Trust-Based Information Filtering for Robust Decentralized Execution of Pre-Trained MARL Policies in UAV Swarms

ERNESTS RUDZĪTIS, University of Twente, The Netherlands

Multi-Agent Reinforcement Learning (MARL) enables complex drone swarm behaviors; however, the mission success is hindered by unreliable communication. Existing robustness solutions often require integration during training or significant configuration, limiting flexibility. This paper introduces Trust-Based Information Filtering (TIF) system that enhances pre-trained MARL policies during decentralized execution. The post-hoc TIF system equips each agent with a mechanism to assess message trustworthiness using learned spatio-temporal expectations from normal operations. This dynamic self-configuration eliminates the need for attack data or policy retraining. Evaluated in UAV formation control under various communication unreliability scenarios, TIF demonstrates a measurable improvement in operational resilience. This validates the prototype of effective, lightweight, post-hoc filtering approach, signaling that robustness can be layered onto existing MARL policies without costly retraining.

Additional Key Words and Phrases: UAV Swarms, Multi-Agent Reinforcement Learning (MARL), Trust Mechanism, Robust Communication, Information Filtering, Decentralized Execution, Pre-trained Policies

1 INTRODUCTION

1.1 Background and Context

Swarms of Unmanned Aerial Vehicles (UAVs), a type of Multi-Agent System (MAS) [4], are a rapidly advancing frontier. Coordinated by Multi-Agent Reinforcement Learning (MARL), these swarms show potential for complex tasks such as search, rescue and defense [9, 11, 12].

Mission success for MARL-based UAV swarms depends heavily on inter-agent communication quality and reliability. Cooperative MARL policies often rely on exchanged messages (e.g., positions, velocities, formation intentions) to communicate and achieve coherent group behavior. In practice, communication channels can be noisy [17], sensors providing data readings can malfunction, and in adversarial scenarios, communication can be intentionally manipulated by compromised agents or extrinsic foes [23]. This reliance introduces a significant vulnerability, potentially causing mission failure or unsafe operations [19].

Communication failures pose a great risk to MARL-based UAV swarms. Using trust mechanisms to improve robustness is crucial but underexplored, particularly for pre-trained policies. Existing approaches often integrate countermeasures directly into the MARL training process itself [6]. While effective, this can restrict algorithm choice, increase training complexity, and require costly retraining. Other strategies involve pre-configured protocols like cryptographic methods [7], which require considerable setup effort and may not adapt to unreliability during a mission. Conversely, simple post-hoc outlier detection filters lack the contextual understanding to be effective against subtle or prior unknown disruptions.

1.2 Research Objectives

This research addresses the outlined gap by enhancing the robustness of pre-trained MARL policies with a dynamic, post-hoc trust and filtering system, demonstrated within the context of UAV formation control. The core idea is to enhance pre-trained MARL policies post-hoc by equipping each agent with a decentralized mechanism. This mechanism allows the agent to assess the trustworthiness of incoming messages based on learned expectations within the swarms normal operational context. Crucially, this trust mechanism is configured after the primary MARL policy training is complete, making it readily applicable to existing, pre-trained policies without the need for modification. The system self-configures by learning a baseline from simulated operations and applying a low anomaly threshold to distinguish untrustworthy messages, a lightweight approach that aims to preserve the original policies performance under reliable conditions.

1.3 Chapter overview

This paper first establishes the core problem and research questions (Chapter 2), reviews literature (Chapter 3), and presents the methodology (Chapter 4). Chapter 5 introduces the Trust-Based Information Filtering (TIF) system, which is evaluated in Chapter 6. Finally, Chapter 7 concludes with contributions and future work.

2 PROBLEM STATEMENT

Existing robustness solutions for MARL-based UAV swarms often lack flexibility and contextual awareness. The core problem is therefore developing a decentralized, post-hoc trust mechanism to strengthen pre-trained policies against unreliable communication without costly retraining or requiring specific adversarial data.

2.1 Research Question

The problem statement leads to the following research question:

RQ1: *How can a decentralized, post-hoc trust mechanism be configured via unsupervised learning, based on an agents normal behavior, to enhance the robustness of pre-trained MARL policies against communication unreliability, without requiring policy retraining or significantly impacting nominal performance?*

To address the main research question, the following sub-research questions will be investigated:

(1) *What specific spatio-temporal consistency checks are most indicative of message reliability within the context of MARL-driven UAV formation flying?*
(2) *How effectively can the proposed self-configuration process, using a predefined anomaly threshold, establish a reliable baseline from normal swarm operation data to accurately configure the trust and filtering mechanism?*

TScIT 43, June 29, 2025, Enschede, The Netherlands

(3) *To what extent does the configured post-hoc trust mechanism improve the swarms formation control performance and resilience when subjected to various types of communication unreliability (such as, sensor noise, faulty agents, simulated adversarial messages) compared to the baseline pre-trained policy?*

3 RELATED WORK

Multi-Agent systems are rapidly evolving in many domains, and so is a notable part of research revolved around multi-agent learning by means of reinforcement learning techniques, particularly Multi-Agent Reinforcement Learning (MARL).

There are several research works that aim to integrate trust or robustness into MARL systems. These works fall into various categories. First category involves integrating these mechanisms directly in the MARL training process. The research work by Fung et al. [6] proposes Reinforcement Learning-based Trusted Consensus (RLTC), a reinforcement learning approach where agents explicitly learn trust scores for neighbor agents by means of Q-learning during training phase.

