The impact of zero-dynamics stealthy attacks on control systems: stealthy attack success probability and attack prevention

by

Stephanie Harshbarger

B.S., University of Nebraska at Kearney, 2018

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Computer Science
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Abstract

Many critical infrastructures rely heavily on automated control systems, making them the target of cyber attacks. Vulnerabilities in control systems are especially dangerous, as they directly affect the physical world.

Zero-dynamics stealthy attacks are a subset of False Data Injection Attacks (FDIAs) that are designed specifically to diverge the states of a controlled cyber-physical system, while producing no discernible changes to the system's output – making these attacks theoretically undetectable. While perfect knowledge of the system model should consistently lead to successful and undetectable attacks, in practice the success of zero-dynamics attacks is limited by the attacker's imperfect knowledge of the system parameters and states, as well as by the system's components' physical limitations. The success of such an attack thus relies no longer on the attack remaining undetectable, but rather on the attacker's ability to significantly diverge the states of the system before detection.

This dissertation explores how the probability of zero-dynamics stealthy attack success is affected by the attacker's knowledge of the system's state space model. Using the quadruple-tank process as an experimental testbed, our results show that it is essential for the attacker to learn an accurate state space representation if they want to have a high probability of a successful attack. Moreover, we show that when the limitations of physical components of the system are considered, the attacker is forced to use an especially accurate state space representation to achieve a reasonable probability of success. Utilizing a grey box approach to system identification, we show that even when the attacker is able to learn a state space model close enough to have a high probability of a successful attack, making small improvements to the system's anomaly detector causes the probability of success to drop drastically.

Finally, we study the trade-offs between making the system less susceptible to zero-dynamics attacks and maintaining its controllability, by increasing the sampling time of the

system, thus providing the attacker fewer samples to learn a state space model. Additionally, results are provided, using a three inverter power system model, showing that strategically choosing model parameters in the design phase of the system can prevent the possibility of zero-dynamics stealthy attacks altogether.

The impact of zero-dynamics stealthy attacks on control systems: stealthy attack success probability and attack prevention

by

Stephanie Harshbarger

B.S., University of Nebraska at Kearney, 2018

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Computer Science
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
George Amariucai

Copyright

Abstract

Many critical infrastructures rely heavily on automated control systems, making them the target of cyber attacks. Vulnerabilities in control systems are especially dangerous, as they directly affect the physical world.

Zero-dynamics stealthy attacks are a subset of False Data Injection Attacks (FDIAs) that are designed specifically to diverge the states of a controlled cyber-physical system, while producing no discernible changes to the system's output – making these attacks theoretically undetectable. While perfect knowledge of the system model should consistently lead to successful and undetectable attacks, in practice the success of zero-dynamics attacks is limited by the attacker's imperfect knowledge of the system parameters and states, as well as by the system's components' physical limitations. The success of such an attack thus relies no longer on the attack remaining undetectable, but rather on the attacker's ability to significantly diverge the states of the system before detection.

This dissertation explores how the probability of zero-dynamics stealthy attack success is affected by the attacker's knowledge of the system's state space model. Using the quadruple-tank process as an experimental testbed, our results show that it is essential for the attacker to learn an accurate state space representation if they want to have a high probability of a successful attack. Moreover, we show that when the limitations of physical components of the system are considered, the attacker is forced to use an especially accurate state space representation to achieve a reasonable probability of success. Utilizing a grey box approach to system identification, we show that even when the attacker is able to learn a state space model close enough to have a high probability of a successful attack, making small improvements to the system's anomaly detector causes the probability of success to drop drastically.

Finally, we study the trade-offs between making the system less susceptible to zero-dynamics attacks and maintaining its controllability, by increasing the sampling time of the

system, thus providing the attacker fewer samples to learn a state space model. Additionally, results are provided, using a three inverter power system model, showing that strategically choosing model parameters in the design phase of the system can prevent the possibility of zero-dynamics stealthy attacks altogether.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Words can not express the gratitude to my advisor, Dr. George Amariucai, for his unwavering patience and support. I am extremely grateful for everything he has done to help me succeed. Thank you for being an outstanding mentor and teacher.

I am thankful to my committee members, Eugene Vasserman, Bala Natarajan, and Mitchell Neilsen, for their expertise and guidance throughout my PhD. Their encouragement, advice, and constructive criticism has helped me greatly.

I would like to thank the researchers at the Intelligent Power Electronics at Grid Edge Laboratory at the University of Illinois - Chicago, including Dr. Mohammad Shadmand, Mohsen Hosseinzadehtaher, Amin Y. Fard, and Alireza Zare, for providing the power system models and diagrams in this dissertation.

Last but not least, I am extremely grateful for the support of my family and friends. Thank you to my parents, James and Gina Slayden, for always believing in me and pushing me to be the best version of myself. My appreciation goes out to my parents-in-law, Kyle and Dena Harshbarger, for celebrating every one of my milestones as if it were their own. I am beyond thankful for all of the support I have received over the years.

Dedication

In dedication to my husband, Kolten, who has been a constant source of encouragement throughout graduate school. Without your support this would not be possible.

Chapter 1

Introduction

Critical infrastructure is more reliant than ever on control system automation, while increasing network connectivity presents an opportunity for remote attacks. Throughout the past few years, we have seen numerous cyber attacks affecting critical infrastructure. In May 2021 a ransomware attack on the Colonial Pipeline caused fuel shortages across much of the East Coast (Newman [2021]). An attacker attempted to introduce a toxic amount of chemicals into drinking water at a water treatment plant in Florida in February 2021 by gaining remote access to the plant's controls (Greenberg [2021]). The world's largest meat processor shut down nearly ten plants in the United States after becoming a victim of a ransomware attack (Sanger and Davis [2021]). As the underlying technology of our critical infrastructure advances, an increase in telecommunication devices and remote access to plants gives attackers more opportunities to attack these systems, with potentially catastrophic outcomes.

We investigate zero-dynamics stealthy attacks, a subset of false data injection attacks (FDIAs), on control systems. With a zero-dynamics stealthy attack, the attacker injects data into the control signal of the system in order to drive the states of the system to be unstable while the observations remain the same (Teixeira et al. [2012]). This means that the attack can have a significant impact on the system while remaining undetected, as many anomaly detectors simply look for changes in the observations (Elmrabit et al.

[2020], Hosseinzadehtaher et al. [2020], Inoue et al. [2017]). However, stealthy attacks have several limitations–including that the attacker must have a state space model that accurately describes the system. Harshbarger et al. [2020] shows that when the attacker does not have an accurate state space model, a stealthy attack may not always be successful. Additionally, specific conditions must exist for a stealthy attack to diverge the states of the system without being detected. These include the attack needing to have an unstable eigenvalue and to have fewer measurements in the system than states. Throughout this dissertation, we investigate these limitations on stealthy attacks to determine how probable it is that an attacker can successfully attack the system under realistic conditions, as well as how we can use the requirements for a stealthy attack to protect control systems against them.

1.1 Related Works on Zero-Dynamics Stealthy Attacks

A considerable amount of work investigates the impact and detection of FDIAs (Deng et al. [2016], Aoufi et al. [2020], Mo et al. [2010]). However, this work does not consider a special case of the FDIA – the stealthy attack. The impact of stealthy attacks can be catastrophic, as they are designed to go undetected. Much of the prior work relating to stealthy attacks only considers the attack under perfect conditions. This includes assuming that the attacker already knows the state space representation for the system they are attacking (Teixeira et al. [2012], Mo and Sinopoli [2015], Yang et al. [2013], Pang et al. [2016]).

Previous work has shown that stealthy attacks can have a serious impact on real systems (Ma et al. [2019], Wei et al. [2022], Teixeira et al. [2011]). Dash et al. [2019] shows that stealthy attacks are able to evade the anomaly detector in robotic vehicles. Similarly, the significance of stealthy attacks on unmanned aerial systems is discussed in Kwon et al. [2014]. The impact and limitations of stealthy attacks on water supervisory control and data acquisition (SCADA) systems are considered in Amin et al. [2010]. Bopardikar and Speranzon [2013] provides conditions for which an attacker is not able to mount a stealthy attack on a discrete time linear time-invariant system by increasing the number of observations in the system. Zhang and Venkitasubramaniam [2017] define the optimal attack strategy by

analyzing the trade off between the increase in quadratic cost that an attack causes and the stealthiness of an attack, concluding that a stealthy attack should align with the eigenvalues of the system they are attacking.

As stealthy attacks are specifically designed to bypass anomaly detectors, much of the work relating to stealthy attacks focuses on detection (Kim and Park [2021], Ding et al. [2018], Zhang et al. [2021], Adepu and Mathur [2018]), with machine learning techniques often being utilized in the detection of stealthy attacks (Sayghe et al. [2020], Haque et al. [2020]). Esmalifalak et al. [2014] use a support vector machine (SVM) to learn the normal operating behavior of a power system in order to determine if measurements should be considered abnormal. Ashrafuzzaman et al. [2018] compares deep learning with various machine learning techniques for anomaly detection and show that the deep learning method provides the best results. Multi-Feature Long short-term memory neural network is used by Wang et al. [2022] to detect stealthy attacks in industrial control systems, showing they are able to lessen the computational power needed for a machine learning based detector, as well as allow for flexibility in changing the system model.

Physics-based anomaly detectors are also popular relating to stealthy attack detection. This is where the physical evolution of the states is considered (Giraldo et al. [2018], Azzam et al. [2021], Raman and Mathur [2021]). The following authors show that watermarking, where a known perturbation is applied to the control signal to verify that it has not been tampered with, can be effective in detecting stealthy attacks (Satchidanandan and Kumar [2016], Ferrari and Teixeira [2021], Jhala et al. [2020]). Zhang et al. [2015] considers anomaly detection from a game theory perspective, determining an optimum approach to defend the system. Additionally, with more advanced technology in industrial control systems, an attacker would have numerous points to access the system to mount an attack, including vulnerabilities in the controller's software (Keliris and Maniatakos [2019]) and inadequate security of the communication between sensors (Flick and Morehouse [2010]). A stateful anomaly detector using a cumulative sum statistic is proposed in Urbina et al. [2016], meaning that the anomaly detector takes into consideration the state of the system, rather than the stateless anomaly detector which considers the observations. Further countermeasures for

the control-signal stealthy attack model in the general control system scenario are developed in Urbina et al. [2016], Aoudi et al. [2018].

We have seen that many industrial control systems are vulnerable to stealthy attacks, and power systems, including smart grids, are no exception to these dangers. Power systems are especially difficult to protect, as the equipment is often outdated and difficult to replace or retrofit. Teixeira et al. [2015] analyzes the risk of an attack for each measurement, which was then applied to determining which buses in a power system are the most vulnerable to an attack. Much of the work on stealthy attacks in power systems is restricted to attacks on measurements. In this context, false data injection attacks are considered stealthy if they are within the span of the system's output matrix (Esmalifalak et al. [2014], Dan and Sandberg [2010], Esmalifalak et al. [2011], Rahman et al. [2013], Ashok et al. [2016]), and are undetectable by an anomaly detector based on the weighted least squares state estimation, or by a CUSUM-like system (Kurt et al. [2018]). Teixeira et al. [2010] analyzes the trade-off between the damage from an attack and the accuracy of the attacker's knowledge, in the context of a stealthy measurement attack where the attacker has imperfect knowledge of the system. Chakhchoukh and Ishii [2016] provides an extension of the weighted least squares state estimation anomaly detector by using multiple least trimmed squares state estimators. The purpose of this anomaly detector is to better detect an attack while remaining cost effective to implement. An online anomaly detector for attacks on power systems is proposed in Ashok et al. [2016]. This anomaly detector uses the topology processing, system parameters, and real-time load forecast information to predict the state of the system to compare with the state estimator. Vuković et al. [2011] emphasizes the importance of securing the system at the network layer. This includes implementing a form of data authentication and using multi-path routing of the data.

Due to the limitations of stealthy attacks, in a realistic industrial control system running under normal conditions, it is nearly impossible for a stealthy attack to remain entirely stealthy. The attacker hopes that they can make the attack stealthy for long enough to cause damage to the system, and the system operator hopes that the anomaly detector will catch the imperfections in a stealthy attack before damage is done. Previous work

shows that this turns into a race of creating an attack that can beat the current anomaly detectors (Esmalifalak et al. [2012], Liu et al. [2021], Tian et al. [2021], Liu et al. [2020]) and creating anomaly detectors that can detect these attacks (Esmalifalak et al. [2014], Urbina et al. [2016], Dan and Sandberg [2010], Roy and Debbarma [2022]) before they are able to damage the system. The consequences of stealthy attacks on power systems can range from small service disruptions to the damaging of very expensive hardware. Even when the power grid is well protected by switches, stealthy attacks can be employed to trigger these switches in a coordinated manner, with the potential of causing significant instability, as shown in Liu et al. [2014, 2013]. Overall, stealthy attacks remain extremely dangerous, but can often be thwarted by smart system design and by an increased number of measurements. Previous work has shown the impact of stealthy attacks on various power systems models (Harshbarger et al. [2020], Keliris and Maniatakos [2019], Choeum and Choi [2019]).

1.2 Overview of the Subsequent Chapters

Chapter 2 considers the impact of applying more realistic conditions to stealthy attacks. First, we chose to incorporate a Linear Quadratic Gaussian (LQG) controller, where we make modifications to the traditional equations in order to push the states of the system to a desired value rather than 0. Additionally, we assume that the attacker does not know the state space model of the system they are attacking – so they would have to learn this information prior to calculating the attack. As the attacker has to learn a state space model for the system, the learned model would most likely be an imperfect representation of the system. We run simulations of systems under attack with the attacker having imperfect knowledge of the state matrix A to determine if the attack can remain stealthy. These results are simulated using the Quadruple-Tank Process (QTP) as well as a simple power system model.

In Chapter 3, we analyze the probability that a stealthy attack is successful when the attacker has an imperfect representation of the state space model. Stealthy attacks are undetectable by anomaly detectors that look for changes in the observations of the system;

however, when the attacker has imperfect knowledge of the system, the attack may become detectable. We consider a simple variance-based anomaly detector to represent the minimum anomaly detector a system would have. The detection time is then compared with the time it takes the attacker to diverge the states of the system to a point of system failure, i.e., the system failure time of the attack, in order to calculate the probability of a successful attack. We consider the cases where the attacker has imperfect knowledge of the state matrix A, the input matrix B, and both A and B in order to determine how close the attacker needs to learn the parameters of the system in order to have a high probability of a successful attack. Additionally, we consider that the system being attacked has physical limitations on the amount of energy applied to the system by the control signal. Since stealthy attacks involve the attacker injecting an exponential amount of energy into the control signal, we analyze the impact of the attacker being limited on the magnitude of their attack signal. We also consider how the attacker can decrease the attack signal once they reach the physical limits of the system while remaining undetected.

Chapter 4 considers how an attacker would learn a state space model of the system. We utilize black box and grey box system identification algorithms to learn a state space model of the system using the observations and control signals, which are assumed to be available to the attacker. First, we analyze whether an attacker can learn a state space model close enough to the real system for a stealthy attack to be successful without having any knowledge about the parameters of the system, i.e. using a black box approach to system identification. Next, we consider that it is reasonable to assume that the attacker would know some information about the system they are attacking. Specifically, we assume that the attacker knows the general form of the differential equations used to describe the dynamics of the system, as well as a set of upper and lower bounds for the parameters in these equations. A grey box approach is then used for the attacker to learn a state space model of the system and the probability of the attack being successful is calculated.

We consider making small changes to the system in order to reduce the probability of a successful attack in Chapter 5. First, we consider the trade off between making the system less susceptible to stealthy attacks and maintaining controllability by increasing the sampling

time of the system in order to provide the attacker with fewer samples to learn a state space model of the system in a given amount of time. Next, a three inverter power system model is used to simulate modifying the parameters of the system in the design phase in order to make a stealthy attack impossible. We determine a range of possible values for the parameters of the system that make a stealthy attack impossible. Additionally, we show that applying a stealthy attack to a state space model that only represents a subset of the power system can still cause the entire system to fail.

1.3 Summary of Contributions

This dissertation makes the following contributions:

- We analyze the impact of the attacker having imperfect model information on the Quadruple-Tank Process (QTP) as well as a power system model.
- We determine the probability of an attack being successful with the attacker having varying levels of uncertainty of the system and gain insight on how close an attacker needs to learn a state space model of the system in order for their attack to be successful.
- We analyze the impact that limiting the energy an attacker can inject into the system within realistic bounds has on the probability of an attack being successful, showing that choosing to not overbuild a system can decrease the probability of an attack being successful.
- We utilize system identification to model the attacker learning a state space model and then use this learned model to attack the QTP.
- We show that small improvements to the anomaly detector used can make a big impact on decreasing the probability of an attack being successful.
- We show that increasing the sampling time used to discretize the system in order

to prevent the attacker from learning an accurate state space model can drastically decrease the probability of a successful attack.

- Finally, we show, on a power system model, that making small changes to the parameters of the system in the design phase can eliminate the possibility of a stealthy attack.

1.4 Acknowledgements

This work was supported in part by the U.S. National Science Foundation under grants No. 1527579 and 1619201, as well as by grant No. NPRP12S-0226-190158 from the Qatar National Research Fund (a member of the Qatar Foundation). A part of this work is published in Harshbarger et al. [2020]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Chapter 2

The Impact of Imperfect Model Information on Stealthy Attacks

2.1 Introduction

This chapter considers stealthy attacks under more realistic conditions than previous work. Under more realistic conditions, a control system will have process and measurement noise as well as a controller used to push the states of the system to a desired value. We show the derivation of the equations for a Linear Quadratic Gaussian (LQG) controller that push the states of the system to a desired value rather than 0. Additionally, an attacker would most likely not have the opportunity to start their attack at time $t = 0$s of the system running, so we consider the attacker mounting their attack at some point after the system has reached a steady state. We also consider that the attacker would not know the exact parameters of the system, so they would have to learn a state space model in order to calculate a stealthy attack. In this chapter, we simply add noise to the state matrix A in order to simulate an imperfect A that the attacker learned. Using these more realistic conditions for a stealthy attack, we analyze the impact they have on the stealthiness of the attack. Our results are demonstrated using the Quadruple-Tank Process as well as a simple power system model.

2.2 Problem Setup

2.2.1 System Model

The stealthy attack used here is executed on a discrete-time state space model. The following equations provide the states and outputs:

$$x_{k+1} = Ax_k + Bu_k + v_k, \tag{2.1}$$

$$y_k = Cx_k + w_k. \tag{2.2}$$

Where x_k, y_k, and u_k are the states, observations, and control signal at time k.

The controller used is a Linear Quadratic Gaussian (LQG) controller, which consists of a Kalman Filter and a Linear Quadratic Regulator (LQR). A Kalman Filter is used here for state estimation as well as the anomaly detector. In order to have the ability to drive the states of the system to a setpoint, an LQR is applied. The equations for the Kalman Filter are shown below:

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_k, \tag{2.3a}$$

$$P_{k|k-1} = AP_{k-1|k-1}A^T + V, \tag{2.3b}$$

$$\tilde{y}_k = y_k - C\hat{x}_{k|k-1}, \tag{2.3c}$$

$$S_k = CP_{k|k-1}C^T + W, \tag{2.3d}$$

$$L_k = P_{k|k-1}C^T S_k^{-1}, \tag{2.3e}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k\tilde{y}_k, \tag{2.3f}$$

$$P_{k|k} = (I - L_kC)P_{k|k-1}, \tag{2.3g}$$

where 2.3a is the predicted state estimate, 2.3b is the predicted estimate covariance, 2.3c is the measurement pre-fit residual, 2.3d is the pre-fit residual covariance, 2.3e is the optimal

Kalman gain, 2.3f is the updated state estimate, 2.3g is the updated estimate covariance, and V and W are the covariance matrices for the process and measurement noise.

Some modifications are made to the traditional LQR in order to drive the states of the system to a setpoint other than 0. Using solutions to the matrix Riccatti difference equation in order to obtain the optimal control signal, we get H, G, and T, which are defined using backwards recursion:

$$H_k = H_N + H_{k+1}A - H_{k+1}B(B^T G_{k+1}B + R)^{-1}B^T G_{k+1}A, \tag{2.4}$$

$$G_k = A^T(G_{k+1} - G_{k+1}B(B^T G_{k+1}B + R)B^T G_{k+1})A + Q, \tag{2.5}$$

$$T_k = T_N + A^T T_{k+1} - A^T G_{k+1}B(B^T G_{k+1}B + R)^{-1}B^T G_{k+1}, \tag{2.6}$$

where Q and R are the state cost and input cost matrices and N is the size of the finite horizon. Now, we define the feedback gain matrix K for the LQR shown in (2.7):

$$K_k = (B^T G_{k+1}B + R)^{-1}BG_{k+1}A. \tag{2.7}$$

The updated state estimate from the Kalman Filter and the feedback gain from the LQR are then used to calculate the control signal shown in (2.8):

$$u_k = -K_k\hat{x}_{k|k} + \frac{1}{2}(B^T G_{k+1}B + R)^{-1}(B^T H_{k+1}^T + B^T T_{k+1}). \tag{2.8}$$

A detailed derivation of the equations used in the LQR can be found in Appendix A. Additionally, we initialize $H_N = x_d^T Q$, $G_N = Q$, and $T_N = Qx_d$, where x_d are the desired states of the system.

2.2.2 Attacker Model

Modern industrial control systems are using increasingly many smart sensors that require communication over a network. These additional communication channels provide an attacker with numerous points to attack the system. We investigate stealthy attacks using the

stealthy attack model described in Teixeira et al. [2012].

Stealthy attacks are a subset of false data injection attacks that are specifically designed to go undetected. The attacker injects the attack signal a_k at time k into the control signal u_k to obtain a new state equation

$$x_{k+1} = Ax_k + B(u_k + a_k) + v_k, \tag{2.9}$$

where a_k is chosen such that the output of the system y_k does not change when the attack is added. Specifically, a_k is calculated by

$$a_k = Fz_k, \tag{2.10}$$

where F is chosen to such that $(A + BF)V^* \subseteq V^*$, where V^* is the maximal output-nulling invariant subspace of the system. V^* can be computed using the algorithm provided in Anderson [1975]. Additionally, F is obtained using A, B, and V^* (since $D = 0$) with the following equation:

$$\begin{bmatrix} F_1V \\ F_2V \end{bmatrix} = \begin{bmatrix} V & B \\ 0 & D \end{bmatrix}^{-1} \begin{bmatrix} A \\ C \end{bmatrix} V, \tag{2.11}$$

where V is a matrix whose columns are a basis for V^* and the F used to calculate a stealthy attack is denoted by F_1 in (2.11). z_k is defined by the following recursive equation:

$$z_k = (A + BF)z_{k-1}, \tag{2.12}$$

and z_0 is chosen to be the Perron eigenvector of $A+BF$. For a stealthy attack to diverge the states of the system, the system must be of non-minimum phase. In discrete-time systems, this means that at least one of the zeros of the system must be outside of the unit circle. The zeros of the system correspond to the eigenvalues of $A + BF$, meaning $A + BF$ must have at least one unstable eigenvalue for a stealthy attack to be successful.

