Adversarial Data-Augmented Resilient Intrusion Detection System for Unmanned Aerial Vehicles

Muneeba Asif*, Mohammad Ashiqur Rahman*, Kemal Akkaya§, Hossain Shahriar‡, Alfredo Cuzzocrea¶
*Analytics for Cyber Defense (ACyD) Lab, Florida International University, Miami, FL, USA
§Advanced Wireless and Security (ADWISE) Lab, Florida International University, Miami, FL, USA
‡Center for Cybersecurity, University of West Florida, Pensacola, FL, USA
¶Big Data Engineering and Analytics (iDEA) Lab, University of Calabria, Arcavacata, Rende, Italy
masif004@fiu.edu, marahman@fiu.edu, kakkaya@fiu.edu, hshahriar@uwf.edu, alfredo.cuzzocrea@unical.it

***Abstract*—With the growing adoption of unmanned aerial vehicles (UAVs) across various domains, the security of their operations is paramount. UAVs, heavily dependent on GPS navigation, are at risk of jamming and spoofing cyberattacks, which can severely jeopardize their performance, safety, and mission integrity. Intrusion detection systems (IDSs) are typically employed as defense mechanisms, often leveraging traditional machine learning techniques. However, these IDSs are susceptible to adversarial attacks that exploit machine learning models by introducing input perturbations. In this work, we propose a novel IDS for UAVs to enhance resilience against such attacks using generative adversarial networks (GAN). We also comprehensively study several evasion-based adversarial attacks and utilize them to compare the performance of the proposed IDS with existing ones. The resilience is achieved by generating synthetic data based on the identified weak points in the IDS and incorporating these adversarial samples in the training process to regularize the learning. The evaluation results demonstrate that the proposed IDS is significantly robust against adversarial machine learning-based attacks compared to the state-of-the-art IDSs while maintaining a low false positive rate.**

***Index Terms*—Unmanned aerial vehicles, embedded systems, adversarial attacks, machine learning, generative adversarial networks, intrusion detection systems**

I. Introduction

Unmanned aerial vehicles (UAVs), also known as drones, have garnered significant global interest over the last decade [1] and have become an indispensable technology for various critical missions across civilian [2], commercial [3], and military [4] sectors. This increased success is owed to their successful operations in multidimensional fields such as relief provision in disaster-stricken areas [5], [6], search and rescue operations for missing persons [7], mobile data sensing and relaying in IoT environment [8], border and coastal surveillance for security purposes [9], construction/mining site monitoring [10], as well as in military warfare [11]–[13]. Owing to UAVs' versatility and diverse applications, maintaining end-to-end security and safety and ensuring resiliency in operations is paramount for their effective and precise mission execution. The implications of security breaches in UAVs can range from jeopardizing personal privacy and national interests to enabling industrial espionage, leaking sensitive information, and threatening mission integrity. Ensuring the safety of UAV operations is paramount, and security measures should reinforce rather than undermine this safety. Like many autonomous vehicles, UAVs rely heavily on global positioning system (GPS)-based navigation and control systems. This reliance increases their vulnerability to cyberattacks like GPS jamming and spoofing [14], [15]. GPS jamming disrupts the signals to the UAV, causing a Denial-of-Service (DoS) attack that impedes accurate location determination [16]. In contrast, GPS spoofing transmits false data to mislead the UAV, altering its trajectory [17]. These attacks can significantly compromise the UAV's performance, safety, and reliability, potentially leading to control loss, in-flight collisions, or unauthorized access to sensitive data. Given their severity and increasing occurrence, addressing these threats provides a tangible context to study the robustness of UAV intrusion detection systems.

Extensive research investigations have been conducted to defend against sensor attacks like GPS spoofing and jamming. Some common defense techniques include data fusion and redundancy [18] and cryptography [19]. Others involve signal processing [20], and probabilistic analyses [21]. The most common defense technique is to deploy an intrusion detection system (IDS) to identify and mitigate these attacks. These IDSs often rely on traditional machine learning techniques such as one-class classification [22], [23], where the benign space is learned as the normal and anything outside that space is classified as an anomaly/intrusion, or multi-class classifications, where some of the attack parameters are also known, and different attacks can be categorized. Catering to the non-linear relationship between the features of UAVs, most of the IDS models use neural networks to attain an adequate mapping between the features. Nevertheless, despite their effectiveness in detecting and mitigating attacks, these neural-network-based IDS models remain vulnerable to adversarial attacks [24], [25], designed to exploit models' weaknesses by introducing carefully crafted input perturbations.

Fig. 1 illustrates how adversarial machine learning can exploit IDS vulnerabilities to launch undetected GPS spoofing attacks on UAVs. Attackers can create benign-looking adversarial samples and bypass the IDS, leading to trajectory deviation and potential mission compromise. The ground control station sends precise GPS commands to the UAV, while a

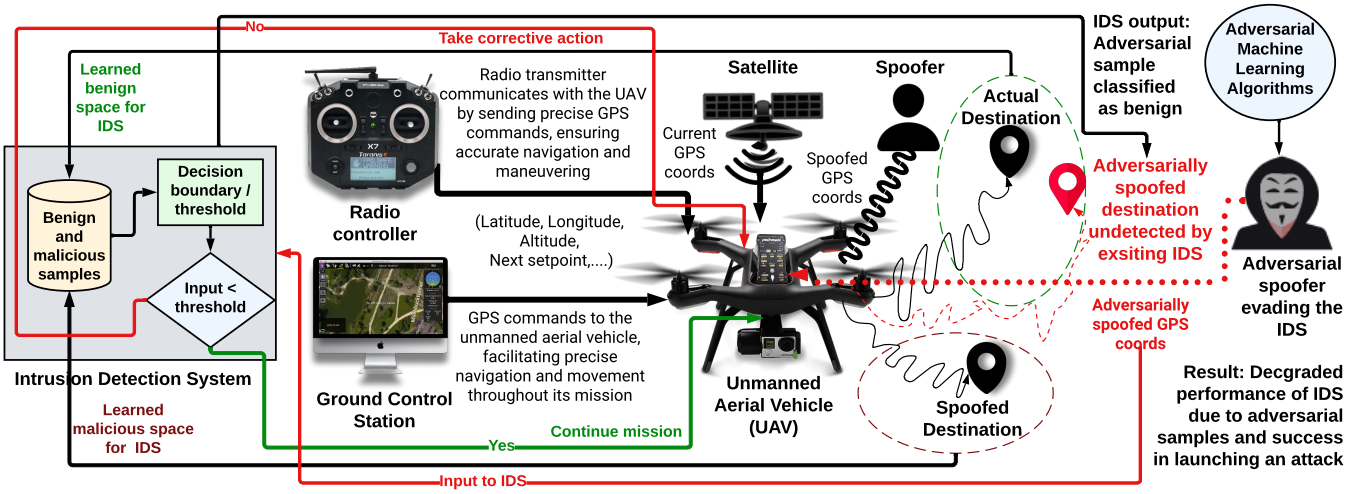


Fig. 1. Impact of Adversarial Machine Learning on IDS in UAV GPS Spoofing: Adversarial machine learning algorithms can generate spoofed GPS coordinates that evade the IDS, leading to undetected attacks and degraded IDS performance in UAV missions.

satellite spoofer generates falsified GPS coordinates to mislead the drone. Owing to the stealthy nature of the attack, the IDS fails to recognize the malicious intent behind the adversarial samples and classifies them as benign, allowing the UAV to continue its mission based on the spoofed coordinates. This highlights the need for more robust and resilient IDS models to counter adversarial attacks and ensure the security and reliability of UAVs.