Second category relates to filtering or modification of communication. Xue et al. [20] propose a two-stage protocol to detect and reconstruct malicious messages. This method focuses on correcting perturbations using a model trained to reverse specific, anticipated manipulations from an adversary. The research work by Sun et al. [16] introduces Ablated Message Ensemble (AME) defensive mechanism, which guarantees the performance of agents when a fraction of communication messages are perturbed. In this work robustness was assured post-hoc by making decisions based on the majority vote from multiple base actions, each generated using a randomly chosen subset of the incoming messages. Mitchell et al. [13] proposed a different approach using Gaussian Processes to model expected message correlations based on agent proximity, allowing inconsistent messages to be identified and down-weighted.

Finally, concepts from adjacent fields like distributed consensus and security are also relevant. The research work by Han et al. [7] on trust for UAV swarms specifically, focuses on achieving secure agreement on specific values using cryptographic protocols.

In summary, existing research addresses MARL robustness via integrated learning methods, post-hoc filtering, and security protocols. This review highlights an opportunity for a prototype focused on dynamically learned trust from observed normal behavior. Such a mechanism, adaptable to pre-trained policies without retraining or threat intelligence, could enhance resilience against general communication unreliability. This paper proposes and investigates such a system.

4 METHODOLOGY

This section details the research methodology and key design choices. The research approach consists of four distinct phases:

(1) Development of a drone simulation environment
(2) Acquisition and integration of a pre-trained Multi-Agent Reinforcement Learning (MARL) policy for the designated task
(3) Design and implementation of the Trust-Based Information Filtering (TIF) system
(4) Evaluation of the TIF systems performance under both nominal and unreliable communication conditions

The primary programming language utilized for this research was Python, with PyTorch [2] and Scikit-learn [15] serving as the core machine learning frameworks. The following subsections elaborate on the specific experimental setup and approach.

4.1 Simulation Environment

Initial exploration considered existing 3D simulation platforms such as RotorPy, a Python-based multi-rotor simulator known for a sophisticated implementation that resembles configurations and sensors of physical drones [5]. However, the complexity and computational demands of such a high-fidelity environment would have hindered rapid prototype iteration within the research scope. As a consequence, to facilitate focus on development and evaluation of the trust mechanism itself, a custom 2D simulation environment was developed.

This 2D environment models agents, hereafter referred to as drones, capable of planar movement. At each step, drones adjust position based on commanded x and y velocities. The environment adheres to the PettingZoo API [18], a standard interface for multi-agent reinforcement learning environments, ensuring compatibility with common MARL frameworks. Drone communication is modeled as an ideal, unrestricted broadcast system, where each drone transmits its internal state to the swarm. A drones observation space combines its local state and received communication.

This architecture introduces two potential failure points. First, a drones own sensors could malfunction, in turn leading it to have an incorrect understanding of its own state. Second, and the primary focus on this research, the communication channel itself can be unreliable. Information that is received from other drones may be corrupted, stale or manipulated during transmission. The Trust-Based Information Filtering (TIF) system, introduced in section 5, is designed to operate at the receiving end, therefore, enabling an agent to primarily assess the trustworthiness of incoming messages from its peers.

The observation space for each drone agent is structured as follows:

(1) Local State and Mission Information: Information derived from the drones own sensors and its assigned mission objectives. This part of the observation is not directly affected by inter-agent communication failures.
   - Its own absolute 2D position (from an onboard positioning system)
   - Its own 2D velocity (from its internal state estimation)
   - The 2D relative position to its designated target formation point
   - The 2D relative position to the overall swarms mission endpoint
(2) Peer Information: Information derived from data broadcast by other drones in the swarm. This channel is the primary source of the unreliability that the TIF system focuses on addressing.
   - Relative 2D positions to all other drones in the swarm (calculated using received position data)

4.2 Pre-trained MARL Policy for Swarm Control

The foundation of this research is highly dependent on a pre-trained MARL policy designed for the control of swarm formation. This section details the characteristics of the MARL archetype, the specific algorithm employed, the task definition, and ultimately the training regimen.

4.2.1 MARL Paradigm Selection. MARL policies can be categorized based on their training and execution approach, namely: Centralized Training with Centralized Execution (CTE), Centralized Training with Decentralized Execution (CTDE), and Decentralized Training with Decentralized Execution (DTE) [1]. While CTDE is often favored for its scalability in larger swarms by mitigating communication overhead (which can grow quadratically with the increase of swarm size) and better handling potential range constraints or latency issues, CTE provides a framework for applications requiring high precision, typically in smaller swarms (2-10 agents).

For this research, a CTDE paradigm with explicit inter-agent communication was adopted. This choice allows training a baseline policy with global information available during training and execution. This facilitates a clearer evaluation of the subsequently introduced Trust-Based Information Filtering (TIF) system, as the focus is on robustness enhancement rather than dealing with inherent limitations.

4.2.2 Algorithm and Architecture. The Multi-Agent Proximal Policy Optimization (MAPPO) algorithm [21] was selected for this research. MAPPO is an on-policy, actor-critic algorithm renowned for its stability and strong performance in cooperate multi-agent tasks.

This algorithm follows CTDE paradigm, as mentioned in Section 4.2.1. During the training phase, a centralized critic has access to the global observation state of the entire drone swarm. This global perspective motivates the critic to learn an accurate value function estimation that accounts for complex inter-agent dynamics, essentially mapping observation state vector to a scalar value. While a single, shared critic network is common in CTDE, this research employed an architecture where each agent has its own individual critic network. This choice, guided by algorithm implementation [22], still adheres to the centralized training principle, as each critic has access to the full global state information during the training phase.

4.2.3 Task Definition. The specific task of the MARL policy was to allow a swarm of three UAVs to achieve and maintain a V-shaped formation during flight. The drone agents objective is to coordinate their movements to form and maintain this predefined geometric pattern.

