ECAI 2025
I. Lynce et al. (Eds.)
© 2025 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA251457

RTF: Reputation-Based Team Formation Approach for Enhancing Dependable Critical Fully Autonomous Multi-UAV BVLOS Operations

Kıvanç Şerefoğlu[a,*], Önder Gürcan[b,c] and Reyhan Aydoğan[a]

[a]Özyegin Üniversity
[b]NORCE - Norwegian Research Center AS
[c]Coral Protocol
ORCID (Kıvanç Şerefoğlu): https://orcid.org/0009-0009-2162-7754, ORCID (Önder Gürcan): https://orcid.org/0000-0001-6982-5658, ORCID (Reyhan Aydoğan): https://orcid.org/0000-0002-5260-9999

Abstract. The development of autonomous multi-unmanned aerial vehicles, or multi-UAVs, represents significant progress in various fields, including the military, business, and civilian sectors. As this innovative field develops, the necessity of creating models that guarantee the fundamental components of security, privacy, dependability, effectiveness, and safety is growing. Based on this observation, this research suggests a reputation-based team formation (RTF) approach that rates each UAV according to the level of services it provides. Such a system is essential for critical multi-UAV Beyond Visual Line of Sight (BVLOS) operations requiring high dependability, effectiveness, safety, and privacy protection. To validate RTF's effectiveness, we focus on a simulation scenario where multi-UAVs are deployed to localize a target in a forest where a fire is spreading. The criticality of this operation is underscored by the need to locate the target before the fire reaches it. Our results demonstrate that RTF successfully enables cooperative UAVs to block malicious ones and significantly reduce their influence, even in the presence of multiple compromised UAVs.

1 Introduction

Autonomous unmanned aerial vehicles (UAVs) are increasingly used in domains such as agriculture, environmental monitoring, disaster response, and defense [4, 12]. As these systems operate in high-stakes settings, their dependability is critical to prevent mission failures, financial losses, or even threats to human safety [16]. In multi-UAV systems, this dependability must extend to robustness, resilience, and fault tolerance under dynamic and uncertain conditions [7, 11, 14]. With the rise of fully autonomous UAV operations [15], one promising paradigm is the UAV marketplace, where UAVs from different—potentially untrusted—owners collaborate in Beyond Visual Line of Sight (BVLOS) missions.

Hence, we anticipate a blockchain and reputation-based system that enables autonomous UAVs to collaborate on specific critical BVLOS missions allowing for various new critical BVLOS missions and establishing a platform for on-demand service provisioning and acquisition driven by incentives. Accordingly, our work extends the reputation-based team formation approach proposed in [2] where a fundamental blockchain-based infrastructure in the UAV marketplace and a basic reputation model have been introduced. While that work assumes a Byzantine Fault Tolerant (BFT) consensus provided by the underlying blockchain-based infrastructure, ensuring that a malicious subset of validators cannot rewrite history, it does not incorporate BFT into the reputation calculation itself. In contrast, in this study, we propose a blockchain-independent reputation system specifically designed to handle Byzantine UAVs without relying on blockchain mechanisms[1]. The reputation system is critical for maintaining high trust and operational integrity in dynamic and uncertain environments.

The proposed reputation model integrates dynamic feedback mechanisms and boosts UAV reliability, safety, and efficiency in complex environments, enabling safer and more scalable autonomous collaborations. To our knowledge, it is the first decentralized, Byzantine-resilient reputation system for critical fully autonomous multi-UAV missions. The contributions of this study are as follows:

1. **Reputation-Based Team Formation:** We introduce an approach to forming *fully autonomous UAV teams* by selecting UAVs from a marketplace, where they are evaluated based on a dynamic reputation system. Unlike traditional IoT Network and UAV literature [10], which typically focus on preassigned or static UAV networks, our method allows users to select UAVs based on their global reputation scores from a marketplace. This enhances the reliability and effectiveness of *multi-UAV BVLOS missions* by ensuring that only the most dependable UAVs are selected for critical tasks.
2. **Global Reputation Score Algorithm:** We develop a novel algorithm to calculate the global reputation scores of UAVs by incorporating recent mission outcomes and dynamically adjusting the weight of previous experiences. We also integrate a blocklist mechanism that identifies and excludes potential Byzantine UAVs—specifically those attempting to manipulate reputation scores by systematically assigning unfairly low ratings. This identification is supported by estimating the expected target detection

* Corresponding Author. Email: kivanc.serefoglu@ozu.edu.tr

[1] The proposed reputation approach has been demonstrated in [13].

time, allowing the system to detect anomalous behavior indicative of malicious intent.

3. **Simulation-based Validation:** We validate the effectiveness and robustness of our reputation system by simulating the system against byzantine faults [8] within UAV teams.

This paper is organized as follows. Section 2 discusses dependable multi-UAV operations, focusing on team formation and mission execution frameworks. Section 3 introduces our reputation-based team formation model for multi-UAV systems. Section 4 provides the experimental evaluation setup of our proposed system using simulations, and Section 5 provides the results of the simulations. Section 6 reviews related work, comparing our approach with existing solutions. Finally, Section 7 concludes the paper, summarizing our findings and their implications for future multi-UAV operations.

2 ENVISIONED SYSTEM

Multi-UAV systems are deployed in missions, which encompass various tasks along with associated conditions or constraints. Coordinating a team of UAVs is frequently more effective than employing a single UAV, particularly for complex assignments. However, the field of multi-UAV system development is relatively in its infancy, requiring substantial research efforts [2]. As UAV technology and autonomy progress, there is an emerging requirement for establishing different levels to delineate advancements in the autonomy of UAV systems. Moreover, as the level of autonomy increases, there is more need for reliable and dependable multi-UAV operations to ensure the success and safety of complex and critical missions.