In this paper, we study the impact of several evasion-based adversarial attacks, like the fast gradient sign method (FGSM) and projected gradient descent (PGD) attack, on the performance of an IDS that secures UAVs against such attacks. Here, we present a novel approach to enhance the performance of IDSs for UAVs by leveraging adversarial learning techniques. By addressing vulnerabilities highlighted by these adversarial attacks, we employ generative adversarial networks (GANs) to generate synthetic data to address these weak points, thus augmenting the resilience of the IDS against attacks like GPS spoofing and jamming. Moreover, we add the adversarial samples as a regularizer in our training process for further augmented resilience. Through extensive experimentation and evaluation, we demonstrate that our proposed approach augments the robustness of IDS models against adversarial attacks. For this, we explore several key research questions. We examine whether our framework can enhance IDS resilience in UAVs against diverse adversarial attacks. We also study if adversarial learning and GAN data can improve this resilience. Additionally, we evaluate different IDS models' performance by analyzing MSE distributions, investigating adversarial augmentation impacts, and studying model stability and sensitivity. Overall, our contributions are threefold:

- We provide an in-depth analysis of the vulnerabilities of existing IDS models for UAVs against adversarial attacks.
- We demonstrate the application of GAN and adding adversarial samples as a regularizer in improving the resilience of the existing IDS by augmenting data for weak points highlighted by adversarial learning.
- Finally, we evaluate the proposed methodology's effectiveness in enhancing detection performance while maintaining low false positives for GPS attacks.

The rest of the paper is organized as follows: Section II discusses existing works, Section III summarizes GPS navigation and adversarial learning-based defenses. Section IV discusses threat model. Section V analyzes impact of adversarial attacks on existing IDSs. Sections VI, VII, and VIII describe model architecture, proposed defense framework, and experimental validation, respectively. Finally, Section IX concludes the paper.

II. Related Work

This section reviews state-of-the-art defenses for UAVs, focusing first on adversarial learning-based approaches and then on existing IDSs for UAVs.

A. Adversarial Learning-based Defenses for UAVs

UAVs' reliance on deep learning makes them vulnerable to adversarial attacks. Thus, robust defenses are crucial for reliable and secure UAV operations. Tian et al. explore the domain of adversarial attacks and defenses in UAVs employing deep learning. They introduce an adversarial training approach to bolster the UAVs' resistance to such attacks while preserving high classification accuracy [26]. Furthermore, Hu et al. introduce a secure estimation algorithm to counter adversarial cyber attacks on UAVs. They establish a mathematical attack model and devise a defense technique ensuring accurate state estimation of the UAV amidst adversarial interference, thereby enhancing its resilience and guaranteeing safe operation [27]. Raja et al. delve into adversarial attacks and defenses for AI-driven UAV infrastructure inspections. They suggest a defense mechanism based on adversarial training, enhancing the AI model's robustness in infrastructure checks and improving

the security of UAV operations in these settings [28]. Moreover, Doyle et al. study the vulnerabilities of UAVs through adversarial learning. They analyze the impact of adversarial attacks on UAVs and propose defense strategies to mitigate these threats, emphasizing the importance of developing more robust and resilient UAVs [29]. Also, McCloskey employs GANs to enrich UAV image classification datasets, showing that this augments classification performance through high-quality synthetic samples. [30]. Furthermore, Guptha et al. propose a GAN-based approach for object detection in UAVs, incorporating fusion technology, which improves the object detection performance of UAVs, making them more robust and reliable for various applications [31].

B. Intrusion Detection Systems for UAVs

As UAVs become integral to various sectors, safeguarding them against cyber threats is crucial. IDSs play a vital role by detecting anomalies like GPS spoofing and jamming in network activity. Basan et al. suggest using entropy to detect changes in traffic patterns caused by DoS attacks on nearby UAVs by training a neural network to detect attacks on neighboring, rather than directly targeted, UAVs [32]. Al-Haija et al. employ UAV-IDS-ConvNet, a deep convolutional neural network, for UAV intrusion detection. They analyze encrypted Wi-Fi traffic from Parrot Bebop, DBPower UDI, and DJI Spark models [33]. Praveena et al. introduce a technique using deep reinforcement learning and the Black Widow Optimization (DRL-BWO) algorithm for UAV network security. This method also integrates an enhanced Deep Belief Network (DBN) for intrusion detection [34]. Recognizing the lack of consistent datasets for UAV IDSs, Whelan et al. introduce MAVIDS, a novelty-based one-class classification system. MAVIDS interfaces with the flight controller to deploy measures like disabling sensors and is tested against GPS attacks [22]. To defend UAVs against cyber-attacks, Mitchell and Chen suggest a specification-based detection method in [35]. This study uses a behavior rule-based UAV-IDS built on defined threat models for various attacks. It aims to optimize UAV security and performance by reducing false positives and negatives. While existing research has examined methods for securing UAVs against sensor attacks such as GPS spoofing and jamming, there has been little investigation into the impact of adversarial attacks on these defenses, nor on the robustness of existing IDSs against such attacks. Our research seeks to address this gap by investigating the impact of adversarial attacks on existing IDSs and identifying their weak points. We then propose a novel approach to augment the training on these weak points to strengthen the IDS and provide a more comprehensive defense.

III. Background

In this section, we discuss GPS-navigation system, adversarial machine learning in UAVs, IDSs, and our motivation.

A. GPS-based Navigation and Flight Control in UAVs

UAVs use GPS navigation for accurate, real-time positioning, velocity, and altitude data. UAVs pinpoint their location in 3D space by connecting to multiple GPS satellites, enabling trajectory adjustments and consistent mission course. The flight controller, a critical component of the UAV navigation system, processes GPS data and other sensor information to regulate the UAV's movements. This controller comprises various algorithms and control loops, ensuring the UAV maintains its desired position, altitude, and orientation. Typically utilizing a proportional-integral-derivative (PID) controller, the position control loop adjusts the desired velocity based on the current and target positions. The velocity control loop modifies UAV thrust to match this desired velocity, leveraging a PID controller to calculate and correct any velocity error. Control loops process GPS and sensor data to maintain the UAV's desired trajectory. The position loop directs the UAV's path, while the velocity loop adjusts speed and responds to external factors like wind/obstacles. This combination ensures accurate and reliable navigation for UAVs.