4.2.4 Reward Function. A composite reward function was designed to guide the drone agents learning process, attempting to balance the mission objective of formation coherence against critical operational constraints like collision avoidance and smooth control. The function is a weighted linear combination of these components, where the collision avoidance penalty is assigned a significantly higher weight (5.0) to prioritize operational safety:

- Target Achievement: To encourage drones to move towards their designated formation points, the reward follows potential-based shaping [14], proportional to the reduction in distance to the target (`prev_error - current_error`). This technique is widely used in reinforcement learning as it provides a more dense reward signal that can help guide the learning process more effectively.
- Velocity Alignment: To promote efficient movement, drones are rewarded for aligning their velocity vector with the direction of their target formation point. This component was intended to encourage direct, purposeful flight paths.
- Formation Cohesion: A penalty was applied based on the average error in relative positioning between a drone and its neighbors. This term was intended to encourage the swarm to move as a coherent unit as a whole, maintaining the intended structure of the V-formation.
- Collision Avoidance: A significant penalty was applied if a drone enters a critical safety distance of another drone or, ultimately, causes collision.
- Control Regularization: A small penalty, proportional to the magnitude of the action, was included to discourage jerky, chaotic movements and promote smoother control.

4.2.5 Training Regimen. The MAPPO policy for the V-shaped formation task assembled by 3 drones was trained for a total of 2 million environment steps. Each training episode was configured to last for a maximum of 200 steps. The training was conducted under ideal communication conditions, without the noise or failures that the TIF system is designed to address.

5 THE PROPOSED TRUST-BASED INFORMATION FILTERING (TIF) SYSTEM

The Trust-Based Information Filtering (TIF) system is an innovative, decentralized mechanism designed to operate post-hoc, enhancing the robustness of pre-trained MARL policies against unreliable communication. This chapter details its architecture, modules, and self-configuration process.

5.1 System Architecture and Design Principles

The TIF system is integrated into each agent independently and does not require a centralized authority, functioning as a layer between incoming communication data and the agents pre-trained MARL policy. The following are the core design principles:

(1) Modularity: The TIF system is decoupled from the MARL policy training process, it purely operates on the features extracted from outputs of a pre-trained policy, requiring no modifications or retraining of the original policy
(2) Self-Configuring: The system autonomously learns to distinguish normal from anomalous communication. It captures normal swarm behavior to create a baseline dataset, then fits an unsupervised model, automatically establishing a decision boundary that flags significant deviations, and unexpected or unseen messages. This process entirely eliminates the need for explicit attack data examples or comprehensive manual parameter tuning.

(3) Generality: While demonstrated in UAV swarm formation, the underlying principles of learning spatio-temporal communication consistencies aim at applicability across various MARL policies and environments

The systems operation relies on a baseline model of normal communication, learned from feature data collected during the pretrained MARL policies standard operation. To establish this baseline, 100 episodes were recorded, each with a maximum of 200 steps, yielding a dataset of approximately 60.000 feature vectors across the swarm. This volume was determined to be sufficient for several reasons. Firstly, the V-shape formation task is well-defined and generally exhibits relatively low variance behavior, meaning its core dynamics can be captured without an excessively large dataset. Secondly, the unsupervised models employed (mentioned in upcoming Section 5.2.2) are known to be sample efficient and do not require the vast amounts of data. Therefore this quantity of data was deemed to be adequate to capture the key operational phases, including initial formation convergence, steady flight, and potential minor corrective maneuvers.

5.2 Trust Assessment Module

The heart of the TIF system is the Trust Assessment Module, whose sole resposibility is to assess the trustworthiness and normality of incoming communication messages from other drone agents.

5.2.1 Feature Engineering for Spatio-Temporal Consistency. To enable the anomaly detection models to identify deviations from expected communication patterns, a set of specific features (or their combinations) are extracted from the incoming observation data. Table 1 details these features, which are organized into five groups to capture different facets of spatio-temporal consistency. These features differ in generality. Generic features (e.g., Temporal Consistency) operate on raw vectors without domain knowledge and are context-agnostic, whereas domain specific features (e.g., Motion Consistency) require a structural understanding of the observation content to compute physically meaningful metrics. This design directly impacts scalability, as the feature vector size for each drone agents decentralized TIF instance may scale linearly $O(N)$ with swarm size, driven by the $O(N)$ requirements of inter-agent features, in contrast to the $O(1)$ size of intra-agent features.

5.2.2 Anomaly Detection and Trust Score. Once the spatio-temporal consistency features (detailed in 5.2.1) are extracted from the incoming exchanged messages (discussed in 4.1), an anomaly detection model is employed to assess whether the current feature vector deviates significantly from patterns observed during normal swarm operations. A crucial preliminary step is the training or fitting of these anomaly detection models using the feature dataset derived from the normal operation data (discussed in Section 5.4.2). This fitting process, which establishes the baseline for 'normal' communication patterns, is generally computationally lightweight and significantly less time-consuming compared to the extensive training required for the base MARL policy.

