



How can we ensure adaptability to
unseen scenarios?

Goals of Safe Separation Assurance

1. Aircraft should arrive to their destinations safely
2. Must prevent safety-critical events from occurring
3. Ensure aircraft stay close to operationally desired speeds

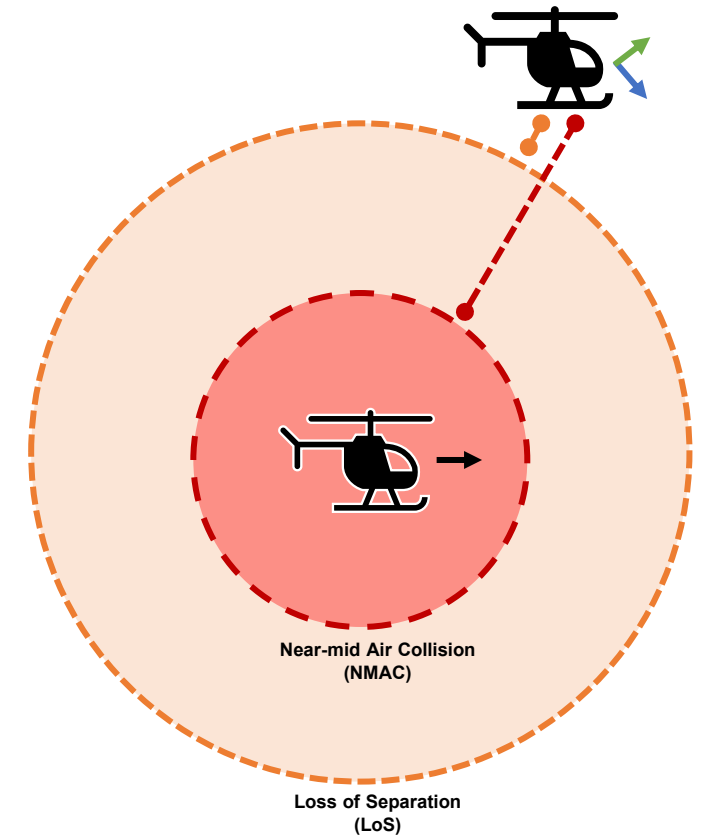
These goals must ideally adapt to unseen scenarios

Formulating Safe Separation MARL

- Multiple aircraft operate in a shared airspace
- Air traffic controller issues speed advisories
 - increase, hold, or decrease current airspeed
- Ego-centric observation space
- Rewards and penalties:
 - Encourage desired behaviors
 - Penalize when safety critical events occur