In this study, we consider the highest level of autonomy, which is full UAV autonomy, where the operator only selects the mission and all other processes of the multi-UAV operation are carried out without help from the operator. We assume that there is a UAV Marketplace System (UMS), as proposed in [2], where various types of UAVs are subscribed by their owners and the mission details (number of UAVs, type of UAV, etc.) are registered by their operators. The UMS plays a crucial role in maintaining operational flow by ensuring that each proposed mission can be executed with the existing UAV fleet. The UMS examines the available UAVs to select the best candidates for the desired mission. This evaluation is based on ranking the suitability of each UAV's global reputation score, and when it was last on mission-to-mission requirements. The highest-grade UAVs (i.e., the most reputed ones) are selected to form the team, optimizing the missions' potential for success. (see Fig. 1).

A mission template is a generalized description of a mission, defined by various parameters including the quantity of UAVs involved, forest density, speed of spread of fire, the mission's start locations and end time, the human target's location, and area size. The mission template is created by human operators. When an operator initiates a mission, based on its dedicated mission template, the UMS first conducts a thorough evaluation to ensure the proposed mission's viability, considering the availability of UAVs that meet the specified criteria and the absence of conflicts with concurrent mission operations. Upon successful validation, the mission is approved, facilitating the subsequent phases of team formation and mission deployment.

UMS uses a reputation-based team formation approach. Basically, it ranks and selects UAVs based on their global reputation scores. The system also integrates a blocklist mechanism that monitors the consistency of trust evaluations over time. UAVs that exhibit repeated anomalies in their evaluation behavior are gradually restricted from participation through a blocking process. This mechanism operates in the background as part of the reputation update cycle, ensuring that UAVs contributing to the degradation of mission reliability are automatically filtered out in subsequent formations. After this process, the UAVs are ready to start the mission. Depending on the calculated flight plans, the UAVs position themselves at their starting positions and then they start executing the mission. Upon completion of the mission, a thorough evaluation is conducted to assess the performance of each UAV. This evaluation informs updates to each UAV's performance records within the UMS. Feedback from this stage is integrated into future team formations, refining the selection algorithm to enhance team compositions for subsequent missions.

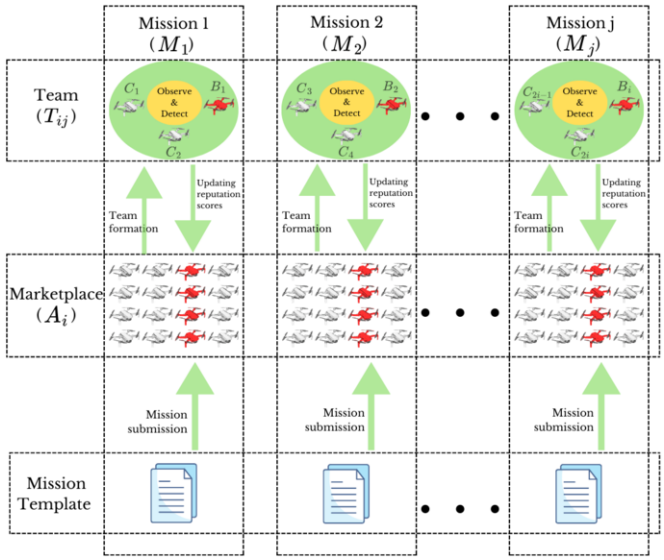

Figure 1. UAV Marketplace System (UMS).

3 REPUTATION-BASED TEAM FORMATION

In this section, we explain our sophisticated reputation-based team formation (RTF) mechanism for critical multi-UAV operations.

3.1 *Reputation Model for Multi-UAVs*

The cornerstone of our approach to establishing a robust and dynamic reputation system for autonomous multi-UAVs lies in the precise calculation of a global reputation score (S_i) for each UAV (i). This score, which is calculated as follows, is crucial for balancing dependability and trust within the system:

$$S_i^t = (1 - \alpha_i) \cdot S_i^{t-1} + (\alpha_i) \cdot P_i^t \quad (1)$$

where S_i^{t-1} represents the UAV's reputation score from previous assessments, and P_i^t is the currently calculated score based on recent mission performances. Note that in our formulation, the range of the scores is between zero and one, and the initial value of each UAV (new UAV in our system) is set to one, allowing newcomers to join without disadvantage, while malicious ones are penalized through reputation updates and the blocklist. The system then relies on their performance reports to update trust over time.

The parameter α_i serves as a weighting factor for each UAV, modulating the impact of new evaluations on the UAV's overall reputation score. Instead of using a constant α_i, we adopt a time-based α_i parameter where the value of α_i declines over time. In the beginning,

the system did not have enough experience; therefore, the current experiences of the agent's performance play a more crucial role. Over time, the system accumulated experiences; hence, the system relies on its current evaluation more by considering the risk of a potential missevaluation of the current performance. That is, we give more weight to the accumulated experience while it can still consider the current performance evaluation in the update process. Acknowledging the evolving nature of UAV collaborations, our system dynamically adjusts the weighting factor (α_i) in accordance with the number of joint missions (n), implementing a decay rate (r), which is 0.02 in our reputation system to modulate the influence of accumulated experience on the reputation score:

$$\alpha_i^n = \alpha_i^{n-1} \cdot (1 - r) \tag{2}$$

Figure 2 shows the relationship between the weighting factor (α_i) and the number of joint missions (n) in a UAV collaboration system.

