Drone Cyber Attack Detection through Machine Learning

Subjects: Computer Science, Cybernetics

Contributor: Zubair Baig, Naeem Syed, Nazeeruddin Mohammad

A smart city comprises a number of integrated components that enable its functioning and rendering of automated services to its citizens. Data generated by smart city components is facilitated for transmission across the ICT network to the next hop devices of the communication network topology. Data collection on a large scale is referred to as data volume. Analytics of collected data is contingent on the level of urgency and the computational capabilities of the controllers, edge devices and the geographic proximity of these to the central cloud platform. For instance, IoT data obtained from curb-side sensors will have to be communicated to the next hop edge device, whereupon further data processing can take place in the cloud. Data volume at a large scale is constantly communicated from resource-constrained devices, including drones, to a central cloud, for processing and subsequent decision making. Artificial Intelligence (AI) plays a significant role in accurate decision making from collected data in a smart city platform.

Keywords: drones ; criminal activity ; machine learning ; cyber attacks

1. Drone Attack Models

Drones, by design, do not comprise a fool-proof end-to-end security solution to reduce design and manufacturing costs. Common security gaps that persist within commercial drones are: vulnerability to firmware manipulation, lack of encryption of static data as well as communicated data (to the ground controller) [1].

Some drone manufacturers provide over-the-air (OTA) firmware updates analogous to mobile phone software patches (updates). Through such practice, vulnerabilities identified in drones post-purchase can be patched to avoid a compromise.

A software or firmware vulnerability in a drone can be exploited by the adversary through the modification of strings of data causing the drone to malfunction. Consequent adversarial actions can include flight trajectory changes and failure to encrypt flight logs [1].

On-device drone data is critical to operations, but also the transit from the drone to a ground-based controller in an unencrypted form could lead to a compromise. If the data are sensitive they could fall into adversarial hands and could subsequently be misused. Legacy network communication protocols would not encrypt drone data by default before it is communicated wirelessly to a ground controller. Additionally, the exposed vulnerabilities through unhardened firmware could lead to the disablement of encryption features, even if they exist on a drone.

In [2], the impact of drone-based threats to its operations have been categorized into the following:

- Unavailability of a UAV;
- Disruption of UAV operations;
- Performance degradation and disconnection with ground controller;
- Misleading GPS information;
- Exposure of confidential UAV information;
- Damage to infrastructure;
- Compromised and misbehaving UAV.

The standard process to weigh in all parameters for a threat model, as proposed in [2] compromise: threat identification, severity estimator, likelihood of an attack, attack ranking and risk scores. The adversarial goals, as listed above, can be mapped to various threats that can be presented to a drone's software/firmware, a ground controller and the central cloud services that a drone connects to.

Additionally, an adversary may also pose a threat by hijacking a communication signal by either disrupting the signal or replacing the same with a malicious one. Consequently, the drone will operate outside its routine mode of operation.

Tran et al. [3] extend the specific operations risk assessment (SORA) [4] to include cybersecurity risks. They have proposed threat and harm extensions to the SORA methodology. Threat extension covers different cybersecurity threats that can occur to drones, which could make drone operations out of control. In other words, the threats are the main reason for drone-related hazards. SORA represents the potential outcomes of the Hazards as Harms. Cyber Harms primarily considers the privacy issues that can occur because of cyberattacks. It also considers physical and digital damages that could occur because of cyberattacks on drones. SORA includes two types of barriers: threat and harm barriers. Threat barriers prevent the Hazard incidence once a threat incident has occurred. Similarly, Harm barriers avert the Harm after a Hazard incident.

1.1. Denial of Service

In [5], a denial of service (DoS) attack was tested against the Parrot ANAFI Drone. The drone was connected to a Wi-Fi access point, that emulates a ground controller. The identified threat was the compromise of the Wi-Fi password, which can foster adversarial attempts to send falsified flight commands. With no knowledge of the password, the adversary may still be able to perform a deauthentication attack and attempt to crack the same. A third option could simply be an attempt to disconnect communication between the drone and the access point (ground controller).

A DoS attack can also be perpetrated through an adversary who is able to disrupt a drone-to-ground controller or ground controller-to-cloud communication channels by flooding the communication channel with a large volume of ficititious network traffic, consequently overwhelming the computational assets, and preventing them from continuing with routine operations. Such an attack can be carried out using a technique such as 'Low Orbit Ion Cannon' with TCP flooding attack enabled on port 80 if the communication between the ground controller and the cloud services adopts the TCP protocol. The Hping3 is command-line software that can also be deployed to carry out a TCP SYN flood attack (DoS) against a drone system asset [5].

A ground controller station can also be compromised by the adversary, which will then send a suspicious signal to mislead a drone or even cause a crash. Through an analysis of the receiver signal strength indicator (RSSI) values and by triangulating the numbers with neighboring drones and their data, the scheme proposed in [6] can prove to be resilient to DoS attacks.

1.2. Hijacking

Disruption of drone operations through modification of software can also be staged at the adversarial machine learning level. The drone quadrotor can be programmed to disrupt drone flight when it is subject to fictitious objects that can be observed through its sensors during flight, causing it to trigger a hazard avoidance routine and thus leading to a variation in flight trajectory. This can also be defined as a hijacking attack.