2.3 Simulation Results on the Quadruple-Tank Process

2.3.1 System Description

Our results are demonstrated using the quadruple-tank process (QTP) Johansson [2000], illustrated in Figure 2.1. The QTP is a system consisting of four interconnected water tanks, where the water levels in the tanks are controlled by the voltage applied to each pump. The following differential equations are used to model the system:

$$\begin{aligned}
\frac{dx_1}{dt} &= -\frac{a_1}{A_1}\sqrt{2gx_1} + \frac{a_3}{A_1}\sqrt{2gx_3} + \frac{\gamma_1 k_1}{A_1}v_1, \\
\frac{dx_2}{dt} &= -\frac{a_1}{A_2}\sqrt{2gx_2} + \frac{a_4}{A_2}\sqrt{2gx_4} + \frac{\gamma_2 k_2}{A_2}v_2, \\
\frac{dx_3}{dt} &= -\frac{a_3}{A_3}\sqrt{2gx_3} + \frac{(1-\gamma_2)k_2}{A_3}v_2, \\
\frac{dx_4}{dt} &= -\frac{a_4}{A_4}\sqrt{2gx_4} + \frac{(1-\gamma_1)k_1}{A_4}v_1,
\end{aligned}$$

where x_i is the water level of tank $i \in \{1, 2, 3, 4\}$, A_i is the cross-section of tank i, a_i is the cross-section of the outlet hole, v_j is the voltage applied to pump $j \in \{1, 2\}$, with flow $k_j v_j$, γ_1, $\gamma_2 \in (0, 1)$ are the flow ratios, and g is the acceleration due to gravity. The outputs of the system y_1, y_2 are the measured water levels in tanks 1 and 2. The linearized, discrete-time state space model is shown below, with a sampling time of 0.5s:

$$A = \begin{bmatrix} 0.975 & 0 & 0.042 & 0 \\ 0 & 0.977 & 0 & 0.044 \\ 0 & 0 & 0.958 & 0 \\ 0 & 0 & 0 & 0.956 \end{bmatrix},$$

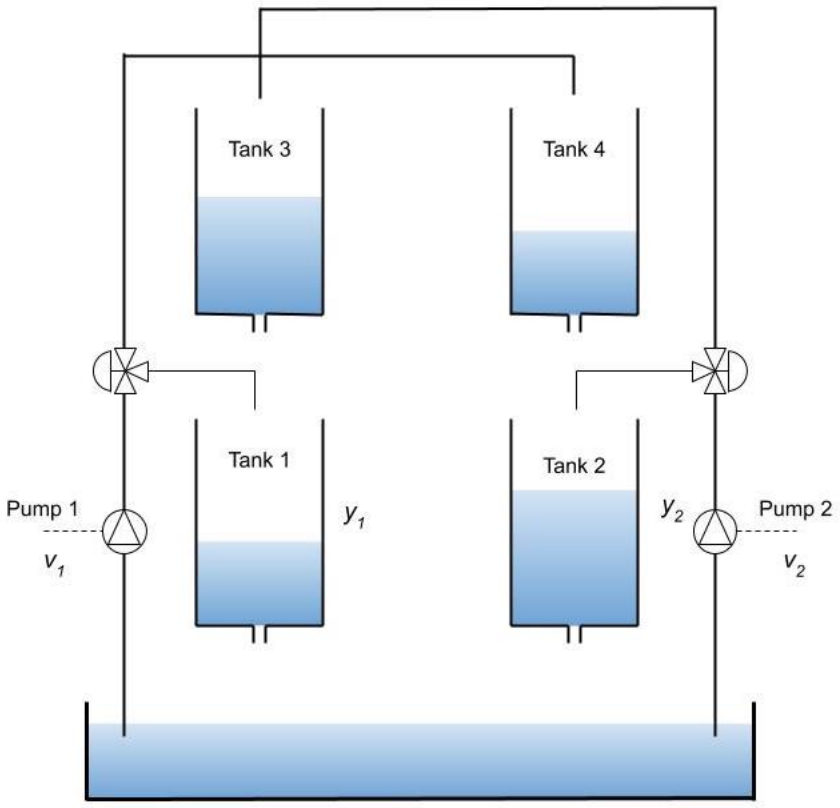


Figure 2.1: The quadruple-tank process: a series of four connected water tanks where the water levels are controlled by the voltage applied to each of the two pumps.

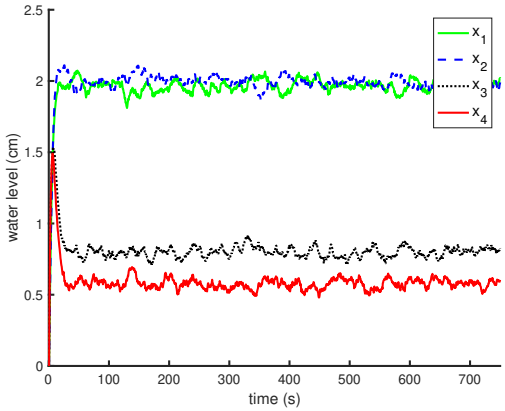
$$B = \begin{bmatrix} 0.0515 & 0 \\ 0 & 0.0447 \\ 0 & 0.0737 \\ 0.085 & 0 \end{bmatrix},$$

$$C = \begin{bmatrix} 0.2 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \end{bmatrix}.$$

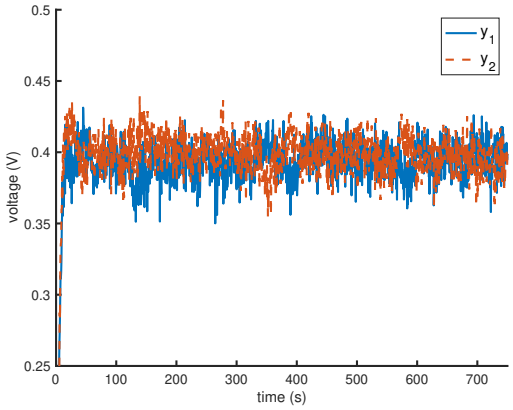
In order to initialize the LQG controller, an initial state and desired state needs to be specified. We let the initial state be $x_0 = \hat{x}_0 = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T$ and the desired state be $x_d = \begin{bmatrix} 2 & 2 & 2 & 2 \end{bmatrix}^T$. Since the water levels in tanks 3 and 4 are not measured and are dependent on the water levels in tanks 1 and 2, it does not matter what we choose for the desired states of tanks 3 and 4.

2.3.2 Simulation Results

We first run the model with the previously described LQG controller and with no attack. Figure 2.2 shows the states and observations of the QTP running under normal operating conditions. We can see that tanks 1 and 2 converge to their desired water levels of 2cm. This will be used as a baseline to compare with the following experiments. The initial state does not matter for this system, as the system will always converge to a steady state. In the next experiment, we want to see if an attack can remain stealthy with the addition of the LQG controller. For the attack to remain stealthy, the output of the system should remain the same in order to avoid triggering an alarm. Figure 2.3 shows the states and observations of the system with the attacker having perfect knowledge of the state space model of the system. The start of the attack is represented with a black vertical line at 250s. We can see that the observations remain the same as the system with no attack, while the water level in tank 3 is rising quickly, and tank 4 is completely empty less than a minute after the attack starts. Since there is no change in the observations of the system, the attack is stealthy.



(a) The states of the QTP, where x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4.

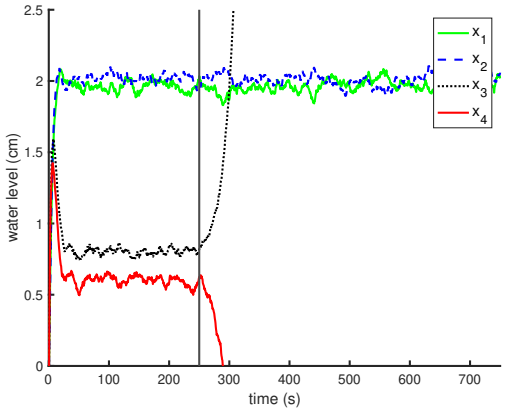


(b) The observations of the QTP, where y_1, y_2 are the voltages in V from the level measurement devices in tanks 1 and 2.

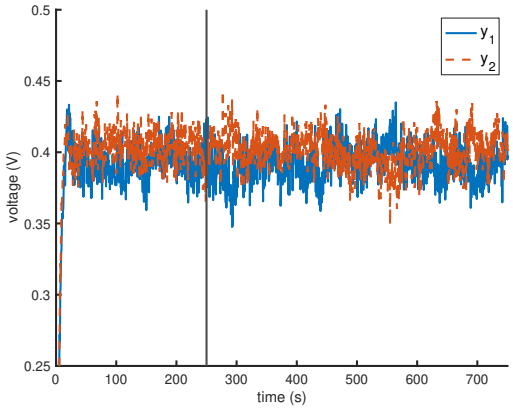
Figure 2.2: The states and observations of the QTP simulation under normal operating conditions. The process and measurement noise added to the system is white noise with a variance of 10^{-4}.

Additionally, we can see that the LQG controller has no impact on the stealthiness of the attack – meaning that a stealthy attack is feasible with the addition of a realistic controller. Note that the attack only runs for 750s because after about 1000s the attack starts to become observable, meaning it would be detected by any anomaly detector, as the states diverge far from the operating points around which the system was linearized. Thus, we only allow the simulation to run for 1000s as to not confuse an attack becoming observable due to the attacker having imperfect knowledge of the system with the attack becoming observable due to the system moving too far away from the operating point.

Next, we demonstrate an attacker with imperfect knowledge of the state matrix A. This usually means that the attacker could choose to allocate some time towards learning better estimates of these values in order to create a more "stealthy" attack. Here, we assume that the attacker is not able to learn these matrices perfectly – specifically, there is some ΔA that describes the attacker's uncertainty with respect to A. We will only consider the case where the attacker knows everything about the system except for the matrix A. We obtain a new state matrix A' such that $A' = A + \Delta A$, where ΔA is zero mean white noise with a standard deviation of 0.1. Thus, ΔA adds noise to all elements of A, including the elements



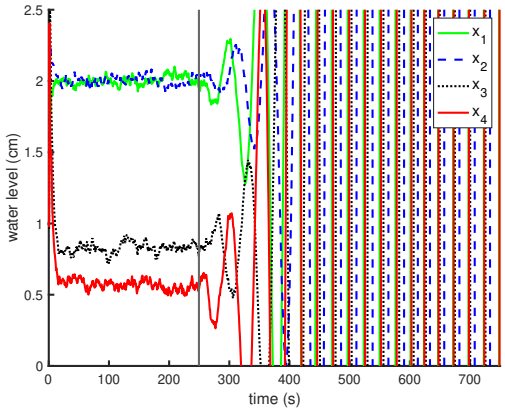
(a) The states of the QTP under attack, where x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4.



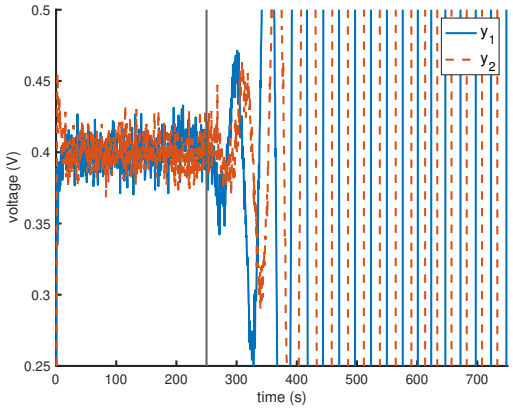
(b) The observations of the QTP under attack, where y_1, y_2 are the voltages in V from the level measurement devices in tanks 1 and 2.

Figure 2.3: The states and observations of the QTP simulation with an attack starting at 250s (denoted by the black vertical line). The process and measurement noise added to the system is white noise with a variance of 10^{-4}. Note that the water height in tank 3 is increasing and the water height in tank 4 is decreasing, while the observations remain the same as the simulation with no attack.

equal to 0. For this case we will also consider the attack starting at 250s, as this is a more realistic scenario. The motivation for this case was to test how quickly the system could become unstable having already reached a steady state. From Figure 2.4, we can see that once the attacker no longer has perfect knowledge of the system it loses its stealth, as the controller's observations start to diverge. We can see that the voltage being applied to the valves is oscillating rapidly causing the amount of water in each tank to also oscillate. This may cause harmful effects. The results are similar for the cases where other system matrices are unknown. For this experiment, the attack began at 250s. The system behaves normally and reaches a steady state before the attack begins. Shortly after the attack is mounted, the states of the system diverge to be unstable and the observations are no longer the same, meaning the attack is detectable. These results show that when an attacker does not know the state space model perfectly, a stealthy attack will not always be successful. Chapter 3 investigates exactly how close the attacker must learn the state space model in order to have a high probability of a successful stealthy attack.



(a) The states of the QTP under attack, with the attacker having imperfect knowledge of A, where x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4.



(b) The observations of the QTP under attack, with the attacker having imperfect knowledge of A, where y_1, y_2 are the voltages in V from the level measurement devices in tanks 1 and 2.

Figure 2.4: The states and observations of the QTP simulation with an attack starting at 250s (denoted by the black vertical line) and the attacker having imperfect knowledge of A, where ΔA is generated by drawing elements in an independent manner from a standard Gaussian distribution. Note that the observations are not the same as the simulation with no attack, so this attack is no longer stealthy. The states and observations oscillate after diverging due to numerical precision, as the attack diverges the states exponentially.

2.4 Simulation Results on a Power System Model

2.4.1 System Description

In addition to the QTP, we consider a stealthy attack on a simple power system model. This is a three inverter model, where the attacker manipulates the active and reactive power set points of the PV inverters. Figure 2.5 shows the three inverter power system model at the point of PV generation.

We assume that the majority of the local loads are supplied from the PV inverters' active and reactive power. In order to remain undetected, an attacker can manipulate the PV inverter's reference currents such that demand-supply of the system does not become unbalanced. As the grid supervisor is concerned with the demand-supply ratios, the attack would be undetectable. The attacker will need to compromise the majority of the PV inverters by causing them to trip due to being outside of the safe operating regions. This

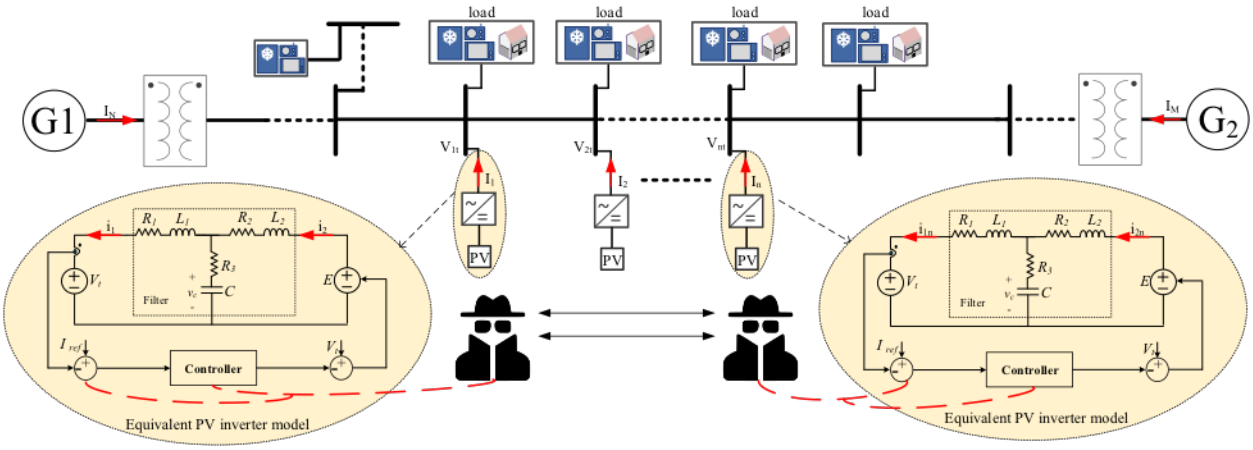


Figure 2.5: Distribution power system with high penetration of PV inverters.

would cause a large unbalance in the demand-supply as well as the three-phase distribution feeder–causing a blackout.

The continuous time state space representation of a single PV inverter system can be found in Equations (2.13) and (2.14).

$$
\begin{bmatrix} \dot{i}^1_{1,p} \\ \dot{i}^1_{1,q} \\ \dot{i}^1_{2,p} \\ \dot{i}^1_{2,q} \\ \dot{v}_{c_p} \\ \dot{v}_{c_q} \\ \dot{x}_{1_p} \\ \dot{x}_{1_q} \\ \dot{x}_{2_p} \\ \dot{x}_{2_q} \end{bmatrix} =
\begin{bmatrix}
\frac{-R_1-R_3}{L_1} & \dot{\theta} & \frac{R_3}{L_1} & 0 & \frac{1}{L_1} & 0 & 0 & 0 & 0 & 0 \\
-\dot{\theta} & \frac{-R_1-R_3}{L_1} & 0 & \frac{R_3}{L_1} & 0 & \frac{1}{L_1} & 0 & 0 & 0 & 0 \\
\frac{R_3-K_P}{L_2} & 0 & \frac{-R_2-R_3}{L_2} & \dot{\theta} & \frac{-1}{L_2} & 0 & \frac{K_R}{L_2} & 0 & 0 & 0 \\
0 & \frac{R_3-K_P}{L_2} & -\dot{\theta} & \frac{-R_2-R_3}{L_2} & 0 & \frac{-1}{L_2} & 0 & \frac{K_R}{L_2} & 0 & 0 \\
\frac{-1}{C} & 0 & \frac{1}{C} & 0 & 0 & \dot{\theta} & 0 & 0 & 0 & 0 \\
0 & \frac{-1}{C} & 0 & \frac{1}{C} & -\dot{\theta} & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & -0.6 & \dot{\theta} & -277.6 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & -\dot{\theta} & -0.6 & 0 & -277.6 \\
0 & 0 & 0 & 0 & 0 & 0 & 512 & 0 & 0 & \dot{\theta} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 512 & -\dot{\theta} & 0
\end{bmatrix}
\begin{bmatrix} i^1_{1,p} \\ i^1_{1,q} \\ i^1_{2,p} \\ i^1_{2,q} \\ v_{c_p} \\ v_{c_q} \\ x_{1_p} \\ x_{1_q} \\ x_{2_p} \\ x_{2_q} \end{bmatrix}
+
\begin{bmatrix}
\frac{-1}{L_1} & 0 & 0 & 0 \\
0 & \frac{-1}{L_1} & 0 & 0 \\
\frac{1}{L_2} & 0 & \frac{K_P}{L_2} & 0 \\
0 & \frac{1}{L_2} & 0 & \frac{K_P}{L_2} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix} V_p \\ V_q \\ I^*_p \\ I^*_q \end{bmatrix}
\tag{2.13}
$$

Parameter	Value
R_1	0.15Ω
R_2	0.03Ω
R_3	0.008Ω
L_1	$5mH$
L_2	$0.25mH$
C	$13.2\mu F$
K_P	0.02×10^7
K_R	83.3
$\dot{\theta}$	120π

Table 2.1: Values for the constants in the state space representation.

$$\begin{bmatrix} y_{eq_p} \\ y_{eq_q} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} i^1_{1,p} \\ i^1_{1,q} \\ i^1_{2,p} \\ i^1_{2,d} \\ v_{c_p} \\ v_{c_q} \\ x_{1_p} \\ x_{1_q} \\ x_{2_p} \\ x_{2_q} \end{bmatrix} \tag{2.14}$$

In this model, $i^k_{j,p}$, $i^k_{j,q}$ are the in-phase components of the current injected in the grid by inverter number k in the dq frame. v_{c_p} and v_{c_q} are the filter capacitor voltages in the dq frame, $\dot{\theta}$ is the frequency inside the PV inverter, and K_P and K_R are the controller gains of the PR controller in the current control loop. The values of these filter parameters can be found in Table 2.1.

The state space representation of the whole system, including three PV inverters, is given

by

$$A_{eq} = \begin{bmatrix} \begin{bmatrix} A_1 \end{bmatrix} & \begin{bmatrix} 0 \end{bmatrix} & \begin{bmatrix} 0 \end{bmatrix} \\ \begin{bmatrix} 0 \end{bmatrix} & \begin{bmatrix} A_2 \end{bmatrix} & \begin{bmatrix} 0 \end{bmatrix} \\ \begin{bmatrix} 0 \end{bmatrix} & \begin{bmatrix} 0 \end{bmatrix} & \begin{bmatrix} A_3 \end{bmatrix} \end{bmatrix},$$

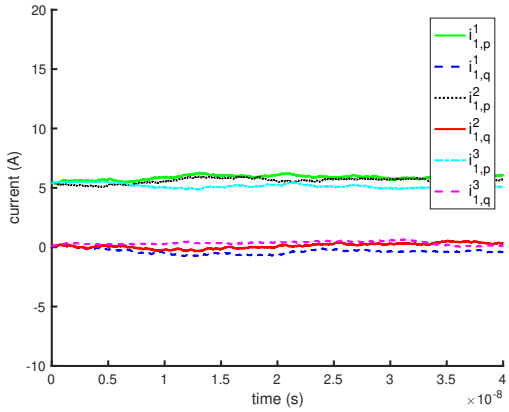
$$B_{eq} = \begin{bmatrix} \begin{bmatrix} B_1 \end{bmatrix} & 0 & 0 \\ 0 & \begin{bmatrix} B_2 \end{bmatrix} & 0 \\ 0 & 0 & \begin{bmatrix} B_3 \end{bmatrix} \end{bmatrix},$$

$$C_{eq} = \begin{bmatrix} \begin{bmatrix} C \end{bmatrix} & \begin{bmatrix} C \end{bmatrix} & \begin{bmatrix} C \end{bmatrix} \end{bmatrix},$$

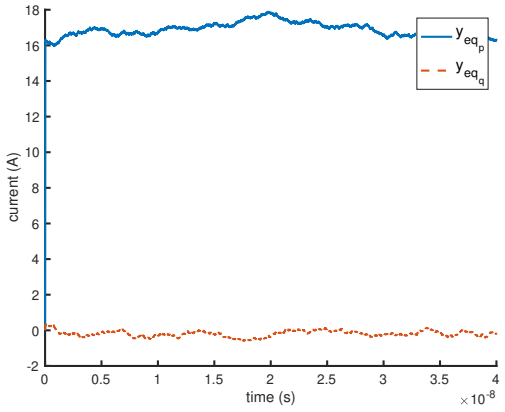
where, $\begin{bmatrix} A_i \end{bmatrix}$, $i = 1, 2, 3$, is the state matrix for each individual inverter. The representation of input and output matrices are done similarly. It can be seen in C_{eq} that only the sum of the currents injected in the grid by all three inverters is observed. Here, $A_{eq} \in \mathbb{R}^{30\times30}$, $B_{eq} \in \mathbb{R}^{30\times30}$, and $C_{eq} \in \mathbb{R}^{2\times30}$. The control input signal of the system is the set of all three PV inverter terminal voltages and all three PV inverter reference currents. This state space model is then transformed to discrete time with a sampling time of $1ps$. This sampling time was chosen in order to speed up the effect of the attack. The states of the system under attack are able to diverge much faster with a smaller sampling time.

2.4.2 Simulation Results

Similarly to the quadruple-tank process, we will first show the system under no attack to use as a baseline to compare with the system under attack. We will then show the system with an attack starting at time 500, and finally, we will show the system under attack starting at time 500 and the attacker having imperfect knowledge of the system. For each experiment, we use $x_0 = \hat{x}_0 = \begin{bmatrix} 5.43 & 0.105 & 5.02 & 0.236 & 352 & 0.719 & 0 & 0 & 0 & 0 \end{bmatrix}^T$. The desired state for each inverter is set to $x_{d_1} = \begin{bmatrix} i_{1_d} & i_{1_q} \end{bmatrix}^T = \begin{bmatrix} 6 & 0.5 \end{bmatrix}^T$, $x_{d_2} = \begin{bmatrix} 7 & 1.5 \end{bmatrix}^T$, $x_{d_3} = \begin{bmatrix} 8 & 2.5 \end{bmatrix}^T$.



(a) The states of the system under normal operating conditions.



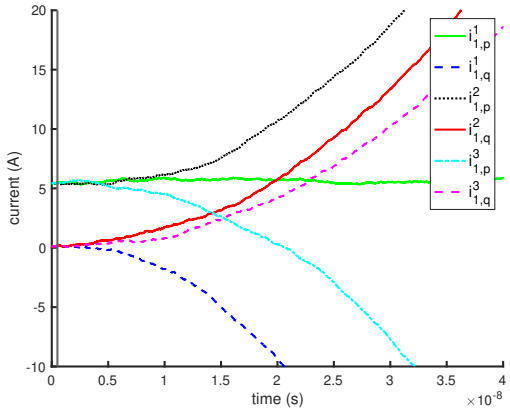
(b) The observations of the system under normal operating conditions.

Figure 2.6: The states and observations of the three inverter power system model under normal operating conditions.

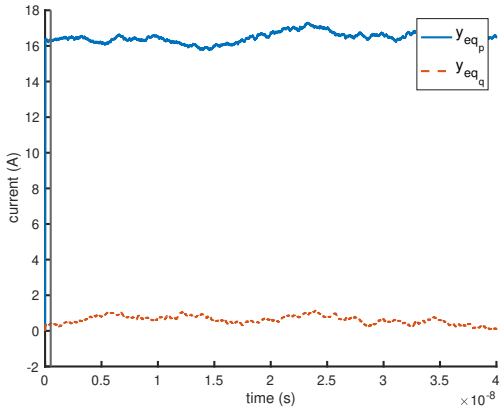
From Figure 2.6, we can see that the states of the system converge to the desired values. The currents fluctuate around the desired values for the states, as these states are not directly controllable. However, it can be seen that the values of these states do not diverge. Only the currents are shown in the graphs, as those are the parameters we are trying to control.

Next, we run the model with an attack starting at time 500. It can be seen in Figure 2.7 that the observations remain the same as the system with no attack, however the currents of the third inverter diverge, demonstrating a stealthy attack on the system.

We can see from Figure 2.8 that when the attacker no longer has perfect knowledge of the system, the observations do not remain the same – the attack is no longer stealthy and would be detected. The attacker's uncertainty of the system is modeled in the same way as the quadruple-tank process – the attacker has perfect knowledge of the system except for the matrix A.