B. Adversarial Learning in IDS-based Defense for UAVs

By leveraging model vulnerabilities, adversaries can use input perturbations (δ_{GPS}) that lead to misclassification, allowing them to bypass the IDS undetected. This is depicted as adding (δ_{GPS}) to the original signal (GPS_{orig}) in Eq. 1.

$$\text{GPS}_{\text{adv}} = \text{GPS}_{\text{orig}} + \delta_{\text{GPS}} \tag{1}$$

Adversarial attacks can subtly alter GPS signals, leading IDS to mislabel benign signals as threats (false positives) or overlook actual threats (false negatives), jeopardizing UAV safety. Attackers craft these signals by optimizing perturbations (δ_{GPS}), as illustrated in Eq. 2, where ε_{GPS} denotes the maximum perturbation limit.

$$\begin{aligned} &\underset{\delta_{\text{GPS}}}{\text{minimize}} \quad \|\delta_{\text{GPS}}\| \\ &\text{subject to:} \quad \text{IDS}(\text{GPS}_{\text{orig}} + \delta_{\text{GPS}}) \neq \text{IDS}(\text{GPS}_{\text{orig}}), \\ &\quad \|\delta_{\text{GPS}}\| \leq \varepsilon_{\text{GPS}}. \end{aligned} \tag{2}$$

A predefined threshold limits the magnitude of adversarial perturbation (δ_{GPS}) applied to the original GPS sample (GPSorig) to form the adversarial GPS sample (GPSadv). This constraint ensures the adversarial perturbation is subtle, evading IDS detection but causing misclassification. The goal is to find an optimal perturbation within this limit to bypass the IDS effectively. Misclassifying benign signals as attacks or vice versa can have severe consequences. If benign signals are labeled as attacks, they can disrupt or disable the UAV's navigation or control system, resulting in crashes or loss of control. False positives can also trigger unnecessary countermeasures, causing operational disruptions and potential harm. If malicious signals are deemed benign, the IDS might not react, enabling undetected attacks that jeopardize safety and potentially compromise safety and security.

C. Research Motivation

Given the expanding role of UAVs in various sectors, securing their GPS-based navigation systems is crucial. While

IDSs have been deployed to safeguard UAVs against cyber-attacks, these systems are still susceptible to adversarial attacks targeting the underlying machine learning models used for detection [36]. Ensuring the security and reliability of UAV navigation systems is critical, and developing robust defense mechanisms is essential for their safety and effectiveness across diverse applications and industries. Adversarial attacks on GPS samples can lead to misclassifications within the system's intrusion detection mechanisms. Such misclassifications could result in false alarms or, more critically, failure to detect actual GPS spoofing or jamming attacks. Consequently, the UAV might rely on tampered GPS data, leading to incorrect navigation decisions, deviation from the intended flight path, or even mission failure. In more extreme cases, adversarial manipulation of GPS data might cause the UAV to collide with obstacles, enter restricted airspace, or suffer damage due to improper altitude and velocity adjustments. Our primary objective is to explore how adversarial attacks impact UAV IDSs that rely on GPS data. We delve deep into the consequences these adversarial attacks pose on both the GPS data and the functioning of the UAV IDS by analyzing the resulting misclassifications, navigation errors, and mission failures to identify vulnerabilities and propose countermeasures.

IV. Threat Model

This section outlines the threat model for adversarial attacks on UAV data and IDSs, detailing adversarial goals, capabilities, and attack methods.

A. Adversarial Goals

The primary goals of adversaries targeting UAVs with GPS spoofing and jamming attacks in the presence of an IDS are as follows:

- Compromise the integrity of the UAV's navigation system by injecting false GPS data, leading to incorrect navigation decisions and deviations from the intended path.
- Evade detection by the UAV's IDS, allowing the attacker to remain undetected and potentially cause harm to the UAV or its mission.
- Disrupt the UAV's mission, causing delays, loss of valuable data, or even mission failure.

B. Adversarial Capabilities

The adversaries are assumed to possess the following capabilities:

- Access to specialized hardware and software tools that enable them to generate and transmit GPS spoofing and jamming signals.
- Knowledge of the UAV's GPS-based navigation system and the machine learning models used by the IDS.
- Ability to create adversarial samples by exploiting the vulnerabilities in the IDS's machine learning model to bypass the detection mechanisms.

Algorithm 1: Adversarial Attack Impact Analysis

Input: Model M, Dataset D, Epsilons ϵ
Output: Accuracies for FGSM and PGD attacks
Initialize attacks: FGSM and PGD
for *each attack* $\mathcal{A}$ *in attacks* **do**
  for *each* ϵ *in epsilons* **do**
    Generate adversarial samples X_{adv} using attack $\mathcal{A}$ with the given ϵ: $X_{adv} \leftarrow \mathcal{A}(M, D, \epsilon)$
    Calculate the model output for the adversarial samples: $O \leftarrow M(X_{adv})$
    Compute mean squared error (MSE) between original and adversarial samples: $\text{MSE}(i) \leftarrow \frac{1}{N}\sum_{n=1}^{N}(D_i^n - O_i^n)^2$
    Update labels based on the computed MSE threshold: $y_{adv_i} \leftarrow \begin{cases} 1, & \text{if MSE}(i) > \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
    Calculate the accuracy of the model on the adversarial dataset: $\text{Acc} \leftarrow \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(y_{adv_i} = y_i)$
    Save and plot the accuracy for the current attack $\mathcal{A}$ and ϵ
  end
end

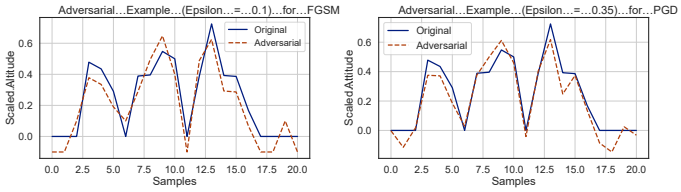
C. Attack Methodology

Adversaries target UAVs using GPS spoofing and jamming, leveraging knowledge of the target system, its GPS navigation, intrusion detection systems, and machine learning models. Using this information, they craft adversarial samples to exploit IDS vulnerabilities, causing misclassifications or bypassing detection. Upon creating adversarial samples, the attacker launches an assault on the UAV's navigation system, leveraging GPS spoofing or jamming signals to undermine its IDS. The attack's success is continuously monitored, and the strategy is adjusted accordingly, involving new adversarial samples or attack methodology alterations. The attacks carried out in this paper are the fast gradient sign method (FGSM) and projected gradient descent (PGD) attack.

FGSM is a white-box attack that creates adversarial samples by perturbing input data based on the loss function's gradient. Essentially, it linearizes the loss and utilizes gradients for maximization. An adversarial sample x_{adv} is generated by adding a small perturbation to the original input x. This perturbation is determined by the sign of the gradient of the loss function J concerning the input x and the true label of the input y_{true}. A small constant ϵ controls the magnitude of the perturbation. FGSM is given by Eq. 3. Here, x_{adv} denotes the adversarial sample, x is the original input, and ϵ limits the perturbation's size. The term $\text{sign}(\nabla_x J(x, y_{true}))$ captures the gradient sign of the loss function J relative to input x, with y_{true} being x's true label.

$$x_{adv} = x + \epsilon \cdot \text{sign}(\nabla_x J(x, y_{true})) \tag{3}$$

PGD is an iterative variant of FGSM and is considered to be more powerful. It performs multiple FGSM-like steps with a smaller step size, and after each step, it projects the perturbed input back into a defined ϵ-ball around the original input.



(a) (b)

Fig. 2. Comparative analysis of the original and adversarial samples generated under (a) fast gradient sign method (FGSM) attack with epsilon = 0.1 and (b) projected gradient descent (PGD) attack with epsilon = 0.35.

