# On the Efficacy of Model-Based Attack Detectors for Unmanned Aerial Systems

Ian Y. Garrett
Virginia Tech
Arlington, Virginia
ianygarrett@vt.edu

Ryan M. Gerdes
Virginia Tech
Arlington, Virginia
rgerdes@vt.edu

## ABSTRACT

Unmanned Aerial Systems (UAS), informally known as drones, are cyber-physical systems (CPS) that operate by remote human control or autonomous control. UAS are increasingly being used in a wide variety of applications, such as search and rescue, delivery of goods, or surveillance. These systems rely on sensors and actuators to evaluate their current state and take further action; due to the reliance on sensors and actuators, it is critical to thwart an adversary's attempt to compromise these elements. Security in these systems rely on detectors to find malicious activity, many of which require models of the system to compare what the readings are versus the expected value. Due to measurement and process noise, it is possible that an adversary may perpetrate undetectable attacks. In this paper we examine the sensitivity of model-based attack detectors to measurement and modeling uncertainty, ultimately showing the weaknesses in relying solely on model-based detectors for attack detection. We demonstrate attacks on a simulation of the Senior Telamaster UAS and evaluate the performance of multiple attack detectors after modifications on various parameters, such as those related to internal factors, e.g., measurement noise, as well as external forces, e.g., wind, ultimately showing that an attacker is able to evade detection due to fundamental limitations in the model-based approach.

## CCS CONCEPTS

• **Security and privacy → Intrusion detection systems**; • **Computer systems organization → Sensors and actuators**;

## KEYWORDS

Unmanned Aerial System; UAS Security; Model-based Detector

**ACM Reference Format:**
Ian Y. Garrett and Ryan M. Gerdes. 2020. On the Efficacy of Model-Based Attack Detectors for Unmanned Aerial Systems. In *Proceedings of the Second ACM Workshop on Automotive and Aerial Vehicle Security (AutoSec '20), March 18, 2020, New Orleans, LA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3375706.3380555


*AutoSec '20, March 18, 2020, New Orleans, LA, USA*

ACM ISBN 978-1-4503-7113-1/20/03…$15.00
https://doi.org/10.1145/3375706.3380555

## 1 INTRODUCTION

Unmanned Aerial Systems (UAS), also known as unmanned aerial vehicles or drones, are cyber-physical systems (CPS) that can function through either remote human control or autonomous control. UAS are increasingly being utilized for a variety of applications, such as providing wireless coverage, search and rescue, delivery of goods, and surveillance [9]. A failure of these CPS to perform in a safe manner may lead to a negative effect on the economy or physical human harm. Whether in human-controlled or autonomous operation a UAS relies on sensor information to properly direct the actuators; thus, sensor and actuator integrity is vital to successful UAS flights. However, UAS have been shown to be susceptible to both sensor and actuator attacks.

A number of previous studies focus on diverse methods for unmanned vehicle attack detection. Some existing methods propose a fault-based approach, detecting an attack by assigning signatures [1], while others focus on model-free approaches, such as using the Mahalanobis distance for anomaly detection [4]. In contrast to these methods, we focus on model-based approaches wherein a system model (i.e., a model describing how the UAS should respond to exogenous and endogenous inputs) is used to create an accurate estimate of the system's state to compare to the sensors' readings of the current state [7]. Model-based detection hinges on the accuracy of the model [3] to determine a precise estimate of the system's current state in order to compare it to the measured readings.

While studies may make the assumption that the model being used is accurate, we instead take the opposite approach: this paper assumes that there are errors allowed in the model that can create an attack surface for an adversary. Weaknesses in model-based approaches have been examined showing that the derivation between the estimated state and real-time values can be leveraged by adversaries to exploit unmanned systems [2], and this work builds upon the stated weaknesses by identifying the effect of modifying model parameters on the model's accuracy and ability to provide estimation for detection.

Specifically, this paper examines the shortcomings of using model-based detection methods for UAS to protect against common attacks, e.g., in actuators [7] or sensors [6]. We improve upon this body of work by examining the effect of modifying various model-based parameters that affect the ability of a model-based detector to uncover an attacker. Furthermore, we approach the evaluation in a systematic manner that highlights the parameters most appropriate for the camouflage of adversarial activity.

