Supplementary Material: Appendix

This supplementary document accompanies the main paper and provides additional implementation details to support reproducibility. The main paper is fully self-contained; these appendices offer extended technical specifications that complement the methodology described therein.

A. Rule-Based Conflict Resolution Policy

This appendix describes the deterministic rule-based policy used for action selection in the multi-agent flight environment. The policy is hand-crafted and does not involve learning or parameter tuning. At each decision step, an agent selects one of three discrete actions: *Decelerate*, *Hold*, or *Accelerate*. Action selection is governed by the ownship's distance to the next waypoint, the presence and relative position of nearby intruders, route alignment, and speed constraints. All rules are evaluated sequentially and are mutually exclusive after speed constraint enforcement.

Decision Rules

The decision rules are partitioned into three sets based on the agent's position relative to bottleneck waypoints, as illustrated in Figure 1. The key parameters governing these rules are:

- d_o^{wp}: Distance from the ownship to its next waypoint
- d_o^{safe}: Safety distance threshold for triggering deconfliction maneuvers

Far from the Next Waypoint ($d_o^{\mathrm{wp}} > d_o^{\mathrm{safe}}$)

When the ownship is sufficiently far from the next waypoint, actions are selected as follows:

- If no intruder is present within the safety distance, the ownship accelerates when its current speed is below the desired speed; otherwise, it maintains its current speed.
- If an intruder is present within the safety distance and is located ahead of the ownship, the ownship decelerates.
- If an intruder is present within the safety distance and is located behind the ownship, the ownship accelerates.

Near the Next Waypoint ($d_o^{\mathrm{wp}} \leq d_o^{\mathrm{safe}}$)

When the ownship is close to the next waypoint, the following rules apply:

- If an intruder is located ahead of the ownship on the same route, the ownship decelerates.
- If an intruder is located ahead of the ownship on a different route, the ownship decelerates.
- If the ownship and a front intruder are within the collision distance threshold, action selection depends on relative speeds: the ownship accelerates if it has a speed advantage, decelerates if it has a speed disadvantage, and randomly selects between acceleration and deceleration when both agents have equal speeds. In this case, the intruder is assigned the opposite action to maintain separation.
- If no intruder is located ahead of the ownship within the safety distance, the ownship accelerates toward the waypoint.

Speed Constraint Enforcement

After an action is selected, speed constraints are enforced as a final step:

- If the chosen action would cause the ownship to violate its minimum or maximum speed limits, the action is overridden and replaced with maintaining the current speed (*Hold*).

B. Example Raw Agent Observation

This appendix presents an example *raw observation record* collected for a single agent at one simulation time step. Listing 1 shows the exact data structure provided to the rule-based policy prior to action selection, including ownship state variables, information about the two closest front intruders, and the resulting action.

Far from the next waypoint:
When the ownship is sufficiently far from the next waypoint, i.e., $d_o^{\text{wp}} > d_o^{\text{safe}}$, actions are selected as follows:

- If no intruder is present within the safety distance, the ownship accelerates when its current speed is below the desired speed; otherwise, it maintains its current speed.
- If an intruder is present within the safety distance and is located ahead of the ownship, the ownship decelerates.
- If an intruder is present within the safety distance and is located behind the ownship, the ownship accelerates.

Near the next waypoint:
When the ownship is close to the next waypoint, i.e., $d_o^{\text{wp}} \leq d_o^{\text{safe}}$, the following rules apply:

- If an intruder is located ahead of the ownship on the same route, the ownship decelerates.
- If an intruder is located ahead of the ownship on a different route, the ownship decelerates.
- If the ownship and a front intruder are within the collision distance threshold, action selection depends on relative speeds: the ownship accelerates if it has a speed advantage, decelerates if it has a speed disadvantage, and randomly selects between acceleration and deceleration when both agents have equal speeds. In this case, the intruder is assigned the opposite action to maintain separation.
- If no intruder is located ahead of the ownship within the safety distance, the ownship accelerates toward the waypoint.

Speed constraint enforcement:
After an action is selected, speed constraints are enforced as follows:

- If the chosen action would cause the ownship to violate its minimum or maximum speed limits, the action is overridden and replaced with maintaining the current speed.

Figure 1. Decision rules for the rule-based policy, organized by ownship proximity to the next waypoint. The policy distinguishes between situations where the ownship is far from the waypoint ($d_o^{\text{wp}} > d_o^{\text{safe}}$) and near the waypoint ($d_o^{\text{wp}} \leq d_o^{\text{safe}}$), with speed constraint enforcement applied as a final override.

The observation record captures all state information necessary for tactical decision-making, including:

- **Ownship state:** Position, velocity, heading, route identifier, distance to next waypoint, and speed constraints
- **Intruder information:** Relative positions, velocities, and route identifiers for the two closest front intruders
- **Collision metrics:** Time-to-collision estimates and Euclidean distances to intruders

```
Ownship info:
  id: A03
  type: Amazon Prime Air - MK30 Model
  lat: 33.137421, lon: -96.861632
  next_wpt_id: WP4
  next_wpt_type: Intersection
  dist_to_nxt_wpt(m): 4759.71
  speed(m/s): 34.98
  min_spd(m/s): 0.0, max_spd(m/s): 41.16
  speed_change_per_second(m/s2): 1.7
  heading(deg): 20.13
  altitude(m): 376.82
  route_id: R_3
  last_action: hold
  num_intruders_ahead: 2
  desired_spd(m/s): 33.44
  time_to_collision_with_intruder1(s): 116.05
  intruder1_on_same_route: True
  did_ownship_have_NMAC: False
  time_to_collision_with_intruder2(s): inf
  intruder2_on_same_route: True
  distance_to_intruder1(m): 1074.77
  distance_to_intruder2(m): 501.82
```

```
First closest front intruder info:
  id: D02
  type: Google X-Wing
  lat: 33.14653, lon: -96.85777
  next_wpt_id: WP4
  next_wpt_type: Intersection
  dist_to_nxt_wpt(m): 3685.01
  speed(m/s): 25.72
  min_spd(m/s): 0.0, max_spd(m/s): 30.87
  speed_change_per_second(m/s2): 1.03
  heading(deg): 20.31
  altitude(m): 347.56
  route_id: R_4
  last_action: hold

Second closest front intruder info:
  id: C04
  type: Amazon Prime Air - MK30 Model
  lat: 33.141682, lon: -96.859853
  next_wpt_id: WP4
  next_wpt_type: Intersection
  dist_to_nxt_wpt(m): 4257.95
  speed(m/s): 34.98
  min_spd(m/s): 0.0, max_spd(m/s): 41.16
  speed_change_per_second(m/s2): 1.7
  heading(deg): 20.24
  altitude(m): 355.92
  route_id: R_3
  last_action: hold

Ownship action: Hold.
```