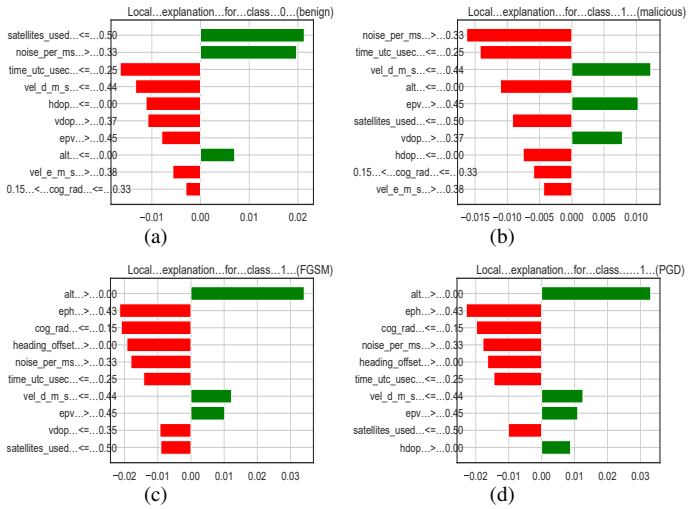
The goal is to optimize the loss within this epsilon-bound. An iterative process generates an adversarial sample in the PGD attack. At each iteration t, the adversarial sample x_{adv}^{t} is updated by adding a perturbation proportional to the sign of the gradient of the loss function J concerning the input x_{adv}^{t} and the true label of the input y_{true}. This perturbation is then projected onto the ϵ-ball around the original input x. A constant α controls the step size for each iteration.

V. Case Study: Analyzing Adversarial Attacks on Autoencoder One-Class Classifier

This section evaluates how an autoencoder classifier identifies GPS attacks on UAVs when faced with FGSM and PGD adversarial challenges, particularly at varying epsilon perturbation levels.

A. Adversarial Example Generation

We generate adversarial samples using the FGSM and PGD attacks with varying epsilon values. To better understand the robustness of the model, we also explore the classifier's response to minute perturbations that closely mimic normal data variations, providing a more comprehensive evaluation of its resilience. We then assess the classifier's performance against each epsilon. This provides insights into the relationship between the magnitude of the adversarial perturbations and the degradation in the classifier's performance. The adversarial samples are shown in Algorithm 1. These adversarial samples significantly impair the IDS's classification accuracy despite their proximity to the original data in the feature space. Specifically, under an FGSM attack with an epsilon parameter set to 0.1, the IDS's accuracy substantially dropped from an initial 93.8% to a mere 73.46%. Similarly, under a PGD attack with an epsilon of 0.35, the accuracy decreased from 93.8% to 77.1%. This is visualized in Fig. 2(a) and 2(b), respectively. This highlights the IDS's vulnerability and sensitivity to adversarial perturbations, where minor alterations can confuse the IDS into misclassifying the data. This emphasizes the importance of incorporating adversarial training or other defensive measures into the IDS to mitigate the risks posed by adversarial attacks. As the epsilon value increases, the adversarial perturbations become more noticeable, but their ability to reduce the accuracy of the IDS also increases. Understanding this trade-off is vital for creating potent adversarial attacks and building strong defenses.

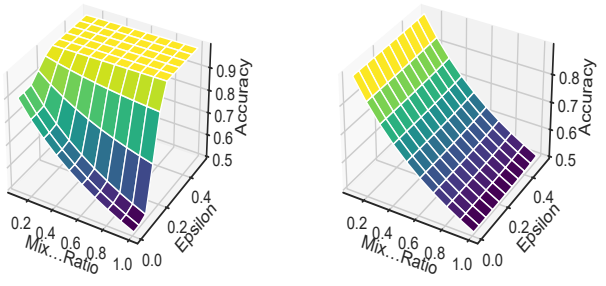


(a) (b) (c) (d)

Fig. 3. Feature Importance Visualization using LIME: (a) Clean Explanation - Class 0 (Benign), (b) Clean Explanation - Class 1 (Malicious), (c) Adversarial Explanation - Class 1 (FGSM), and (d) Adversarial Explanation - Class 1 (PGD). This figure presents the primary features influencing the model's predictions, as obtained using the LIME explanation method.

B. Feature Importance Analysis

This part of the case study examines feature importance within the model's decision-making process. This is achieved by the local interpretable model-agnostic explanations (LIME) method, which provides interpretability by revealing influential features in individual predictions, allowing the identification of key factors. Green indicates features supporting a prediction, while red shows features countering it. Comparing feature importance between clean and adversarial instances helps detect significant shifts and potential vulnerabilities that adversarial attacks can exploit. This can reveal how adversarial attacks might influence or shift the classifier's attention toward or away from specific features. The primary features contributing to the decision-making process for benign instances of the clean data are shown in Fig. 3(a), where, as visualized the most important features are *satellites_used* and *noise per ms*. For malicious instances in the clean dataset, as presented in Fig. 3(b), features like *noise_per_ms* and *time_utc_usec* are more important. These characteristics serve as pivotal decision-makers for the autoencoder classifier. Furthermore, for adversarial data samples under FGSM attack, the distinguishing features as depicted in Fig. 3(c), are *alt* and *eph*. Similarly, adversarial samples under PGD attack as inferred from Fig. 3(d), important features are *alt* and *eph*. The comparison of benign, malicious, and adversarial data reveals significant differences in the importance of certain features, as seen in Fig. 3. This suggests that attackers might target these varying feature significance. For instance, if an adversary recognizes that the IDS gives undue importance to specific features under adversarial conditions, they can craft inputs that manipulate these features, effectively diverting the classifier's attention and making genuine malicious activities harder to detect. By understanding key features, adversaries can refine attacks using methods like PGD for greater impact. It's crucial



(a) (b)

Fig. 4. Comparative Analysis of Classifier Accuracy under (a) FGSM and (b) PGD adversarial attacks, with varying perturbation strength (epsilon) and different proportions of adversarial samples (mixed ratio).

to build models that can resist these tactics, especially in security-focused applications like UAV IDS.

C. Impact of Mixed Ratios and Epsilon

To delve deeper into the impact of FGSM and PGD attacks on our classifier, we tested its performance on datasets with different proportions of clean and adversarial samples. By analyzing accuracy across varied mixed ratios and perturbation strengths (epsilon), we gauged how the classifier responds to different adversarial threat levels. This study is vital to grasp the IDS's resilience and pinpoint which adversarial conditions most affect its accuracy.

Under an FGSM attack, for lower epsilon values (weaker adversarial perturbations), the model's accuracy remains relatively high even as the mixed ratio increases. The model's accuracy decreases as the epsilon value increases, particularly for higher mixed ratios. Under a PGD attack, the model's accuracy is more sensitive to epsilon changes in the PGD case. The accuracy decreases more drastically as the epsilon value increases, particularly for higher mixed ratios. The model performs worse under a PGD attack than an FGSM attack. The performance of the autoencoder-based on-class classifier model under FGSM and PGD attacks with variation in mixed ratio and epsilon can be seen in Fig. 4(a) and 4(b), respectively. The classifier's performance is generally worse under PGD attacks compared to FGSM attacks. This is likely due to the iterative nature of PGD attacks, which generate more diverse and challenging adversarial samples. Fig. 4 shows the classifier's decreasing accuracy with rising epsilon values and mixed ratios, especially under PGD attacks compared to FGSM. This underscores the importance of robust adversarial training, especially against iterative attacks like PGD.