## 2 SYSTEM AND THREAT MODELS

UAS require communication with the system's sensors and actuators to measure the current state since there is no user on-board

and able to use other senses to determine if the instruments are faulty; this reliance on sensor and actuator accuracy leads to vulnerability. We hypothesize that the allowed difference between the sensor measurement and the detector's threshold can vary between parameters and detectors, which ultimately gives more allowance to an adversary to conduct an attack. If the adversary is able to read the parameters, or modify the parameters, then the attacks can be strategically timed to minimize likelihood of detection.

## 2.1 Threat model and assumptions

The goal of the adversary is to perform attacks using covert methods, e.g. false data injection (FDI) to manipulate sensor measurement, that ultimately cause the UAS to perform actions that are outside of its desired path. In order for the attack to be successfully covert it must be undetected, which requires any sensor measurement deviations to fall within the acceptable boundaries of the chosen detector. In addition to an attack not being detected as a false negative, it is assumed that the detector is adjusted for false positive rates so that an attack could occur within the accepted noise range of the detector as long as it stays within the range of the rate. The primary motivation of the adversary is to remain undetected so they are searching for the set of parameters that allow for the greatest detection error.

We assume the adversary has exploited the UAS prior to the attack and is capable of performing false data injection to affect the sensors. Furthermore, the adversary has implanted the device so that there is continual access that allows system value monitoring; the adversary is assumed to have the ability to read the UAS control inputs, as well as the model parameters, and can use them to create the UAS state estimation that is used for anomaly detection. We further assume that the UAS operators regularly modify the UAS parameters in an attempt to prevent an adversary from adapting attacks to specific parameters.

The adversary is limited by only being able to conduct the chosen attack as well as view, but not able to modify, the set parameters. Therefore, the adversary is a passive persistent threat focused on searching for a time when the optimal set of parameters are in place to conduct the attack; thus, they will remain dormant on the UAS otherwise. The adversary is further constrained by being unable to jam the detectors, and does not have access to the system logs so the attack cannot be masked, but can only attempt to remain hidden. Therefore, the attack will be read by the UAS and passed to the detectors; it is up to the adversary to execute it when the optimal set of parameters are in place to achieve their goal.

## 2.2 UAS Attack

The attack focused on disrupting the UAS position, such as in the case of an FDI on a GPS sensor, so that the UAS path spirals instead of maintaining a constant circular flight pattern. The FDI can be performed by modifying the sensor readings in such a way that the UAS sends control inputs according to its believed position compared to the actual position. Figure 1 shows three UAS paths with north/south (in meters) on the y-axis and east/west (in meters) on the x-axis: the blue represents the ground truth (high-fidelity), the green represents the estimated path (low-fidelity with an Extended Kalman Filter), and the red represents the path during an attack. Since many UAS applications require position accuracy, the adversary is motivated to conduct the position-based attack to reduce the effectiveness of the UAS, e.g. a delivery could be sent to a different address. While a drastic shift could be easily detected, small modifications over a period of time could allow the UAS to slowly divert from its expected path.

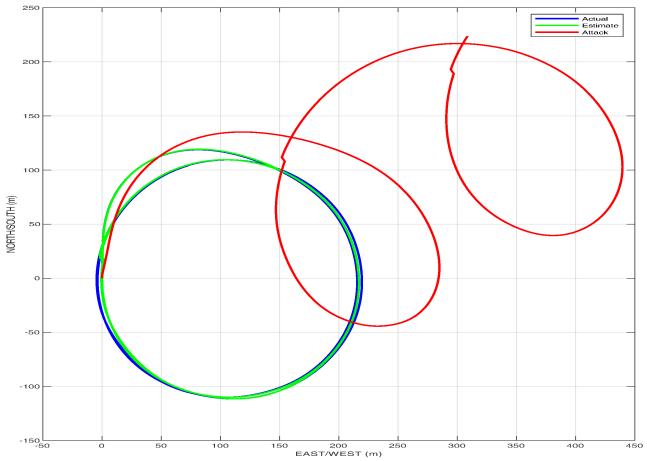


**Figure 1: UAS Normal Flight Path Vs Attack**

# 3 EXPERIMENTAL DESIGN AND EVALUATION

To conduct the experiment we used a high-fidelity model to create a simulation that represented the ground truth of a Senior Telamaster UAS in flight [7], a low-fidelity model to create the state-estimate used by the detectors, two different types of detectors, and made modifications to four parameters. The output of the experiment was the equal error rate (EER), which was determined by finding the intersection of the false acceptance rate and the false rejection rate, of each of the detectors as well as the average absolute error, which was the difference between the high and low-fidelity model.