The TIF systems core untrustworthy message detection mechanism was selected by means of a comparative evaluation of three models representing distinct theoretical approaches. As detailed in Section 6.4, the Local Outlier Factor (LOF) demonstrated superior performance in identifying communication anomalies, which was subsequently used for all final system evaluations. The candidate models explored were:

(1) Autoencoder (AE): This neural network is trained to reconstruct normal feature vectors based on the principles of nonlinear principal component analysis [8]. The reconstruction error is then normalized to produce a continuous trust score in the range `[0,1]`, where value of `0` signifies complete distrust (high error) and `1` signifies complete trust (low error).
(2) Isolation Forest (IF): This ensemble tree algorithm isolates anomalies by randomly partitioning the feature space [10]. It directly classifies instances as normal or anomalous, and its output is treated as a binary trust score (`0` for untrusted/anomalous, contrary `1` for trusted/normal)
(3) Local Outlier Factor (LOF): This density-based algorithm identifies outliers by measuring the local deviation of a given data

Table 1. Overview of Spatio-Temporal Features for Trust Assessment

Feature / Group	Description
Temporal Consistency (Generic)	
Magnitude of Change	Overall change between current and previous observation vectors (Euclidean norm); provides a general sense of state transition stability
Component-wise Change	Vector of the differences for each element in the observation; detects abrupt shifts or stale data
Inter-Agent Consistency (Generic)	
Pairwise Differences	Comparison of an agents observation to all others in the swarm at the same timestamp; provides a general sense of proximal similarity
Summary Statistics	Mean, max, and min of pairwise differences to identify swarm consensus outliers
Motion Consistency (Specific)	
Velocity Magnitude	Physical plausibility on the reported speed of the agent
Position Consistency	Comparison of actual reported position change with that predicted by previous velocity
Formation-Aware (Specific)	
Distance from Centroid	Agents distance from the geometric center of the swarm formation
Velocity Alignment	Checks for agents velocity vector being aligned with the groups overall movement
Anomaly Pattern (Mixed Generality)	
Observation Variance	Statistical variance of the observation vector
Extreme Value Ratios	Identifies physically implausible ratios between components (e.g., position vs velocity)

Table 2. Ranking of feature configurations by average F1-Score and Accuracy. Configurations are combinations of feature groups (or individual features themselves): T (Temporal), M (Motion), I (Inter-Agent), F (Formation), A (Anomaly).

Configuration	Groups	Avg. F1-Score	Avg. Accuracy
comprehensive	T, M, I, F	0.999±0.002	0.999±0.001
temporal_only	T	0.997±0.003	0.998±0.002
spatial_temporal	T, M	0.922±0.155	0.948±0.104
spatial_aware	T, M, F	0.917±0.167	0.944±0.111
temporal_inter_agent	T, I	0.814±0.292	0.868±0.209
formation_motion	F, M	0.790±0.380	0.839±0.293
motion_only	M	0.784±0.433	0.822±0.356
full_suite	T, M, I, F, A	0.699±0.469	0.797±0.310
anomaly_patterns_only	A	0.647±0.374	0.703±0.338
inter_agent_only	I	0.623±0.061	0.645±0.055
spatial_inter_agent	I, F	0.615±0.070	0.659±0.063
formation_only	F	0.549±0.202	0.611±0.152

point with respect to its neighbors [3]. Similar to IF, its output is interpreted as a binary trust score

The final output of this stage is a binary trust assessment for the incoming message. This assessment is subsequently used by the Information Filtering Logic (Section 5.3) to determine how to process the message, particularly if it is deemed untrustworthy.

5.3 Information Filtering Logic

Based on the trust assessment provided by the Trust Assessment Module (5.2.2), the Information Filtering Logic (IFL) module determines the final processed observation data to be passed to the drone agents pre-trained MARL policy for action selection. If data contained within a message is assessed to be trusted, the message is passed directly and unaltered to the MARL policy, however, if a message is considered to be untrustworthy, indicating a potential communication anomaly or manipulation, the IFL attempts to first recover or reconstruct a plausible observation rather than discarding the information, which could lead to policy inaction or reliance on overly stale data. Two simple and computationally lightweight recovery heuristics were implemented and compared. These were chosen to represent distinct fundamental strategies, one being based on smoothing (averaging), while the other on projection (trending). A comparative analysis in Section 6.5.2 evaluates their relative performance. The two strategies, namely:

(1) Historical Average Recovery: This strategy smooths out sudden, anomalous spikes or drops, by replacing the observation with the component-wise average of its own vectors from a recent history window, assuming the recent past provides a reasonable estimate of the current state
(2) Trend Extrapolation Recovery: This strategy projects forward momentum. It uses the last two trusted historical observations to establish a linear trend, which is then extrapolated one step forward to replace the untrustworthy data

The choice of recovery method can highly influence the systems resilience and behavior under different types of communication failures. The observation history for each agent is maintained to support these recovery mechanisms. The output of this filtering and potential recovery process is the observation vector ultimately fed to the drone agents MARL policy.

5.4 Data-Driven Self-Configuration of TIF Parameters

A key characteristic of the TIF system is its data-driven self-configuration capability. This process tunes the parameters of the Trust Assessment Module by analyzing data collected from normal swarm operations, thereby adapting the system to the specific communication patterns and inherent variability of the pre-trained MARL policy and its operational environment.

5.4.1 Data Collection from Normal Swarm Operations. The foundation of the self-configuration process is a dataset representative of normal system behavior. As previously outlined, this involves collecting data from the pre-trained MARL policy operating under ideal, reliable communication conditions. For this research, 100 episodes, each with a maximum of 200 steps, were recorded (discussed in Section 5.1). This dataset contains extracted spatio-temporal features (discussed in Section 5.2.1), forming an applied baseline of trustworthy communication.

5.4.2 Parameter Initialization and Threshold Setting. The primary objective of self-configuration stage is to dynamically set the anomaly detection models parameters to distinguish untrustworthy communications from normal variations. This process aims to maximize detection sensitivity while crucially minimizing any negative impact on the pre-trained MARL policy performance under nominal (ideal) conditions, therefore preserving baseline operational effectiveness.

To achieve this without complex hyperparameter tuning and align with the goal of a lightweight system, a unified thresholding strategy guided by a `contamination` parameter was adopted. This standard hyperparameter specifies the expected amount of outliers in the training data. For all models, this value was set to 0.05 (5%). This choice aligns with the core methodological assumption, that the baseline data, collected under ideal conditions, is overwhelmingly benign, but may contain a tiny portion of infrequent operational variations. This effectively sets the sensitivity for all of the models in a consistent approach, instructing them to flag the 5% most unusual samples. The specific application of this principle varies slightly by model.