VI. Model Architecture and Dataset for IDS

We employ a deep learning autoencoder for one-class classification in an IDS to discern GPS spoofing and jamming attacks by learning standard GPS data patterns.

A. Model Architecture

The autoencoder model consists of two primary components: the encoder and the decoder. The encoder maps the input GPS data (x) into a lower-dimensional latent space representation (z) as can be seen in Eq. 4. At the same time, the decoder reconstructs the original input data from the latent space representation as shown in Eq. 5. The autoencoder is trained to minimize the difference between the input and reconstructed data, thereby learning the normal behavior.

$$\text{Encoder:} \quad z = f_\theta(x) \tag{4}$$

$$\text{Decoder:} \quad x' = g_\phi(z) \tag{5}$$

In our autoencoder framework, we use the encoder f_θ and decoder g_ϕ, parametrized by θ and ϕ. The goal is to reduce the reconstruction error to improve output fidelity to the input. We utilize a deep feed-forward neural network in our architecture, designed to learn the normal behavior of GPS data. Normal GPS data refers to the patterns we observe when a UAV operates without external malicious influence, while abnormal data indicates potential tampering or adversarial attacks. The input layer has 21 neurons, corresponding to the UAV data features, followed by three dense layers with 40, 20, and 40 neurons, all using the rectified linear unit (ReLU) activation function. The output layer uses 21 neurons and a linear activation to reconstruct input and minimize error, capturing the UAV data's essence.

Exploitation of autoencoder-based IDS by adversarial learning: Adversarial learning generates perturbed versions of input called adversarial samples to manipulate the model's outputs. Techniques like FGSM can be used to craft adversarial samples that exploit the model's architecture, leading to compromised performance. Our autoencoder-based IDS for UAVs can be vulnerable to adversarial learning in two ways. Attackers can generate adversarial samples in the input space, deceiving the autoencoder into reconstructing false data as normal UAV data, misclassifying attacks as benign, and allowing attackers to bypass defense mechanisms. This approach targets the input layer and the encoder function, exploiting their susceptibility to slight perturbations in the input data. Secondly, an attacker could generate adversarial samples in the latent space representation (z), inducing a large reconstruction error in the autoencoder. This tactic would cause the IDS to misclassify benign data as an attack, leading to false alarms and potentially undermining the system's credibility. By exploiting the decoder function and the reconstruction process, the attacker can manipulate the autoencoder's internal representations and mislead the UAV while disrupting the IDS's ability to detect attacks accurately.

B. UAV Dataset for IDS Development and Evaluation

We utilize the UAV Attack Dataset [37], an open-access dataset, to detect GPS spoofing and jamming attacks on UAVs, investigate the impact of adversarial attacks on the IDS, and finally, evaluate our proposed framework. This dataset is a comprehensive collection of flight logs designed for studying GPS spoofing and jamming attacks on UAVs. This open-access dataset includes data from benign flights and flights with GPS interference. It provides various features, such as

latitude, longitude, velocity, heading, and GPS quality indicators. It enables researchers to develop and evaluate IDSs specifically tailored for UAVs, offering valuable insights into UAV behavior under different flight conditions.

From the wide array of features in the dataset, we chose features such as *evh, time_utc_usec, lat, lon, heading, z_deriv, vz, ax, hdop, vel_m_s, q[2], jamming_indicator, vel_e_m_s, and noise_per_ms* for training our autoencoder-based IDS. These features cover UAV flight dynamics, GPS signal quality, potential interference indicators, and telemetry data, comprehensively representing the UAV's state during flight. The autoencoder detects GPS spoofing or jamming patterns by observing an increased reconstruction error. This higher error enables the one-class classification-based IDS to accurately differentiate between normal and anomalous data, offering a reliable detection mechanism for GPS interference. We merged sensor data based on timestamps during data preprocessing and handled missing values using linear interpolation. The remaining gaps were filled using forward and backward filling. We removed irrelevant columns and assigned labels to different flight types. Selected features were standardized for consistent input to the autoencoder-based IDS. The preprocessed data were used to develop and evaluate the IDS, specifically targeting GPS spoofing and jamming detection. After selecting the relevant features for our analysis, we applied the *MinMaxScaler* to the data, excluding the *'Label'* column. The scaled data was split into training and testing sets, using a fixed random state for result reproducibility. This ensured a consistent input format for the model, enabling accurate evaluation of its performance in detecting GPS spoofing and jamming attacks on UAVs.

VII. Proposed GAN and Adversarial Samples-based Defense

The framework enhances UAV IDS robustness by generating data with InfoGAN/ WGAN, training autoencoders, and using adversarial samples as model regularizers, as follows.

A. Generation of New Data Points

Let $\mathcal{D}$ be the original dataset consisting of input samples $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$, where each $\mathbf{x}_i \in \mathbb{R}^d$ represents a feature vector, and $\mathbf{Y} = y_1, y_2$ denotes the corresponding benign or malicious labels. Let M be the autoencoder model, which reconstructs the input samples $\mathbf{X}$ by learning a latent representation space. The model learning parameter is the mean squared error (MSE) between the original and reconstructed input. Let T be the threshold for the autoencoder-based one-class classifier to identify the anomaly points. To generate new data points for these points, we train a combination of InfoGAN and WGAN, consisting of a generator G and a discriminator D, using the original dataset $\mathcal{D}$. The generator generates synthetic data points resembling the original data, while the discriminator differentiates between the original and generated data. We compute the MSE for each input sample $\mathbf{x}_i$ using the autoencoder model M and identify the weak points as samples with MSE close to or above the threshold T.

Algorithm 2: Proposed Defense Methodology

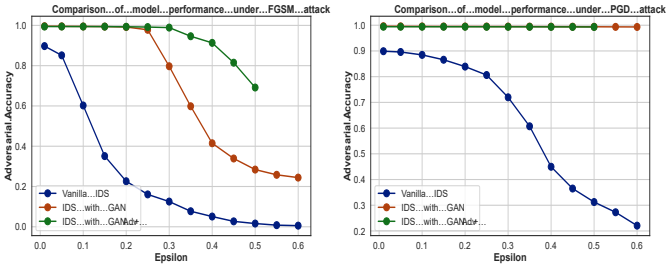
Input: Original data $\mathcal{D}$, Autoencoder model M
Output: Improved autoencoder model M_{improved}
Train InfoGAN with WGAN loss using $\mathcal{D}$ and M;
Generate new data points using InfoGAN;
Combine $\mathcal{D}$ and generated data;
Split data into training set $\mathcal{D}_{\text{train}}$ and validation set $\mathcal{D}_{\text{val}}$;
Train WGAN with gradient penalty using $\mathcal{D}_{\text{train}}$;
for $i \leftarrow 1$ *to* N **do**
  Update the WGAN discriminator using real and fake samples;
  Update the WGAN generator using the discriminator's feedback;
  Train M using reconstructed samples of $\mathcal{D}_{\text{train}}$;
  Evaluate M on $\mathcal{D}_{\text{val}}$;
  if *performance of* M *on* $\mathcal{D}_{val}$ *improves* **then**
    Update M_{improved} with M;
  end
  else
    Break the loop;
  end
end
Output: Improved autoencoder model M_{improved}

B. Training the Autoencoder with the Generated Data Points

In this step, we retrain the autoencoder model using the generated data points to improve its understanding of the data distribution, especially in the weak areas of the input space. We combine the original dataset $\mathcal{D}$ with the generated data points. The combined dataset is then split into training and validation sets, denoted as $\mathcal{D}$train and $\mathcal{D}$val, respectively. We initialize the autoencoder model M_{improved}. Next, we train M_{improved} using the combined dataset $\mathcal{D}_{\text{train}}$, encouraging the model to minimize both the reconstruction error (MSE) and the adversarial perturbations present in the generated data until the performance of the model on the validation set stops improving. We evaluate M_{improved} on the validation set $\mathcal{D}_{\text{val}}$ to monitor its performance.