## 3.1 Setup

Since the high-fidelity model represents the ground truth, the model does not take into account measurement noise; however, the model does take into account external disturbances in the form of wind. The model uses a PID controller to simulate a steady-sate circular flight path. The state-space of this model takes into account the roll, pitch, yaw, as well as the associated Euler angles, body-axis linear velocities, x-position, y-position, height, actuator states, and states pertaining to the Dryden turbulence model.

To represent the model-based detectors, a discrete-time linear time-invariant low-fidelity model of the UAS estimated the position of the UAS. The low-fidelity model took into account not only external disturbances, but also measurement noise from the sensors. Since the model was steady-state, an EKF was used to create a more accurate estimate.

To assess the changes to detection ability, we measured metrics from two detectors, a residual detector and the CUmulative SUM (CUSUM), as well as various model parameters: measurement noise, wind, sample rate, and state-space matrices.

### 3.2 Detectors

Our baseline detection method was to measure the residual between the sensor reading's current state and the estimated state [5] to determine if that residual exceeds a certain threshold. If at time $t$ the residual is greater than the threshold, then an alarm is raised. To find the alarm value $A(t)$, the detector is defined by

$$A(t) = \begin{cases} 1, & \text{if } |r_t| > \tau_t \\ 0, & \text{if } |r_t| \leq \tau_t \end{cases} \quad (1)$$

where $A(t) = 0$ is when the alarm is not triggered and $A(t) = 1$ is when the alarm is sounded, given residual $r_t$ and threshold $\tau_t$. The threshold can be tuned to allow for an acceptable ratio between the false acceptance rate and false rejection rate.

The second detector evaluated was the CUSUM, which allows for more robust detection than that of using just the residual at a single timestep [8]. Instead of simply checking the residual at a single time, CUSUM evaluates the sum of residuals over a sequence and triggering the alarm if that exceeds the threshold. The CUSUM detector is defined by

$$A(t) = \begin{cases} 1, & \text{if } S_{t-1} > \tau_t \\ S_t = \max(0,\ S_{t-1} + |r_t| - b_t), & \text{if } S_{t-1} \leq \tau_t \end{cases} \quad (2)$$

where sequence $S$, which accumulates residuals $r$, is monitored for malicious activity. The bias $b$ allows for further tuning of the acceptance and rejection rate of the detector.

### 3.3 Measured Parameters

To evaluate the potential effect of differing values on a model, we selected various parameters to modify. The first parameter is the measurement noise that is accounted for in the low-fidelity model. This parameter accounts for any error in the readings as a result of imperfection within the sensors; due to a number of causes, e.g. weather's effect on a GPS reading, no sensor can be expected to report 100% accuracy at every moment. The measurement noise is measured in standard deviations from a normal distribution; thus, the baseline measurement has no measurement noise while the modified parameter measures one standard deviation.

The second parameter captured is the wind, an external disturbance, that is accounted for in the model. The baseline measurement accounts for no wind, however the modified parameter accounts for an external disturbance coming from three directions: north, east, and down. The wind parameter is measured in meters per second and in this study is assumed to be a constant force. This study focuses on having only a constant wind for an external disturbance, however other sources of external disturbance could include forces from an adversary, e.g. an air cannon to manipulate the UAS physical state.

The third parameter, the sample rate, is measured in hertz (hz) and represents the samples per second of the model. Varying the sample rate can either create a more accurate model at the cost of higher resources, in the case of a high sample rate, or a less accurate model, in the the case of a low sample rate. Three levels of sample rate were tested: 13 hz, 25 hz, and 50 hz. The 25 hz sample rate represented the baseline measurement.

The state-space matrices that make up the foundation of the model can also be altered to create a more or less accurate model. The measurement of this parameter is represented by how different from the baseline model the modification is made. Thus, the baseline parameters are measured at 100% and any modifications, e.g. 90%, mean they are a scaled value of the baseline.