For the Autoencoder (AE), the model is first trained on the normal operation features. The `contamination` parameter is then used to automatically set a decision threshold at the 95th percentile of the reconstruction errors observed when applied to the training data.

For the Isolation Forest (IF) and Local Outlier Factor (LOF), the contamination parameter is passed directly to the models during the fitting process. It internally informs their algorithms on how to set their decision boundaries for classification purposes. Additionally, for the LOF model, the `n_neighbors` hyperparameter was set to 20, a standard value that defines the neighborhood size for local density estimation.

6 RESULTS AND DISCUSSION

To evaluate the robustness and effectiveness of the proposed TIF system, a series of experiments were conducted. The baseline pretrained MARL policy was subjected to various communication unreliability scenarios, both with and without the TIF system applied.

6.1 Evaluation Setup and Unreliability Scenarios

To evaluate the TIF systems robustness enhancement, communication unreliability was introduced into the simulation environment. This was simulated through three primary modes affecting the messages received by an agent:

(1) Message Freezing: A drone agent receives stale information from a peer, simulating a scenario of a replay attack or a connection discrepancy (where the last known value is used)
(2) Message Offset: A consistent error is added to reported values, simulating a compromised agent or a sensor with a persistent bias
(3) Random Noise Injection: Gaussian noise is added to the transmitted contents, simulating channel noise or minor sensor inaccuracies

6.2 Computational Overhead

TIFs overhead includes one-time model fitting and per-message inference cost. As outlined in the original implementation of LOF [3], computational efficiency greatly hinges on the underlying data structure used for k-nearest neighbor search. Standard implementations, including the one used in this research, employ tree-based index structures, resulting in a training complexity of approximately $O(N \log N)$ on the baseline feature data and inference complexity of $O(\log N)$ per message during deployment. The subsequent message recovery step, which involves a simple historical average or trend calculation, is computationally trivial $O(1)$ and adds negligible latency.

6.3 Effectiveness of Spatio-Temporal Consistency Checks

This section presents an empirical analysis to determine which spatio-temporal features (or their combinations) are most effective, directly addressing sub-research question (1). To perform this evaluation, a features effectiveness was measured by its ability to contribute or enable the Trust Assessment Module to correctly discern between trustworthy and untrustworthy messages. This discrimination performance, quantified by F1-score and accuracy (Table 2), served as the primary criterion for selecting the optimal feature combinations. The core assumption is that features that are better at this discrimination task will, in turn, provide the foundation for a more robust overall system in its final mission of improving formation control.

6.3.1 Overall Feature Performance. The analysis (Table 2) revealed that feature combinations incorporating `temporal_only` consistently achieved the highest F1-scores and accuracy. The `comprehensive` group achieved a near-perfect average F1-score of 0.999 (±0.002) and accuracy of 0.999 (±0.001) across all compromise types. The unaccompanied `temporal_only` feature also performed exceptionally well, achieving an average F1-score of 0.997 (±0.003) and accuracy of 0.998 (±0.002), demonstrating that even a single well-chosen temporal feature can be effective.

6.3.2 Analysis of Key Feature Groups and Generality. While the initial design aimed for features as generic as possible without explicit knowledge of the observation content each agent possesses, some feature types inherently required structural understanding of the observation vector (for example, to identify position or velocity components). The most effective and truly generic features turned out to be `temporal_only` and `inter_agent_only`. `temporal_only` proved critically important by assessing changes between a drone agents current and previous flattened observations, effectively detecting sudden shifts, stagnations (like message freezing), or erratic jumps. `Inter_agent_only` operated solely on differences between flattened observation vectors from different drone agents and required no component semantics. While its standalone performance of 0.623 F1 was moderate, it was nevertheless retained to be part of the final `comprehensive` feature group, which achieved the highest overall performance.

Other feature types, while valuable, necessitated explicit knowledge about the observations velocity or position components. While `motion_only` showed decent standalone discriminative potential (0.784 F1), its addition to `temporal_only` in the `spatial_temporal` configuration (0.922 F1) actually resulted in a decrease in performance compared to `temporal_only` alone (0.997 F1). This suggests that its specific details might introduce noise or redundancies that negatively impact the highly effective `temporal_only` baseline in certain combinations. `formation_only` features showed the lowest standalone effectiveness (0.549 F1 for `formation_only`), indicating their primary value was in providing contextual enhancement rather than direct indication of untrustworthy communication. `anomaly_patterns_only` features showed moderate performance and could sometimes introduce noise.

6.3.3 Performance by Specific Compromise Type. Evaluation across specific compromise types revealed that `noise`, `random` and `offset` were generally easier to detect, with many combination of characteristics achieving F1 scores near 1.000. The simulated `freeze` compromise proved to be the most challenging for many combinations.

6.4 Effectiveness of the Self-Configuration Process

This section addresses the second sub-research question (2), which investigates the effectiveness of the self-configuration process. As detailed in Section 5.4.2, the TIF systems does not perform any complex optimization search for its hyperparameters, rather it follows a lightweight approach and relies on the premise of abundance of normal swarm operation data. Furthermore, it learns a baseline from pure operational data and applies a predefined contamination factor of 0.05 to configure the anomaly detection models. For algorithms like Local Outlier Factor and Isolation Forest, this hyperparameter directly informs the model during the fitting phase, allowing it to determine its own internal decision threshold. For the Autoencoder, the contamination value is used post-training to calculate a threshold based on the 95th percentile of reconstruction errors observed on the normal baseline data. This evaluation, therefore, works out

whether this practical and efficient method is sufficient to configure the various anomaly detection methods for effective performance.