C. Incorporating the Adversarial Samples for Regularization

The IDS aims to learn a reliable representation of benign data to detect anomalies. Adversarial samples exploit weaknesses in the learned representation. Using adversarial samples as a regularizer involves including them in training to enhance the autoencoder's robustness. The model learns a more resilient representation by minimizing the reconstruction error for both benign data and adversarial samples. This incorporation helps prioritize the aspects of data that are less susceptible to manipulation or perturbations. Let $\mathcal{X}$ be the input space consisting of input samples $\mathbf{x}i \in \mathcal{X}$, and let $\mathbf{X}$adv denote the set of adversarial samples generated from $\mathcal{X}$. The autoencoder model M aims to reconstruct the input samples $\mathbf{X}$ by learning a latent representation space. The autoencoder is trained using a modified optimization objective, incorporating



(a) (b)

Fig. 5. Adversarial accuracy of all three IDS models against (a) FGSM and (b) PGD attacks at various epsilon values, demonstrating an increase in model resilience with GAN and adversarial augmentation.

a reconstruction loss term, comparing original input samples with their reconstructions by incorporating the adversarial samples as a regularizer shown in Eq. 6. To enhance robustness, a regularization term based on adversarial samples is introduced. By penalizing adversarial reconstructions, the model becomes less perturbation-sensitive and more resilient.

$$\min_{M} \frac{1}{N} \sum_{i=1}^{N} \text{ReconstructionLoss}(\mathbf{x}_i, M(\mathbf{x}_i)) + \lambda \cdot \text{RegularizationTerm}(\mathbf{X}_{\text{adv}}, M) \tag{6}$$

The ReconstructionLoss$(\mathbf{x}_i, M(\mathbf{x}i))$ is discrepancy between original sample $\mathbf{x}i$ and autoencoder's output. $\mathbf{X}$adv is the adversarial sample set, while RegularizationTerm$(\mathbf{X}$adv$, M)$ evaluates the autoencoder's response to perturbations.

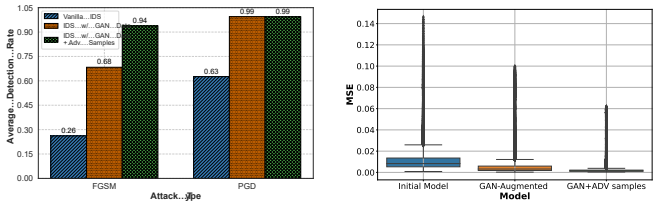
VIII. Evaluation

Our research aimed to evaluate the resilience of a UAV IDS using GANs and adversarial learning. The goal was to maintain high performance under all scenarios.

A. Research Questions and Evaluation Metrics

Our framework involves augmenting the resilience of IDS against adversarial attacks by incorporating data from GANs and adversarial learning. Hence, we have three IDS models to evaluate the efficacy of our framework: the baseline IDS (which we refer to as the vanilla IDS), the IDS with GAN-data augmentation, and finally, the IDS incorporating adversarial samples in the learning. This leads us to investigate the following research questions (RQs). To experimentally validate the performance of our framework and to answer these questions, we make use of the following well-known metrics: (1) *Accuracy*, the proportion of correctly classified samples to the total number of samples; (2) *False Positive Rate*, the proportion of negative samples incorrectly classified as positive to the total number of negative samples; (3) *Mean Squared Error (MSE)*, the average of the squared differences between the predicted values and the true values; and (4) *R2 score*, the proportion of the variance in the dependent variable that the independent variables can explain.

RQ1: Can the proposed framework make the existing IDS for UAVs resilient against different adversarial attacks?



(a) (b)

Fig. 6. Comparative (a) detection rates under FGSM and PGD attacks, (b) MSE values across all three models, demonstrating enhanced model fit and resilience against adversarial attacks with the integration of GAN and adversarial learning.

RQ2: Is the integration of adversarial learning as a regularization measure necessary to enhance the resilience of IDS against adversarial attacks beyond what GAN data augmentation alone achieves?

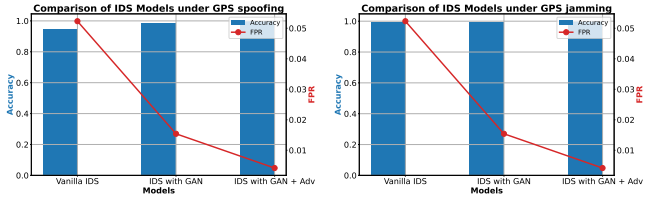
RQ3: How do the distribution and central tendency of MSE differ among the three models, and what does this reveal about their comparative performance in the context of IDSs?

RQ4: How does adversarial augmentation affect the IDS's performance against actual GPS spoofing/jamming attacks?

RQ5: How do the stability and sensitivity of the models vary with different perturbation magnitudes?

B. Results and Discussion

To answer **RQ1**, we tested all three IDS models under two adversarial attacks, FGSM and PGD, and recorded the adversarial accuracy for a range of epsilon values. As shown in Fig. 5(a) and Fig. 5(b), the adversarial accuracy for the initial IDS model dropped significantly with the increase in epsilon values for both the FGSM and PGD attacks. The accuracy dipped to 0.016042 for FGSM and 0.220658 for PGD attacks at epsilon values of 0.50 and 0.60, respectively. Upon augmentation with GAN, there was a substantial increase in the IDS model's resilience against both FGSM and PGD attacks. The adversarial accuracy of the GAN-augmented IDS model remained well above 0.99 for both attack methods for epsilon values up to 0.25. However, the adversarial accuracy declined more noticeably for higher epsilon values, particularly for FGSM attacks, falling to 0.244198 and 0.993354 at an epsilon of 0.60 for FGSM and PGD attacks. Furthermore, the adversarially regularized IDS model improved resilience against adversarial attacks. The adversarial accuracy remained notably stable across attack methods and all epsilon values, with a minimum accuracy of 0.691458 for FGSM and 0.993262 for PGD attacks at an epsilon of 0.50. These results show that the proposed framework enhances the resilience of the existing IDS for UAVs against adversarial attacks. The GAN augmentation significantly boosts the adversarial accuracy of the model. Still, including adversarial samples as a regularization measure provides the most comprehensive resilience, maintaining high accuracy even against stronger adversarial attacks. Regarding **RQ2**, we wanted to determine if integrating adversarial learning was necessary to enhance the IDS resilience against adversarial attacks. The results seem to support this hypothesis. As seen in Fig. 6(a), for the vanilla



(a) (b)

Fig. 7. Comparative performance analysis in (a) GPS spoofing and (b) jamming attacks detection. Plots highlight significant improvements in accuracy and reductions in false positives by integrating GAN and adversarial learning.