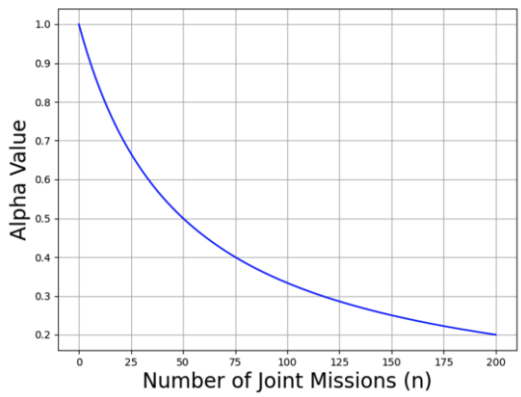

Figure 2. Alpha Value vs. Number of Joint Missions.

The computation of the performance scores (P_i^t) for the current mission is a critical component of our model, integrating evaluations from all participating agents ($j \in A$). However, some UAV agents may aim to mislead the system. That is, they might give low scores for reliable agents but high scores for malicious agents. To avoid such cases, the evaluation of the unreliable agent, whose global score is less than a threshold (Θ), which is 0.5 for our system, is dismissed in the global reputation score calculation as formulated below:

$$P_i^t = \frac{\sum_{j:S_j>\Theta}(S_j \cdot E_j^i)}{\sum_{j:S_j>\Theta} Sj} \tag{3}$$

The evaluation E_j^i represents how agent j assesses the performance of agent i during a mission, based on the accuracy of the shared information provided by agent i. The value of E_j^i is determined as follows:

$$E_j^i = \begin{cases} 1.0 & \text{if the information is accurate} \\ 0.5 & \text{if the information is inaccurate but the mission does not fail} \\ 0.0 & \text{if the information is inaccurate and the mission fails} \end{cases} \tag{4}$$

The rationale for this evaluation method is to recognize that the mission can still succeed despite receiving incorrect information. Thus, incorrect information is tolerated to an extent if the overall mission is successful, which is reflected by setting E_j^i to 0.5 in such cases. However, if the incorrect information leads to mission failure, the impact is more severe, resulting in E_j^i being 0.

This evaluation metric ensures that the integrity and reliability of the mission reports are upheld. By multiplying E_j^i by the global reputation score of agent j (S_j), the formula ensures that more reliable agents have a greater influence on the performance score of Agent i. This approach discourages malicious reporting and promotes accurate evaluations, thus maintaining the integrity of the scoring mechanism. Including a reliability criterion underscores our commitment to fostering a dependable and secure multi-UAV environment.

To detect and eliminate Byzantine UAVs, we also implement a block list mechanism based on a consistency check between anticipated and actual performance. This check is triggered only when the mission is completed successfully. The UAV agent is marked as suspect when its observed actual performance is significantly deviated. Any UAV flagged as suspect in two different missions is added to a *block list*, excluding it from future team formations to safeguard the integrity of the system.

3.2 Team Formation using RTF

The selection of UAVs for each mission is guided by criteria considering global reputation scores derived from the performance of UAVs in previous missions, as discussed in Section 3.1. The system prioritizes UAVs with higher reputation scores to reflect their reliability and effectiveness. UAVs listed in the *block list*—due to repeated suspicious behavior - are automatically excluded from the selection process to ensure the integrity of the mission. If more than one UAV has the same reputation score, the system selects the one considering the number of missions completed by each UAV, favoring those with fewer missions to ensure balanced utilization in the marketplace.

In cases where reputation scores and the number of missions are the same, the system selects UAVs based on the timeliness of their last missions and prefers those that have not been on a mission for longer. In this way, all UAVs in the system have been on a mission. This structured approach to UAV selection ensures that the UAVs deployed for each mission are optimally suited to meet the specific challenges of the mission, balancing performance metrics with operational demands to achieve the best possible results.

3.3 Case Study

We define multiple UAV marketplaces, each with a specific proportion of Byzantine UAVs (detailed in [13]). The simulation environment includes fire-based search and rescue missions, where UAVs must locate a stationary human target before the fire reaches the target's location (Figure 3). UAVs follow a sweeping search strategy and engage in cooperative information sharing and verification during missions. When a UAV detects the target, it broadcasts the coordinates to the rest of the team. Upon receiving this message, other UAVs navigate to the specified location to verify the accuracy of the information. If the target is indeed at the reported location, the information is marked as accurate, and those UAVs also propagate this information. Otherwise, the UAVs disregard the information and return to their previous positions to continue the search. We assume a strong adversary, where Byzantine UAVs can coordinate to carry out attacks such as False Feedback Attacks [5, 10], Collusion Attacks [9, 10]. In other words, Byzantine UAVs can collaborate, giving themselves high reputation scores while assigning low scores to the other UAVs. Also, they provide false coordination information of target [17] Byzantine UAVs may provide highly inaccurate target locations to mislead other UAVs, causing the team to waste time searching for the real target. To avoid easy detection, Byzantine UAVs do

not always behave maliciously; instead, they alternate between cooperative and deceptive behavior based on a predefined error rate, allowing them to gain the trust of honest UAVs while simultaneously undermining their reputations. These interactions serve as the basis for peer performance evaluations (E_j^i) at the end of each mission. Missions follow the template

$$M = \{n, d, t, p, s, h\}, \tag{5}$$

where n is number of UAVs, d forest density, t fire spread rate (ticks), p fire origin, s grid size, and h target position. All missions employ a sweeping search pattern from west to east, descending one row at a time. The UAVs complete the mission successfully by locating the stationary human target before the fire reaches it (Figure 4).