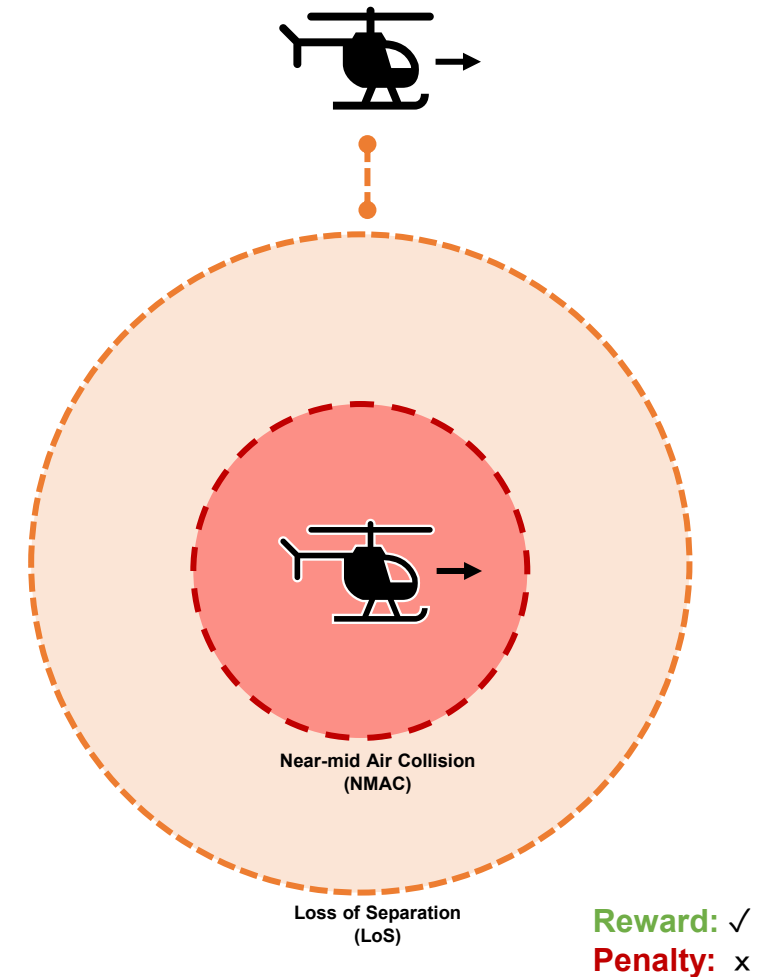
Egocentric Observation Space

- Critical to ensure adaptability to unseen scenarios
- **Ownship information:**
 - Airspeed
 - Deviation from desired operational speeds
- **Intruder information:**
 - Distance to NMAC and LoS boundary
 - Relative bearing
 - Loss of separation indicator function
 - Relative velocity (**radial** and **tangential**)

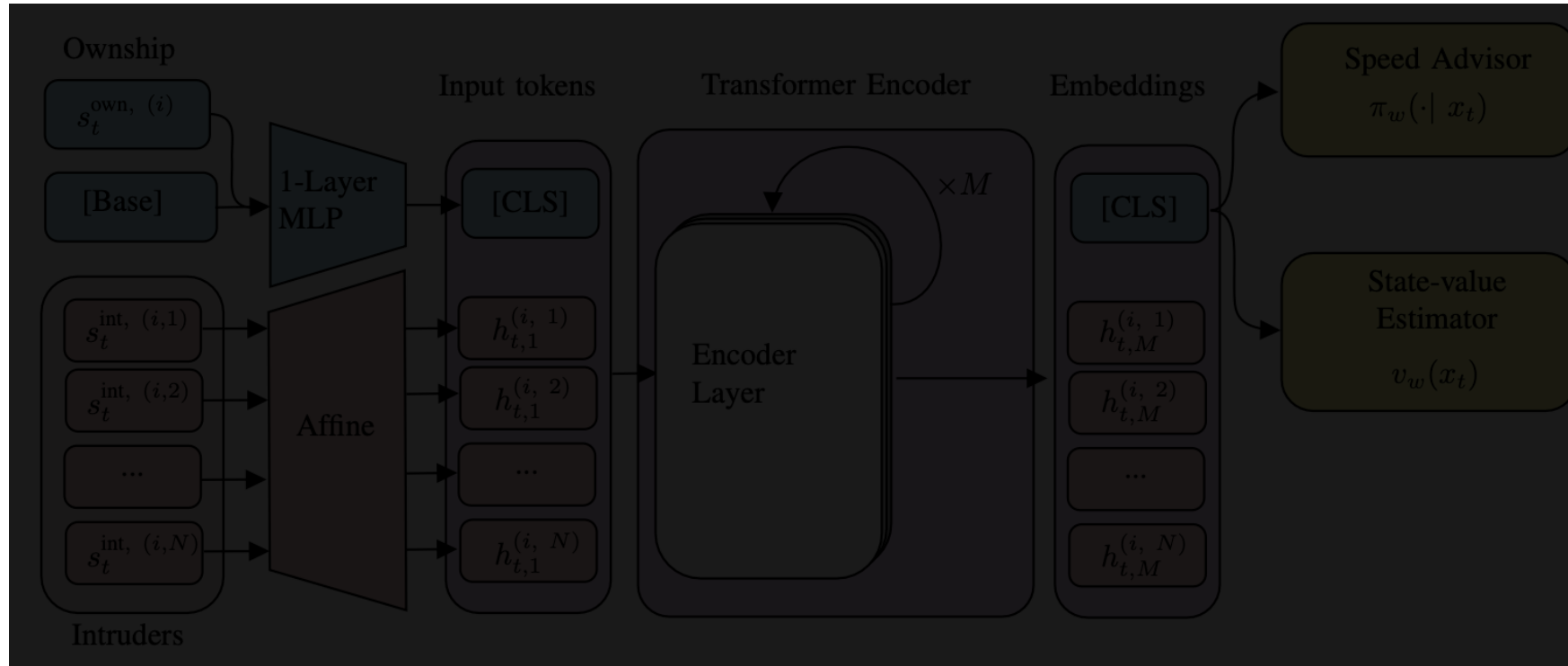


Rewards and Penalties

- Three regimes:
 - **No conflict:**
 - reward proximity to desired speed
 - **In conflict:**
 - penalize time to intrusion and/or distance to NMAC
 - **In NMAC:**
 - Large negative constant penalty