The final value is a measurement of the error between the model and the ground truth values, which is shown as the average absolute error value. This value represents how effective the parameters are at creating an accurate estimate; the smallest average absolute error is from a highly accurate model.

## 4 RESULTS AND DISCUSSION

The results seek to quantify the effect of a model's parameters on its detection ability using two different detectors. Changes to the parameters were examined by making parameter modifications, then finding the equal error rate of each of the detectors. Table 1 shows each of the four parameters with the corresponding equal error rate associated with the type of detector as well as the average absolute error between the estimate and actual.

Neither detector outperformed the other detector in all experimental categories, although interestingly the residual detector generally performed with higher accuracy in these tests than the CUSUM detector. We believe that even though the CUSUM detector is more robust, the simple nature of the circular flight path allowed the residual detector to more easily determine deviations from the expected path. Given a more variable and dynamic flight path, we believe the CUSUM detector would prove to be generally more effective. Nevertheless, the CUSUM detector performed significantly better than the residual detector in the case of measurement noise, which is a more realistic case than the absence of measurement noise.

The introduction of an external disturbance increased the accuracy of the residual detector while worsening the effectiveness of the CUSUM detector. One factor that may have led to this result is that the external disturbance was in the form of a constant wind, which would allow the residual detector to perform well compared to a variable disturbance. Restricting the tests to a constant force allowed for insight into the effect of applying an external disturbance in general, however future studies should examine the effect of both constant and variable disturbances.

Broadly speaking, tests within the same parameter category with different sample rates performed relatively similarly; the equal error rates and the average absolute errors were within similar ranges compared to the errors of other parameter categories. This highlights that while the sample rate of the model can have a slight effect, any changes would not significantly assist the adversary in finding a window of opportunity to conduct a covert attack. Moreover, this result means that sample rates can be modified and optimized after implementation without introducing unnecessary amount of risk.

The modification to the state-space matrices created the most collective drastic effect on the detectors. The residual detector performed similarly to when measurement noise was introduced, however the CUSUM detector performed significantly worse. The primary factor leading to this was that any modifications to the state-space matrices inherently change the behavior of the model. While this may seem intuitive, differing state-space matrices can

| Detectors | | Parameters | | | | |
|---|---|---|---|---|---|---|
| Residual EER | CUSUM EER | Measurement Noise | Wind (m/s) | Sample Rate | State-space Matrices | Avg Absolute Error |
| 2.31% | 5.25% | None | None | 25 hz | 100% | 3.89 |
| 2.44% | 5.99% | None | None | 50 hz | 100% | 3.85 |
| 2.27% | 4.95% | None | None | 13 hz | 100% | 3.88 |
| 1.89% | 8.71% | None | 2.5 N /2.5 E/ 2.5D | 25 hz | 100% | 11.49 |
| 1.99% | 9.09% | None | 2.5 N /2.5 E/ 2.5D | 50 hz | 100% | 11.54 |
| 1.86% | 8.45% | None | 2.5 N /2.5 E/ 2.5D | 13 hz | 100% | 11.41 |
| 13.24% | 1.05% | 1 Std Dev | None | 25 hz | 100% | 18.32 |
| 13.30% | 2.71% | 1 Std Dev | None | 50 hz | 100% | 18.61 |
| 12.58% | 3.45% | 1 Std Dev | None | 13 hz | 100% | 18.50 |
| 12.18% | 24.58% | None | None | 25 hz | 90% | 13.82 |
| 12.86% | 22.51% | None | None | 50 hz | 90% | 14.29 |
| 11.96% | 25.57% | None | None | 13 hz | 90% | 13.74 |

**Table 1: Model-Based Detectors and Parameters**

be a highly likely source of disparity between the models in a model-based detector as models have to adapt to any changes to the behavior of the system. By referring to the baseline parameters as 100%, this study assumes the low-fidelity state-space matrices are the best representation of the ground truth, yet the assumption does not stand true when applied to a physical system when compared to a simulation. Thus, if the detector and the state-space matrices are not completely in sync, an attack is more likely to be undetected.

Interestingly, the average absolute error was not directly correlated to the detectors performance; while the baseline case performed more accurately with the smallest average absolute error, the different modification categories saw mixed results. The highest error was in the case of introducing measurement noise, which is plausible since the other categories involved a more constant offset from the baseline. The measurement noise was more variable and therefore provided more opportunities for error.