To implement blocklist mechanism, we estimate the mission completion time by calculating the shortest direct travel time (using Euclidean distance [3] and UAV speed) to determine the first detection. Then, we simulate the sweep-pattern movement of the remaining UAVs to estimate their positions at the moment of notification. From these positions, we compute the time required for each UAV to reach the target, considering their sensor ranges. The latest arrival among them defines the expected mission duration. If the actual mission time closely matches the expected time (i.e., within a small tolerance margin of 5%) and a UAV has assigned low trust scores (≤ 0.5) to teammates, it is flagged as *suspicion*. However, if the actual mission time significantly exceeds the expected value, no UAV is flagged; this condition indicates the likely presence of at least one Byzantine agent, which justifies possible low evaluations.

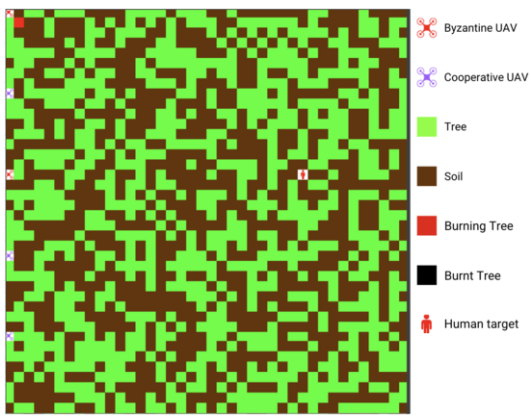

Figure 3. An example mission where search for human target.

4 EXPERIMENTAL EVALUATION

Here, we provide a basic proof-of-concept simulation model for the feasibility and efficacy of the proposed solution. To validate our approach, we implemented a proof-of-concept simulation on the SCOP Framework[2]. We use the UAV Marketplace System (UMS) model from our prior work [13], in which each marketplace comprises cooperative and we assume that a marketplace is fixed during missions, i.e., no UAV is added or removed.

4.1 Simulation Setup

Marketplaces are categorized by their level of difficulty based on the presence of Byzantine UAVs. Marketplaces with a higher percentage

[2] SCOP Framework, https://scop-framework.netlify.app, last accessed 06/05/2025.

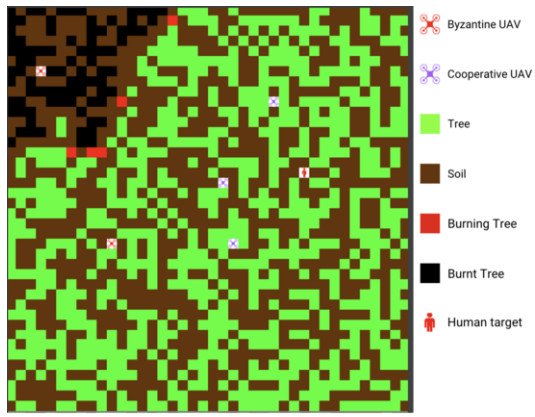

Figure 4. An example of successful mission.

of Byzantine UAVs present more challenging conditions, requiring robust strategies to mitigate potential sabotage or disruptions. Four distinct UAV marketplaces have been established (as in [13]), each tailored to specific operational requirements. Different marketplaces with varying Byzantine UAV ratios are used to evaluate the reputation systems. These marketplaces present distinct levels of difficulty, allowing the reputation systems to be tested under diverse conditions based on the proportion of Byzantine UAVs.

- **MarketPlace%0**: Equipped with a total of 12 UAVs, this marketplace includes 12 cooperative UAVs, of which 0 are Byzantine UAVs, maintaining a Byzantine ratio of 0%. Hence, there are no error rates to report for Byzantine UAVs in this MarketPlace.
- **MarketPlace%25**: Equipped with a total of 12 UAVs, this marketplace includes 9 cooperative UAVs, of which 3 are Byzantine UAVs, maintaining a Byzantine ratio of approximately 25%. The error rates for the Byzantine UAVs are as follows: one with an error rate of 0.703, another with an error rate of 0.354, and a third with an error rate of 0.793.
- **MarketPlace%33**: This marketplace includes a total of 12 UAVs, including 8 cooperative and 4 Byzantine UAVs, resulting in a Byzantine ratio of about 33%. The error rates for the Byzantine UAVs are as follows: 0.407, 0.747, 0.665, and 0.637.
- **MarketPlace%42**: This marketplace consists of a total of 12 UAVs, including 7 cooperative UAVs and 5 Byzantine UAVs, resulting in a Byzantine ratio of approximately 42%. The error rates for the Byzantine UAVs are as follows: 0.298, 0.359, 0.299, 0.377, and 0.512.

4.2 Simulations

In this section, the simulation environment used to obtain the mission results is described. To ensure a valid comparison between different systems, a consistent mission template was applied across all simulations. A single mission was defined with the following parameters:

$$M = \{n = 5,\ d = 0.5,\ t = 8,\ p = (1, 1),\ s = 40 \times 40,\ h = (17, 29)\}, \tag{6}$$

The target coordinates (17,29) were chosen because the fire reaches this location at 280 ticks, creating a time-critical scenario. Despite the interference of unpredictable Byzantine UAVs, the cooperative UAVs still have enough time to reach the target if they function effectively, though they may not always succeed due to the Byzantine UAVs' potential to disrupt their performance.