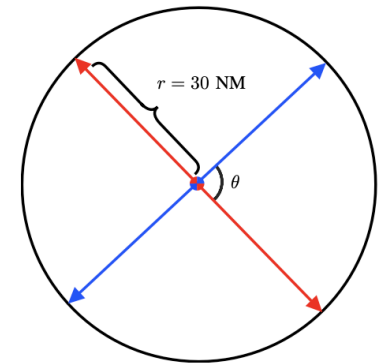


Model Architecture

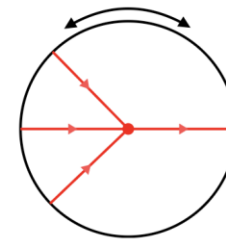


Training Setup

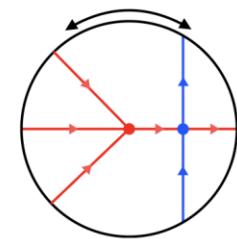
- Proximal Policy Optimization (PPO) trains the model within a stable “trust region”
- Simulator implemented in Bluesky
- Randomized airspace structure:
 - Route entry and exit points
 - Intersection angle
 - Traffic density



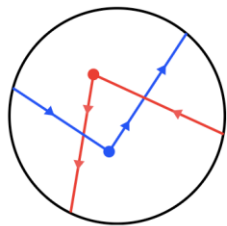
Training



Case (a)



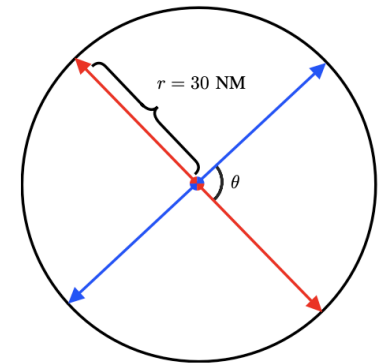
Case (b)



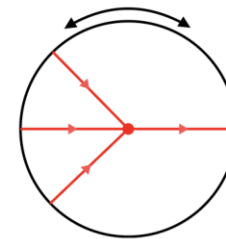
Case (c)

Simulation Conditions

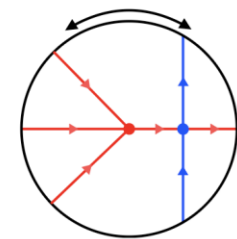
- Aircraft with VTOL capabilities
- Aircraft termination:
 - Safely exits the route
 - Exceeds predefined maximum arrival time
 - NMAC with another aircraft
- We benchmark:
 - Pure attention
 - Transformers of **one**, **two**, and **three** layers



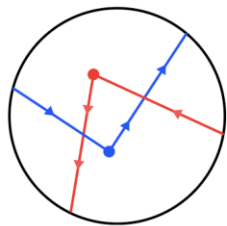
Training



Case (a)