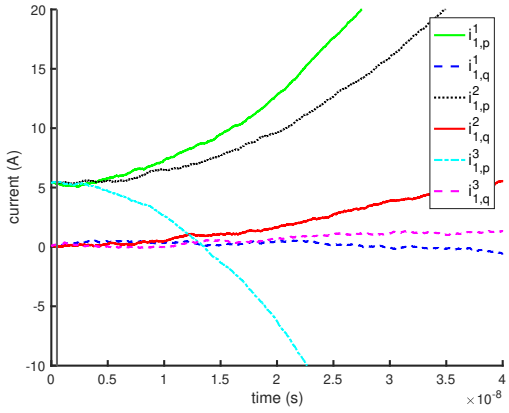


(a) The states of the system under attack.

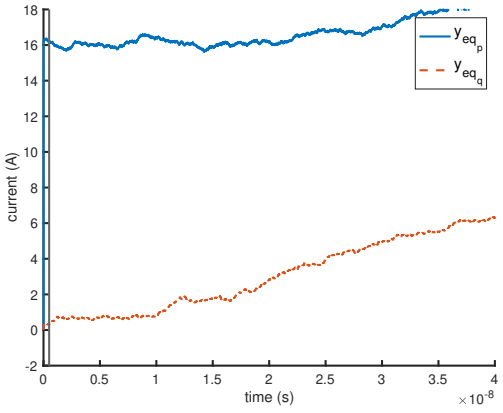


(b) The observations of the system under attack.

Figure 2.7: The states and observations of the three inverter power system model with an attack starting at 0.5ns (denoted by the black vertical line). Note that the currents for the third inverter are diverging, while the observations remain the same as the simulation with no attack.



(a) The states of the system under attack with the attacker having imperfect knowledge of A.



(b) The observations of the system under attack with the attacker having imperfect knowledge of A.

Figure 2.8: The states and observations of the three inverter power system model with an attack starting at 0.5ns (denoted by the black vertical line) with the attacker having imperfect knowledge of A. Note that the observations diverge, which would cause the attack to be detectable.

Chapter 3

The Probability of a Successful Stealthy Attack with Imperfect Model Information

3.1 Introduction

In this chapter, we consider the probability of a stealthy attack being successful with varying levels of the attacker's uncertainty in the state space representation, as well as what causes an attack to be successful or unsuccessful. In order to calculate the probability of an attack being successful, we first need to define how the attacker's uncertainty in the state space model is represented, and what type of anomaly detector is used. The attacker's uncertainty in the state space model is represented by simply perturbing the values of the state and input matrices A and B, respectively. Additionally, we use a windowed χ^2 variance test to detect changes in the variance of the observations. We want to determine how close to the actual system the attacker needs to learn the state space model in order to have a high probability of the attack being successful, as this information can be used to attempt to limit the attacker's ability to learn an accurate state space representation. Additionally, we consider the impact that limiting the energy the attacker can inject into the system has on

the probability of the attack being successful. Realistically, the systems would have physical limits on the amount of control signal energy that it can tolerate without causing damage to the physical components, and may even have relays or other ways to protect the hardware from operating outside of its limits. Thus, we consider the impact of halting the attack when these physical limits are reached.

3.2 Problem Setup

3.2.1 Perturbations of the State Space Matrices

We assume that the attacker does not have perfect knowledge of the system matrices. This is because the attacker would have to know very specific information about every component of the system in order to calculate an accurate state space model of the system. In reality, the attacker would have to learn this information using the previous control signals and observations. We use A' and B' to represent the attacker's versions of the system's A and B matrices. We assume that the attacker knows C perfectly, as they should know what states are not being measured in order to diverge these states with an attack. To simulate the attacker having imperfect knowledge of the state space model, we perturb the state and input matrices by adding element-wise noise to A and B with the noise having a standard deviation equal to a percentage of the actual values of the matrix value. The values in A are perturbed by ΔA giving us $A' = A + \Delta A$, where ΔA is a $n \times n$ matrix of independent normal data with mean 0 and component-wise standard deviations $\sigma_A \times A$. Using the QTP specified in Section 2.3, if we perturb the values of A with noise having a standard deviation of 5% of the actual values of A, then we could obtain an A' shown by (3.1), where σ_A is the percentage of A, that is the standard deviation of the noise added to A and $\odot$ is element-wise multiplication. The right-hand-side matrix is generated by drawing elements in an independent manner from a standard Gaussian distribution. This perturbation is similar to the attacker's imperfect knowledge of A in Section 2.3; however, in this chapter the standard deviation of the noise added to A and B is proportional to their actual values.

$$
\begin{aligned}
A' &= A + \Delta A \\
&= \begin{bmatrix} 0.975 & 0 & 0.042 & 0 \\ 0 & 0.977 & 0 & 0.044 \\ 0 & 0 & 0.958 & 0 \\ 0 & 0 & 0 & 0.956 \end{bmatrix} \\
&\quad + 0.05 \times \begin{bmatrix} 0.975 & 0 & 0.042 & 0 \\ 0 & 0.977 & 0 & 0.044 \\ 0 & 0 & 0.958 & 0 \\ 0 & 0 & 0 & 0.956 \end{bmatrix} \odot \begin{bmatrix} 1.992 & -0.146 & -0.304 & -2.168 \\ 0.457 & 1.330 & 0.465 & -0.281 \\ 0.614 & -1.333 & -0.117 & 1.713 \\ -0.525 & 0.758 & -0.733 & -0.797 \end{bmatrix} \\
&= \begin{bmatrix} 1.072 & 0 & 0.041 & 0 \\ 0 & 1.042 & 0 & 0.043 \\ 0 & 0 & 0.952 & 0 \\ 0 & 0 & 0 & 0.918 \end{bmatrix}
\end{aligned} \tag{3.1}
$$

3.2.2 Anomaly Detector

When the attacker does not have an accurate state space model of the system, a stealthy attack is not always successful. We investigate the probability that the attacker is successful as the attacker's uncertainty in the state space model increases. A stealthy attack is defined as successful if the attacker is able to diverge the states to a point that causes an abrupt change in the system model. In the QTP this means that at least one tank is empty or overflows, before the attack is detected. For the anomaly detector, we use a windowed χ^2 variance test where the test statistic is given by

$$
T = (n-1)\frac{s^2}{\sigma^2}, \tag{3.2}
$$

where n is the size of the window, s^2 is the sample variance of the measurement pre-fit residual from the Kalman filter over the previous n time steps and σ^2 is the innovation covariance provided by the Kalman Filter covariance matrix S at the current time. This anomaly detector is applied to the simulation in Figure 2.4. The results of the χ^2 anomaly detector, with a false alarm rate of $\alpha = 0.5\%$ are shown in Figure 3.1. In this graph, a χ^2 index below the critical value, depicted by the green horizontal line in the figure, means that the observations are considered normal behavior, and a χ^2 index above the critical

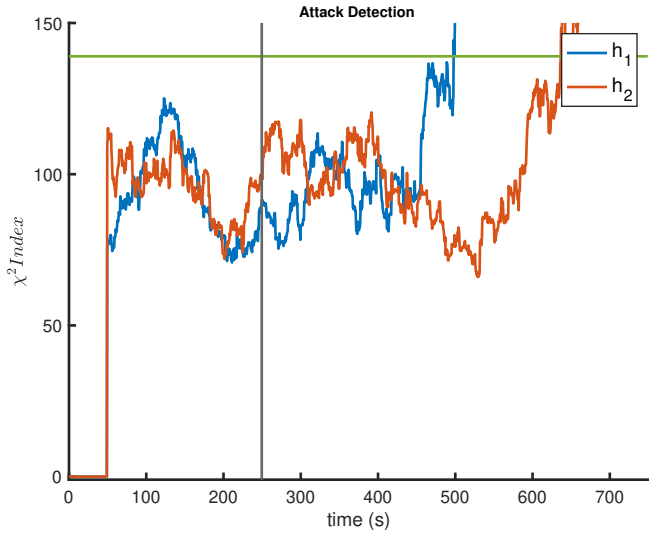


Figure 3.1: The χ^s variance test results of the QTP simulation with an attack starting at 250s (denoted by the black vertical line) and the attacker having imperfect knowledge of A. The green horizontal line signifies the critical value for the variance test. The standard deviation of the noise added to A is 5% of the values of A. The false alarm rate α is set to 0.5%, meaning there would be about 5 false alarms for every 1000 time steps, and the sliding window is 100 time steps. We can see that the first anomaly is detected at 498.5s.

value means the observations are considered abnormal and an anomaly is detected. The first anomaly is detected at time 498.5s. This means the attack was detected 94s after the attacker emptied tank 4, thus the attack is considered successful. A limitation of this attack is that it is only run for 500s, meaning the attacker only has a few minutes to empty or overflow a tank. This time restriction is due to the fact that the large divergence causes an abrupt change to the system model, thus making the previously accurate state space model no longer an accurate representation of the system. The time the attack can be run will vary depending on the specific system and state space model.

3.2.3 Probability of a Successful Attack

In order to determine the percentage of successful attacks, we ran $10,000$ simulations of the system under attack with a fixed σ_A, and calculated the difference between the detection time and system failure time for each simulation. When the difference is greater than or equal to 0, the attack is considered to be successful. There are a few special cases to consider

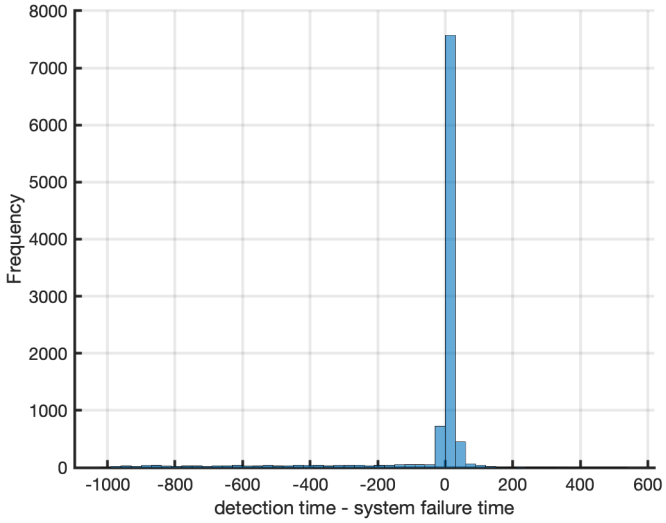


Figure 3.2: The difference between the detection time and system failure time over 10,000 simulations. We define the probability that the attack is successful as the percentage of simulations where the difference between the detection time and success time is greater than 0. We let $\sigma_A = 0.25$, $\alpha = 0.5\%$, and the process and measurement noise variance is equal to 10^{-4}.

when calculating the difference:

1. the attacker diverges the states within $1,000$ time steps of the attack starting, but is never detected, which we define as a successful attack;

2. the attack is detected before the attacker is able to diverge the states, which we define as an unsuccessful attack;

3. the attacker is not able to diverge the states within $1,000$ time steps and the attack is not detected, which we define as an unsuccessful attack.

Figure 3.2 shows a histogram of the difference between the detection time and system failure time for $10,000$ simulations. Note that when the difference is greater than 0, the attack is considered to be successful.

3.3 Simulation Results

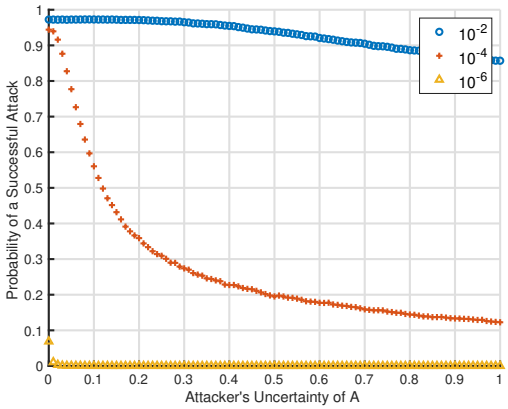
The difference between the detection time and success time of an attack is calculated for 10,000 simulations for the attacker's uncertainty in the state space model ranging from 0 to 1. We investigate the impact the attacker's uncertainty in A, B, or A and B has on the probability of the attack being successful.

3.3.1 Attacker's Imperfect Knowledge of A

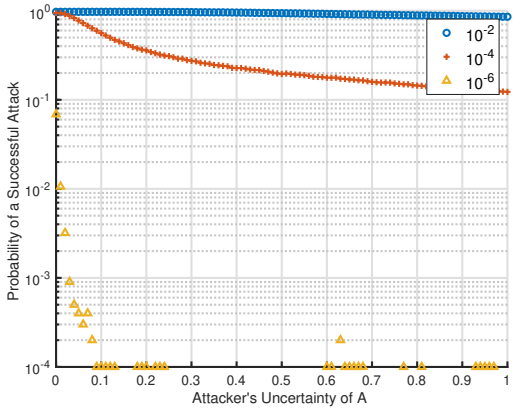
After calculating the difference between the detection time and system failure time for 10,000 simulations for a specific σ_A, the attacker's uncertainty of A, we calculate the probability that the attacker is successful by finding the percentage of simulations where the difference between the detection time and system failure time is greater that 0. We repeat this process for σ_A ranging from 0 to 1 at an interval of 0.01. Because process and measurement noise and the false alarm rate α have a significant impact on the anomaly detector, we vary these parameters in order to determine how close the attacker would need to learn A in order to have a specific successful attack probability.

Figure 3.3 shows the scatter plots of the probability of the attack being successful with varying process and measurement noises vs. σ_A. We can see that the attacker is most successful with the highest amount of noise in the system, and least successful with the lowest amounts of process and measurement noise. It is expected that more process and measurement noise corresponds to a higher probability of the attack being successful, and lower noise corresponding to a lower probability of the attack being successful, as we are using a variance based anomaly detector. This could suggest that it would be beneficial for a system to try to minimize the process and measurement noise in order to decrease the probability of a stealthy attack being successful. However, this comes at a cost, as the attacker would more accurately be able to learn the state space model with lower process and measurement noise.

Figure 3.4 shows the scatter plots of the probability of the attack being successful with a varying false alarm rate vs. σ_A. We can see that varying α does not have much of an impact



(a) Linear y-axis scale.



(b) Logarithmic y-axis scale; zero values omitted.

Figure 3.3: The probability of successful attack for varying process and measurement noise vs. the attacker's uncertainty in A, where the variance of the process and measurement noise added to the system varies from 10^{-6} to 10^{-2}. Where the blue circles represent noise with a variance of 10^{-2} being added to the system, the red plus signs represent noise with a variance of 10^{-4}, and the yellow triangles represent noise with a variance of 10^{-6}. Additionally, the false alarm rate α is set to 0.5%. We can see the more noisy the system is, the higher the probability that the attacker will have a successful attack.

on the probability of a successful attack. This is because the attacker is able to empty a tank very quickly, as the desired water heights of tanks 1 and 2 are 2cm and tanks 3 and 4 converge to water heights of about 1cm and 0.5cm, respectively. However, we believe that significantly increasing α could allow the detector to identify the attack in time. This also comes with a trade off, because the highest value of $\alpha = 1\%$ for these simulations would correspond to nearly $2,000$ false alarms in a 24 hour period. This means that even a false alarm rate of 1% would likely be too large for this system. Overall, it seems that modifying the variance of the process and measurement noise is the easiest and most realistic way to minimize the probability of an attacker having a successful attack. We can see that when the variance of the process and measurement noise is 10^{-4} and $\alpha = 0.5\%$, the attack has a 10% probability of being successful when the attacker learns A within 100% of its actual values. Even when the attacker does not know an accurate representation of the system, they still have a chance of the attack being successful. Additionally, this probability could be increased if the attacker made a few improvements to the attacker model. For example,

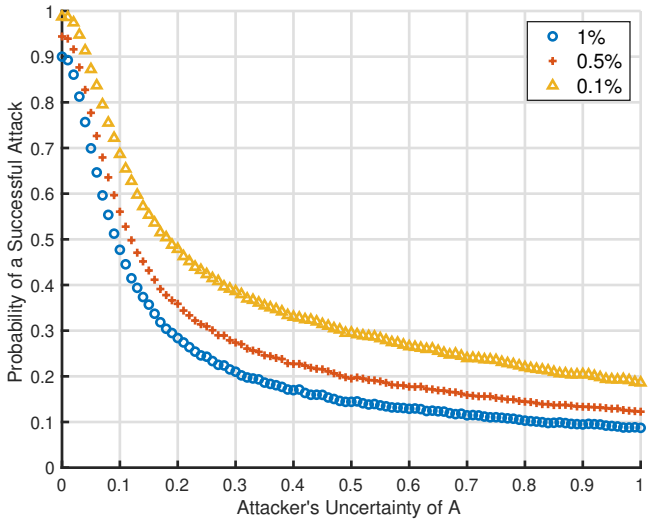
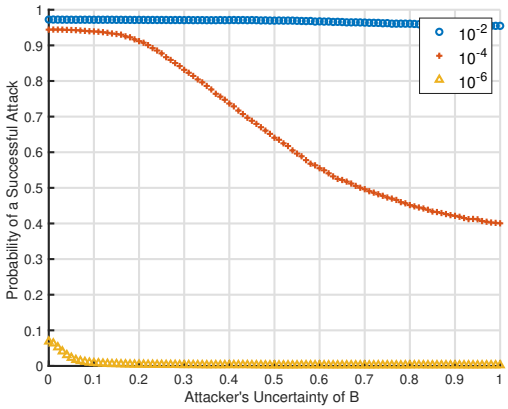


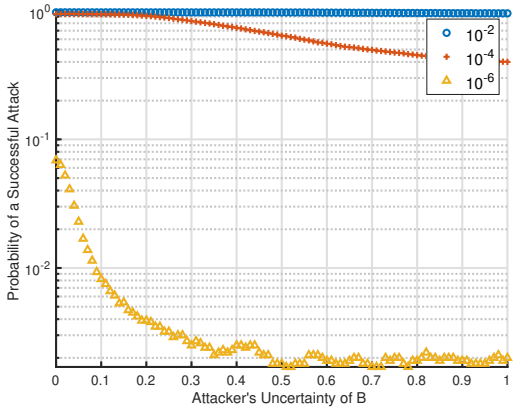
Figure 3.4: The probability that the attack is successful for varying false alarm rates α for the χ^2 variance test vs. the attacker's uncertainty in A. The false alarm rate α varies from 0.1% to 1%. Where the blue circles represent $\alpha = 1\%$, the red plus signs represent $\alpha = 0.5\%$, and the yellow triangles represent $\alpha = 0.1\%$. Additionally, the variance of the process and measurement noise added to the system is 10^{-4}. We can see that changing α does not have much of an impact on the attacker's probability of success.

the attacker could stop the attack once they saw that it would not be successful due to the observations diverging. They could then make some adjustments to their learned state space representation and try the attack again. Applying the attack in this manner would allow the attacker to not raise any alarms while they perfected the state space model needed for a successful stealthy attack. Thus, we will need to investigate measures that can be taken in order to minimize how accurately an attacker can learn the system.

Focusing on the case where the measurement and process noise of the system is 10^{-4} and $\alpha = 0.005$, we can see that the probability of a successful attack significantly decreases as the attacker's uncertainty in A increases. This means that it is very beneficial for the attacker to learn A as accurately as possible in order to improve their chances of a successful stealthy attack.



(a) Linear y-axis scale.



(b) Logarithmic y-axis scale; zero values omitted.

Figure 3.5: The probability of successful attack for varying process and measurement noise vs. the attacker's uncertainty in B. The variance of the process and measurement noise added to the system varies from 10^{-6} to 10^{-2}. Where the blue circles represent noise with a variance of 10^{-2} being added to the system, the red plus signs represent noise with a variance of 10^{-4}, and the yellow triangles represent noise with a variance of 10^{-6}. Additionally, the false alarm rate α is set to 0.5%. We can see the more noisy the system is, the higher probability the attacker will have a successful attack.

3.3.2 Attacker's Imperfect Knowledge of B

Similarly, the previous simulations are run with the attacker having imperfect knowledge of B. Figure 3.5 shows the scatter plots of the probability of the attack being successful with varying process and measurement noises vs. σ_B. These results are very similar to the attacker's probability of a successful attack when A is not perfectly known. The attacker is the most successful with larger magnitudes of process and measurement noise, and we can see that how close the attacker learns B does not greatly impact the attacker's probability of success. Additionally, we can see that even when the attacker has a very inaccurate estimate of B, they are still successful with their attack about 40% of the time. This could tell the attacker that if they have a limited amount of time to learn the state space model, they should focus on learning A accurately, as having an inaccurate estimate of A causes the probability of a successful attack to decrease much more than when having an inaccurate estimate of B.

The false alarm rate α is varied in Figure 3.6. We can see that as α increases, the

attacker's probability of a successful attack decreases. Even in the worst case for the attacker, when $\alpha = 1\%$ and the anomaly detector is the most sensitive to changes in the variance of the measurements, the attacker is still able to mount a successful attack about 30% of the time. Similar to the results in Section 3.3.1, α does not have much of an impact on the probability of the attacker having a successful attack. Realistically, α is not a parameter that can be increased in order to catch the rapidly diverging attack, as it would result in a disproportionate amount of false alarms.

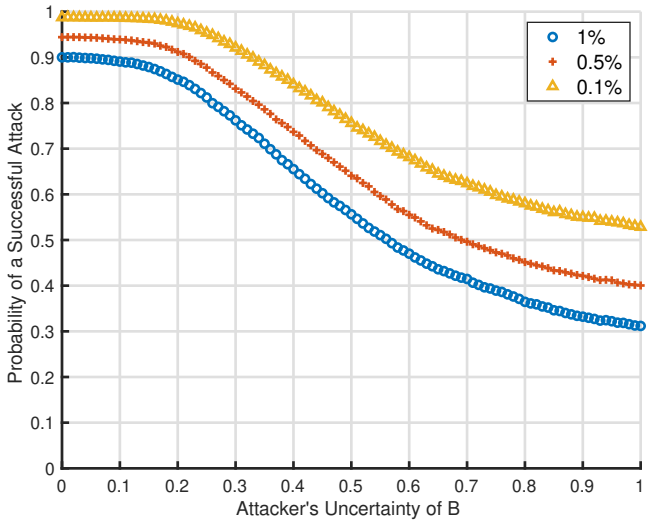
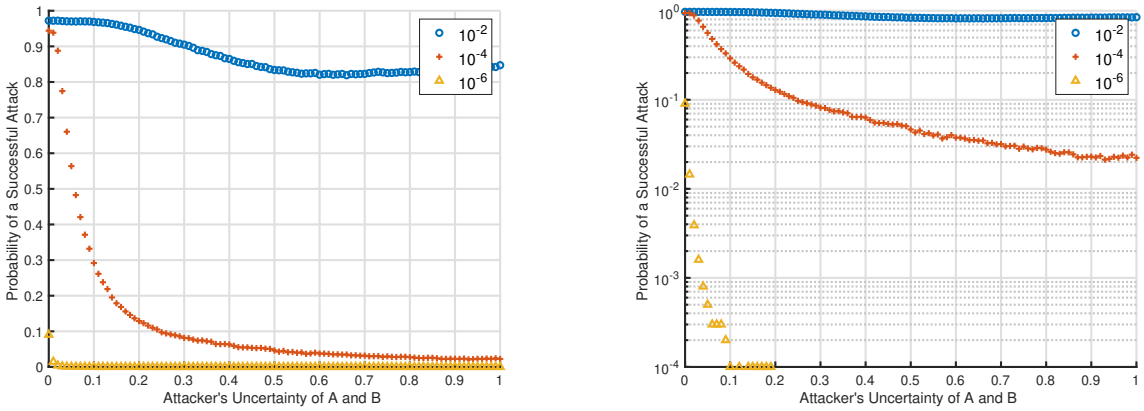


Figure 3.6: The probability that the attack is successful for varying false alarm rates α for the χ^2 variance test vs. the attacker's uncertainty in B. The false alarm rate α varies from 0.1% to 1%. Where the blue circles represent $\alpha = 1\%$, the red plus signs represent $\alpha = 0.5\%$, and the yellow triangles represent $\alpha = 0.1\%$. Additionally, the variance of the process and measurement noise added to the system is 10^{-4}. We can see that changing α does not have much of an impact on the attacker's probability of success.

3.3.3 Attacker's Imperfect Knowledge of A and B

We have observed the impact of the attacker having imperfect knowledge of A and the impact of the attacker having imperfect knowledge of B. However, the attack would realistically have to learn both A and B. Figure 3.7 shows the probability of a stealthy attack being successful for varying process and measurement noise. We can see that the attack is successful with a probability of 80% when the variance of the process and measurement noise added to the



(a) Linear y-axis scale.

(b) Logarithmic y-axis scale; zero values omitted.