The results strongly indicate that this heuristic based configuration approach is not only sufficient, but also promisingly effective for the specific task of enhancing MARL-driven UAV swarm formation control.

In order to select the optimal untrustworthy message detection model, a comparative evaluation test was conducted. The performance of each model was assessed across all 12 feature configurations to identify its peak potential. The Local Outlier Factor (LOF) emerged as the clear winner, achieving a top F1-score of 0.999 on the `comprehensive` feature group. This peak performance was considerably higher than the best result from the Autoencoder (AE), which reached 0.964 (also on the `comprehensive` feature group), and far surpassed the Isolation Forests (IF) peak score of 0.670.

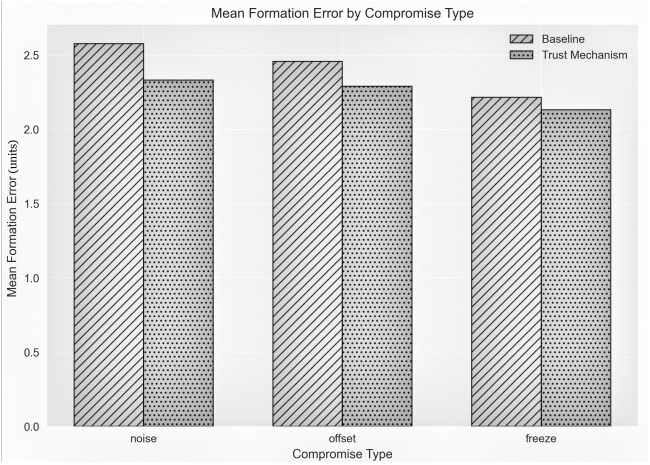

Fig. 1. Mean formation error of the swarm formation with the baseline policy versus the policy enhanced by TIF system. Results are averaged across three distinct communication compromise types: noise, offset, and freeze. The TIF system consistently reduces formation error in all scenarios. (Lower is better).

6.5 Robustness Enhancement Evaluation

This section directly addresses the third sub-research question (3) by evaluating the extent to which the configured post-hoc Trust-Based Information Filtering (TIF) system improves swarm formation control and resilience compared to the baseline policy. The following results were generated under the TIF systems optimal configuration, as determined by the analyses in the preceding sections. Specifically, it employs the Local Outlier Factor (LOF) model for trust assessment, which demonstrated the most consistent performance (Section 6.4), and utilizes the `comprehensive` spatio-temporal feature group, which proved most effective at identifying untrustworthy communication (Section 6.3.1). All performance improvements are averaged over thousands of simulated episodes to ensure statistical significance. These findings confirm that the TIF systems provides a considerable and measurable enhancement to robustness under various communication unreliability scenarios.

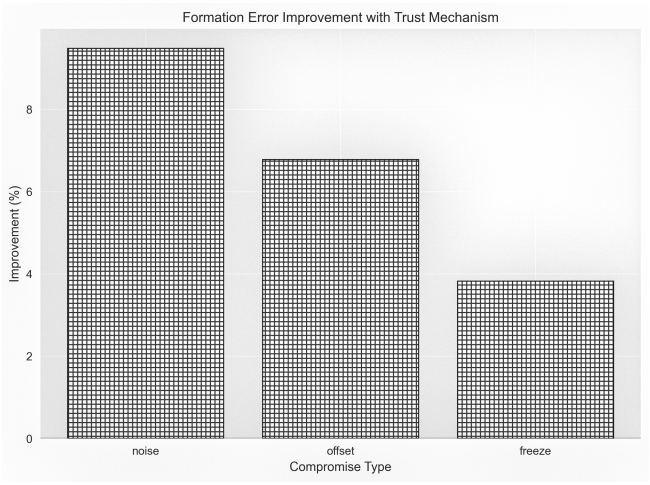

Fig. 2. Percentage improvement in mean formation error achieved by the TIF system, categorized by compromise type. The system shows the highest effectiveness against sensor noise (9.5% improvement) and its lowest against message freezing (3.8% improvement).

When evaluating the magnitude of these improvements, it is crucial to consider them within the specific context of this research. First and foremost, the TIF system is designed as a lightweight, post-hoc module that requires no to very little modification or costly retraining of the base MARL policy. Therefore, any performance gain represents a highly efficient performance enhancement. Secondly, in the domain of cooperative UAV swarms, formation cohesion can be directly linked to mission effectiveness and operational safety, that is, even incremental reductions in formation error may substantially decrease the risk of collision and improve the quality of such coordinated tasks. Finally, the consistent performance improvement over the baseline policy across multiple failure types demonstrates a tangible enhancement in overall system resilience. The following results should be interpreted through this lens.

6.5.1 Overall Performance Enhancement. Across all tested compromise types and rates, the TIF demonstrated a noticeable enhancement in swarm resilience. It achieved an overall reduction in mean formation error of 6.8%.

6.5.2 Performance Across Unreliability Scenarios. The effectiveness of the TIF system varies depending on the nature of the communication failure, as illustrated in Figure 1 and quantified in Figure 2.

Analysis shows that the system is most effective against sensor noise, achieving a substantial 9.5% improvement. This holds because the 'historical average' recovery method is well-suited to smoothen out high-frequency, random perturbations. Against consistent offset errors, the system provides a 6.8% improvement, aligning with the overall average. However, the system proved to be least effective against message freezing, yielding a lower improvement of 3.8%. This reduced effectiveness is likely because frozen (stale) messages do not immediately violate consistency checks if the swarms state changes slowly, making them challenging to detect.