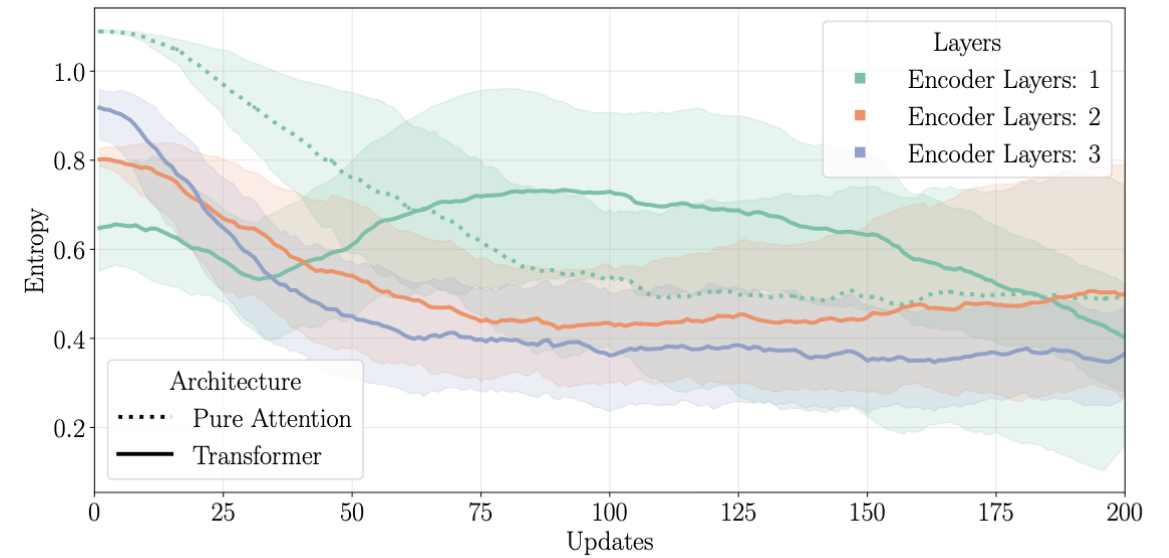
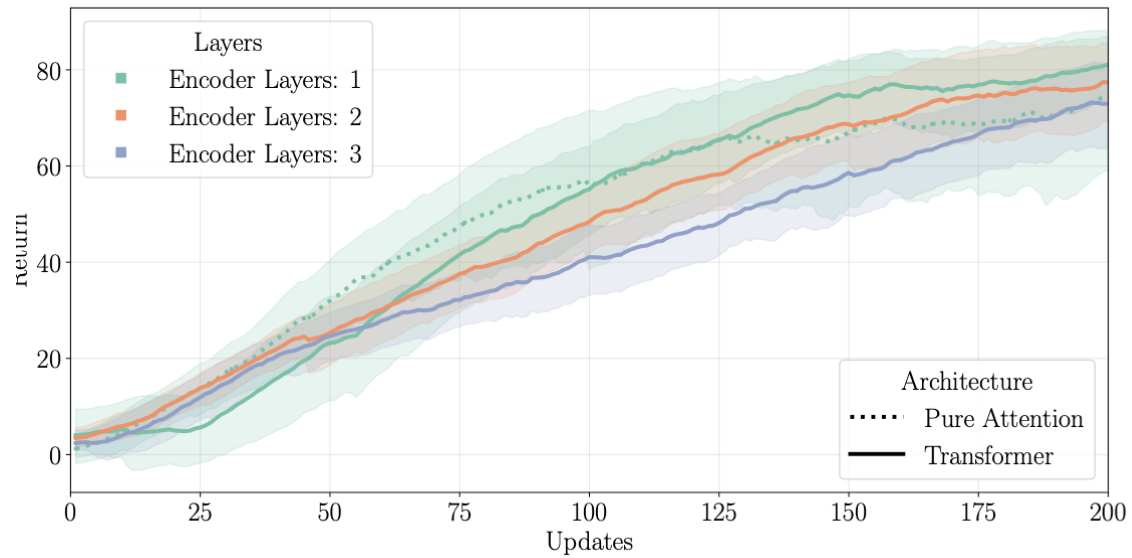


Case (b)



Case (c)