Figure 3.7: The probability of successful attack for varying process and measurement noise vs. the attacker's uncertainty in both A and B. The variance of the process and measurement noise added to the system varies from 10^{-6} to 10^{-2}. Where the blue circles represent noise with a variance of 10^{-2} being added to the system, the red plus signs represent noise with a variance of 10^{-4}, and the yellow triangles represent noise with a variance of 10^{-6}. Additionally, the false alarm rate α is set to 0.5%. We can see the more noisy the system is, the higher probability the attacker will have a successful attack.

system is 10^{-2}. On the other end, the attack is successful with a probability of nearly 0% when the variance of the process and measurement noise is 10^{-6}. Additionally, we can see that in the middle case of the variance of the process and measurement noise being 10^{-4}, it is imperative that the attacker learns the state space model as accurately as possible. When the attacker knows the state space model perfectly, they are successful with a probability of about 95%. The probability of the attack being successful drops suddenly to about 30% when the attacker is able to learn the state space model within 10% of its actual values.

Compared to varying the process and measurement noise of the system, varying α has a much less significant impact on the probability of a stealthy attack being successful, shown in Figure 3.8. In each case of varying the false alarm rate, we can see that how accurately the attacker is able to learn the state space model has a large impact on the probability of the attack being successful–especially in the cases where the attacker is able to learn a model very close to the actual state space model. Overall, we can see that it is very beneficial for the attacker to learn an accurate state space model.

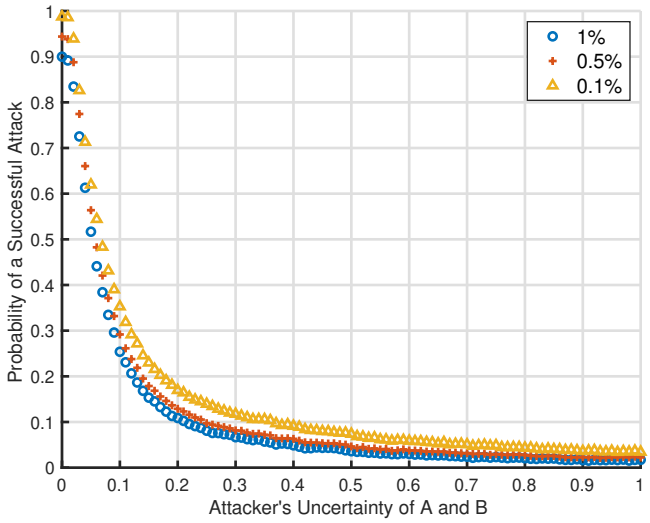


Figure 3.8: The probability that the attack is successful for varying false alarm rates α for the χ^2 variance test vs. the attacker's uncertainty in both A and B. The false alarm rate α varies from 0.1% to 1%. Where the blue circles represent $\alpha = 1\%$, the red plus signs represent $\alpha = 0.5\%$, and the yellow triangles represent $\alpha = 0.1\%$. Additionally, the variance of the process and measurement noise added to the system is 10^{-4}. We can see that changing α does not have much of an impact on the attacker's probability of success.

Since stealthy attacks are exponential, the desired states of the system do not impact the probability of a successful attack. Figure 3.9 shows the probability of a stealthy attack being successful with $x_d = \begin{bmatrix} 2 & 2 & 2 & 2 \end{bmatrix}^T$, $\begin{bmatrix} 10 & 10 & 10 & 10 \end{bmatrix}^T$, and $\begin{bmatrix} 18 & 18 & 18 & 18 \end{bmatrix}^T$. The variance of the process and measurement noise equal to 10^{-4} and the false alarm rate $\alpha = 0.5\%$. This desired state is chosen such that the desired water levels are right in the middle of the tanks, as the tanks are 20cm tall. We can see that these results are nearly identical to the case of $\alpha = 0.5\%$ in Figure 3.4. Thus, we can conclude that the desired state of the system does not have an impact on the probability of a stealthy attack being successful.

Additionally, we want to determine what properties of the learned state space representation causes an attack to be either successful or unsuccessful.

We believe that the main eigenvalue and eigenvector of $A' + B'F'$ have the biggest impact on the success of a stealthy attack, so these are the parameters that we will consider. When analyzing the impact of the main eigenvector, we will consider the dot product between the main eigenvector of $A + BF$ and $A' + B'F'$, which we will denote as θ. Additionally, we

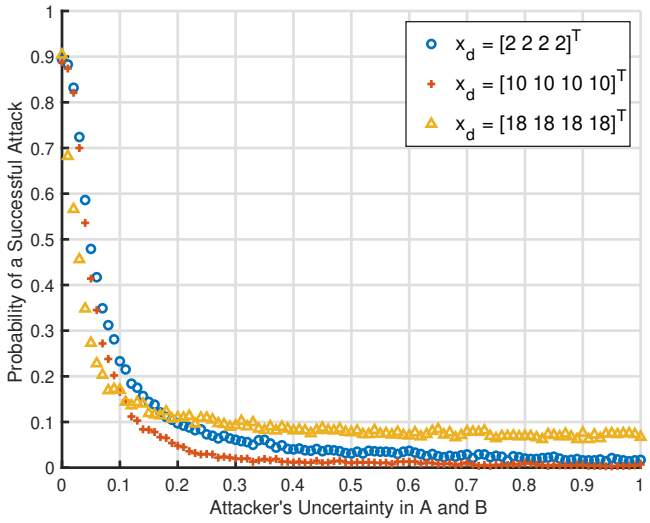


Figure 3.9: The probability that the attack is successful for varying false alarm rates α for the χ^2 variance test vs. the attacker's uncertainty in both A and B. The desired state is set to $x_d = \begin{bmatrix}10 & 10 & 10 & 10\end{bmatrix}^T$. The false alarm rate $\alpha = 0.5\%$ and the variance of the process and measurement noise added to the system is 10^{-4}. We can see that changing the desired state does not have a significant impact on the probability of a successful attack.

will consider the case where the variance of the process and measurement noise added to the system is 10^{-4} and the false alarm rate $\alpha = 0.5\%$, shown in Figures 3.3 and 3.8. The simulations that fall under the case where the attack has no impact on the system, i.e. case 3 in Section 3.2.3, can be categorized into two groups:

1. The main eigenvalue of $A' + B'F'$ is in the unit circle OR

2. $\theta = 0$, meaning the main eigenvector of $A' + B'F'$ is orthogonal to the main eigenvector of $A + BF$.

These results make sense, as the attack signal will not diverge if there are no eigenvalues outside of the unit circle. Additionally, when the main eigenvector of $A' + B'F'$ is orthogonal to the main eigenvector $A + BF$, the attack signal is not going in the correct direction to excite the unstable eigenvalue of the attack, meaning the attack will not be able to diverge the states of the system.

Determining the differences between the cases where the attacker is detected before damage is done to the system and where the attacker is able to diverge the states of the system to

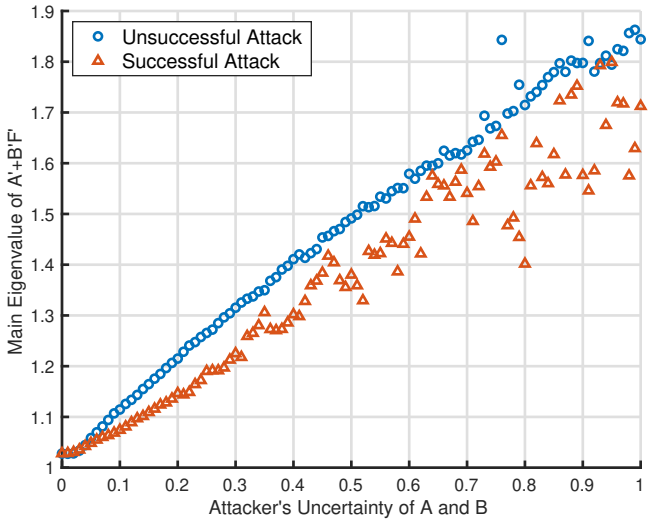


Figure 3.10: The main eigenvalue of $A' + B'F'$ vs. the attacker's uncertainty in both A and B. The variance of the process and measurement noise is 10^{-4} and the false alarm rate α is 0.5%. The blue circles represent the case where the attack is detected before the attacker is able to cause damage to the system and the red triangles represent the case where the attacker is able to diverge the states of the system to a point of system failure before the attack is detected. We can see that the main eigenvalue is lower for the cases where the attack is successful.

a point of system failure before the attack is detected is not as straightforward. Figures 3.10 and 3.11 show the mean of the main eigenvalues of $A'+B'F'$ and θ for $10,000$ simulations for varying uncertainties. We can clearly see that the attack is successful with lower eigenvalues and with higher values of θ. This makes sense, as with a lower main eigenvalue, the states of the system will diverge slower causing the attack to be harder to detect. Also, when the direction of the attack with imperfect A and B matrices is closer to the direction of the actual A and B matrices, we would expect the attack to be less detectable, as more of the attack is in the nullspace of the original system. Thus, it is imperative for the attacker to not only learn an accurate state space representation of the system, but to learn a system whose eigensystem aligns with the real system. This is further shown in Figure 3.12, where we can see that, as the attacker's uncertainty in A and B increases, the percentage of simulations where the attack has no effect on the states of the system and the percentage of simulations where the attack is detected before causing system failure increase, while the percentage of successful attacks decreases. The high percentage of simulations that have no effect on

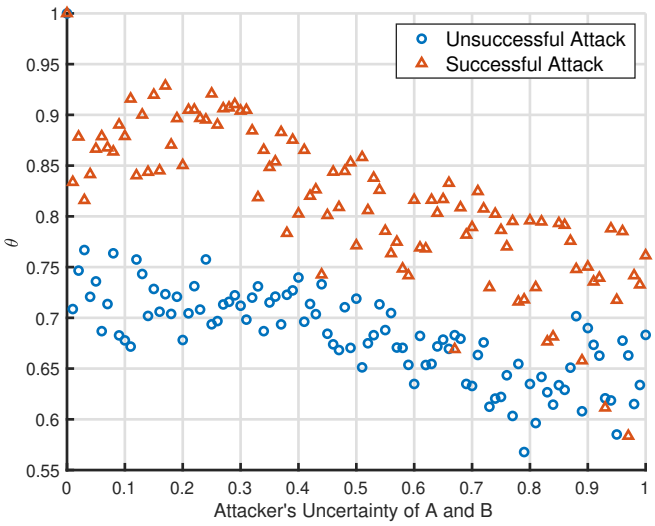


Figure 3.11: The dot product between the main eigenvector of $A+BF$ and $A'+B'F'$, θ, vs. the attacker's uncertainty in both A and B. The variance of the process and measurement noise is 10^{-4} and the false alarm rate α is 0.5%. The blue circles represent the case where the attack is detected before the attacker is able to cause damage to the system and the red triangles represent the case where the attacker is able to diverge the states of the system to a point of system failure before the attack is detected. We can see that θ is higher for the cases where the attack is successful.

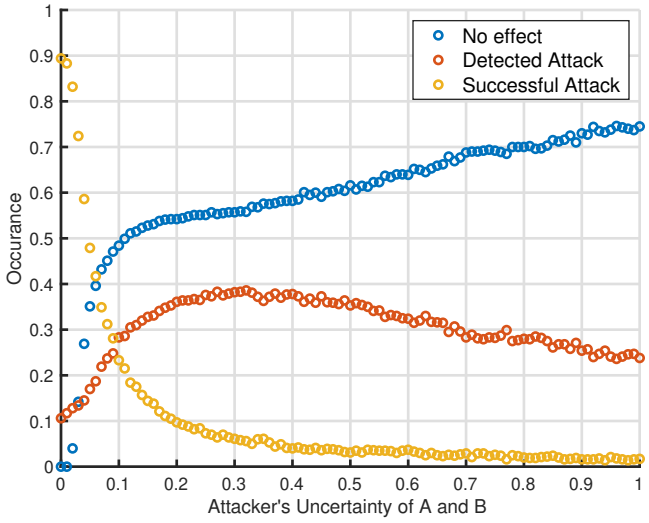


Figure 3.12: The percentage of occurrences of an attack that has no effect on the states of the system, are detected, and are successful vs. the attacker's uncertainty in A and B.

the states of the system could be used to the advantage of the attacker, who could attempt numerous attacks and observe the impact on the system while remaining undetected.

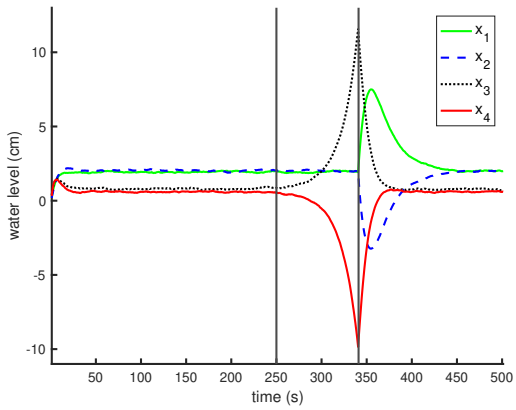
3.4 Stealthy Threshold Attack

We have assumed that the attack can inject an infinite amount of energy into the system. However, in order to make a stealthy attack more realistic, we must consider that the magnitude of the attack signal would be bounded due to the physical limitations of the system. We investigate the impact that a finite energy stealthy attack has on the probability of the attack being successful.

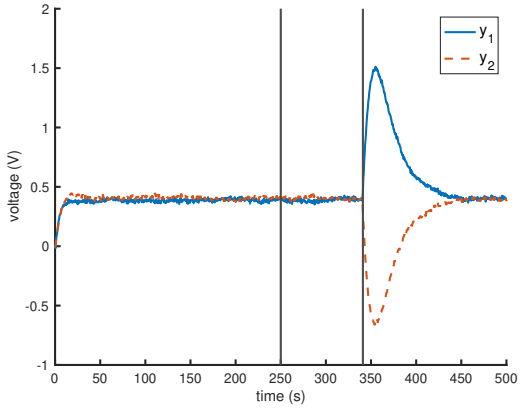
3.4.1 Limited Energy Stealthy Attacks

Assuming that the attacker can only inject a finite amount of energy into the control signal, we must limit the attack signal a_k to be within a specified threshold. A stealthy attack can only remain undetected while the attacker is actively inserting energy into the control signal. If the attacker reaches the physical limits the control signal can tolerate before they are able to significantly diverge the states of the system, the attack will be unsuccessful. Figure 3.13 shows the states and observations of a stealthy attack where the attacker has limited energy that they can inject into the control signal. We choose to limit the attack signal to 10V, as when the system is running under normal operating conditions, the maximum control signal is about 4V, thus it is realistic to assume that the pumps on the QTP would be able to handle a maximum of 14V before damaging the system (when the attack passes the physical damage threshold then it becomes detectable).

Running simulations with the same setup as Section 3.2, we can analyze the attacker's probability of success when they only have a finite amount of energy to inject into the control signal. Figure 3.14 shows the probability of a stealthy attack being successful with the attacker having an increasing uncertainty in the matrix A and with the attack signal limited to 10V. We can see that the probability of a successful attack is slightly lower when



(a) The states of the QTP under attack, with the attacker limited to injecting 10V into the control signal, where x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4.



(b) The observations of the QTP under attack, with the attacker limited to injecting 10V into the control signal, where y_1 and y_2 are the voltages in V from the level measurement devices in tanks 1 and 2.

Figure 3.13: The states and observations of the QTP simulation with an attack starting at 250s and ending at 341s when the attacker reaches the threshold for the control signal, with black vertical lines signifying the beginning and end of the attack. We can see that the attack will be detected as soon as the attacker stops injecting energy into the control signal.

the attacker is limited in the energy they can inject into the control signal; however, the results are not significantly impacted. This means that even if the system has physical limitations on the control signal, a successful stealthy attack is still possible.

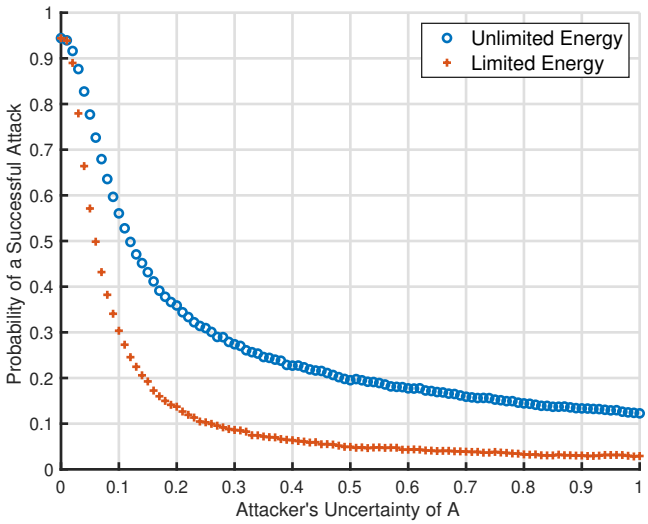


Figure 3.14: Probability of a successful stealthy attack (y axis) and uncertainty σ_A (x axis) when the attacker has imperfect knowledge of A. Additionally, the attacker is limited to injecting less than 10V into the control signal. The variance of the process and measurement noise added to the system is 10^{-4} and the false alarm rate α is set to 0.5%.

Chapter 4

The Effect of a Learned State Space Model on the Success of a Stealthy Attack

4.1 Introduction

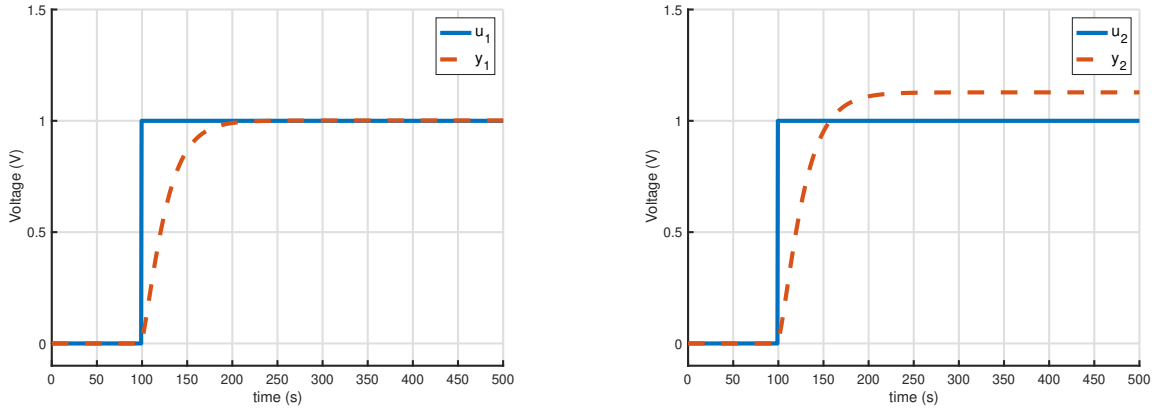
This chapter considers the problem of the attacker having to learn a state space representation of the system in order to calculate a stealthy attack. Realistically an attacker would not already have a state space model for the system, thus they would have to use the control signal and observations of the system in order to calculate a model. We utilize system identification by using input and output data to learn a mathematical model of the system. Both black box and grey box approaches to system identification are considered in this chapter. The black box approach assumes that the attacker has no prior knowledge about the system and will have to rely solely on the inputs and outputs of the system in order to calculate a state space representation, while the grey box approach assumes the attacker knows some information about the system. This information could include the values of certain parameters, or bounds on these parameters, etc. In this chapter, we use the MATLAB system identification algorithms which are implemented to minimize a cost function that considers

the difference between the predicted output and the actual output of the system. A gradient descent algorithm is then used to find the state space representation where a minimum cost occurs. We first consider the probability of an attacker mounting a successful stealthy attack using the black box approach. Then, we assume that the attacker knows the differential equations that represent the dynamics of the system they are attacking, as well as upper and lower bounds for these parameters. The probability of a stealthy attack being successful is then calculated using the grey box approach. These results are shown using the QTP.

4.2 Black Box Approach

4.2.1 Problem Setup

We first consider the case where the attacker has no information about the system, i.e. the black box approach. As mentioned in Chapter 2, the attacker has access to the control signals and observations of the system, thus they can use this information to calculate a state space representation using a system identification algorithm. We want to determine if the attacker is able to mount a successful stealthy attack without having any prior knowledge of the system that would help produce a state space representation.



(a) Step input and response, u_1 and y_1 respectively.

(b) Step input and response, u_2 and y_2 respectively.

Figure 4.1: The step input and response used for the black box approach with no process and measurement noise.

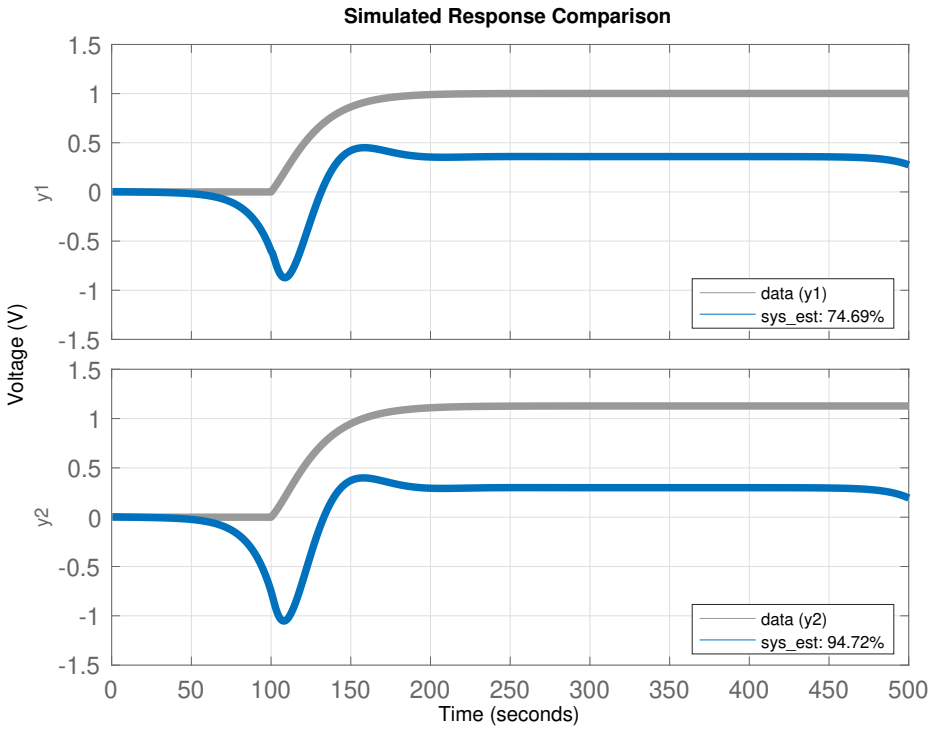
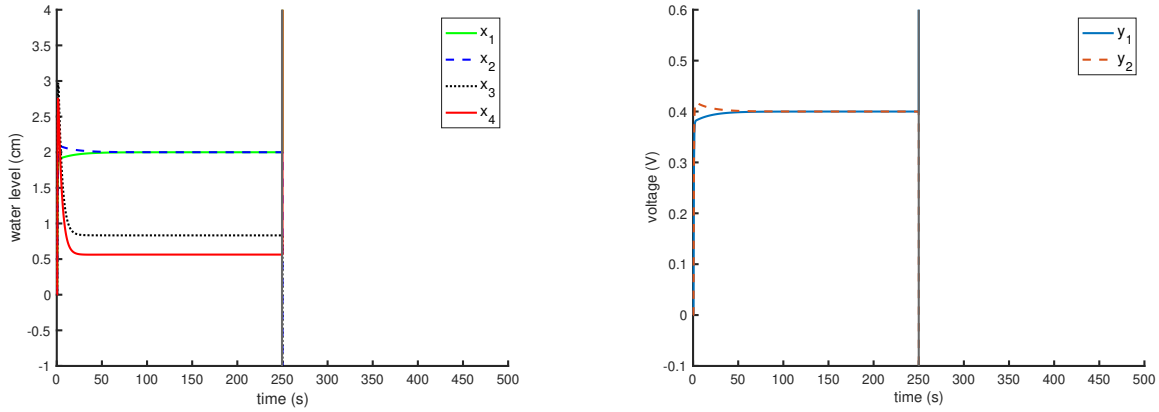


Figure 4.2: A comparison of the actual output of the system, represented by the grey line, with the learned output of the system, represented by the blue line. There is no process and measurement noise added to the system, and we are estimating a fourth order model using a black box approach. We can see that the normalized root mean square error (NRMSE) between the actual output and learned output is 74.69% and 94.72% for y_1 and y_2, respectively.

4.2.2 Results

An estimate for a state space representation of the QTP is calculated and a stealthy attack is applied to the system under various conditions. We first consider the attacker learning the system from a step input and response with no process and measurement noise, shown in Figure 4.1. In addition to learning a state space representation, the attacker will also have to determine the order of the system they want to learn. First, we consider the attacker learning a fourth order system, as this is the same as the QTP. Even though this is a black box approach to system identification, we assume that the attacker knows the order of system they should estimate. Figure 4.2 shows a comparison of the actual output of the system with the learned output of the fourth order system and the attacker using the step



(a) The states of the QTP under attack, with the attacker learning the state space model, where x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4.