Although most parameters in our framework can be adjusted, for the purpose of illustrating the algorithms' performance, we kept

the environment parameters constant across all simulations. This allowed the focus to be on testing the algorithms' ability to select cooperative UAVs or handle Byzantine UAVs. The behavior of the Byzantine UAVs, which acted either cooperatively or exhibited Byzantine behavior in a random manner, served as a significant test for the algorithms. This randomness, driven by an error rate that determines the percentage of time Byzantine UAVs act maliciously, helped evaluate how well each system could identify and select cooperative UAVs under unpredictable conditions.

Each simulation consists of a marketplace where the UAVs of this marketplace are selected for 50 consecutive missions under the same configuration (see Equation 6, i.e. the target located at the same position in the same forest and the fire ignites from the same position) and using a specific reputation algorithm. With this setting, we can compare different reputation algorithms in a fair way. For each algorithm, the simulations are run 10 times and an average performance measure is calculated. The repetition was necessary to ensure accurate results due to the random actions of the Byzantine UAVs, which could impact the performance of the algorithms.

The simulations were executed using three different reputation systems: Reputation Trust Framework (RTF), Multi-Domain Blockchain-Based Routing (MDBR) [1], EigenTrust [6], and a Random selection system. In the MDBR [1], EigenTrust [6], and RTF systems, UAVs were selected for missions by the UAV Marketplace System (UMS) based on their respective reputation scores.

1. **MDBR Algorithm:** The algorithm calculates local reputation using a Bayesian formula based on correct and incorrect information. It then computes a weighted value from the global reputations of interacting UAVs. These local values are multiplied by the weighted values and summed to get the global reputation. We also removed the time decay function from MDBR, as UAVs wait in the marketplace when not on missions, making the decay unsuitable for this scenario.
2. **EigenTrust Algorithm:** The algorithm calculates local trust values based on the difference between correct and incorrect messages exchanged between UAVs. These normalized values follow the process from Algorithm 2 [6]. The trust values are iteratively updated over 100 iterations or until changes fall below a threshold (1e-6), with an alpha value (0.85) to balance global and local trust, continuing until global trust values stabilize.
3. **Random Algorithm:** The algorithm assigns UAVs to missions using a random selection process. UAVs are selected without considering their reputation or any past performance, making the assignment completely random.

We removed the time decay function from MDBR to adapt the algorithm for the simulation because the time decay function is not suitable for this scenario. UAVs are not always engaged in missions; they wait in the marketplace when not selected for a mission.

5 RESULTS AND DISCUSSION

In evaluating the performance of UAVs, finish time is a key metric, as it represents the time required for the UAVs to either locate or fail to locate a target. This metric is especially important in time-sensitive, critical missions, such as rescuing individuals during a forest fire, where delays can lead to severe consequences. Therefore, comparing finish times offers valuable insights into the efficiency of UAVs, particularly in urgent missions where timely intervention is critical. The intervals for evaluating finish times were determined based on the distribution of mission completion ticks. The optimal result for this mission occurs within 100 ticks; hence, the first interval is set between 0 and 100 ticks. Subsequent intervals were defined based on the distribution of finish ticks across the simulations.

Figure 5 presents the average number of completed missions under different Byzantine UAV ratios, evaluated over 10 trials and grouped by mission finish time intervals. The marketplaces are characterized by 0%, 25%, 33%, and 45% Byzantine UAV presence. The x-axis shows the finish time intervals (in ticks), while the y-axis denotes the average number of missions completed in each interval. The results demonstrate a clear degradation in performance as the percentage of Byzantine UAVs increases. In the 0% Byzantine marketplace, nearly all missions are completed efficiently within the 0–100 tick range, highlighting the optimal functioning of the system when no adversarial agents are present. As the Byzantine ratio rises to 25% and 33%, there is a noticeable decline in the number of missions completed in the 0–100 tick range. Instead, more missions begin to spill over into the 101–150 and 151–200 intervals, reflecting delays introduced by the influence of Byzantine UAVs. However, the majority of missions are still eventually completed. The most severe degradation is observed at the 45% Byzantine level. Here, the number of missions completed in the 0–100 tick interval drops significantly. A substantial number of missions are delayed to later time intervals, and a non-negligible portion of missions fail altogether. This trend indicates that as Byzantine presence approaches nearly half of the UAV population, mission success rates are severely compromised, both in timeliness and reliability.

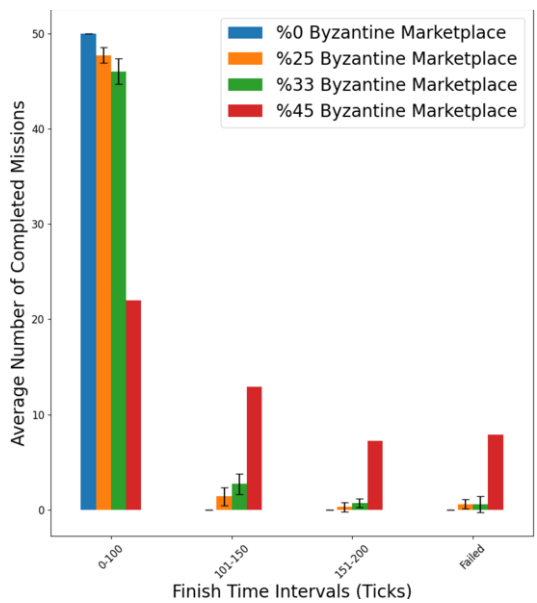

Figure 5. Impact of Byzantine UAVs on the success of the RTF algorithm

Figure 6 presents a comparison of mission finish times across four algorithms: RTF, EigenTrust, Random, and MDBR, in the MarketPlace %25. As shown in the figure, most algorithms handle the presence of Byzantine UAVs effectively. The RTF algorithm continues to demonstrate robust performance, with the majority of missions being completed within the 0-100 tick range and minimal failures. This reinforces its adaptability even as Byzantine presence becomes more disruptive. MDBR follows with relatively strong performance, although a noticeable number of missions shift into the 101-150 and even 151-200 tick intervals. This suggests some delays in mission completion, likely due to intermittent misidentification of Byzantine agents. In contrast, the EigenTrust and Random algorithms suffer

more visibly. Both show fewer timely completions and greater dispersion across delayed intervals and failed missions. The high variance in EigenTrust's performance indicates sensitivity to adversarial behavior, while Random's results confirm its vulnerability in strategic environments.