Training Curves



Experimental Results

Environment	Density	Pure Attention			1-Layer Transformer			2-Layer Transformer			3-Layer Transformer		
		NMACs	Speed Adher.	LoS Time	NMACs	Speed Adher.	LoS Time	NMACs	Speed Adher.	LoS Time	NMACs	Speed Adher.	LoS Time
Training	≤5	0.008 ± 0.125	0.653 ± 0.175	108.338 ± 527.907	0.000 ± 0.000	0.843 ± 0.182	119.300 ± 570.666	0.000 ± 0.000	0.819 ± 0.170	190.182 ± 813.784	0.011 ± 0.145	0.868 ± 0.132	92.371 ± 461.763
	6-10	0.080 ± 0.437	0.578 ± 0.205	1771.115 ± 4060.318	0.008 ± 0.124	0.709 ± 0.212	1119.789 ± 2463.339	0.002 ± 0.067	0.646 ± 0.235	2215.736 ± 4111.380	0.049 ± 0.329	0.726 ± 0.202	1017.803 ± 2379.765
	11+	0.337 ± 0.927	0.371 ± 0.218	11727.380 ± 14699.918	0.000 ± 0.000	0.490 ± 0.206	4059.592 ± 5150.277	0.000 ± 0.000	0.418 ± 0.217	6455.796 ± 6595.136	0.091 ± 0.417	0.539 ± 0.218	3161.449 ± 4144.970
	Average	0.081 ± 0.453	0.583 ± 0.211	2191.441 ± 6422.559	0.005 ± 0.097	0.736 ± 0.223	1016.157 ± 2585.519	0.001 ± 0.052	0.685 ± 0.241	1874.085 ± 4020.136	0.040 ± 0.294	0.756 ± 0.206	898.693 ± 2344.195
Case (a)	≤5	0.000 ± 0.000	0.671 ± 0.110	3.524 ± 43.188	0.000 ± 0.000	0.828 ± 0.081	0.515 ± 5.325	0.000 ± 0.000	0.789 ± 0.128	3.973 ± 19.792	0.000 ± 0.000	0.858 ± 0.082	131.431 ± 804.446
	6-10	0.001 ± 0.045	0.645 ± 0.138	55.615 ± 326.128	0.000 ± 0.000	0.759 ± 0.075	13.360 ± 145.403	0.001 ± 0.045	0.696 ± 0.152	97.871 ± 650.170	0.010 ± 0.143	0.781 ± 0.071	387.216 ± 1531.111
	11+	0.014 ± 0.166	0.560 ± 0.169	650.686 ± 1550.735	0.002 ± 0.070	0.572 ± 0.111	78.825 ± 192.645	0.000 ± 0.000	0.542 ± 0.175	744.886 ± 2412.845	0.063 ± 0.363	0.660 ± 0.075	1544.828 ± 3230.180
	Average	0.005 ± 0.097	0.622 ± 0.152	225.269 ± 916.817	0.001 ± 0.037	0.712 ± 0.123	30.327 ± 158.063	0.001 ± 0.037	0.659 ± 0.175	273.182 ± 1412.005	0.025 ± 0.230	0.750 ± 0.095	713.279 ± 2217.769
Case (b)	≤5	0.000 ± 0.000	0.677 ± 0.114	0.000 ± 0.000	0.000 ± 0.000	0.839 ± 0.066	0.000 ± 0.000	0.000 ± 0.000	0.814 ± 0.094	3.889 ± 18.900	0.000 ± 0.000	0.868 ± 0.047	4.791 ± 25.383
	6-10	0.002 ± 0.066	0.593 ± 0.148	156.961 ± 509.802	0.000 ± 0.000	0.700 ± 0.080	37.113 ± 113.309	0.000 ± 0.000	0.641 ± 0.129	206.468 ± 863.310	0.019 ± 0.194	0.731 ± 0.082	249.403 ± 895.354
	11+	0.023 ± 0.216	0.496 ± 0.159	1245.955 ± 1747.369	0.000 ± 0.000	0.551 ± 0.079	215.158 ± 408.550	0.000 ± 0.000	0.482 ± 0.125	962.233 ± 2023.816	0.177 ± 0.594	0.599 ± 0.081	1271.186 ± 2338.114
	Average	0.010 ± 0.141	0.558 ± 0.159	555.713 ± 1250.620	0.000 ± 0.000	0.652 ± 0.110	96.785 ± 268.506	0.000 ± 0.000	0.590 ± 0.151	462.197 ± 1418.890	0.073 ± 0.390	0.688 ± 0.105	598.839 ± 1626.276
Case (c)	≤3	0.000 ± 0.000	0.535 ± 0.210	149.869 ± 527.145	0.000 ± 0.000	0.771 ± 0.179	223.657 ± 679.895	0.000 ± 0.000	0.759 ± 0.150	306.918 ± 802.779	0.000 ± 0.000	0.774 ± 0.153	217.779 ± 659.005
	4-6	0.014 ± 0.166	0.537 ± 0.179	1415.154 ± 1793.011	0.003 ± 0.078	0.688 ± 0.109	1480.423 ± 1883.388	0.000 ± 0.000	0.667 ± 0.104	1996.577 ± 2150.875	0.009 ± 0.135	0.678 ± 0.108	1602.484 ± 2009.602
	7+	0.018 ± 0.191	0.522 ± 0.179	3525.313 ± 3123.919	0.009 ± 0.132	0.614 ± 0.093	3626.186 ± 2983.981	0.000 ± 0.000	0.600 ± 0.085	4420.622 ± 3195.953	0.018 ± 0.186	0.617 ± 0.083	3681.700 ± 3049.316
	Average	0.012 ± 0.154	0.534 ± 0.185	1485.924 ± 2133.126	0.003 ± 0.082	0.692 ± 0.132	1569.346 ± 2183.789	0.000 ± 0.000	0.676 ± 0.122	2002.420 ± 2447.275	0.009 ± 0.131	0.687 ± 0.125	1653.816 ± 2273.653
Overall	Average	0.027 ± 0.256	0.574 ± 0.181	1114.587 ± 3556.437	0.002 ± 0.066	0.698 ± 0.156	678.154 ± 1817.575	0.001 ± 0.032	0.652 ± 0.181	1152.971 ± 2676.009	0.037 ± 0.279	0.720 ± 0.143	966.157 ± 2173.696