## 5 CONCLUSION AND FUTURE WORK

The evidence from this study points to the idea that there is an inherent weakness within model-based detectors that can be exploited by an adversary to evade attack detection; furthermore, this weakness can be systematically assessed. Ultimately, the results suggest that there is room for error in the assumption that a model is sufficiently accurate in a model-based detection system. The primary limitation of this study is the reliance on a simulation, however our study provides the basis for which a more thorough assessment on a physical platform may be conducted. This study is the first step in creating a more holistic assessment of the effect of parameters on the ability of a model-based detector to perform.

Future work will focus on examining an expanded number of detectors on the current simulation, other UAS simulations, as well as a physical UAS; expanding the experiments to a physical UAS will provide more variable results than a simulation and thus strengthen findings. Furthermore, future work will focus on a greater range of parameters categories and modifications since this study focused primarily on one modification per category. One factor that may lead to interesting results is introducing variable parameters, such as in the case of external disturbances, e.g. wind that changes directions. Additionally, this study focused on systematically toggling parameter categories, but did not look into the effect of combining modifications with each other; we would like to expand this work in the future by examining the effect of combinations.

## ACKNOWLEDGMENT

This work was supported in part by the Department of Energy under grant No. DE-EE0008453. The authors thank Devaprakash Muniraj of Virginia Tech for providing insight with the UAS models.

## REFERENCES

[1] M. J. Daigle, X. D. Koutsoukos, and G. Biswas. 2009. A Qualitative Event-Based Approach to Continuous Systems Diagnosis. *IEEE Transactions on Control Systems Technology* 17, 4 (July 2009), 780–793. https://doi.org/10.1109/TCST.2008.2011648

[2] Pritam Dash, Mehdi Karimibiuki, and Karthik Pattabiraman. 2019. Out of Control: Stealthy Attacks Against Robotic Vehicles Protected by Control-based Techniques. In *Proceedings of the 35th Annual Computer Security Applications Conference (ACSAC '19)*. ACM, New York, NY, USA, 660–672. https://doi.org/10.1145/3359789.3359847

[3] Michael Hofbaur, Johannes Köb, Gerald Steinbauer, and Franz Wotawa. 2007. Improving Robustness of Mobile Robots Using Model-based Reasoning. *Journal of Intelligent and Robotic Systems* 48 (01 2007), 37–54. https://doi.org/10.1007/s10846-006-9102-0

[4] R. Lin, E. Khalastchi, and G.A. Kaminka. 2010. Detecting anomalies in unmanned vehicles using the Mahalanobis distance. *2010 IEEE International Conference on Robotics and Automation, Robotics and Automation (ICRA), 2010 IEEE International Conference on* (2010), 3038 – 3044. http://login.ezproxy.lib.vt.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edseee&AN=edseee.5509781&site=eds-live&scope=site

[5] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli. 2010. False data injection attacks against state estimation in wireless sensor networks. In *49th IEEE Conference on Decision and Control (CDC)*. 5967–5972. https://doi.org/10.1109/CDC.2010.5718158

[6] D. Muniraj and M. Farhood. 2017. A framework for detection of sensor attacks on small unmanned aircraft systems. In *2017 International Conference on Unmanned Aircraft Systems (ICUAS)*. 1189–1198. https://doi.org/10.1109/ICUAS.2017.7991465

[7] Devaprakash Muniraj and Mazen Farhood. n.d.. Detection and mitigation of actuator attacks on small unmanned aircraft systems. *CONTROL ENGINEERING PRACTICE* 83 (n.d.), 188 – 202. http://login.ezproxy.lib.vt.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edswsc&AN=000456903600016&site=eds-live&scope=site

[8] C. Murguia and J. Ruths. 2016. Characterization of a CUSUM model-based sensor attack detector. In *2016 IEEE 55th Conference on Decision and Control (CDC)*. 1303–1309. https://doi.org/10.1109/CDC.2016.7798446

[9] Hazim Shakhatreh, Ahmad H. Sawalmeh, Ala Al-Fuqaha, Zuochao Dou, Eyad Almaita, Issa Khalil, Noor Shamsiah Othman, Abdallah Khreishah, and Mohsen Guizani. 2019. Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges. *IEEE Access* 7 (2019), 48572–48634. https://doi.org/10.1109/access.2019.2909530