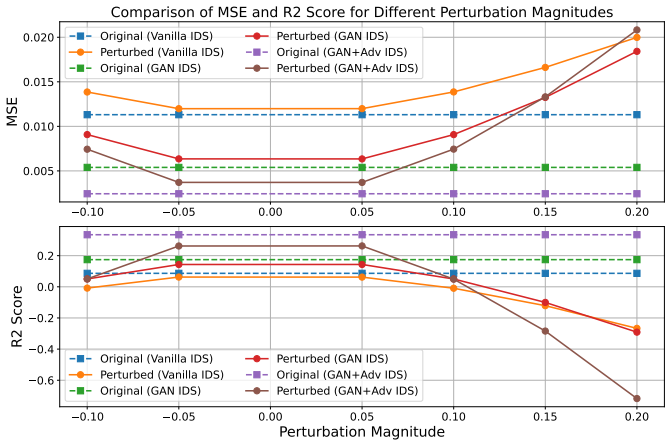


Fig. 8. Sensitivity analysis of all three models under different perturbation magnitudes. Improved stability and resilience of the augmented and adversarially regularized models is seen as compared to the baseline model.

IDS, the adversarial accuracy was notably low with FGSM (0.261409) and somewhat better with PGD (0.626006). For the GAN-augmented IDS, there was a significant improvement in adversarial accuracy. GAN data augmentation alone led to a detection rate of 0.994800 for PGD attacks, a significant improvement. However, adding adversarial learning to GAN data augmentation further increased FGSM attack detection rates to 0.937810 and maintained a high PGD detection rate of 0.993904. Thus, both strategies effectively bolster IDS resilience against adversarial attacks.

To answer **RQ3**, we calculated the MSE for each model, analyzed their distributions, and observed the following results. The initial model had MSE values ranging approximately from 0.008927 to 0.007868. This suggests a moderately good fit of the model to the data; however, there is still considerable room for improvement. The GAN-augmented model improved performance, with MSE values ranging from 0.006594 to 0.002299. This demonstrates that adding GAN-augmented data led to a better fit of the model to the data. The model trained with GAN data augmentation and adversarial samples demonstrated the best performance. MSE values were significantly lower, approximately from 0.002309 to 0.002104. This indicates a superior fit of the model to the data compared to the previous two models. The lower MSE values in Fig. 6(b) show the model's effectiveness in reconstructing input data, highlighting its resilience against adversarial attacks.

To address **RQ4**, we need to ensure that enhancing the resilience of IDS does not come at the cost of an increased false positive rate (FPR). Thus, we computed the performance of the models in detecting GPS spoofing and jamming attacks. For GPS spoofing attacks, as shown in Fig. 7(a), the baseline IDS had an accuracy of 0.9476 and an FPR of 0.0523. GAN augmentation improved this decent performance, where accuracy significantly increased to 0.9845, and FPR reduced to 0.0154. This indicates that the GAN augmentation effectively enhanced the model's ability to identify GPS spoofing attacks accurately. Finally, the IDS with both GAN augmentation and adversarial learning achieved the best performance, with an accuracy of 0.9957 and the lowest FPR of 0.0042. This highlights the effectiveness of integrating GAN and adversarial learning in improving IDS performance against spoofing attacks. For GPS jamming attacks, as seen in Fig. 7(b), the accuracies of all three models were quite high, ranging from 0.9942 to 0.9977. Moreover, the models had low FPRs, with the last model having the lowest value of 0.0023. The GAN-augmented and adversarially trained models exhibit higher accuracy and lower FPRS, making them more reliable for real-world applications.

In response to **RQ5**, we conducted a sensitivity analysis on the three models to ascertain their stability and sensitivity against different perturbation magnitudes. Key metrics utilized for this analysis were MSE and R2 Score, with an increase in MSE after perturbation indicating a performance decline, while a decrease in R2 Score indicating a lesser fit of the model to the perturbed data. Upon applying perturbations, the vanilla IDS demonstrated a decline in performance, with an increase in MSE from 0.0113 to 0.0138 and a decrease in the R2 Score from 0.0865 to -0.008. This suggests a lack of stability and higher sensitivity to changes. The GAN-augmented IDS also experienced an increase in MSE from 0.0054 to 0.0091, implying a decline in performance post-perturbation. However, the R2 Score for this model was less affected, with a change from 0.1744 to 0.0506, indicating that this model was more resilient to perturbations than the baseline model. The adversarially regularized IDS showed the best performance, having the smallest increase in MSE from 0.0024 to 0.0074, and its R2 Score remained relatively stable, going from 0.3344 to 0.0506. Fig. 8 shows that GAN augmentation and adversarial samples enhance the IDS models' stability against various perturbation levels.

IX. Conclusion

In this work, we have highlighted the vulnerabilities of current IDS for UAVs against GPS spoofing and jamming attacks and proposed a framework using GANs and adversarial-sample-based regularization. Under FGSM and PGD adversarial attacks, the detection rates for our improved IDS are 93.78% and 99.39%, respectively, outperforming the baseline rates of 26.14% and 62.6%. Additionally, our resilient IDS demonstrated an accuracy of 99.57% against GPS spoofing, substantially better than the conventional IDS accuracy of 94.76%. Importantly, the false positive rate was also reduced to 0.42% compared to the previous 5.23%. This approach

enhances the IDS's resilience, improves accuracy, reduces false positives against spoofing/jamming attacks, and remains robust against adversarial perturbations. In future research, we will examine our framework's suitability for adversarial attacks on UAVs, explore techniques like deep reinforcement learning, and study adaptability to other domains.

X. Acknowledgement

This work is partially supported by the National Security Agency (NSA) under Award H98230-22-1-0327 and the National Science Foundation (NSF) under Grant No. 1946442, 2100115, and 2209638. Any opinions, findings, conclusions, or recommendations in this document are those of the author(s) and do not necessarily reflect the views of NSA or NSF.

References

[1] Syed Agha Hassnain Mohsan, Nawaf Qasem Hamood Othman, Yanlong Li, Mohammed H Alsharif, and Muhammad Asghar Khan. Unmanned aerial vehicles (uavs): practical aspects, applications, open challenges, security issues, and future trends. *Intelligent Service Robotics*, pages 1–29, 2023.

[2] Drones for civilian apllications. https://umsskeldar.aero/capabilities/drones-for-civilian-applications/, Apr 2023.

[3] James Rennie. Commercial uses and applications of drones. https://www.auav.com.au/articles/what-is-a-drone-used-for/, May 2022.

[4] David Glade. Unmanned aerial vehicles: Implications for military operations. https://apps.dtic.mil/sti/pdfs/ADA425476.pdf, Jul 2000.

[5] Sharifah Mastura Syed Mohd Daud, Mohd Yusmiaidil Putera Mohd Yusof, Chong Chin Heo, Lay See Khoo, Mansharan Kaur Chainchel Singh, Mohd Shah Mahmood, and Hapizah Nawawi. Applications of drone in disaster management: A scoping review. *Science Justice*, 62(1):30–42, 2022.