(b) The observations of the QTP under attack, with the attacker learning the state space model, where y_1 and y_2 are the voltages in V from the level measurement devices in tanks 1 and 2.

Figure 4.3: The states and observations of the QTP simulation with an attack starting at 250s, with black vertical lines signifying the beginning of the attack. The attacker used the input and output data of the system to estimate a fourth order state space representation of the QTP using a black box approach. There is no process and measurement noise added to the system. We can see that the attack is detectable, as the observations diverge immediately.

input and response with no process and measurement noise to learn a mathematical model of the system. The normalized root mean square error (NRMSE) between the actual output and learned output is 74.69% and 94.72% for y_1 and y_2, respectively. We can see that the attacker is not able to learn a state space representation that accurately represents the real system; however, we want to determine if this is close enough for the attacker to mount a successful stealthy attack. The states and observations under attack using the learned fourth order state space model are shown in Figure 4.3. We can see that the attacker is not successful, as the observations diverge immediately after the attack begins, thus the attacker will learn a third order model next to determine if an estimate of a different order model is close enough to the real system to mount a successful attack.

Figure 4.4 shows a comparison between the actual output of the system and the learned output of the third order system. The NRMSE between the actual output and learned output is 8% and 8.82% for y_1 and y_2, respectively. This is a much better fit than the fourth order model. Now, we want to determine if the better fitting third order model is capable of

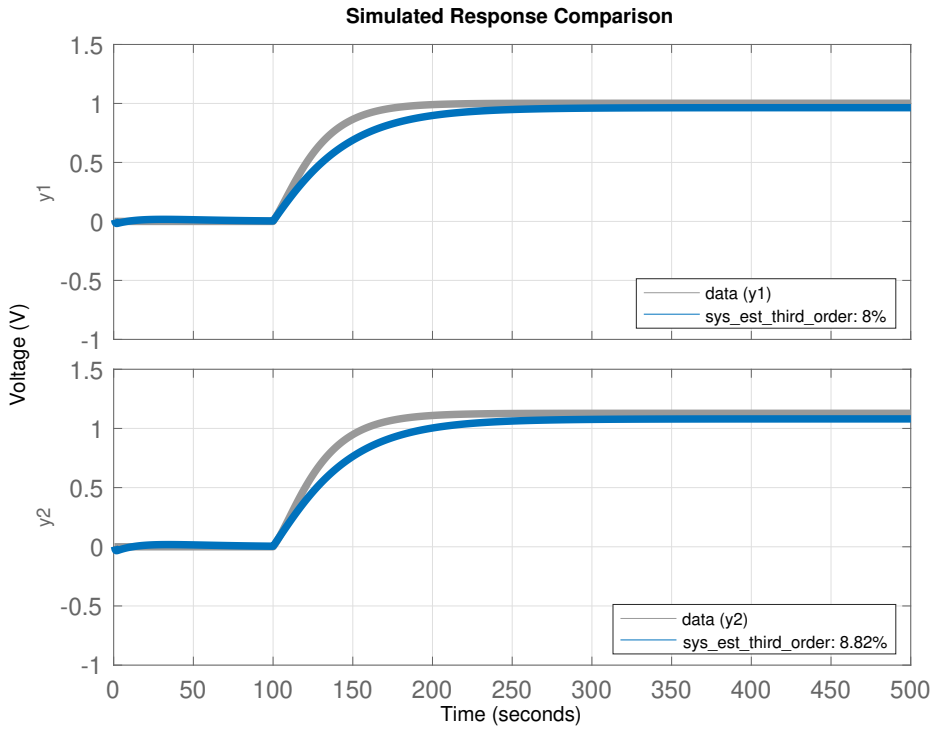
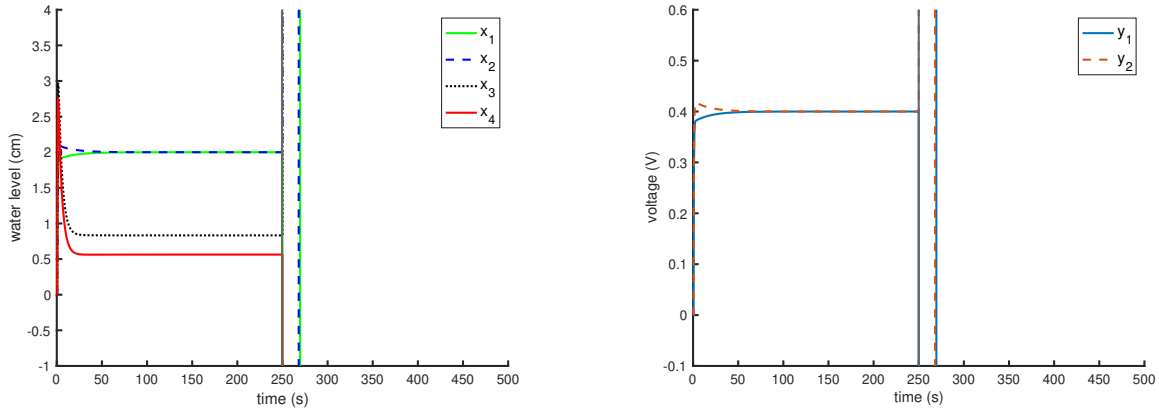


Figure 4.4: A comparison of the actual output of the system, represented by the grey line, with the learned output of the system, represented by the blue line. There is no process and measurement noise added to the system and we are estimating a third order model using a black box approach. We can see that the normalized root mean square error (NRMSE) between the actual output and learned output is 8% and 8.82% for y_1 and y_2, respectively.

producing a successful stealthy attack. The states and observations of the QTP under attack using the learned third order model are shown in Figure 4.5. We can see that even though the third order model matches the real system better than the fourth order model, this is still not enough for a stealthy attack to be successful, as the observations of the system diverge immediately after the attack begins. We have seen that a stealthy attack is not successful when the attacker learns the model using a step input and response with no process and measurement noise, thus we will incorporate process and measurement noise into the output of the system to simulate more realistic learning conditions.

Next, we consider the attacker learning the system from a step input and response with the output of the system having process and measurement noise with a variance of 10^{-4}, shown in Figure 4.6. The attacker then uses this input and output data to learn a fourth



(a) The states of the QTP under attack, with the attacker learning the state space model, where x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4.

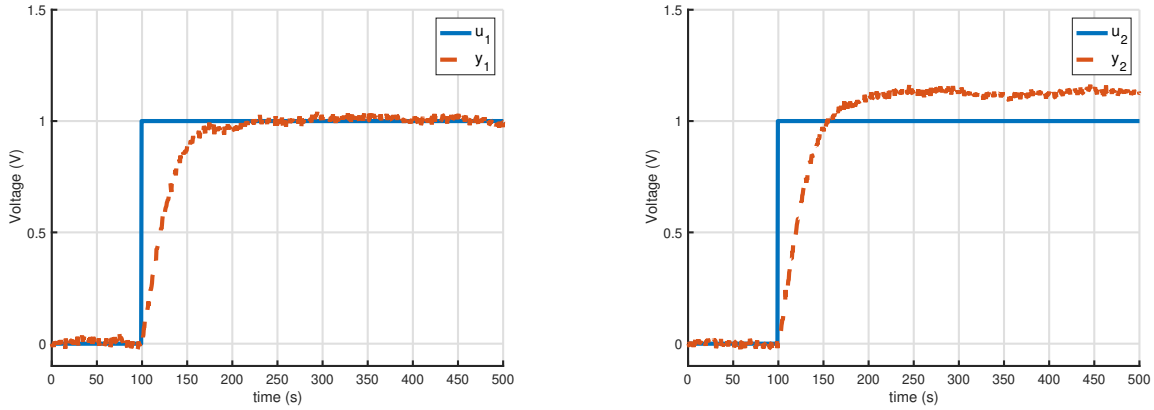
(b) The observations of the QTP under attack, with the attacker learning the state space model, where y_1 and y_2 are the voltages in V from the level measurement devices in tanks 1 and 2.

Figure 4.5: The states and observations of the QTP simulation with an attack starting at 250s, with black vertical lines signifying the beginning of the attack. The attacker used the input and output data of the system to estimate a third order state space representation of the QTP using a black box approach. There is no process and measurement noise added to the system. We can see that the attack is detectable, as the observations diverge immediately.

order state space representation of the QTP. Figure 4.7 shows a comparison between the actual output of the system and the learned output of the fourth order system with process and measurement noise added to the output. The NRMSE between the actual output and the learned output is 74.87 and 95.15 for y_1 and y_2, respectively. These results are not much different than the learned fourth order model with no process and measurement noise.

Figure 4.8 shows a comparison between the actual output of the system and the learned output of the third order system with process and measurement noise added to the output. The NRMSE between the actual output and the learned output is 5.09% and 5.51% for y_1 and y_2, respectively. These results are similar to the case of the learned third order system with no process and measurement noise; however, we still simulate the states and observations under attack to determine if a stealthy attack is successful, shown in Figure 4.9. We can see that the observations diverge immediately after the attack begins, meaning the attack would be detected, thus unsuccessful.

We have seen that a stealthy attack is not successful when the attacker learns a state



(a) Step input and response, u_1 and y_1, respectively.

(b) Step input and response, u_2 and y_2, respectively.

Figure 4.6: The step input and response used for the black box approach, with the variance of the process and measurement noise added to the system equal to 10^{-4}.

space representation from a step input and step response. Next, we will consider having the attacker use 500s of input and output data from the system running, rather than simulating a step input and response. We first consider the attacker learning a fourth order state space model, shown in Figure 4.11. The NRMSE between the actual output and the learned output is 3.36% and 3.53 for y_1 and y_2, respectively. Thus, the attacker is not able to learn a model that accurately represents the real system. We want to determine if there are any conditions that produce a successful stealthy attack when using the black box approach. Thus, we will consider the attacker learning a third order model, as that is the order of system that matches the real system best with the black box approach. Figure 4.12 shows a comparison between the actual output of the system and the learned output of the third order system running under normal conditions. The NRMSE between the actual output and the learned output is 2.81 and 2.59 for y_1 and y_2, respectively. This is the closest match to the real system that we have seen in this chapter. Now, we want to determine if this is enough for the attacker to mount a successful attack. Figure 4.13 shows the states and observations of the system under attack with the attacker using the learned third order model. We can see that the observations diverge immediately after the attack begins, thus the stealthy attack is unsuccessful. However, the rate of divergence of the observations is much smaller than the

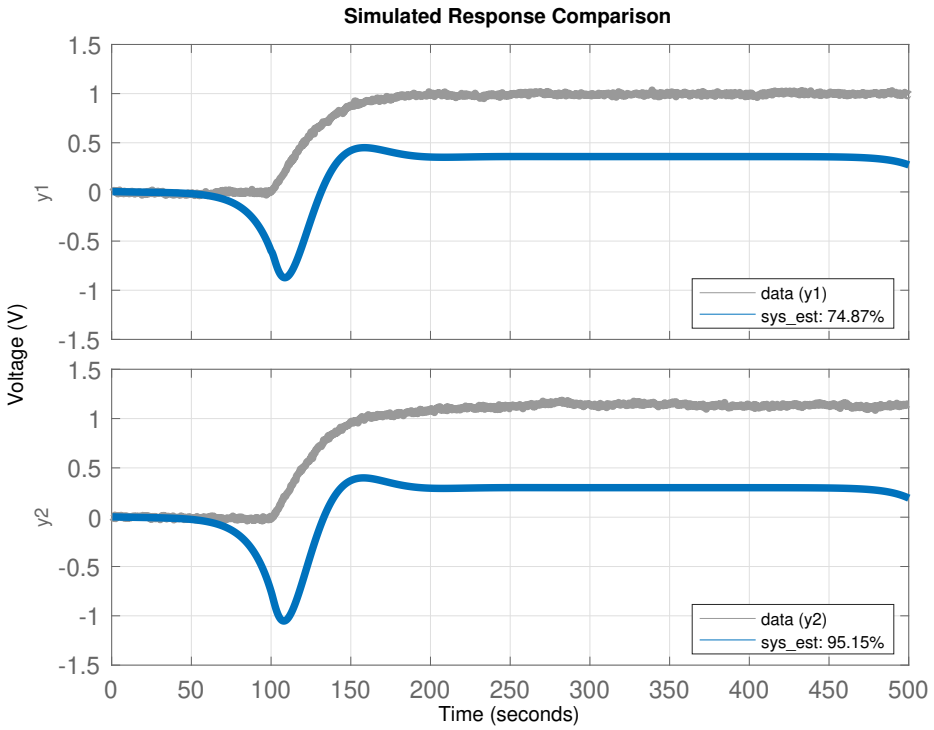


Figure 4.7: A comparison of the actual output of the system, represented by the grey line, with the learned output of the system, represented by the blue line. The variance of the process and measurement noise is 10^{-4} and we are estimating a fourth order model using a black box approach. We can see that the normalized root mean square error (NRMSE) between the actual output and learned output is 74.87% and 95.15% for y_1 and y_2, respectively.

previous cases we have seen, meaning the attacker is closer to mounting a successful stealthy attack. From the results of these simulations, we believe that the black box approach is not the best way for the attacker to learn a state space representation that would produce a successful attack. Additionally, we believe the attacker would realistically have some knowledge about the system they are attacking, thus we will consider a grey box approach to system identification next.

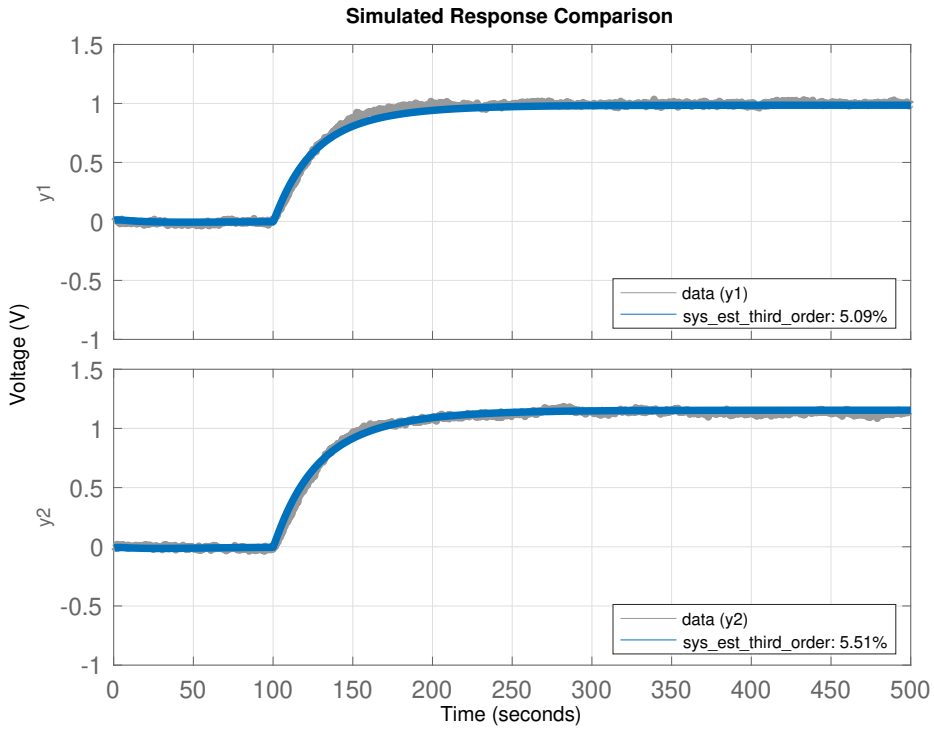
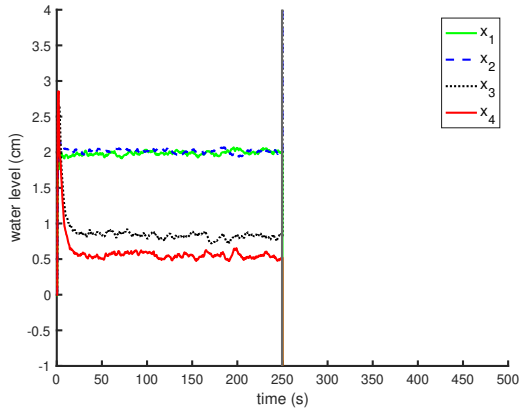


Figure 4.8: A comparison of the actual output of the system, represented by the grey line, with the learned output of the system, represented by the blue line. The variance of the process and measurement noise is 10^{-4} and we are estimating a third order model using a black box approach. We can see that the normalized root mean square error (NRMSE) between the actual output and learned output is 5.09% and 5.51% for y_1 and y_2, respectively.

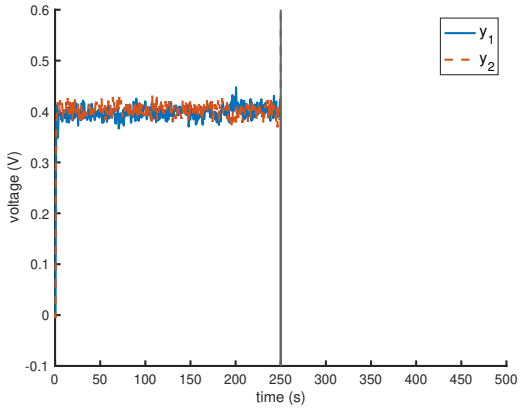
4.3 Grey Box Approach

4.3.1 Problem Setup

We consider that the attacker knows the differential equations that represent the dynamics of the system they are attacking; however, they may not know the exact values of the parameters in these equations. The attacker may know upper and lower bounds that the parameters must remain within, or they could know the values of some, but not all, of the parameters. For the simulations in this chapter, we must first have a state space representation of the QTP in terms of the parameters in the differential equations. We begin by using the differential equations for the state space model of the QTP found in Section 2.3. In order to linearize this system, we need to calculate the Jacobian matrix $J_{f(x)}$ shown in (4.1).



(a) The states of the QTP under attack, with the attacker learning the state space model, where x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4.

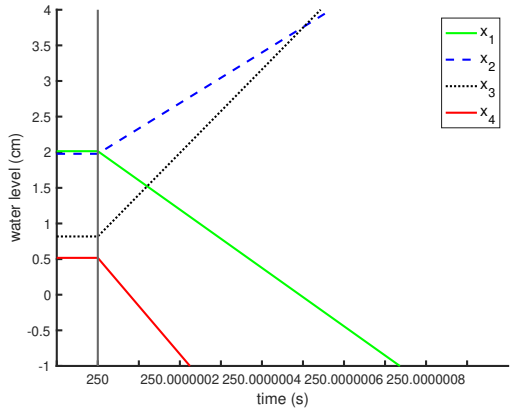


(b) The observations of the QTP under attack, with the attacker learning the state space model, where y_1 and y_2 are the voltages in V from the level measurement devices in tanks 1 and 2.

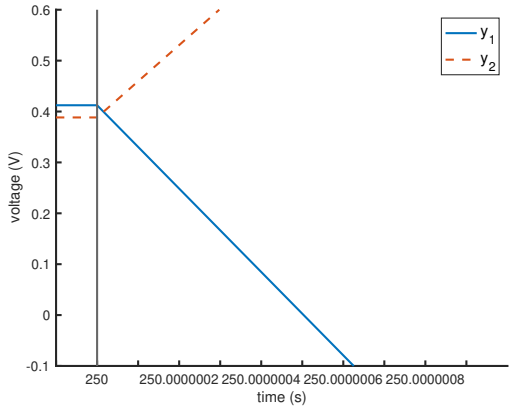
Figure 4.9: The states and observations of the QTP simulation with an attack starting at 250s, with black vertical lines signifying the beginning of the attack. The attacker used the input and output data of the system to estimate a third order state space representation of the QTP using a black box approach. The variance of the process and measurement noise added to the system is 10^{-4}. We can see that the attack is detectable, as the observations diverge immediately.

$$J_{f(x)} = \begin{bmatrix} \frac{-a_1 g}{A_1\sqrt{2gx_1}} & 0 & \frac{a_3 g}{A_1\sqrt{2gx_3}} & 0 \\ 0 & \frac{-a_1 g}{A_2\sqrt{2gx_2}} & 0 & \frac{a_4 g}{A_2\sqrt{2gx_4}} \\ 0 & 0 & \frac{-a_3 g}{A_3\sqrt{2gx_3}} & 0 \\ 0 & 0 & 0 & \frac{-a_4 g}{A_4\sqrt{2gx_4}} \end{bmatrix} \tag{4.1}$$

From the differential equations in Section 2.3 and (4.1), we can get state space matrices A, B, and C in terms of system parameters, shown in (4.2), where h_{i_0}, $i \in 1, 2, 3, 4$, are the operating points for the height of the water in the tanks.



(a) The states of the QTP under attack, with the attacker learning the state space model, where x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4.



(b) The observations of the QTP under attack, with the attacker learning the state space model, where y_1 and y_2 are the voltages in V from the level measurement devices in tanks 1 and 2.