Listing 1. Raw observation snapshot for a single agent at one simulation time step. This record is provided to the rule-based policy for action selection and subsequently transformed into a natural-language prompt for LLM training.

C. Example Prompt for Action Recommendation

This appendix illustrates the prompt format used for LLM training and inference. The raw observation data (Appendix B) is transformed into a structured natural-language prompt comprising two components:

1. **System Prompt:** Defines the model's operational role as a tactical deconfliction assistant, specifying the decision context and expected response format.
2. **User Prompt:** Describes the current local traffic situation in natural language, including ownship state, intruder information, and relevant spatial relationships.

This translation process converts low-level simulator states into human-readable descriptions that emphasize relative relationships, safety-relevant constraints, and decision context. As a result, the LLM is encouraged to infer tactical reasoning patterns rather than merely learning numerical correlations.

Prompt Structure

Figure 2 presents a complete example prompt constructed from raw state information. The prompt uses qualitative descriptors (e.g., "very safe," "very long") derived from the numerical state values to facilitate natural-language reasoning.

System Prompt:
You are an airspace tactical deconfliction assistant. At each time step, an ownship agent is approaching a bottleneck waypoint, such as merging or intersections points, where other agents (intruders) are approaching as well. Based on the information of the ownship and intruders, the ownship should take an action to avoid collisions. The ownship agent only has access to the information of the front intruders, but there might be other intruders behind the ownship. Your task is to help the ownship aircraft avoid collisions with front intruder aircraft by suggesting appropriate speed adjustments. The ownship cannot unnecessarily decelerate since it might occlude the airspace for other agents behind it. Your response should start with 'The recommended action is: ' followed by one of the actions: Decelerate, Hold, or Accelerate.

User Prompt:
Given the information of the ownship and intruders as follows:
- Ownship:
  - Speed is medium (12.86 m/s), where minimum possible speed is 0.0 m/s and maximum possible speed is 30.87 m/s.
  - Speed is lower than the desired speed.
  - Speed is not minimum and is not maximum.
  - Distance to the next waypoint is very long (3299.46 m).
  - There are two intruders ahead.

- Front Intruder 1:
  - The euclidean distance to the ownship is very safe (1774.78 m).
  - The intruder is not on the same route as the ownship.
  - The intruder distance to the next waypoint is very long (1665.31 m).
  - The intruder distance to the next waypoint is significantly different than the ownship.
  - The intruder speed is 25.21 m/s.
  - The intruder is moving at a moderately higher speed compared to the ownship.

- Front Intruder 2:
  - The euclidean distance to the ownship is very safe (1548.81 m).
  - The intruder is on the same route as the ownship.
  - The intruder distance to the next waypoint is very long (1750.64 m).
  - The intruder distance to the next waypoint is significantly different than the ownship.
  - The intruder speed is 25.21 m/s.
  - The intruder is moving at a moderately higher speed compared to the ownship.

Based on the above information, what actions should the ownship take? (Decelerate/Hold/Accelerate)
Your response should start with 'The recommended action is: ' followed by one of the actions: Decelerate, Hold, or Accelerate.

Figure 2. Example prompt for tactical deconfliction at a single time step. The system prompt establishes the model's role and constraints, while the user prompt provides a structured description of the current traffic situation. Qualitative descriptors are derived from numerical thresholds to support natural-language reasoning.

Response Format

The expected response format is shown in Figure 3. The model is trained to produce a brief, structured response beginning with "The recommended action is:" followed by one of the three discrete actions: *Accelerate*, *Hold*, or *Decelerate*.

Answer:
The recommended action is: Accelerate.

Figure 3. Target response format corresponding to the prompt in Figure 2. The constrained response format ensures consistent parsing during both training and closed-loop inference.

Prompt Design Considerations

Several design choices guide the prompt engineering process:

- **Qualitative descriptors:** Numerical values are converted to qualitative categories (e.g., distance "very safe" vs. "critical") to align with human reasoning patterns and reduce sensitivity to exact numerical values.
- **Relative comparisons:** Intruder information emphasizes relative quantities (e.g., "moving at a moderately higher speed compared to the ownship") rather than absolute values, supporting transferable reasoning across diverse traffic configurations.
- **Constrained output format:** The response format is strictly specified in both the system prompt and the closing instruction, ensuring consistent parsing during evaluation and deployment.
- **Safety emphasis:** The system prompt explicitly frames the task in terms of collision avoidance and airspace safety, priming the model toward conservative, safety-oriented decisions.

The prompt format is kept consistent across training and inference to ensure behavioral stability. This consistency is critical for maintaining alignment between the fine-tuned model's behavior and the human-aligned supervisory signals encoded in the training dataset.