A key design choice validated by the experiments was the message recovery strategy. The 'historical average' recovery method consistently outperformed 'trend extrapolation', proving on average 13% more effective at reducing formation error. Under its optimal configuration, the TIF system was capable of achieving a maximum improvement of 33.6% over the baseline, highlighting its potential in specific scenarios.

Furthermore, the systems performance was noteworthy even as the percentage of compromised message rate increased. It provided an overall 8.3% improvement at 10% compromise rate, which diminished slightly to 6.1% and 6% at compromise rates of 20% and 30%, respectively. This demonstrates that while the systems relative effectivenes decreases as communication quality degrades, it can continue providing valuable protective benefit.

7 CONCLUSION

This research set out to address the critical vulnerability of MARL-based UAV swarms to communication unreliability. The core objective was to design and validate a mechanism to enhance pretrained policies in a post-hoc, decentralized manner, without costly retraining or specific attack data. This work concludes that a decentralized, post-hoc trust and filtering mechanism, configured through unsupervised learning on normal operational data, can effectively and efficiently enhance the robustness of pretrained MARL UAV swarm policies, validated under context of drone swarm formation control. This is achieved by equipping each agent with a Trust-Based Information Filtering (TIF) system that leverages carefully engineered spatio-temporal features to discern and mitigate unreliable communication, thereby preserving mission performance without sacrificing the original policies integrity or requiring significant reconfiguration.

7.1 Key Findings and Contributions

To support the aforementioned conclusion, the research yielded several key findings corresponding to the initial research questions.

First, in investigating which spatio-temporal checks are most indicative of message reliability (sub-research question 1), the analysis revealed that temporal consistency features are considerably effective. A model relying solely on the temporal consistency of an agents reported observation history achieved near-perfect discrimination (0.997 F1-score). While the `comprehensive` feature group provided the highest observed performance, this finding underscores that even substantially simple, context-agnostic temporal checks can form the foundation of a highly robust system.

Second, the study validated that the proposed self-configuration process is effective for establishing a reliable operational baseline (sub-research question 2). By leveraging unsupervised models like the Local Outlier Factor (LOF) and applying a predefined, low anomaly threshold (a contamination factor of `0.05`) on normal operational data, the TIF system can be efficiently configured without hyperparameter tuning or labeled attack data.

Third, the evaluation process demonstrated that the optimal configured TIF system provides a measurable improvement in swarm resilience (sub-research question 3). Across all tested unreliability scenarios (noise, offset and freeze), the TIF system achieved an overall 6.8% reduction in mean formation error compared to the baseline policy. The system proved to be the most effective against high-frequency sensor noise (9.5% improvement) and demonstrated consistent, valuable protection results even as the rate of compromised messages increased, confirming its tangible contribution to operational robustness.

7.2 Significance of the Work

The significance of this research is twofold. Practically, it offers a modular, 'plug-and-play' solution that lowers the barrier of deploying robustly enhanced MARL systems. Stakeholders can enhance the reliability of existing pretrained policies without investing in costly and time consuming retraining cycles. Scientifically, this work presents a successful prototype for post-hoc trust mechanism in multi-agent environment, demonstrating that robust behavior enhancement can be layered on top of, rather than integrated within, the learning process.

7.3 Limitations and Future Research Directions

This study, while a strong proof-of-concept, has several limitations that open avenues for future research. First, the evaluation was conducted in a custom 2D simulation environment. Future work should validate the TIF system or its principles in high-fidelity 3D environments (for instance, RotorPy) and ultimately on physical UAV hardware to assess its performance with real-world physics and communication latencies.

Secondly, the TIF system was built upon the assumption of static baseline of normal behavior, configured once from an initial dataset. However, in very long duration missions or dynamically changing environments, the swarms expected 'normal' operational patterns might gradually shift. The current system cannot adapt to such shifts and might eventually misclassify legitimate behaviors. Future work could address this by incorporating online learning or a sort of periodic re-calibration mechanisms, allowing the trust system to adapt over time.

Furthermore, the system was tested against relatively simple communication issues like message freezing, offsets, and random Gaussian noise. A crucial next step could be to evaluate resilience against more sophisticated and adaptive adversaries that may attempt to strategically mimic normal behavior patterns. Such adversaries might not use random noise but instead inject intentional colored noise, that is, subtle, but temporally correlated signals designed to strategically imitate plausible flight behavior or exploit the system dynamics.

Finally, the TIF system relies on simple recovery strategies ('historical average' and 'trend extrapolation'). Future iterations could explore more advanced reconstruction techniques, such as those based on generative models (for example, Variational Autoencoders or GANs), to reconstruct more plausible replacement data for deemed to be untrustworthy messages. Investigating the systems scalability and performance in larger, more complex swarm configurations also remains a key area for future exploration.

ACKNOWLEDGMENTS

The author would like to thank Alex Chiumento for their guidance and feedback.

REFERENCES

[1] Christopher Amato. 2025. An Initial Introduction to Cooperative Multi-Agent Reinforcement Learning. arXiv:2405.06161 [cs.LG] https://arxiv.org/abs/2405.06161

[2] Jason Ansel, Edward Yang, Horace He, Natalia Gimelshein, Animesh Jain, Michael Voznesensky, Bin Bao, Peter Bell, David Berard, Evgeni Burovski, Geeta Chauhan, Anjali Chourdia, Will Constable, Alban Desmaison, Zachary DeVito, Elias Ellison, Will Feng, Jiong Gong, Michael Gschwind, Brian Hirsh, Sherlock Huang, Kshiteej Kalambarkar, Laurent Kirsch, Michael Lazos, Mario Lezcano, Yanbo Liang, Jason Liang, Yinghai Lu, CK Luk, Bert Maher, Yunjie Pan, Christian Puhrsch, Matthias Reso, Mark Saroufim, Marcos Yukio Siraichi, Helen Suk, Michael Suo, Phil Tillet, Eikan Wang, Xiaodong Wang, William Wen, Shunting Zhang, Xu Zhao, Keren Zhou, Richard Zou, Ajit Mathews, Gregory Chanan, Peng Wu, and Soumith Chintala. 2024. PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. In *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*. ACM. https://doi.org/10.1145/3620665.3640366