Green: Highest performance
Orange: Lowest performance

Experimental Results

Environment	Density	Pure Attention			1-Layer Transformer			2-Layer Transformer			3-Layer Transformer		
		NMACs	Speed Adher.	LoS Time	NMACs	Speed Adher.	LoS Time	NMACs	Speed Adher.	LoS Time	NMACs	Speed Adher.	LoS Time
Training	≤5	0.008 ± 0.125	0.653 ± 0.175	108.338 ± 527.907	0.000 ± 0.000	0.843 ± 0.182	119.300 ± 570.666	0.000 ± 0.000	0.819 ± 0.170	190.182 ± 813.784	0.011 ± 0.145	0.868 ± 0.132	92.371 ± 461.763
	6-10	0.080 ± 0.437	0.578 ± 0.205	1771.115 ± 4060.318	0.008 ± 0.124	0.709 ± 0.212	1119.789 ± 2463.339	0.002 ± 0.067	0.646 ± 0.235	2215.736 ± 4111.380	0.049 ± 0.329	0.726 ± 0.202	1017.803 ± 2379.765
	11+	0.337 ± 0.927	0.371 ± 0.218	11727.380 ± 14699.918	0.000 ± 0.000	0.490 ± 0.206	4059.592 ± 5150.277	0.000 ± 0.000	0.418 ± 0.217	6455.796 ± 6595.136	0.091 ± 0.417	0.539 ± 0.218	3161.449 ± 4144.970
	Average	0.081 ± 0.453	0.583 ± 0.211	2191.441 ± 6422.559	0.005 ± 0.097	0.736 ± 0.223	1016.157 ± 2585.519	0.001 ± 0.052	0.685 ± 0.241	1874.085 ± 4020.136	0.040 ± 0.294	0.756 ± 0.206	898.693 ± 2344.195
Case (a)	≤5	0.000 ± 0.000	0.671 ± 0.110	3.524 ± 43.188	0.000 ± 0.000	0.828 ± 0.081	0.515 ± 5.325	0.000 ± 0.000	0.789 ± 0.128	3.973 ± 19.792	0.000 ± 0.000	0.858 ± 0.082	131.431 ± 804.446
	6-10	0.001 ± 0.045	0.645 ± 0.138	55.615 ± 326.128	0.000 ± 0.000	0.759 ± 0.075	13.360 ± 145.403	0.001 ± 0.045	0.696 ± 0.152	97.871 ± 650.170	0.010 ± 0.143	0.781 ± 0.071	387.216 ± 1531.111
	11+	0.014 ± 0.166	0.560 ± 0.169	650.686 ± 1550.735	0.002 ± 0.070	0.572 ± 0.111	78.825 ± 192.645	0.000 ± 0.000	0.542 ± 0.175	744.886 ± 2412.845	0.063 ± 0.363	0.660 ± 0.075	1544.828 ± 3230.180
	Average	0.005 ± 0.097	0.622 ± 0.152	225.269 ± 916.817	0.001 ± 0.037	0.712 ± 0.123	30.327 ± 158.063	0.001 ± 0.037	0.659 ± 0.175	273.182 ± 1412.005	0.025 ± 0.230	0.750 ± 0.095	713.279 ± 2217.769
Case (b)	≤5	0.000 ± 0.000	0.677 ± 0.114	0.000 ± 0.000	0.000 ± 0.000	0.839 ± 0.066	0.000 ± 0.000	0.000 ± 0.000	0.814 ± 0.094	3.889 ± 18.900	0.000 ± 0.000	0.868 ± 0.047	4.791 ± 25.383
	6-10	0.002 ± 0.066	0.593 ± 0.148	156.961 ± 509.802	0.000 ± 0.000	0.700 ± 0.080	37.113 ± 113.309	0.000 ± 0.000	0.641 ± 0.129	206.468 ± 863.310	0.019 ± 0.194	0.731 ± 0.082	249.403 ± 895.354