Quadrotors typically have a return to home (RTH) state, that may be induced by the adversary so as to cause abandonment of the next waypoint. Typically, a drone that runs low on battery or is disconnected from the ground controller would have the RTH state activated. However, the authors in [7] have identified four threat models, namely, jamming of the communication channel, obscuring the main sensor/mirror, duplicating the target image and disrupting the object tracker (hide or change object makeup). Subject to these threats, the drone can not simply be forced to change to an RTH state, but can also be misconfigured and forced to reach an incorrect waypoint. Deliberate attempt to present to the drone's vision sensors images, reflections or other attack vectors (visual) can be adopted by the adversary to cause such disruptions [7].

1.3. GPS Signal Jamming/Spoofing

GPS spoofing entails the presentation of inaccurate geo-location/coordinates to a UAV in-flight. GPS blueprints are widely available and can therefore easily be spoofed [8]. Such a malicious attack can be perpetrated simply by setting up a power amplifier and a transmission antenna that relays RF signals to a target. As drones are programmed to accept unencrypted GPS signals, spoofing of the data is a feasible attack and can lead to catastrophic results for the target drone. A varied

coordinate signal could cause the drone to change its flight path and trajectory and may lead to a crash, with affected entities, including human citizens of the smart city, power lines, vehicles in commute and other ground objects [9].

Table 1 summarizes security property compromised, threat types and their impact analysis.

Table 1. Attack impact analysis.

Level	Resources Affected	Motive	Impact
Hardware	**Drone, ground controller**	**Sabotage, data exfiltration**	**Flight path, crash**
Firmware	**Drone, ground controller**	**Process and state tampering, Incapacitation**	**Crash, data loss**
Software	**Drone, ground controller, edge devices**	**Sabotage, compromise**	**Process disruption, Flight path/services crippled**
Network	**Communication channel**	**Data exfiltration, modification**	**Flight disruption, access denial**
Process	**Drone, ground controller, edge, cloud**	**Software malfunction**	**Flight disruption, data loss**

2. IDS Design

Drone-based attacker tactics against heterogeneous smart city ICT platforms are reliant upon the ability of the threat actor to intrude the drone firmware and/or the ground controller with the intent to either divert the same to an unwarranted locale, to cause it to crash or to modify its trajectory and make it observe and report phenomena back to a rogue command and control center (C2). Traditionally, intrusion detection for cyber physical systems can be categorized into signature-based and anomaly-based. For the former, the detection system has to be pre-configured with signatures of known attacks, which can subsequently be matched with live/observed drone data, possibly at the ground controller unit. Anomaly-based systems, on the other hand, are trained to identify normal/legitimate data flow, including network traffic which refers to routine drone behavior. Deviation from the norm is subsequently tagged as an attack. Contemporary intrusion detection system design comprises robust machine learning techniques that allow for correlation of multiple data streams that can emanate from the drone, ground controllers and edge nodes of a drone control system. Popular techniques that can be adopted for detecting adversarial attempts to penetrate the drone system include support vector machines (SVMs), deep learning, and extreme learning machines (ELMs) [10]. To place an intrusion detection system as an overlay on a ground controller would entail significant computing and storage usage.

The concept of training an intrusion detection system with known or malicious signatures, for subsequent detection, can be adopted without modification albeit the data itself will require massaging and preparation for presentation to the intrusion detection system during the training as well as the testing phases.

Whilst an intrusion detection system can be placed in a high performance device, including a ground controller to detect adversarial attempts to compromise a drone, the ability of such a system to function on resource-constrained devices, such as drones, is a challenge.

3. Machine Learning for Intrusion Detection

According to the Capgemini report, 61% of organizations confirm that they will not be able to identify critical threats without Artificial Intelligence [11]. AI-driven cybersecurity controls can detect a malicious attack before it achieves its malicious goals, predict future attacks through intelligent forecasting based on analysis of empirical data and present mechanisms for automated response to threats through generation and rendering of software patches to digital assets [11].

Machine learning can be defined as a category of Artificial Intelligence, wherein the notion of mathematical modeling of data is adopted to train the machine learning classifier. The classifier subsequently is subject to test data, which it classifies based on its developed capability during the training phase. Broadly, machine learning classifiers can be categorized into the following four categories:

- Supervised learning—the data presented to the machine learning classifier is labeled as per its class definition. For instance, in a *drone attack* the label can be placed for those data samples (rows of data) that represent an attack vector. Similarly, routine drone flight data can be categorized as *normal*, which can serve as the second label for data samples. During the testing phase, data samples are presented to the trained classifier (model) without labels, and

performance measurements of the classifier are measured through comparison of its classification outcomes to the actual class lables of the test dataset.

- Unsupervised learning—the data presented to the classifier are unlabeled, and the classification procedure in itself follows the process of clustering similar data samples into a given cluster and through differentiation at the inter-cluster levels.