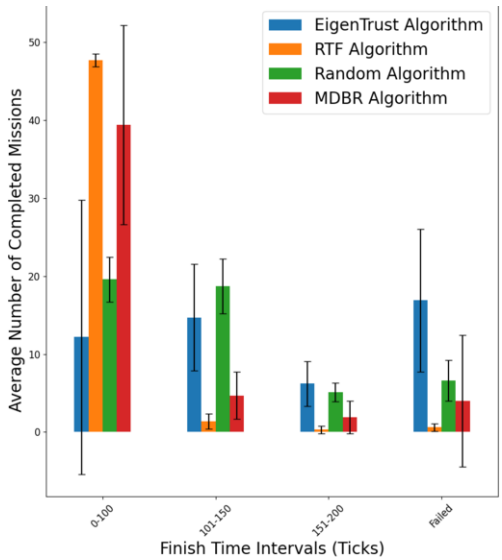

Figure 6. 25% Byzantine UAVs. Comparison of finish times between Random, EigenTrust, MDBR, and RTF algorithms.

As the proportion of Byzantine UAVs increases to 33%, the impact on algorithm performance becomes even more pronounced, as shown in Figure 7. Compared to the 25% scenario in Figure 6, the gap between robust and fragile algorithms further widens. The RTF algorithm still maintains its superiority, with most missions successfully completed within the 0-100 tick range. Despite the increased adversarial influence, RTF shows minimal degradation and continues to avoid Byzantine UAVs effectively, highlighting its resilience under more hostile conditions. In contrast, both EigenTrust and MDBR algorithms exhibit clear signs of being manipulated by Byzantine UAVs due to the impact of false feedback. EigenTrust, in particular, completes very few missions in the optimal time range and has a high failure count. This pattern suggests that Byzantine UAVs are able to deceive the EigenTrust scoring mechanism, leading to poor trust assessments and subsequent poor UAV selection. MDBR, while somewhat better than EigenTrust, also experiences increased delays and failures, likely due to its inability to consistently filter out malicious agents. The growing standard deviations in their performance further support this instability. Furthermore, for critical missions - such as search and rescue— the performance of the other algorithms cannot be accepted due to their vulnerability to Byzantine manipulation. RTF stands out as the most reliable algorithm, particularly when the percentage of Byzantine UAVs is up to 33%. Beyond this threshold, the other algorithms suffer significant performance degradation, making RTF the most suitable choice for ensuring reliable and timely mission completion under such conditions.

Figure 8 shows the performance of four algorithms—RTF, EigenTrust, Random, and MDBR—in the MarketPlace 42%. As shown in the figure, all algorithms are significantly affected by the increasing density of Byzantine UAVs, and the decline in performance is now more severe and widespread across all finish time intervals.

RTF, which previously held the strongest position, now shows a considerable drop in efficiency. Fewer missions are completed within

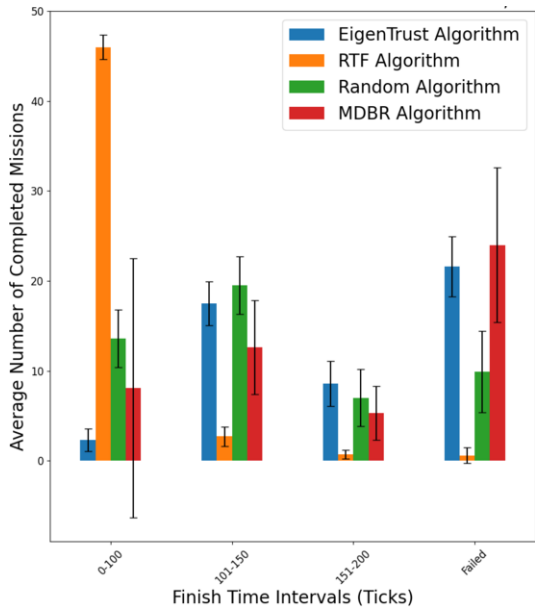

Figure 7. 33% Byzantine UAVs. Comparison of finish times between Random, EigenTrust, MDBR, and RTF algorithms.

the optimal 0-100 tick range, and more are delayed into the 101-150 and 151-200 intervals. Additionally, the failure rate increases, highlighting that even the most robust reputation-based algorithm becomes vulnerable at this level of Byzantine interference. This instability is also reflected in the standard deviation of RTF's performance. As seen in the figure, the blocklist mechanism sometimes functions effectively by identifying and excluding Byzantine UAVs, while at other times it fails to detect them. This inconsistency stems from the design of the mechanism itself: it requires the mission's actual completion time to be close to the expected completion time in order to validate trust updates. However, when Byzantine UAVs are abundant, they can deliberately manipulate the mission duration—causing it to significantly exceed expectations. In such cases, the mechanism is not triggered, and the Byzantine UAVs are not blocked, which negatively impacts the algorithm's reliability. EigenTrust and MDBR continue to exhibit clear symptoms of manipulation. Both algorithms experience widespread mission delays and high failure rates, particularly due to their heavy reliance on trust metrics, which are easily corrupted by false feedback from Byzantine agents. MDBR, in particular, shows a worrying shift toward the later intervals and failed missions, with increasingly unstable performance reflected in the large standard deviations. The Random algorithm, despite being non-strategic, once again performs relatively better under these extreme conditions. It completes a competitive number of missions in the earlier intervals and avoids the steep failure rates seen in reputation-based methods. Since Random does not depend on manipulated reputation values, it is less affected by Byzantine deception.