	11+	0.023 ± 0.216	0.496 ± 0.159	1245.955 ± 1747.369	0.000 ± 0.000	0.551 ± 0.079	215.158 ± 408.550	0.000 ± 0.000	0.482 ± 0.125	962.233 ± 2023.816	0.177 ± 0.594	0.599 ± 0.081	1271.186 ± 2338.114
	Average	0.010 ± 0.141	0.558 ± 0.159	555.713 ± 1250.620	0.000 ± 0.000	0.652 ± 0.110	96.785 ± 268.506	0.000 ± 0.000	0.590 ± 0.151	462.197 ± 1418.890	0.073 ± 0.390	0.688 ± 0.105	598.839 ± 1626.276
Case (c)	≤3	0.000 ± 0.000	0.535 ± 0.210	149.869 ± 527.145	0.000 ± 0.000	0.771 ± 0.179	223.657 ± 679.895	0.000 ± 0.000	0.759 ± 0.150	306.918 ± 802.779	0.000 ± 0.000	0.774 ± 0.153	217.779 ± 659.005
	4-6	0.014 ± 0.166	0.537 ± 0.179	1415.154 ± 1793.011	0.003 ± 0.078	0.688 ± 0.109	1480.423 ± 1883.388	0.000 ± 0.000	0.667 ± 0.104	1996.577 ± 2150.875	0.009 ± 0.135	0.678 ± 0.108	1602.484 ± 2009.602
	7+	0.018 ± 0.191	0.522 ± 0.179	3525.313 ± 3123.919	0.009 ± 0.132	0.614 ± 0.093	3626.186 ± 2983.981	0.000 ± 0.000	0.600 ± 0.085	4420.622 ± 3195.953	0.018 ± 0.186	0.617 ± 0.083	3681.700 ± 3049.316
	Average	0.012 ± 0.154	0.534 ± 0.185	1485.924 ± 2133.126	0.003 ± 0.082	0.692 ± 0.132	1569.346 ± 2183.789	0.000 ± 0.000	0.676 ± 0.122	2002.420 ± 2447.275	0.009 ± 0.131	0.687 ± 0.125	1653.816 ± 2273.653
Overall	Average	0.027 ± 0.256	0.574 ± 0.181	1114.587 ± 3556.437	0.002 ± 0.066	0.698 ± 0.156	678.154 ± 1817.575	0.001 ± 0.032	0.652 ± 0.181	1152.971 ± 2676.009	0.037 ± 0.279	0.720 ± 0.143	966.157 ± 2173.696

Green: Highest performance
Orange: Lowest performance

Recap and Conclusion

Key Ideas:

- Polar, ego-centric state representation
- Speed incentivized reward
- Transformer with ownship conditioned **[CLS]** token for variable intruder sets
- Training over diverse geometries and densities for better generalization

Findings:

- Single layer encoder transformer gives best safety and speed adherence and outperforms a pure attention baseline

Future Work

- Improve reliability and generalizability to reduce NMAC and its variants
- Complexify problem:
 - Add heading or altitude maneuvers
 - Define more complex route networks and denser traffic
- Refine reward or curriculum for further enhanced performance on unstructured airspace

Thank you, Any questions and comments?

Arsyi Aziz

Graduate Research Assistant, Intelligent Aerospace Systems Lab

PhD Student, Department of Computer Science

The George Washington University