Figure 4.10: A zoomed-in version of 4.9

$$A = \begin{bmatrix} \frac{-a_1 g}{A_1\sqrt{2gh_{1_0}}} & 0 & \frac{a_3 g}{A_1\sqrt{2gh_{3_0}}} & 0 \\ 0 & \frac{-a_1 g}{A_2\sqrt{2gh_{2_0}}} & 0 & \frac{a_4 g}{A_2\sqrt{2gh_{4_0}}} \\ 0 & 0 & \frac{-a_3 g}{A_3\sqrt{2gh_{3_0}}} & 0 \\ 0 & 0 & 0 & \frac{-a_4 g}{A_4\sqrt{2gh_{4_0}}} \end{bmatrix} \tag{4.2}$$

$$B = \begin{bmatrix} \frac{\gamma_1 k_1}{A_1} & 0 \\ 0 & \frac{\gamma_2 k_2}{A_2} \\ 0 & \frac{(1-\gamma_2)k_2}{A_3} \\ \frac{(1-\gamma_1)k_1}{A_4} & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} k_c & 0 & 0 & 0 \\ 0 & k_c & 0 & 0 \end{bmatrix}$$

The values of the parameters in (4.2) are shown in Table 4.1. Throughout this chapter we assume that the only parameter that the attacker knows perfectly is the acceleration due

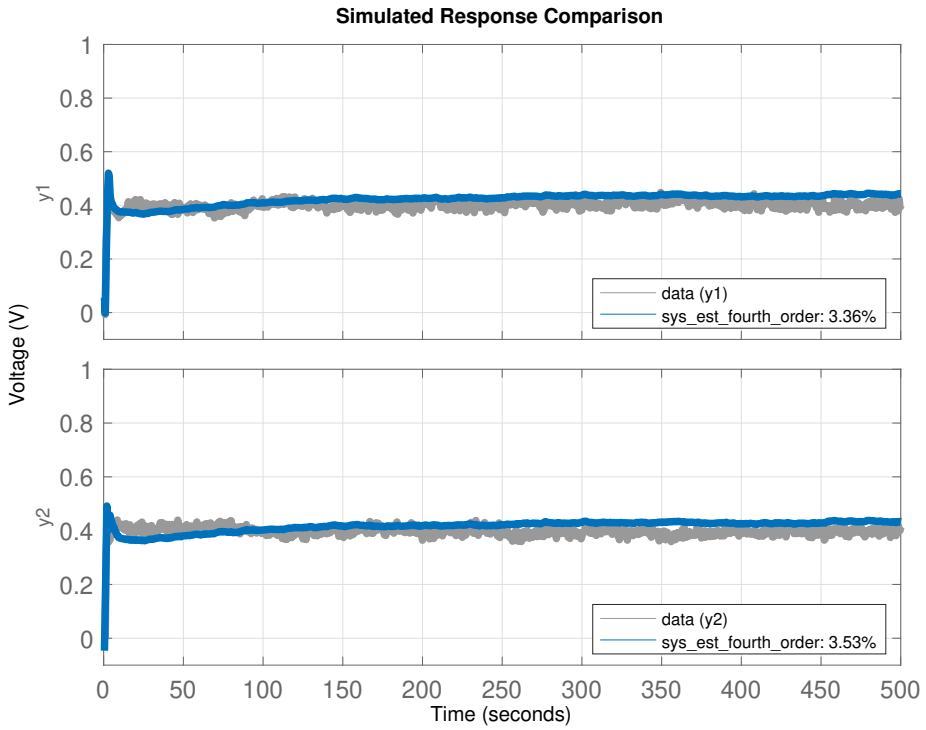


Figure 4.11: A comparison of the actual output of the system, represented by the grey line, with the learned output of the system, represented by the blue line. The variance of the process and measurement noise is 10^{-4} and we are estimating a fourth order model using a black box approach. This comparison is for the inputs and outputs of the system running under normal conditions. We can see that the normalized root mean square error (NRMSE) between the actual output and learned output is 3.36% and 3.53% for y_1 and y_2, respectively.

to gravity $g = 9.81\frac{\mathrm{m}}{s^2}$. The attacker's uncertainty in the remaining parameters is modeled by assuming the attacker knows upper and lower bounds for the parameters. We will consider bounds of the parameters set to the actual value of the parameter $\pm$ $\sigma\times$ the actual value of the parameter, where σ is the attacker's uncertainty in the parameters of the state space representation. For example, (4.3) shows the upper and lower bounds for a_1 when $\sigma = 0.1$.

$$a_1' \in [a_1 - \sigma * a_1, a_1 + \sigma * a_1] \tag{4.3}$$

$$a_1' \in [0.0639, 0.0781]$$

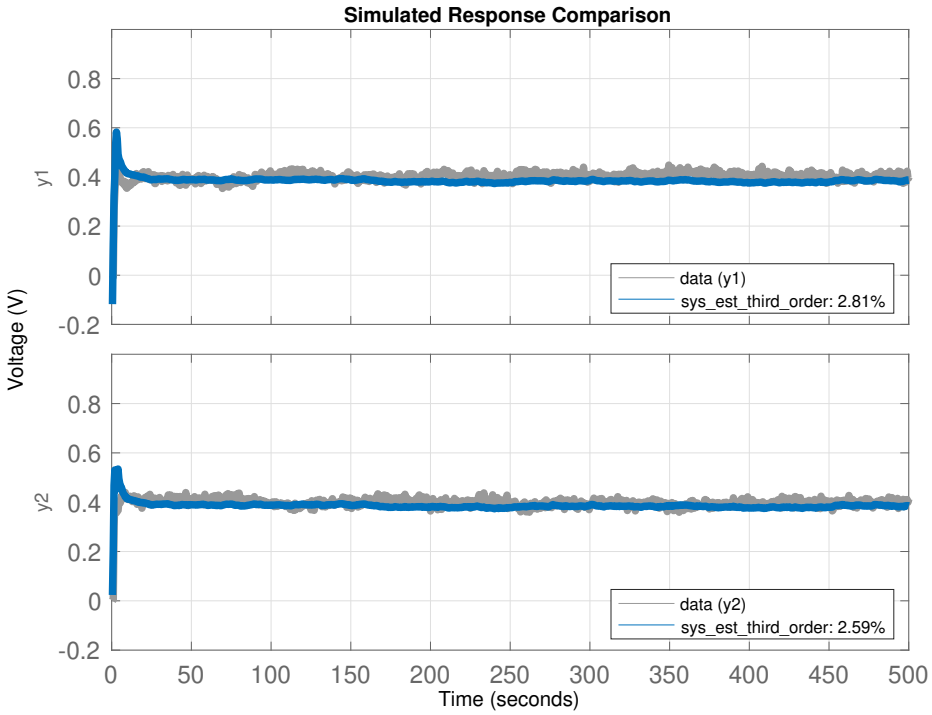
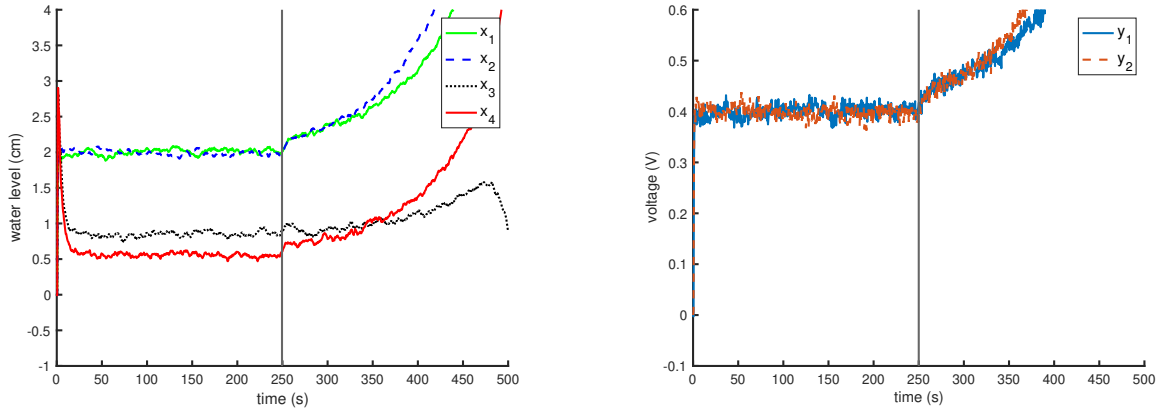


Figure 4.12: A comparison of the actual output of the system, represented by the grey line, with the learned output of the system, represented by the blue line. The variance of the process and measurement noise is 10^{-4} and we are estimating a third order model using a black box approach. This comparison is for the inputs and outputs of the system running under normal conditions. We can see that the normalized root mean square error (NRMSE) between the actual output and learned output is 2.81% and 2.59% for y_1 and y_2, respectively.

The attacker uses the input and output data of the system from $t = 0$ to $t = 500$s to learn a state space representation of the system. Additionally, the attacker also knows upper and lower bounds for each parameter. Each parameter is then initialized to a random value in this range for the system identification algorithm. We use the same uncertainty σ for every parameter in the state space model, besides the acceleration due to gravity g. Figure 4.14 shows the output of the actual system and the output of the learned system for each output y_1 and y_2. The normalized root mean square error (NRMSE) between the actual output and the learned output is also shown. The NRMSE for y_1 is 1.44% and is 1.68% for y_2. This tells us that the learned state space representation behaves similarly to the actual system. Additionally, we can see from Figure 4.15 that increasing the learning time for the attacker



(a) The states of the QTP under attack, with the attacker learning the state space model, where x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4.

(b) The observations of the QTP under attack, with the attacker learning the state space model, where y_1 and y_2 are the voltages in V from the level measurement devices in tanks 1 and 2.

Figure 4.13: The states and observations of the QTP simulation with an attack starting at 250s, with black vertical lines signifying the beginning of the attack. The attacker used the input and output data of the system running under normal conditions to estimate a third order state space representation of the QTP using a black box approach. The variance of the process and measurement noise added to the system is 10^{-4}. We can see that the attack is detectable, as the observations diverge after the attack begins.

to 1000s does not impact how close they are able to learn the system.

The learned state space representation from Figure 4.14 is used to calculate a stealthy attack. We can see the states and observations of the QTP under attack with the learned state space representation in Figure 4.16. The attacker is able to quickly empty tank 4 at $t = 378.5$ seconds, about 2 minutes after the attack began. We can see that the attack is not perfectly stealthy, as the observations diverge. However, the attack is not detected until $t = 417$s, about 40s after the attacker is able to empty tank 4. This means that the attack is successful.

4.3.2 Results

We want to determine the probability of an attack being successful when the attacker has to learn a state space representation of the system. The attacker uses the inputs and outputs of the system running under normal operating conditions to learn a state space representation,

Parameter	Value
a_1, a_3	0.071cm^2
a_2, a_4	0.057cm^2
A_1, A_3	28cm^2
A_2, A_4	32cm^2
h_{1_0}	12.6cm
h_{2_0}	13cm
h_{3_0}	4.8cm
h_{4_0}	4.9cm
γ_1	0.43
γ_2	0.34
k_1	$3.14\frac{\text{cm}^3}{\text{Vs}}$
k_2	$3.29\frac{\text{cm}^3}{\text{Vs}}$
k_c	$0.5\frac{\text{V}}{\text{cm}}$

Table 4.1: Values for the parameters in the state space representation of the quadruple-tank process.

as described in Section 4.3.1. Similarly to Section 3, we calculate the difference between the system failure time and detection time of the system for $10,000$ simulations with σ varying from 0 to 1 in increments of 0.01. Figure 4.17 shows the probability of a successful attack where the variance of the process and measurement noise added to the system is 10^{-4} and the false alarm rate $\alpha = 0.5\%$. We can see that even when the attacker knows the parameters of the system within 100% of their actual values, they are successful about 50% of the time. Additionally, the attacker's uncertainty in the parameters of the state space model do not have as much of an impact on the probability of success as in Figure 3.7. This tells us that the attacker would be able to mount a successful stealthy attack only with knowing loose bounds for the parameters of the system.

Similarly to Section 3.3.3, we want to determine what causes an attack to be successful or unsuccessful. We consider the main eigenvalue of $A' + B'F'$ and the dot product between the main eigenvector of $A + BF$ and $A' + B'F'$, denoted θ. Previously, we saw distinct groupings between the main eigenvalue of $A' + B'F'$ and θ when the attack is successful or unsuccessful, shown in Figures 3.10 and 3.11. We show the mean of the main eigenvalues of $A' + B'F'$ for $10,000$ simulations for the attacker's uncertainty ranging from $[0, 1]$ for

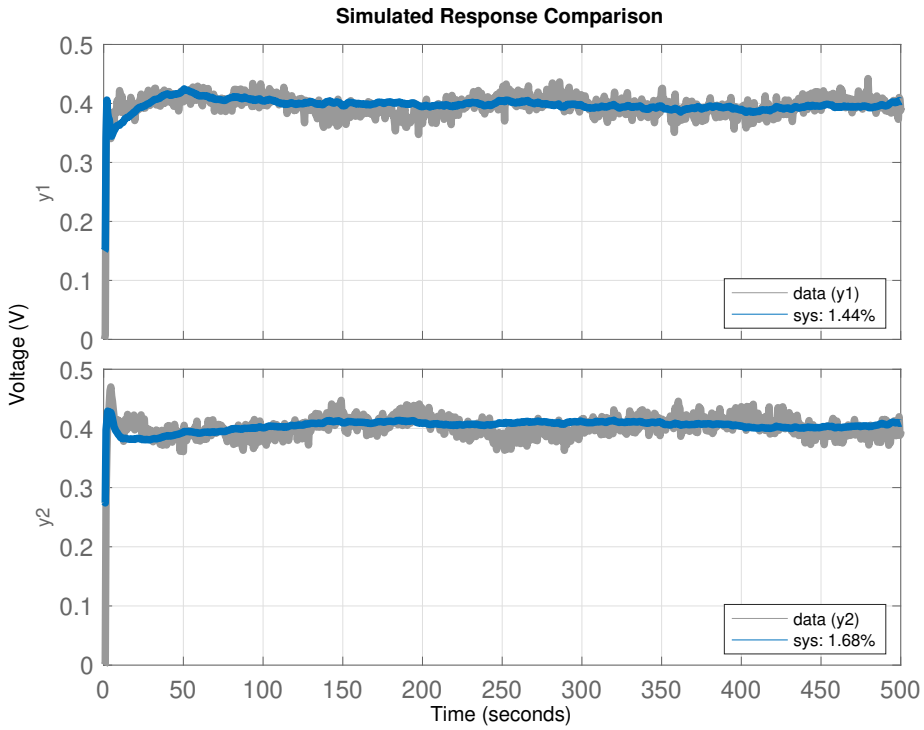


Figure 4.14: A comparison of the actual output of the system, represented by the grey line, with the learned output of the system, represented by the blue line, estimated using a grey box approach. We can see that the normalized root mean square error (NRMSE) between the actual output and learned output is 1.44% and 1.68% for y_1 and y_2, respectively.

the cases where the stealthy attack is successful or unsuccessful. These results are shown in Figure 4.18. We can see that there is not much of a difference between the cases where the attack is successful or unsuccessful. These results are very different than the case where the attacker's uncertainty in the state space representation is modeled by adding noise to the system rather than learning a model of the system. Additionally, Figure 4.19 shows the mean of the dot product θ between the main eigenvector of $A+BF$ and $A'+B'F'$, for $10,000$ simulations for the attacker's uncertainty ranging from $[0,1]$. We can see that there is also not much of a difference between the cases where the attack is successful and those where the attack is unsuccessful. Additionally, θ is close to 0 regardless of whether the attack is successful or not. Since there is not a distinct grouping between the cases where the attack is successful or unsuccessful, we believe that the system failure and detection times are much closer together when the attacker learns a state space model of the system. This leads us to

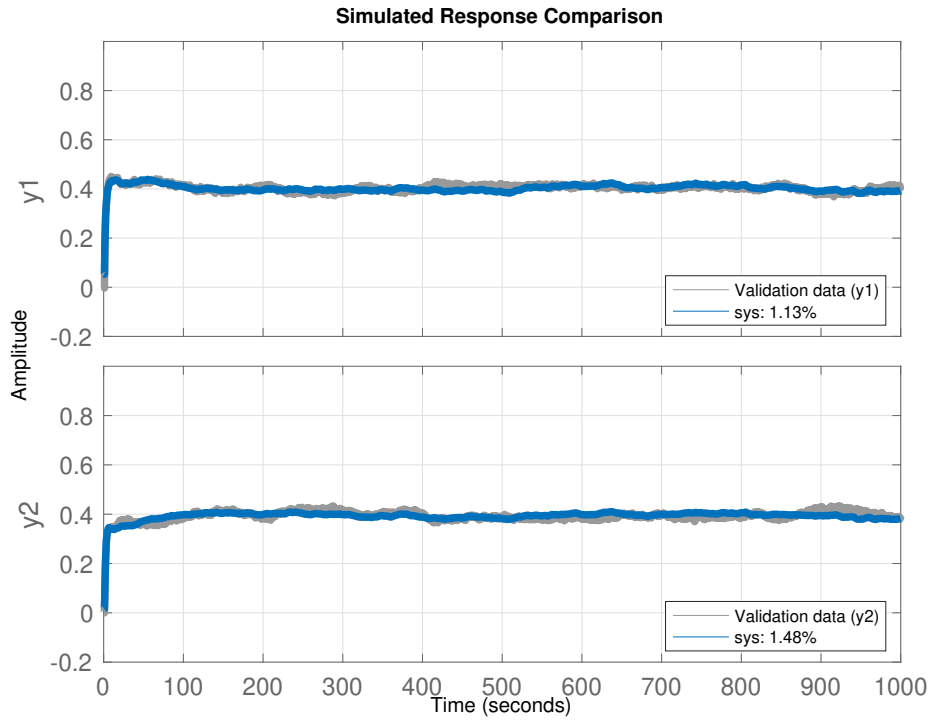
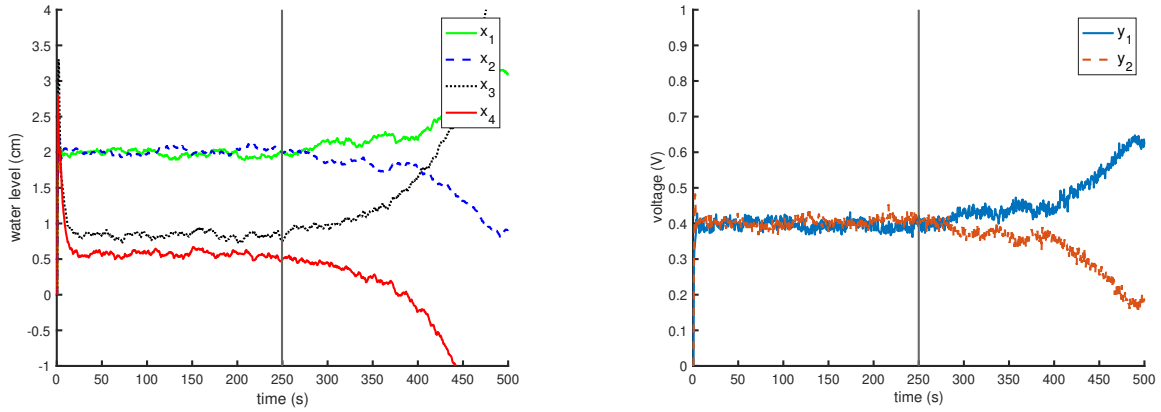


Figure 4.15: A comparison of the actual output of the system, represented by the grey line, with the learned output of the system, represented by the blue line, estimated using a grey box approach and 1000s of input data. We can see that the normalized root mean square error (NRMSE) between the actual output and learned output is 1.44% and 1.68% for y_1 and y_2, respectively.

believe that a small increase in the sensitivity of the anomaly detector will cause the attacker to have a much lower probability of a successful attack. Figure 4.20 shows three percentages of simulations for varying uncertainty in the system; 1: the percentage of attacks that have no effect on the states of the system, 2: the percentage of attacks that are detected before they cause system failure, and 3: the percentage of attacks that are successful. We can see that as the attacker's uncertainty in A and B increases, the percentage of simulations that are detected and that have no effect increase, while the percentage of simulations that have no effect on the states of the system decrease. This is what we expected, as Figure 4.18 shows the main eigenvalue of $A' + B'F'$ increasing which means the states of the system will diverge faster. This paired with the fact that the dot product between the main eigenvector of $A + BF$ and $A' + B'F'$ is decreasing, meaning less energy from the attack is going in the



(a) The states of the QTP under attack, with the attacker learning the state space model, where x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4.

(b) The observations of the QTP under attack, with the attacker learning the state space model, where y_1 and y_2 are the voltages in V from the level measurement devices in tanks 1 and 2.

Figure 4.16: The states and observations of the QTP simulation with an attack starting at 250s, with black vertical lines signifying the beginning of the attack, where the attacker learns the model using a grey box approach. The attack is successful, as the attacker is able to empty tank 4 before the attack is detected.

nullspace of C, makes the attack more detectable.

We want to determine if making small changes to the anomaly detector can better protect systems against stealthy attacks. We consider modifying the false alarm rate to be $\alpha = 1\%$, meaning 1% of the time there will be a false alarm. Figure 4.21 shows the probability of a stealthy attack being successful with the modified value for α. We can see that the probability of a successful attack is significantly less, simply by increasing the false alarm rate of the anomaly detector. Previously in Figure 3.8, we saw that when the attacker's uncertainty in the system was modeled by adding noise to the A and B matrices, increasing or decreasing the false alarm rate did not have a significant impact on the probability of an attack being successful. However, when the attacker has to actually learn a state space representation, the eigensystems of the cases where the attacks are successful or unsuccessful are much closer. Thus, small changes to the anomaly detector have a large impact on the probability of an attack being successful, as the system failure time and detection time of the attack are much closer together. This means that being able to detect an attack as little as a second earlier can prevent the attacker from being successful. These results tell

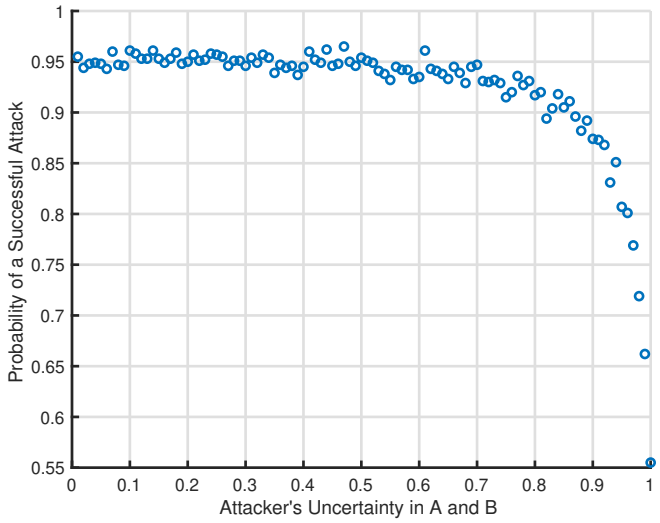


Figure 4.17: The probability that the attack is successful vs. the attacker's uncertainty in the parameters in the state space representation. The variance of the process and measurement noise added to the system is 10^{-4} and the false alarm rate α is set to 0.5%.

us that having a better anomaly detector can significantly lower the probability of an attack being successful. Throughout this dissertation, we simply consider a variance based anomaly detector; however, future systems should consider implementing a more advanced detector, as we have shown that every second counts when detecting anomalies within a system.

We also consider the impact that changing the QTP model has on our results relating to the probability of a successful attack. Figure 4.22 shows that increasing the cross-section of the tanks A_i from $A_1, A_3 = 28$ cm^2 and $A_2, A_4 = 32$ cm^2 to $A_1, A_3 = 40$ cm^2 and $A_2, A_4 = 50$ cm^2 causes a slightly lower probability of success. This is expected, as we are increasing the volume of the tank, meaning that more water is needed to reach the threshold for system failure.

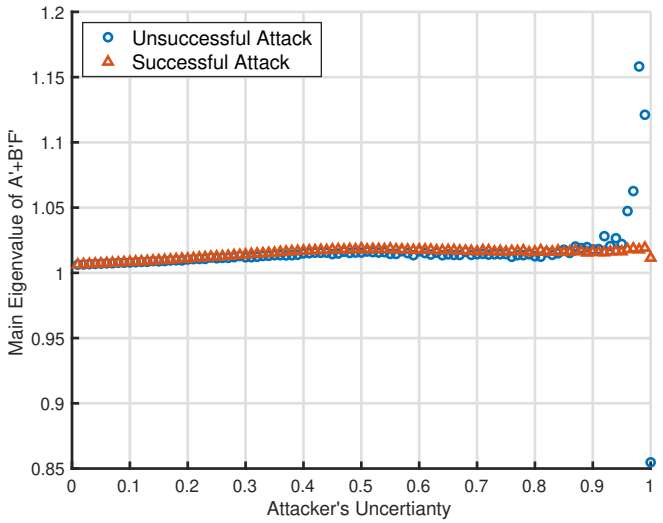


Figure 4.18: The main eigenvalue of $A' + B'F'$ vs. the attacker's uncertainty in both A and B. The variance of the process and measurement noise is 10^{-4} and the false alarm rate α is 0.5%. The blue circles represent the case where the attack is detected before the attacker is able to cause damage to the system and the red triangles represent the case where the attacker is able to diverge the states of the system to a point of system failure before the attack is detected. We can see that there is not much of a difference between the cases where the attack is successful or unsuccessful.

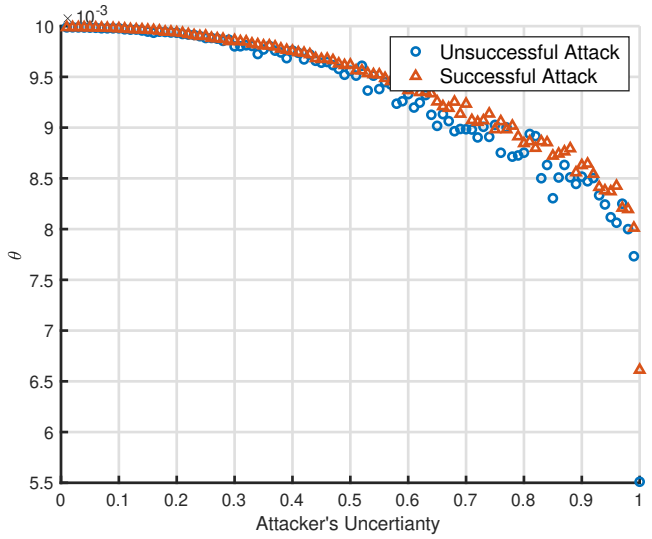


Figure 4.19: The dot product between the main eigenvector of $A + BF$ and $A' + B'F'$, θ, vs. the attacker's uncertainty in both A and B. The variance of the process and measurement noise is 10^{-4} and the false alarm rate α is 0.5%. The blue circles represent the case where the attack is detected before the attacker is able to cause damage to the system and the red triangles represent the case where the attacker is able to diverge the states of the system to a point of system failure before the attack is detected. We can see that θ is close to zero for the cases where the attack is successful or unsuccessful.

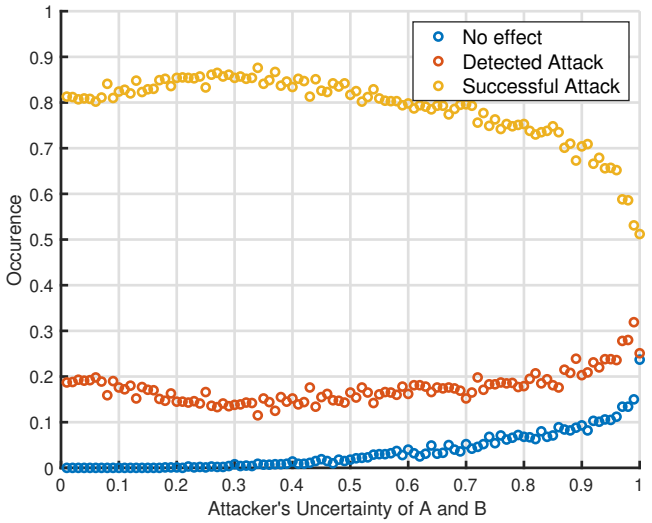


Figure 4.20: Three percentages of simulations for varying uncertainty in the system; 1: the percentage of attacks that have no effect on the states of the system, 2: the percentage of attacks that are detected before they cause system failure, and 3: the percentage of attacks that are successful vs. the attacker's uncertainty in A and B.