- Reinforcement learning—the concept is based upon producing a 'rewarding function', which produces an optimal or a near-optimal classification of data samples, without the dependence upon labels or supervision. Typical reinforcement learning algorithms adopt Markov decision models to assess input data samples to attain the highest cumulative reward, when the classification is performed. This concept can be combined with supervised learning (for labeled data samples) to enhance the overall accuracy of the classifier.

Popular machine learning classifiers for drone applications include Naive Bayes, support vector machines (SVMs) and random classifiers. Random forest classifiers are ensemble-based classifiers known for their robustness in image classification. In [12], random forest classifiers are adopted for classification of images captured by a drone to identify vegetation in remote sensing fields. The first step of the classification is to adopt a bootstrap strategy wherein, nearly two-thirds of the training data samples are consumed to produce a decision tree. The remainder data samples are named out-of-bag data, which are subsequently used for inner cross validation of the trained random forest decision tree model, for accuracy.

Support vector machines (SVMs) are supervised machine learning algorithms that belong to the family of linear classifiers. The objective of the SVM training algorithm is to build a hyperplane in an *N*-dimension space that maximizes the margin between the two classes of data. Hyperplanes are typically decision boundaries that enable the distinguishing of data points of one class from another.

The Naive Bayes (NB) classifier is a simple probabilistic technique that is based on the concept of the Bayes theorem. It functions by assigning a posterior probability to a data sample for belonging to a class, *Y*, based on the a priori training of the classification algorithm on a dataset. Attributes are assumed to be independent of each other, i.e., feature dependence is not a criteria for training and testing. NB classifiers are known for their high accuracy in classifying string data.

Supervised learning techniques have been previously adopted to identify cyber threats in a drone system. In [13], a framework is presented for the classification of drone data into malicious or normal through adoption of standard drone data for presenting to a post-incident machine learning classifier, in order to infer a cyber criminal activity as part of digital forensic investigations.

References

1. Salamh, F.E.; Karabiyik, U.; Rogers, M. A Constructive DIREST Security Threat Modeling for Drone as a Service. J. Digit. Forensics Secur. Law 2021, 16. Available online: https://commons.erau.edu/cgi/viewcontent.cgi?article=1695&context=jdfsl (accessed on 26 May 2022).
2. Singh, K.; Verma, A.K. Threat modeling for multi-UAV adhoc networks. In Proceedings of the TENCON 2017—2017 IEEE Region 10 Conference, Penang, Malaysia, 5–8 November 2017; pp. 1544–1549.
3. Tran, T.D.; Thiriet, J.M.; Marchand, N.; El Mrabti, A. A Cybersecurity Risk Framework for Unmanned Aircraft Systems under Specific Category. J. Intell. Robot. Syst. 2022, 104, 1–15.
4. Specific Operations Risk Assessment (SORA). 2021. Available online: https://www.eurocockpit.be/sites/default/files/2019-01/SORA_ECA_Position_Paper_19_0128_F.pdf (accessed on 26 May 2022).
5. Feng, J.; Tornert, J. Denial-of-Service Attacks Against the Parrot ANAFI Drone. Batchelor Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2021.
6. Chibi, N.; El Ghazi, H.; Fihri, W. Drone cyber-attack: An intrusion detection technique based on RSSI and trilateration. In Proceedings of the Third International Conference on Transportation and Smart Technologies, Tangier, Morocco, 27–28 May 2021; pp. 42–45.
7. Doyle, M.; Harguess, J.; Manville, K.; Rodriguez, M. The vulnerability of UAVs: An adversarial machine learning perspective. In Geospatial Informatics XI; SPIE: Bellingham, WA USA, 2021.

8. Pardhasaradhi, B.; Cenkeramaddi, L.R. GPS Spoofing Detection and Mitigation for Drones Using Distributed Radar Tracking and Fusion. IEEE Sens. J. 2022, 22, 11122–11134.
9. Yaacoub, J.-P.; Noura, H.; Salman, O.; Chehab, A. Security analysis of drones systems: Attacks, limitations, and recommendations. Internet Things 2020, 11, 100218.
10. You, I.; Yim, K.; Sharma, V.; Choudhary, G.; Chen, I.R.; Cho, J.H. On IoT misbehavior detection in Cyber physical systems. In Proceedings of the 2018 IEEE 23rd Pacific Rim International Symposium on Dependable Computing (PRDC), Taipei, Taiwan, 4–7 December 2018; pp. 189–190.
11. How Sailpoint, AI and Machine Learning Are Improving Cybersecurity. SailPoint. 2022. Available online: https://www.sailpoint.com/topics/ai-machine-learning/?gclid=EAIaIQobChMIgOj71d7U-AIVlGSLCh2UqQssEAAYASAAEgJYtvD_BwE (accessed on 26 May 2022).
12. Feng, Q.; Liu, J.; Gong, J. UAV Remote Sensing for Urban Vegetation Mapping Using Random Forest and Texture Analysis. Remote Sens. 2015, 7, 1074–1094.
13. Baig, Z.; Khan, M.; Mohammad, N.; Brahim, G.B. Drone forensics and machine learning: Sustaining the investigation process. Sustainability 2022, 14, 4861.