In summary, at the 42% Byzantine level, reputation-based algorithms like RTF, EigenTrust, and MDBR struggle to maintain operational efficiency. Their performance is severely degraded by false feedback and manipulation, leading to delays and mission failures. Meanwhile, the Random algorithm—though simple and non-adaptive—proves to be less affected by Byzantine deception. Despite its sharp decline, RTF still performs slightly better than Random in terms of overall mission completions, suggesting it retains a minimal level of resilience even under heavy adversarial pressure.

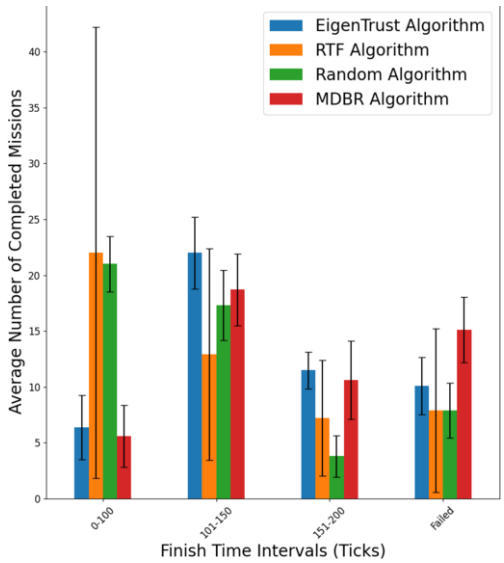

Figure 8. 42% Byzantine UAVs. Comparison of finish times between Random, EigenTrust, MDBR, and RTF algorithms.

6 RELATED WORK

The advancement of fully autonomous unmanned aerial vehicle (UAV) systems has significantly impacted fields ranging from surveillance to environmental monitoring, necessitating robust frameworks for team dynamics and mission integrity. This section reviews existing literature on the mechanisms of autonomous collaboration and reputation systems among UAVs, emphasizing their contributions to operational efficiency and trustworthiness. By situating our research within this context, we aim to highlight the challenges and solutions identified in the realm of UAV teamwork, underpinning our project's objectives and methodology.

Buzcu et al. introduce reputation-based teamwork for UAV, focusing on BVLOS missions and utilizing blockchain to ensure trust, privacy, and reliability in multi-UAV collaborations [2]. Their approach involves forming UAV teams based on reputation scores calculated from previous mission performances, enabling efficient and trustworthy team formation. However, their reputation model primarily focuses on a basic UAV marketplace scenario and its applicability in limited mission contexts. In our work, we extend their foundational model by developing a more advanced and dynamic reputation system. Our system introduces real-time performance evaluations and employs Byzantine fault tolerance, ensuring the system can withstand attempts by compromised UAVs to disrupt mission execution. Additionally, we developed a more complex simulation environment, simulating diverse UAV behaviors and incorporating privacy-preserving mechanisms to secure mission data. This advanced reputation system not only enhances mission reliability but also supports more scalable and complex UAV collaborations in critical operations.

Wardega et al. introduce a decentralized blocklist protocol for achieving Byzantine resilience in multi-robot systems, where robots issue accusations based on observed misbehavior [17]. In their case study, the Byzantine UAVs move to nearby locations and intentionally transmit false information about the target's location, thereby misleading cooperative robots and causing the swarm to follow incorrect paths. Inspired by their approach, we modeled a similar behavior in our simulation. In our system, the Byzantine UAVs are programmed to move to nearby positions and deliberately send inaccurate data about the target's location to the cooperative UAVs, with the goal of disrupting the mission. In addition to this, our system incorporates a reputation system, which the Byzantine UAVs attempt to manipulate. Specifically, they collaborate by giving each other high ratings while giving low ratings to cooperative UAVs. This manipulation distorts the reputation scores, weakens trust in the group, and increases the impact of the faulty UAVs on the mission.

Aldaej et al. propose a reputation system in their Multidomain blockchain-based intelligent routing framework, utilizing Bayesian theory to estimate local reputation in UAV-IoT networks [1]. Their system ensures secure cross-domain routing by managing local and global reputations using blockchain, enhancing trust and security across multiple domains. Kamvar et al.'s EigenTrust algorithm similarly computes global reputation by aggregating local trust scores through transitive trust and eigenvector computations, aiming to reduce malicious activity in peer-to-peer networks [6]. While both approaches are effective in general UAV and peer-to-peer environments, our proposed reputation algorithm is more suitable for critical mission scenarios, especially those with limited time constraints. By integrating real-time feedback and Byzantine fault tolerance, our system ensures faster adaptation and more reliable performance, making it particularly robust in handling situations where rapid decision-making and timely coordination are essential for mission success.