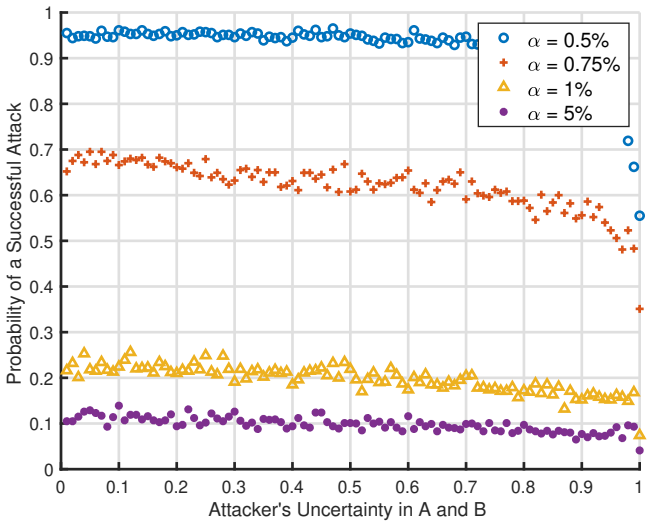


Figure 4.21: The probability that the attack is successful vs. the attacker's uncertainty in the parameters in the state space representation. The variance of the process and measurement noise added to the system is 10^{-4}. The false alarm rate α varies from 0.5% to 5%. The blue circles represent $\alpha = 0.5\%$, the red plus signs represent $\alpha = 0.75\%$, the the yellow triangles represent $\alpha = 1\%$, and the purple circles represent $\alpha = 5\%$. We can see that a small change in the anomaly detector causes the attack to have a much smaller probability of success.

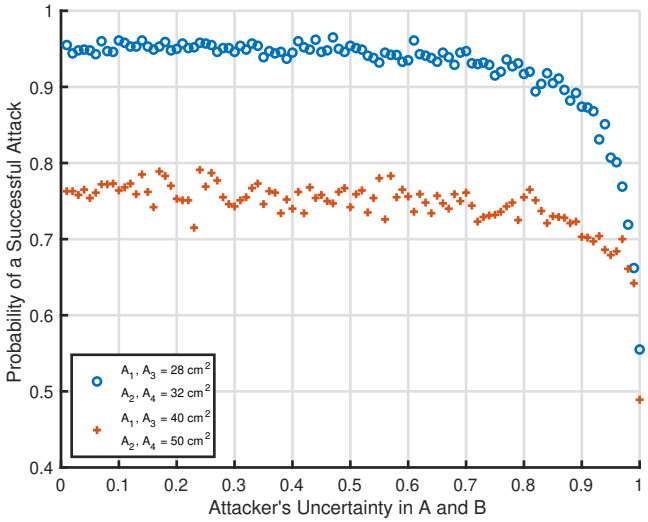


Figure 4.22: The probability that the attack is successful vs. the attacker's uncertainty in the parameters in the state space representation. The variance of the process and measurement noise added to the system is 10^{-4} and the false alarm rate α is set to 0.5%. The cross-section of tanks A_i varies from $A_1, A_3 = 28$ cm^2 and $A_2, A_4 = 32$ cm^2, shown by to blue circles to $A_1, A_3 = 40$ cm^2 and $A_2, A_4 = 50$ cm^2. We can see that when the cross-sections of the tanks are larger, the probability of success is slightly lower.

Chapter 5

Limiting the Impact of Stealthy Attacks

5.1 Introduction

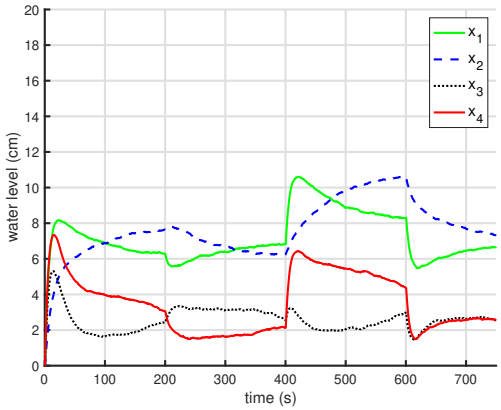
Throughout the previous Chapters we have shown that even when we place limitations on the attacker, such as limiting the energy they can inject into the control signal and assuming they would have to learn a state space representation of the model in order to calculate an attack, the attacker still has a large probability of their attack being successful. Thus, it is imperative that we investigate ways to limit the impact, and even eliminate the possibility of stealthy attacks entirely. In this chapter we first investigate how to limit the impact of stealthy attacks. We consider the trade off between making the system less susceptible to stealthy attacks and maintaining its controllability, by increasing the sampling time of the system–providing the attacker with fewer samples to learn a state space model. These results are demonstrated using the QTP. Next, we consider how to completely eliminate the opportunity for a stealthy attack. We begin with a state space representation for a 3 inverter power system model and assume that some of the parameters in this system have leeway when choosing these parameters in the design phase. Then, we determine values for these parameters closest to their original values that push the main eigenvalue of the attack model

to be stable, meaning the attack will have no impact on the system.

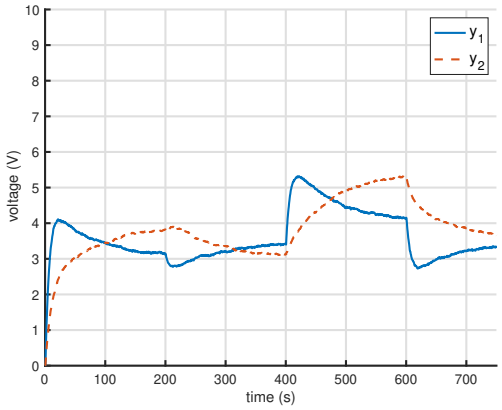
5.2 Simulation Results on the Quadruple-Tank Process

Previously, we have shown that an attacker is able to utilize a grey box approach to system identification in order to learn a state space representation of the system and successfully mount a stealthy attack. We want to determine how to prevent an attacker from learning a state space representation that is close enough to the actual system in order to successfully attack the system. The setup to the simulations in this section are similar to Chapter 4. We give the attacker a time interval of 250s to learn a state space representation of the system and 500s to mount the attack. Additionally, the attacker's uncertainty σ_{AB} in A and B is calculated in the same manner as in Chapter 4 – by assuming the attacker knows upper and lower bounds for the system parameters. However, in contrast to the previous sections, we change the desired states of the system x_d every 200s in order to test the capabilities of the controller.

We begin by determining how much the sampling time used to discretize the state space model T_s can be increased while maintaining acceptable performance of the system. Figures 5.1, 5.2, and 5.3 show the states and observations of the system running under normal conditions with $T_s = 0.5, 5, 50$s, respectively. Additionally, the desired states of the system x_d change every 200s in order to more accurately model the behavior of a real system. We can see that in Figures 5.1, 5.2 the controller is effectively able push the states of the system to their desired values while maintaining acceptable water levels in each of the four tanks, i.e. having a water height greater than zero and less than 20cm–the height of the tank. However, in Figure 5.3, we can see that the controller is not able to properly control the system, as the states and observations drop below 0 which is not an acceptable water level for this system. This is expected, as the controller only has 4 time steps to get the states to

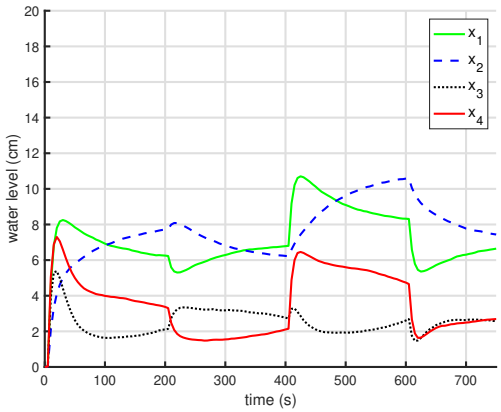


(a) x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4, respectively.

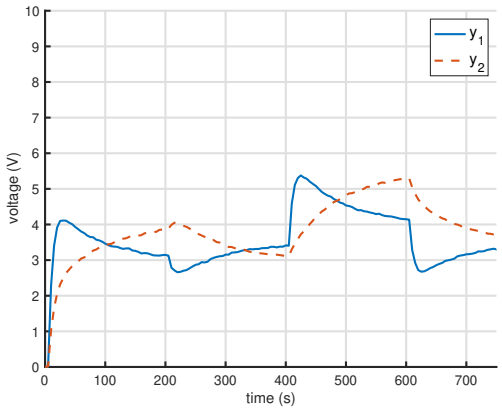


(b) y_1 and y_2 are the voltages in V from the level measurement devices in tanks 1 and 2, respectively.

Figure 5.1: The states and observations of the QTP under normal conditions, with the state space model discretized with a sampling time $T_s = 0.5$s and the desired states of the system x_d changing every 200s.

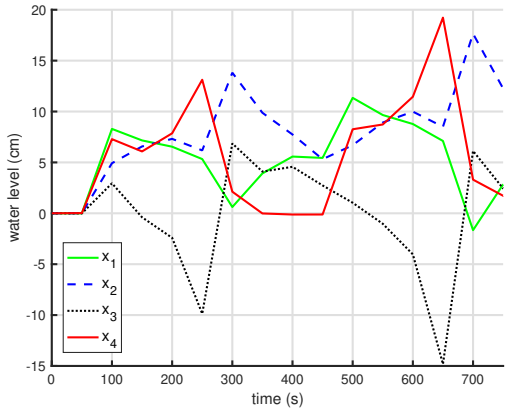


(a) x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4, respectively.

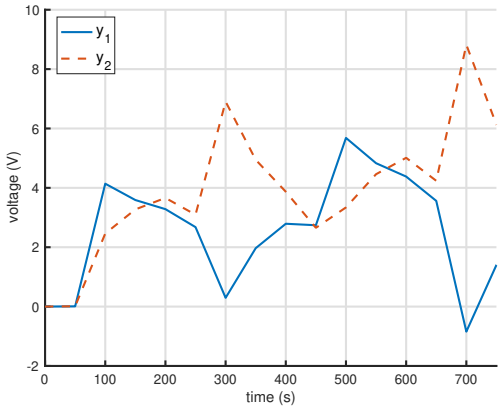


(b) y_1 and y_2 are the voltages in V from the level measurement devices in tanks 1 and 2, respectively.

Figure 5.2: The states an observations of the QTP under normal conditions, with the state space model discretized with a sampling time $T_s = 5$s and the desired states of the system x_d changing every 200s.



(a) x_1, x_2, x_3, and x_4 are the water heights in cm of tanks $1, 2, 3$, and 4, respectively.



(b) y_1 and y_2 are the voltages in V from the level measurement devices in tanks 1 and 2, respectively.

Figure 5.3: The states an observations of the QTP under normal conditions, with the state space model discretized with a sampling time $T_s = 50$s and the desired states of the system x_d changing every 200s.

their desired values before x_d switches. Thus, throughout the rest of this section we will use $T_s = 5$s as the maximum sampling time needed to maintain controllability of the system. Additionally, Figure 5.4 shows the NRMSE between the desired state and the actual state of the system vs. the sampling time used to discretize the state space model. We can see, as Figures 5.1, 5.2, and 5.3 show, that as T_s increases, there is a larger error between the desired states and the actual states, meaning the controller is performing worse.

We give the attacker 250s to learn a state space representation of the system, meaning when T_s increases, the attacker has fewer samples to learn a state space model. However, since the desired states of the system change every 200s, the attacker has more diversified data to learn from. Figure 5.5 shows the actual output of the system compared with the learned output of the system with $T_s = 5$s. The NRMSE between the actual output and the learned output is 13.41% and 16.53% for y_1 and y_2, respectively. This is a much better estimation than the comparable case shown in Figure 4.14 where the desired states remain constant. Additionally, we consider that providing the attacker with more diverse data to learn a state space model of the system could help them learn a more accurate model. However, Figure 5.6 shows this is not the case, as the NRMSE between the actual output

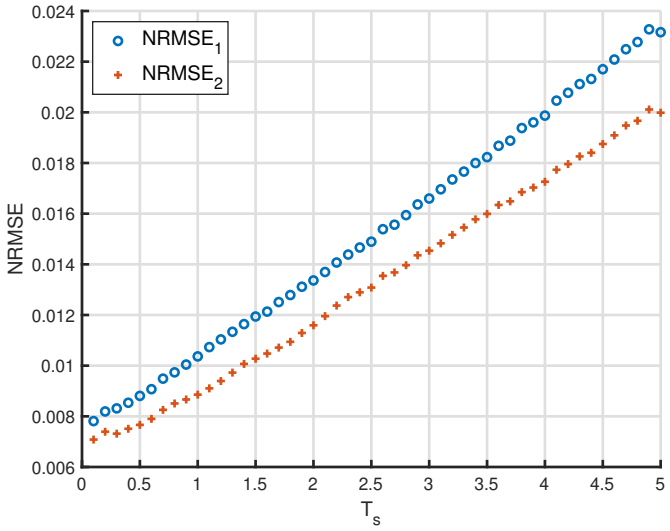


Figure 5.4: The NRMSE between the desired state of the system for states 1 and 2 vs. the sampling time used to discretize the state space model.

of the system and the learned output of the system is comparable to that shown in Figure 4.14 where the desired states remain constant.

Increasing the sampling time used to discretize the state space model of the system comes with pros and cons to the attacker. On one hand, changing the desired states gives the attacker more diverse data to learn from, but on the other hand the attacker has fewer time steps within the 500s given for the attack to diverge the states of the system. We calculate the probability of stealthy attack being successful for $T_s = 0.5, 1,$ and 5s, shown in Figure 5.7. The probability of a successful attack is close to 1 for the case of $T_s = 0.5$s when the attacker knows relatively close bounds for the parameters of the system. However, we see the probability of a successful attack drop drastically when the sampling time is increased to $T_s = 1$ and $5s$. The probability does not change much between the cases of $T_s = 1$ and 5s, suggesting that simply increasing the sampling time to $T_s = 1$s is enough to drastically reduce the probability of an attack being successful. Thus, we can conclude that $T_s = 1$s is completely within the bounds of maintaining controllability. Additionally, merely doubling the sampling time used to discretize the system is enough make the probability of a successful attack three times lower.

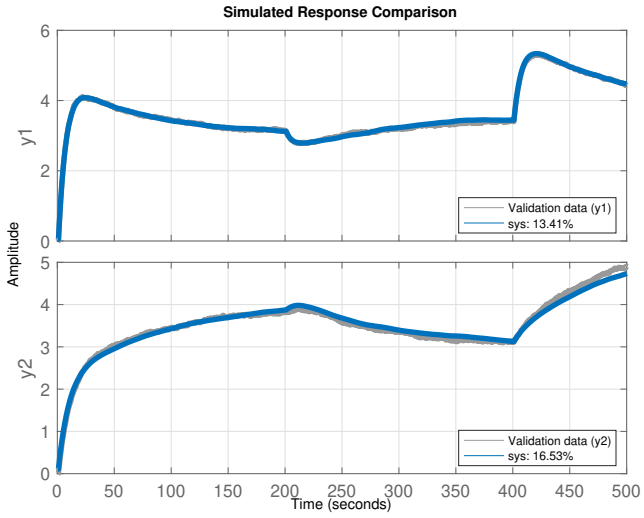


Figure 5.5: A comparison of the actual output of the system, represented by the grey line, with the learned output of the system, represented by the blue line and $T_s = 5$s. We can see that the normalized root mean square error (NRMSE) between the actual output and learned output is 13.41% and 16.53% for y_1 and y_2, respectively.

5.3 Simulation Results on a Power System Model

The impact of stealthy attacks on PEDGs can be catastrophic. However, stealthy attacks have many limitations–including that the state space model must be of non minimum phase, meaning there must be an unstable eigenvalue of $A+BF$. This is because the attack is made in the direction of the main eigenvector of $A+BF$, and without an unstable eigenvalue the system will not diverge when it is attacked. We want to protect PEDGs from a class of attacks, i.e. false data injection attacks, and more specifically stealthy attacks. Protecting a PEDG from such attacks includes making stealthy attacks impossible, making the attack detectable, or minimizing the impact that the attack can make on the system.

We consider preventing stealthy attacks in the design phase by making small changes to the parameters of the system in order to push all of the eigenvalues of $A+BF$ inside of the unit circle.

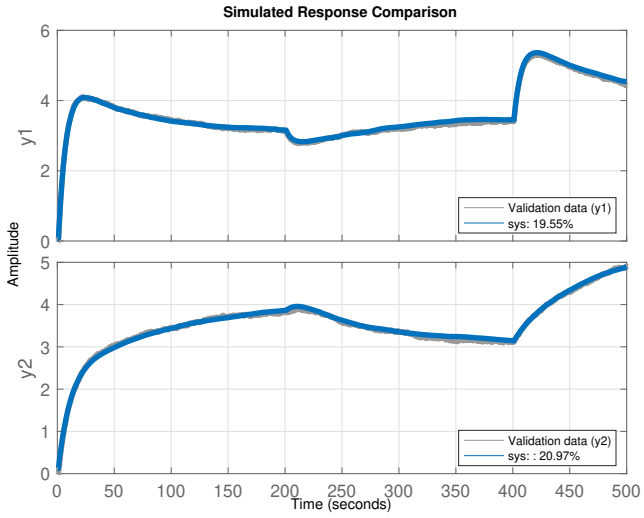


Figure 5.6: A comparison of the actual output of the system, represented by the grey line, with the learned output of the system, represented by the blue line and $T_s = 0.5$s. We can see that the normalized root mean square error (NRMSE) between the actual output and learned output is 19.55% and 20.97% for y_1 and y_2, respectively.

5.3.1 Model Description

We consider a cluster of power-electronics-dominated grids (CPEDGs). This system supplies the local loads and supports the AC bus frequency and voltage. Figure 5.8 shows a CPEDG, where the system consists of a three-phase grid-following inverter and a synchronous generator. Additionally, a MATLAB SIMULINK model of the system, as shown in the *Local control* portion of Figure 5.8, is used to simulate the system running under normal conditions as well as under attack. The zero-dynamics stealthy attack relies on an accurate state space model of the system. Such a state space model is calculated by linearizing the ordinary differential equations that describe the dynamics of the system about an operating point. We first calculate a continuous time system and convert it to discrete time using a sampling time of 10μs. A detailed derivation of the continuous time system and input matrices, A

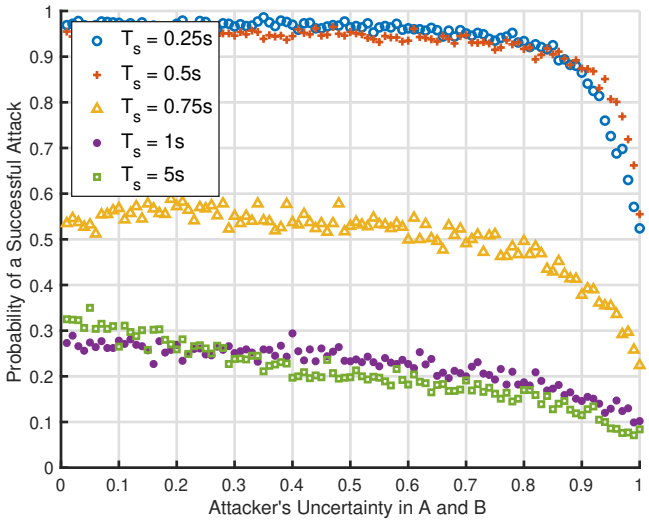


Figure 5.7: The probability that the attack is successful for varying sampling times T_s used to discretize the state space model for the χ^2 variance test vs. the attacker's uncertainty in both A and B. The sampling time T_s varies from 0.5s to 5s. The blue circles represent $T_s = 0.25$s, the red plus signs represent $T_s = 0.5$s, the the yellow triangles represent $T_s = 0.75$s, the purple circles represent $T_s = 1$s, and the green squares represent $T_s = 5$s. Additionally, the variance of the process and measurement noise added to the system is 10^{-4}. We can see that changing T_s has a considerable impact on the attacker's probability of a successful attack.

and B, is shown in Appendix B. The output matrix C is defined by

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5.1}$$

The state variables are defined by

$$\hat{x} = \begin{bmatrix} v_c & i_q & i_d & v_q^f & v_d^f & i_{Lq} & i_{Ld} \end{bmatrix}^T, \tag{5.2}$$

where v_c is the DC-line voltage of the inverter, i_q and i_d are the q and d components of inverter-side current, v_q^f and v_d^f are the q and d components of the filter's capacitor voltage, and i_{Lq} and i_{Ld} are the q and d components of the grid-side current at the point of common

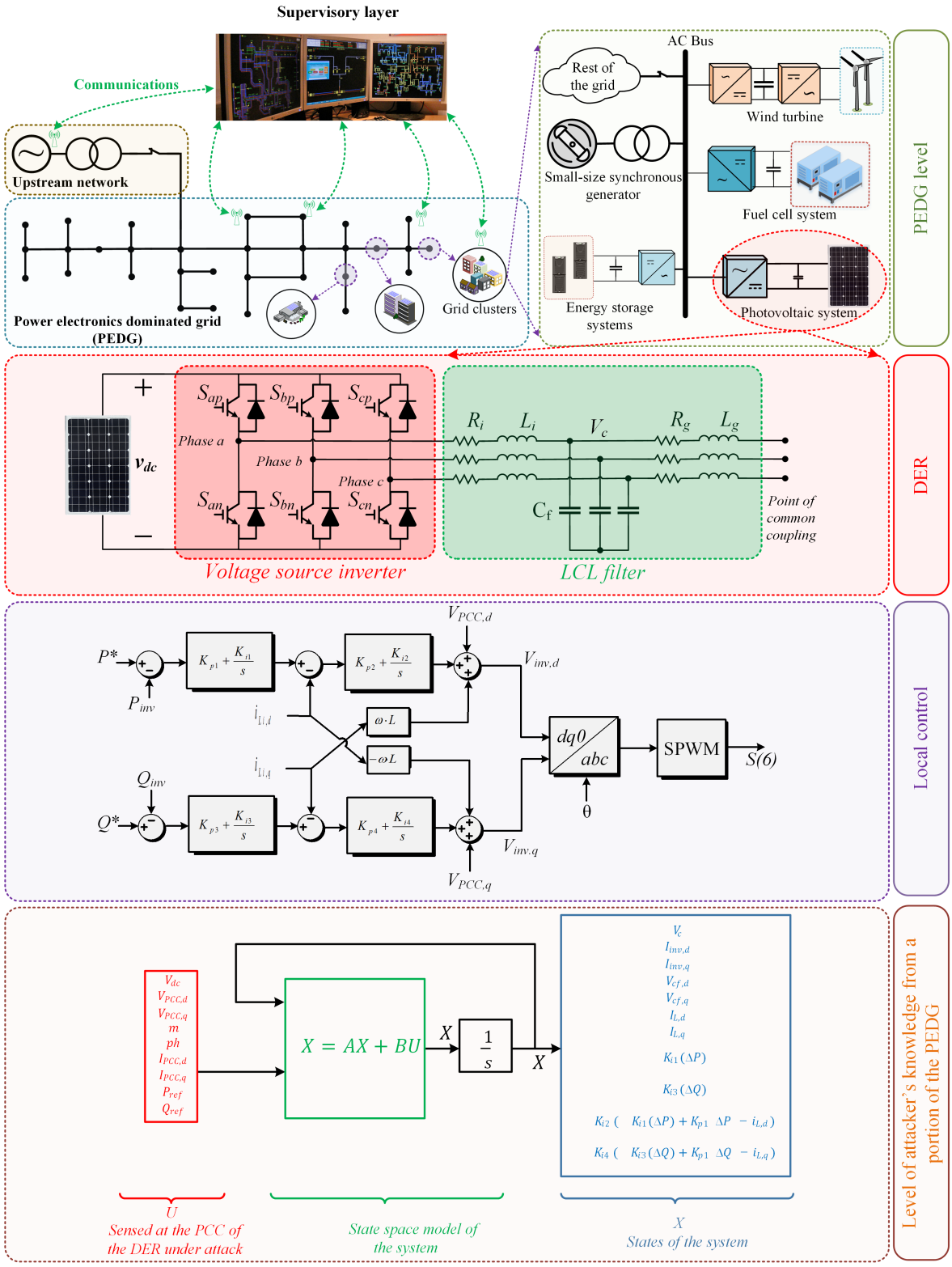


Figure 5.8: A generic overview of the power-electronics-dominated grid (PEDG) consisting of grid clusters with high penetration of renewable resources.

coupling. The control signal is defined by

$$\hat{u} = \begin{bmatrix} v_{dc} & v_q^g & v_d^g \end{bmatrix}^T, \tag{5.3}$$

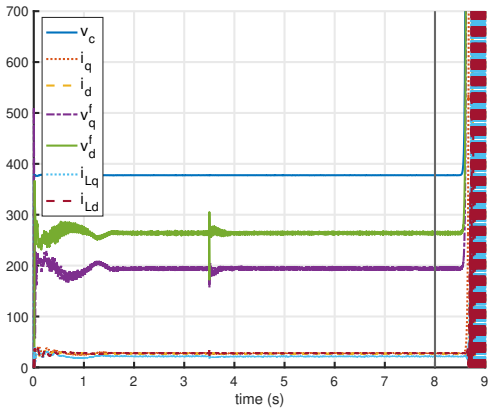
where v_{dc} is the inverter input voltage, v_q^g and v_d^g are the filter capacitor voltage transferred to the d-q frame.