[6] Jennifer Mary. Drones for disaster relief. https://beta.nsf.gov/news/drones-disaster-relief.

[7] Snapshot: First responders assess drones for search and rescue missions. https://www.dhs.gov/science-and-technology/news/2020/04/02/snapshot-first-responders-assess-drones-search-and-rescue-missions, Apr 2020.

[8] Yassine Yazid, Imad Ez-Zazi, Antonio Guerrero-González, Ahmed El Oualkadi, and Mounir Arioua. Uav-enabled mobile edge-computing for iot based on ai: A comprehensive review. *Drones*, 5(4), 2021.

[9] ADF Staff. Drones prove worth in maritime surveillance, security. https://adf-magazine.com/2022/12/drones-prove-worth-in-maritime-surveillance-security/, Dec 2022.

[10] Construction amp; mining: Aerial surveying: Georeferenced: Orthomosaic. https://bst.aero/construction-mining/, Jun 2021.

[11] How will drones impact the future of military warfare? https://www.zenadrone.com/drones-impact-the-future-of-military-warfare/, Oct 2022.

[12] Jonathan Marcus. Combat drones: We are in a new era of warfare - here's why. https://www.bbc.com/news/world-60047328, Feb 2022.

[13] Andrea Gilli. Drone warfare: An evolution in military affairs. https://www.ndc.nato.int/news/news.php?icode=1754, Oct 2022.

[14] CG Leela Krishna and Robin R Murphy. A review on cybersecurity vulnerabilities for unmanned aerial vehicles. In *2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*, pages 194–199. IEEE, 2017.

[15] Harshad Sathaye, Martin Strohmeier, Vincent Lenders, and Aanjhan Ranganathan. An experimental study of {GPS} spoofing and takeover attacks on {UAVs}. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 3503–3520, 2022.

[16] Renato Ferreira, João Gaspar, Pedro Sebastião, and Nuno Souto. Effective gps jamming techniques for uavs using low-cost sdr platforms. *Wireless Personal Communications*, 115:2705–2727, 2020.

[17] Seong-Hun Seo, Byung-Hyun Lee, Sung-Hyuck Im, and Gyu-In Jee. Effect of spoofing on unmanned aerial vehicle using counterfeited gps signal. *Journal of Positioning, Navigation, and Timing*, 4(2):57–65, 2015.

[18] Tao Zhang and Quanyan Zhu. Strategic defense against deceptive civilian gps spoofing of unmanned aerial vehicles. In *Decision and Game Theory for Security: 8th International Conference, GameSec 2017, Vienna, Austria, October 23-25, 2017, Proceedings*, pages 213–233. Springer, 2017.

[19] Shenqing Wang, Jiang Wang, Chunhua Su, and Xinshu Ma. Intelligent detection algorithm against uavs' gps spoofing attack. In *2020 IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS)*, pages 382–389. IEEE, 2020.

[20] Tala Talaei Khoei, Shereen Ismail, and Naima Kaabouch. Dynamic selection techniques for detecting gps spoofing attacks on uavs. *Sensors*, 22(2):662, 2022.

[21] Elena Basan, Alexandr Basan, Alexey Nekrasov, Colin Fidge, Nikita Sushkin, and Olga Peskova. Gps-spoofing attack detection technology for uavs based on kullback–leibler divergence. *Drones*, 6(1):8, 2021.

[22] Jason Whelan, Abdulaziz Almehmadi, and Khalil El-Khatib. Artificial intelligence for intrusion detection systems in unmanned aerial vehicles. *Computers and Electrical Engineering*, 99:107784, 2022.

[23] Jason Whelan, Thanigajan Sangarapillai, Omar Minawi, Abdulaziz Almehmadi, and Khalil El-Khatib. Novelty-based intrusion detection of sensor attacks on unmanned aerial vehicles. In *Proceedings of the 16th ACM symposium on QoS and security for wireless and mobile networks*, pages 23–28, 2020.

[24] Olakunle Ibitoye, Omair Shafiq, and Ashraf Matrawy. Analyzing adversarial attacks against deep learning for intrusion detection in iot networks. In *2019 IEEE global communications conference (GLOBECOM)*, pages 1–6. IEEE, 2019.

[25] Han Qiu, Tian Dong, Tianwei Zhang, Jialiang Lu, Gerard Memmi, and Meikang Qiu. Adversarial attacks against network intrusion detection in iot systems. *IEEE Internet of Things Journal*, 8(13):10327–10335, 2020.

[26] Jiwei Tian, Buhong Wang, Rongxiao Guo, Zhen Wang, Kunrui Cao, and Xiaodong Wang. Adversarial attacks and defenses for deep-learning-based unmanned aerial vehicles. *IEEE Internet of Things Journal*, 9(22):22399–22409, 2021.

[27] Qie Hu, Young Hwan Chang, and Claire J Tomlin. Secure estimation for unmanned aerial vehicles against adversarial cyber attacks. *arXiv preprint arXiv:1606.04176*, 2016.

[28] Ashok Raja, Laurent Njilla, and Jiawei Yuan. Adversarial attacks and defenses toward ai-assisted uav infrastructure inspection. *IEEE Internet of Things Journal*, 9(23):23379–23389, 2022.

[29] Michael Doyle, Josh Harguess, Keith Manville, and Mikel Rodriguez. The vulnerability of uavs: an adversarial machine learning perspective. In *Geospatial Informatics XI*, volume 11733, pages 81–92. SPIE, 2021.

[30] Benjamin J McCloskey. Using generative adversarial networks to augment unmanned aerial vehicle image classification training sets. 2023.

[31] Nageswara Guptha M, YK Guruprasad, Yuvaraja Teekaraman, Ramya Kuppusamy, and Amruth Ramesh Thelkar. Generative adversarial networks for unmanned aerial vehicle object detection with fusion technology. *Journal of Advanced Transportation*, 2022, 2022.

[32] Elena Basan, Maria Lapina, Nikita Mudruk, and Evgeny Abramov. Intelligent intrusion detection system for a group of uavs. In Ying Tan and Yuhui Shi, editors, *Advances in Swarm Intelligence*, pages 230–240, Cham, 2021. Springer International Publishing.

[33] Qasem Abu Al-Haija and Ahmad Al Badawi. High-performance intrusion detection system for networked uavs via deep learning - neural computing and applications. https://link.springer.com/article/10.1007/s00521-022-07015-9, Feb 2022.

[34] V. Praveena, A Vijayaraj, Chinnasamy Ponnusamy, Ali Ihsan, Roobaea Alroobaea, Saleh Yahya, and Muhammad Raza. Optimal deep reinforcement learning for intrusion detection in uavs. *Computers, Materials and Continua*, 70:2639–2653, 09 2021.

[35] Robert Mitchell and Ray Chen. Adaptive intrusion detection of malicious unmanned air vehicles using behavior rule specifications. *IEEE transactions on systems, man, and cybernetics: systems*, 44(5):593–604, 2013.

[36] Afnan Alotaibi and Murad A. Rassam. Adversarial machine learning attacks against intrusion detection systems: A survey on strategies and defense. *Future Internet*, 15(2), 2023.

[37] Jason Whelan, Thanigajan Sangarapillai, Omar Minawi, Abdulaziz Almehmadi, and Khalil El-Khatib. Uav attack dataset. *IEEE Dataport*, 2020.