Ogunbunmi et al. provide a comprehensive survey of reputation systems in UAV networks, categorizing them into centralized, decentralized, and hybrid models [10]. While their survey addresses various UAV applications, it does not specifically focus on the unique challenges of BVLOS operations and fully autonomous UAV teamwork in mission-critical, time-sensitive environments. Our approach differs by introducing an algorithm-based reputation system that incorporates real-time feedback and Byzantine fault tolerance, ensuring rapid decision-making and adaptability in BVLOS missions and fully autonomous operations. This makes our system particularly well-suited for critical mission scenarios, addressing gaps in the current literature by focusing on trust and reliability in dynamic contexts.

7 CONCLUSIONS

We introduced a reputation-based team formation mechanism (RTF) for enhancing the dependable critical fully autonomous multi-UAV BVLOS operations. Consequently, UAVs can securely evaluate each other's performance and form dependable teams capable of conducting BVLOS missions efficiently and safely. RTF mitigates risks associated with unreliable UAVs and fosters a competitive environment where UAV operators are incentivized to maintain high-performance standards to enhance their reputation scores. It has been shown that the proposed global reputation score algorithm significantly improves the reliability of UAV operations by dynamically adapting to the performances of UAVs over time.

Moreover, the simulation results corroborate the effectiveness of reputation-based team formation in enhancing mission efficiency and robustness against byzantine faults. These results highlight the potential of the proposed system to revolutionize UAV operations by enabling more trustworthy collaborations among diverse UAV entities. In conclusion, this paper not only advances the field of autonomous UAV operations but also sets the stage for future innovations that could fundamentally transform how UAVs are deployed and managed in complex operational environments.

References

[1] A. Aldaej, M. Atiquzzaman, T. A. Ahanger, and P. K. Shukla. Multi-domain blockchain-based intelligent routing in uav-iot networks. *Computer Communications*, 205:158–169, 2023. doi: 10.1016/j.comcom.2023.04.016. URL https://www.elsevier.com/locate/comcom.

[2] B. Buzcu, M. Özgün, Önder Gürcan, and R. Aydoğan. Fully autonomous trustworthy unmanned aerial vehicle teamwork: A research guideline using level 2 blockchain. *IEEE Robotics & Automation Magazine*, 2023.

[3] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, New York, 2nd edition, 2001.

[4] R. Gupta, A. Kumari, S. Tanwar, and N. Kumar. Blockchain-envisioned softwarized multi-swarming UAVs to tackle COVID-19 situations. *IEEE Network*, 35(2):160–167, 2021. doi: 10.1109/MNET.011.2000439.

[5] S. Gyawali, Y. Qian, and R. Hu. Machine learning and reputation based misbehavior detection in vehicular communication networks. *IEEE Transactions on Vehicular Technology*, 69(8):8871–8885, 2020.

[6] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web*, pages 640–651. ACM, 2003.

[7] D. Kucherov, S. Dolgikh, I. Miroshnichenko, and O. Poshyvailo. Protocol and logical models for robustness and survivability in autonomous uav systems. In *2023 13th International Conference on Dependable Systems, Services and Technologies, DESSERT 2023*, 2023. doi: 10.1109/DESSERT61349.2023.10416491.

[8] L. Lamport, R. Shostak, and M. Pease. The byzantine generals problem. *ACM Trans. Program. Lang. Syst.*, 4, 1982. doi: 10.1145/357172.357176.

[9] E. Meamari, H. Guo, C. Shen, and J. Hur. Collusion attacks on decentralized attribute-based encryption: analyses and a solution. *arXiv*, 2020.

[10] S. Ogunbunmi, Y. Chen, E. Blasch, and G. Chen. A survey on reputation systems for uav networks. *Drones*, 8(6):253, 2024. doi: 10.3390/drones8060253.

[11] E. Ordoukhanian and A. M. Madni. Engineering resilience into multi-uav systems. *Procedia Computer Science*, 153:9–16, 2019. doi: 10.1016/j.procs.2019.05.050.

[12] E. Politi, I. Panagiotopoulos, I. Varlamis, and G. Dimitrakopoulos. A survey of UAS technologies to enable beyond visual line of sight (BVLOS) operations. In *Proc. of the 7th International Conf. on Vehicle Technology and Intelligent Transport Systems*, pages 505–512, 2021. ISBN 978-989-758-513-5.

[13] K. Serefoglu, Önder Gürcan, and R. Aydogan. Uav marketplace simulation tool for bvlos operations. In Y. Vorobeychik, S. Das, and A. Nowé, editors, *24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025)*. International Foundation for Autonomous Agents and Multiagent Systems, 2025.

[14] M. Tahavori and A. Hasan. Fault recoverability for nonlinear systems with application to fault tolerant control of uavs. *Aerospace Science and Technology*, 107, 2020. doi: 10.1016/j.ast.2020.106282.

[15] M. A. Tahir, I. Mir, and T. U. Islam. A review of uav platforms for autonomous applications: Comprehensive analysis and future directions. *IEEE Access*, 11:52540–52554, 2023. doi: 10.1109/ACCESS.2023.3273780.

[16] D. Wanner, H. A. Hashim, S. Srivastava, and A. Steinhauer. Uav avionics safety, certification, accidents, redundancy, integrity, and reliability: a comprehensive review and future trends. *Drone Systems and Applications*, 12:1–23, 2024. doi: 10.1139/dsa-2023-0091.

[17] K. Wardega, M. von Hippel, R. Tron, C. Nita-Rotaru, and W. Li. Byzantine resilience at swarm scale: A decentralized blocklist protocol from inter-robot accusations. In *Proc. of the 22nd Int. Conf. on Aut. Agents and Mul. Systems (AAMAS)*, London, UK, May 29-June 2 2023.