[3] Markus Breunig, Peer Kröger, Raymond Ng, and Joerg Sander. 2000. LOF: Identifying Density-Based Local Outliers. *ACM Sigmod Record* 29, 93–104. https://doi.org/10.1145/342009.335388

[4] Juan C. Burguillo. 2018. *Multi-agent Systems*. Springer International Publishing, Cham, 69–87. https://doi.org/10.1007/978-3-319-69898-4_5

[5] Spencer Folk, James Paulos, and Vijay Kumar. 2023. RotorPy: A Python-based Multirotor Simulator with Aerodynamics for Education and Research. *arXiv preprint arXiv:2306.04485* (2023).

[6] Ho Long Fung, Victor-Alexandru Darvariu, Stephen Hailes, and Mirco Musolesi. 2024. Trust-based Consensus in Multi-Agent Reinforcement Learning Systems. arXiv:2205.12880 [cs.MA] https://arxiv.org/abs/2205.12880

[7] Pengbin Han, Xinfeng Wu, and Aina Sui. 2024. DTPBFT:A dynamic and highly trusted blockchain consensus algorithm for UAV swarm. *Computer Networks* 250 (2024), 110602. https://doi.org/10.1016/j.comnet.2024.110602

[8] Mark A Kramer. 1991. Nonlinear principal component analysis using autoassociative neural networks. *AIChE journal* 37, 2 (1991), 233–243.

[9] Rui Li and Hongzhong Ma. 2020. Research on UAV Swarm Cooperative Reconnaissance and Combat Technology. In *2020 3rd International Conference on Unmanned Systems (ICUS)*. 996–999. https://doi.org/10.1109/ICUS50048.2020.9274902

[10] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation Forest. In *2008 Eighth IEEE International Conference on Data Mining*. 413–422. https://doi.org/10.1109/ICDM.2008.17

[11] Vincenzo Lomonaco, Angelo Trotta, Marta Ziosi, Juan de Dios Yáñez Ávila, and Natalia Díaz-Rodríguez. 2018. Intelligent Drone Swarm for Search and Rescue Operations at Sea. arXiv:1811.05291 [cs.CY] https://arxiv.org/abs/1811.05291

[12] Mingyang Lyu, Yibo Zhao, Chao Huang, and Hailong Huang. 2023. Unmanned aerial vehicles for search and rescue: A survey. *Remote Sensing* 15, 13 (2023), 3266.

[13] Rupert Mitchell, Jan Blumenkamp, and Amanda Prorok. 2020. Gaussian Process Based Message Filtering for Robust Multi-Agent Cooperation in the Presence of Adversarial Communication. arXiv:2012.00508 [cs.RO] https://arxiv.org/abs/2012.00508

[14] Andrew Y. Ng, Daishi Harada, and Stuart J. Russell. 1999. Policy Invariance Under Reward Transformations: Theory and Application to Reward Shaping. In *Proceedings of the Sixteenth International Conference on Machine Learning (ICML '99)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 278–287.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[16] Yanchao Sun, Ruijie Zheng, Parisa Hassanzadeh, Yongyuan Liang, Soheil Feizi, Sumitra Ganesh, and Furong Huang. 2022. Certifiably Robust Policy Learning against Adversarial Communication in Multi-agent Systems. arXiv:2206.10158 [cs.LG] https://arxiv.org/abs/2206.10158

[17] Andrew S. Tanenbaum and David J. Wetherall. 2011. *Computer Networks* (5th ed.). Prentice Hall. 306 pages.

[18] Jordan Terry, Benjamin Black, Nathaniel Grammel, Mario Jayakumar, Ananth Hari, Ryan Sullivan, Luis Santos, Rodrigo Perez, Caroline Horsch, Clemens Dieffendahl, Niall Williams, and Yashas Lokesh. 2021. PettingZoo: Gym for Multi-Agent Reinforcement Learning. In *Advances in Neural Information Processing Systems*. https://doi.org/10.48550/arXiv.2009.14471

[19] Bei Xu, Guanghan Bai, Yun'an Zhang, Yining Fang, and Junyong Tao. 2022. Failure analysis of unmanned autonomous swarm considering cascading effects. *Journal of Systems Engineering and Electronics* 33, 3 (2022), 759–770. https://doi.org/10.23919/JSEE.2022.000069

[20] Wanqi Xue, Wei Qiu, Bo An, Zinovi Rabinovich, Svetlana Obraztsova, and Chai Kiat Yeo. 2022. Mis-spoke or mis-lead: Achieving Robustness in Multi-Agent Communicative Reinforcement Learning. arXiv:2108.03803 [cs.LG] https://arxiv.org/abs/2108.03803

[21] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2022. The Surprising Effectiveness of PPO in Cooperative, Multi-Agent Games. arXiv:2103.01955 [cs.LG] https://arxiv.org/abs/2103.01955

[22] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2025. on-policy. GitHub repository. https://github.com/marlbenchmark/on-policy Main branch, commit `de66d7a4b23fac2513f56f96f73b3f5cb96695ac`. Accessed between May and June 2025..

[23] Xiang Zheng, Xingjun Ma, Shengjie Wang, Xinyu Wang, Chao Shen, and Cong Wang. 2024. Toward Evaluating Robustness of Reinforcement Learning with Adversarial Policy. In *2024 54th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. 288–301. https://doi.org/10.1109/DSN58291.2024.00038