The novelty of applying a stealthy attack to this PEDG is the fact that the state space model only represents a subset of the whole system. Additionally, the state space model of the inverters requires input from the synchronous generator, shown in Fig. 5.8. This means that as the control signal from the circuit simulation of the PEDG changes, the state space model of the inverters change. However, in order to calculate a stealthy attack, we need a time invariant state space representation. Here, we calculate the state space model using the steady state values from the circuit simulation.

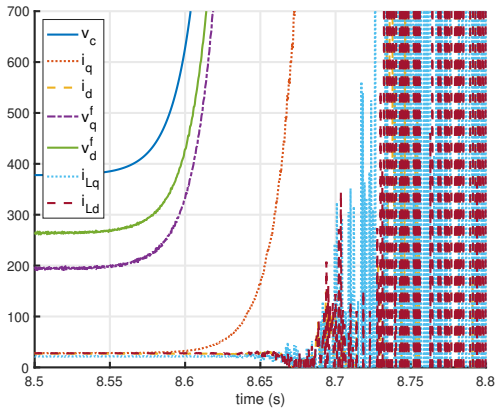
5.3.2 Results

We want to calculate the values for parameters of the state space model that make a stealthy attack impossible. Figures 5.9 and 5.10 show the states and observations of the system with an attack starting at 8s. We can see that the states of the system diverge about 0.65s before the observations diverge, meaning the attack is successful. This simulation will be used as a baseline to compare with the state space model that eliminates a stealthy attack.

In order to calculate the values of parameters that prevent a stealthy attack from being possible, we consider a range of possible values for the constant parameters, found in Table B.1, in the state space model. In this Chapter, we simplify the problem by modifying two parameters at a time. First, we considering modifying L_1 and L_2, shown in Figure 5.11. The desired values for L_1 and L_2 are 0.001H and 0.0005H, respectively. We chose to consider $L_1 \in [0.0001, 0.1]$ and $L_2 \in [0.00005, 0.05]$ to ensure all realistic options for L_1 and L_2 were considered. A state space representation was then calculated for all combinations of L_1 and L_2 in these ranges in intervals of 0.00001. For each state space representation,

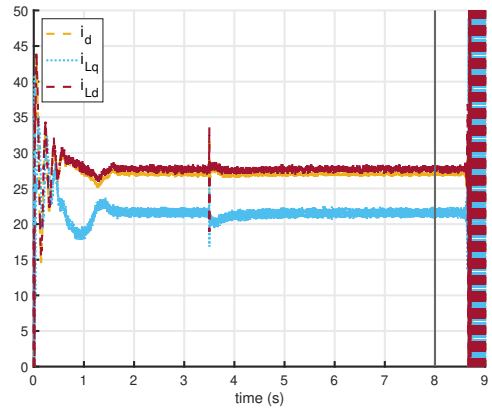


(a) The states of the system with an attack starting at 8s.

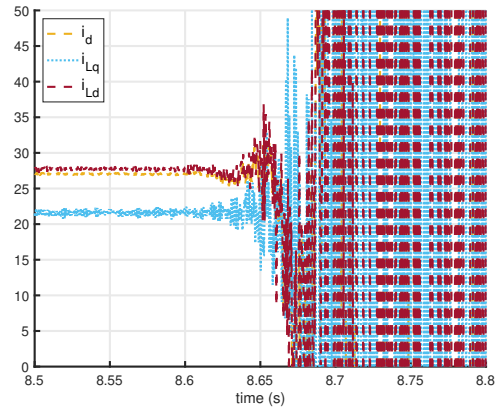


(b) The states of the system under attack–a zoomed-in graph of Fig. 5.9a showing the divergence of the states.

Figure 5.9: System states with an attack starting at 8s, denoted by the black vertical line. Note that the states being attacked diverge significantly around 8.55s. The states and observations oscillate after diverging due to numerical precision, as the attack diverges the states exponentially.



(a) The observations of the system with an attack starting at 8s.



(b) The observations of the system under attack–a zoomed-in graph of Fig. 5.10a showing that the attack becomes observable around 8.65s.

Figure 5.10: System observations with an attack starting at 8s, denoted by the black vertical line. Note that observations begin to diverge about 0.1s after the states begin to diverge. The states and observations oscillate after diverging due to numerical precision, as the attack diverges the states exponentially.

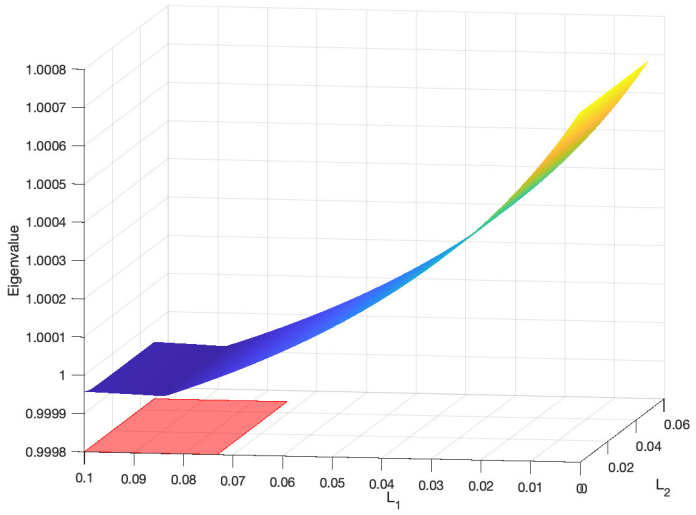


Figure 5.11: The main eigenvalue of $A + BF$ with varying values for L_1 and L_2. A stealthy attack is impossible for $L_1 \in [0.0732, 0.1]$ and $L_2 \in [0.00005, 0.05]$, shown by the red rectangle.

the main eigenvalue of $A + BF$ is calculated, and when this eigenvalue is in the unit circle a stealthy attack is not possible. The values that make a stealthy attack impossible are $L_1 \in [0.0732, 0.1]$ and $L_2 \in [0.0005, 0.05]$, thus we choose $L1 = 0.0732$H and $L_2 = 0.0005$H, as these are the closest to their desired values.

Additionally, we consider changing the DC link capacitor and the inverter side inductor, C_1 and L_1, in order to prevent a stealthy attack. We consider $C_1 \in [0.000018, 0.18]$ and $L_1 \in [0.0001, 0.1]$, where the desired values of C_1 and L_1 are 0.0018F and 0.001H, respectively. From Figure 5.12, we can see that $C_1 \in [0.000018, 0.18]$ and $L_1 \in [0.0732, 0.1]$ make all eigenvalues of $A + BF$ inside the unit circle, meaning a stealthy attack is impossible. We choose $C_1 = 0.0018$F and $L_1 = 0.0732$H, as these are the closest to their desired values.

Figure 5.13 shows the states an observations of the system under attack with the modified values of C_1 and L_1 chosen above. We can see that the attack is unsuccessful, as there is no change in the states of the system. This means that simply choosing alternative values for parameters in the design phase of a system can completely eliminate the possibility of a stealthy attack.

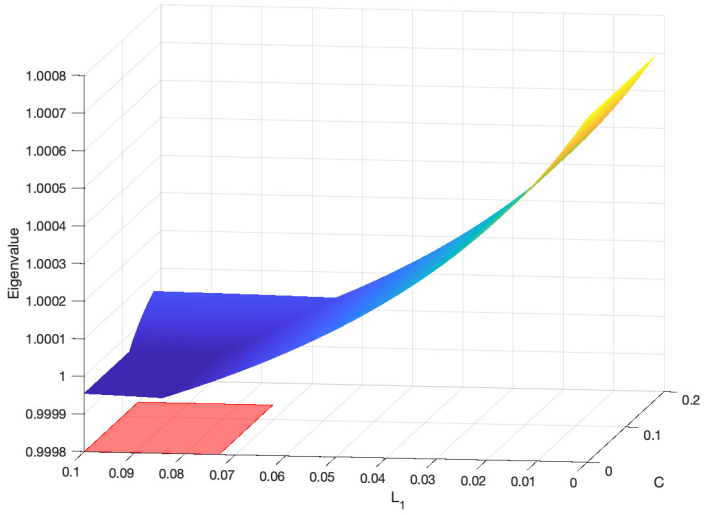
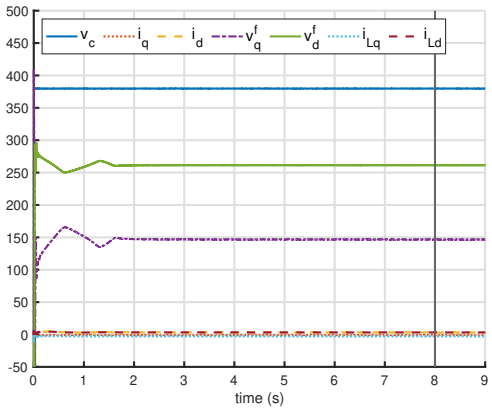
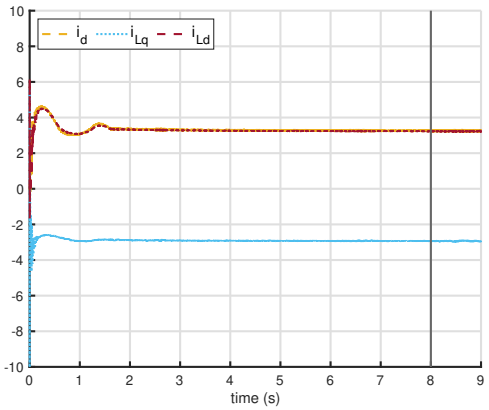


Figure 5.12: The main eigenvalue of $A + BF$ with varying values for C_1 and L_1. A stealthy attack is impossible for $C_1 \in [0.000018, 0.139818]$ and $L_1 \in [0.0732, 0.1]$, shown by the red rectangle.



(a) The states of the system with an attack starting at 8s.



(b) The observations of the system with an attack starting at 8s.

Figure 5.13: The states and observations of the system with an attack starting at 8s, denoted by the black vertical line, with modified values for C_1 and L_1. Note that the states do not diverge, as some of the system parameters were slightly modified in order to make the system minimum phase.

Chapter 6

Conclusion

Stealthy attacks have the potential to cause considerable damage to control systems, thus it is important to understand the feasibility and impact of such attacks. A successful attack on control systems can cause physical damage to the system, or in the case of a power system, a blackout. The main motivation of this dissertation is to show methods that can be used to limit the probability of a stealthy attack being successful even under realistic conditions and plausible attacker models (2.2.2). Additionally, we investigated what can be done to limit the impact of stealthy attacks, including eliminating the possibility of a stealthy attack being successful.

6.1 Review of the Contributions

Chapter 2 considered the impact of incorporating more realistic conditions to the system being attacked. We integrated a Linear Quadratic Gaussian (LQG) controller into the system, as well as reformulated a few of the equations for the LQG to push the states of the system to a desired value rather than zero. Contrary to previous work, we assumed that the attacker would not know or have access to a state space model of the system – thus they would have to learn this information. As the attacker had to learn a state space representation, we assumed that the learned model would not perfectly match the system. We investigated the

impact on a stealthy attack of the attacker having an imperfect state space representation of the system. We demonstrated on the Quadruple-Tank Process (QTP), as well as on a simple power system model, that a stealthy attack can be detectable when the assumption that the attacker knows a state space model of the system is removed.

In Chapter 3, we considered how varying the attacker's uncertainty in the state space model affects the probability of an attack being successful. Additionally, we analyzed the impact of changing the process and measurement noise added to the system as well as the false alarm rate of the anomaly detector, concluding that increasing the process and measurement noise is the best method to minimize the probability of a successful attack. We considered placing further limitations on the attacker by limiting the amount of energy they are able to inject into the control signal, showing that it is necessary to not overbuild a system, as having the physical limits of the system as close to the system requirements limits the probability of a successful attack.

Applying system identification algorithms to model the attacker learning a state space representation of the system is considered in Chapter 4. We first analyze if an attacker can successfully attack the system using a black box approach to system identification–meaning they do not know any information about the system, they are only able to use the inputs and outputs of the system to learn. We concluded the attacker was not able to learn an accurate enough state space model of the system in order to successfully mount an attack. Thus, we then loosened the restrictions on the attacker and assumed it is realistic for them to have some prior knowledge about the system. We assumed that the attacker knows upper and lower bounds for the parameters used to calculate the state space model.

Chapter 5 considered what can be done to limit the impact and prevent stealthy attacks. We first considered the trade off between making the system less susceptible to stealthy attacks and maintaining its controllability by increasing the sampling time used to discretize the state space model of the system. Increasing the sampling time of the system provides the attacker with fewer samples to learn a state space representation of the system as well as fewer samples to diverge the states of the system once the attack begins. We showed that the sampling time only needed to be doubled in order to drastically reduce the probability of

an attack being successful, thus this approach is an acceptable method to limit the impact of a stealthy attack. Additionally, we considered making modifications to the parameters of a three inverter power system model in order to eliminate the possibility of a stealthy attack. We were able to show that making small changes to multiple parameters of the system had the power to make stealthy attacks fail.

6.2 Limitations and Future Work

Future work could include designing a machine learning model to learn a stealthy attack rather than an attacker having to understand the mathematics behind a stealthy attack and then learning a state space model of the system. Reinforcement learning could be used for the attack to diverge the states of the system while remaining stealthy.

Additionally, investigating the impact of stealthy attacks in a lab environment on physical system rather than simulations would be an interesting extension of this work.

We believe another way to extend this work could include investigating more of the mathematical theory behind stealthy attacks. This could include determining exactly how close at attacker needs to learn a state space model in order to avoid various types of anomaly detectors.

6.3 Practical Insights

We believe the following methods should be considered in order to limit the impact and prevent stealthy attacks:

1. Limit overbuilding the system whenever possible. When the system allows for more energy than it requires, a stealthy attack has a much higher probability of being successful (3.14).

2. Make small improvements to the anomaly detector, as detecting an attack a few seconds earlier can make the attack unsuccessful (4.21).

3. Increase the sampling time used to discretize the system in order to prevent an attacker from learning and delay the impact of the attack (5.7).

4. Make small changes to the parameters of the system during the design phase to eliminate the possibility of a stealthy attack (5.11, 5.12).

Bibliography

Lily Hay Newman. Colonial Pipeline paid a $5M ransom–and kept a vicious cycle turning, May 2021. URL https://www.wired.com/story/colonial-pipeline-ransomware-payment/.

Andy Greenberg. A hacker tried to poison a Florida city's water supply, officials say, February 2021. URL https://www.wired.com/story/oldsmar-florida-water-utility-hack/.

David E. Sanger and William P. Davis. Ransomware disrupts meat plants in latest attack on critical U.S. business, June 2021. URL https://www.nytimes.com/2021/06/01/business/meat-plant-cyberattack-jbs.html.

André Teixeira, Iman Shames, Henrik Sandberg, and Karl H Johansson. Revealing stealthy attacks in control systems. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1806–1813. IEEE, 2012.

Nebrase Elmrabit, Feixiang Zhou, Fengyin Li, and Huiyu Zhou. Evaluation of machine learning algorithms for anomaly detection. In *International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, pages 1–8. IEEE, 2020.

Mohsen Hosseinzadehtaher, Ahmad Khan, Mohammad B Shadmand, and Haitham Abu-Rub. Anomaly detection in distribution power system based on a condition monitoring vector and ultra-short demand forecasting. In *IEEE CyberPELS*, pages 1–6. IEEE, 2020.

Jun Inoue, Yoriyuki Yamagata, Yuqi Chen, Christopher M Poskitt, and Jun Sun. Anomaly detection for a water treatment system using unsupervised machine learning. In *IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 1058–1065. IEEE, 2017.

Stephanie Harshbarger, Mohsen Hosseinzadehtaher, Bala Natarajan, Eugene Vasserman, Mohammad Shadmand, and George Amariucai. (a little) ignorance is bliss: The effect of imperfect model information on stealthy attacks in power grids. In *IEEE Kansas Power and Energy Conference (KPEC)*, pages 1–6. IEEE, 2020.

Ruilong Deng, Gaoxi Xiao, Rongxing Lu, Hao Liang, and Athanasios V Vasilakos. False data injection on state estimation in power systems—attacks, impacts, and defense: A survey. *IEEE Transactions on Industrial Informatics*, 13(2):411–423, 2016.

Souhila Aoufi, Abdelouahid Derhab, and Mohamed Guerroumi. Survey of false data injection in smart power grid: Attacks, countermeasures and challenges. *Journal of Information Security and Applications*, 54:102518, 2020.

Yilin Mo, Emanuele Garone, Alessandro Casavola, and Bruno Sinopoli. False data injection attacks against state estimation in wireless sensor networks. In *49th IEEE Conference on Decision and Control (CDC)*, pages 5967–5972. IEEE, 2010.

Yilin Mo and Bruno Sinopoli. On the performance degradation of cyber-physical systems under stealthy integrity attacks. *IEEE Transactions on Automatic Control*, 61(9):2618–2624, 2015.

Qingyu Yang, Jie Yang, Wei Yu, Dou An, Nan Zhang, and Wei Zhao. On false data-injection attacks against power system state estimation: Modeling and countermeasures. *IEEE Transactions on Parallel and Distributed Systems*, 25(3):717–729, 2013.

Zhong-Hua Pang, Guo-Ping Liu, Donghua Zhou, Fangyuan Hou, and Dehui Sun. Two-channel false data injection attacks against output tracking control of networked systems. *IEEE Transactions on Industrial Electronics*, 63(5):3242–3251, 2016.

Rongkuan Ma, Peng Cheng, Zhenyong Zhang, Wenwen Liu, Qingxian Wang, and Qiang Wei. Stealthy attack against redundant controller architecture of industrial cyber-physical system. *IEEE Internet of Things Journal*, 6(6):9783–9793, 2019.

Xingxing Wei, Ying Guo, and Jie Yu. Adversarial sticker: A stealthy attack method in the physical world. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

André Teixeira, György Dán, Henrik Sandberg, and Karl H Johansson. A cyber security study of a scada energy management system: Stealthy deception attacks on the state estimator. *IFAC Proceedings Volumes*, 44(1):11271–11277, 2011.

Pritam Dash, Mehdi Karimibiuki, and Karthik Pattabiraman. Out of control: Stealthy attacks against robotic vehicles protected by control-based techniques. In *Proceedings of the 35th Annual Computer Security Applications Conference*, pages 660–672, 2019.

Cheolhyeon Kwon, Weiyi Liu, and Inseok Hwang. Analysis and design of stealthy cyber attacks on unmanned aerial systems. *Journal of Aerospace Information Systems*, 11(8): 525–539, 2014.

Saurabh Amin, Xavier Litrico, S Shankar Sastry, and Alexandre M Bayen. Stealthy deception attacks on water SCADA systems. In *Proceedings of the 13th ACM international conference on Hybrid systems: computation and control*, pages 161–170, 2010.

Shaunak D Bopardikar and Alberto Speranzon. On analysis and design of stealth-resilient control systems. In *2013 6th International Symposium on Resilient Control Systems (ISRCS)*, pages 48–53, 2013.

Ruochi Zhang and Parv Venkitasubramaniam. Stealthy control signal attacks in linear quadratic gaussian control systems: detectability reward tradeoff. *IEEE Transactions on Information Forensics and Security*, 12(7):1555–1570, 2017.

Sangjun Kim and Kyung-Joon Park. A survey on machine-learning based security design for cyber-physical systems. *Applied Sciences*, 11(12):5458, 2021.

Derui Ding, Qing-Long Han, Yang Xiang, Xiaohua Ge, and Xian-Ming Zhang. A survey on security control and attack detection for industrial cyber-physical systems. *Neurocomputing*, 275:1674–1683, 2018.

Dan Zhang, Qing-Guo Wang, Gang Feng, Yang Shi, and Athanasios V Vasilakos. A survey on attack detection, estimation and control of industrial cyber–physical systems. *ISA transactions*, 116:1–16, 2021.

Sridhar Adepu and Aditya Mathur. Distributed attack detection in a water treatment plant: Method and case study. *IEEE Transactions on Dependable and Secure Computing*, 18(1): 86–99, 2018.

Ali Sayghe, Yaodan Hu, Ioannis Zografopoulos, XiaoRui Liu, Raj Gautam Dutta, Yier Jin, and Charalambos Konstantinou. Survey of machine learning methods for detecting false data injection attacks in power systems. *IET Smart Grid*, 3(5):581–595, 2020.

Nur Imtiazul Haque, Md Hasan Shahriar, Md Golam Dastgir, Anjan Debnath, Imtiaz Parvez, Arif Sarwat, and Mohammad Ashiqur Rahman. Machine learning in generation, detection, and mitigation of cyberattacks in smart grid: A survey. *arXiv preprint arXiv:2010.00661*, 2020.

Mohammad Esmalifalak, Lanchao Liu, Nam Nguyen, Rong Zheng, and Zhu Han. Detecting stealthy false data injection using machine learning in smart grid. *IEEE Systems Journal*, 11(3):1644–1652, 2014.

Mohammad Ashrafuzzaman, Yacine Chakhchoukh, Ananth A Jillepalli, Predrag T Tosic, Daniel Conte de Leon, Frederick T Sheldon, and Brian K Johnson. Detecting stealthy false data injection attacks in power grids using deep learning. In *14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pages 219–225. IEEE, 2018.

Jiexi Wang, Yingxu Lai, and Jing Liu. Stealthy attack detection method based on multi-feature long short-term memory prediction model. *Future Generation Computer Systems*, 137:248–259, 2022.

Jairo Giraldo, David Urbina, Alvaro Cardenas, Junia Valente, Mustafa Faisal, Justin Ruths, Nils Ole Tippenhauer, Henrik Sandberg, and Richard Candell. A survey of physics-based

attack detection in cyber-physical systems. *ACM Computing Surveys (CSUR)*, 51(4):1–36, 2018.

Mazen Azzam, Liliana Pasquale, Gregory Provan, and Bashar Nuseibeh. Grounds for suspicion: physics-based early warnings for stealthy attacks on industrial control systems. *arXiv preprint arXiv:2106.07980*, 2021.

MR Gauthama Raman and Aditya P Mathur. A hybrid physics-based data-driven framework for anomaly detection in industrial control systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021.

Bharadwaj Satchidanandan and Panganamala R Kumar. Dynamic watermarking: Active defense of networked cyber–physical systems. *Proceedings of the IEEE*, 105(2):219–240, 2016.

Riccardo MG Ferrari and André MH Teixeira. Detection of cyber-attacks: A multiplicative watermarking scheme. In *Safety, Security and Privacy for Cyber-Physical Systems*, pages 173–201. Springer, 2021.

Kumarsinh Jhala, Parth Pradhan, and Balasubramaniam Natarajan. Perturbation-based diagnosis of false data injection attack using distributed energy resources. *IEEE Transactions on Smart Grid*, 12(2):1589–1601, 2020.

Ming Zhang, Zizhan Zheng, and Ness B Shroff. A game theoretic model for defending against stealthy attacks with limited resources. In *International Conference on Decision and Game Theory for Security*, pages 93–112. Springer, 2015.

Anastasis Keliris and Michail Maniatakos. ICSREF: A framework for automated reverse engineering of industrial control systems binaries. In *Network and Distributed Systems Security Symposium (NDSS)*, 2019.

Tony Flick and Justin Morehouse. *Securing the smart grid: Next generation power grid security*. Elsevier, 2010.

David I Urbina, Jairo A Giraldo, Alvaro A Cardenas, Nils Ole Tippenhauer, Junia Valente, Mustafa Faisal, Justin Ruths, Richard Candell, and Henrik Sandberg. Limiting the impact of stealthy attacks on industrial control systems. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 1092–1105, 2016.

Wissam Aoudi, Mikel Iturbe, and Magnus Almgren. Truth will out: Departure-based process-level detection of stealthy attacks on control systems. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 817–831, 2018.

Andre Teixeira, Kin Cheong Sou, Henrik Sandberg, and Karl Henrik Johansson. Secure control systems: A quantitative risk management approach. *IEEE Control Systems Magazine*, 35(1):24–45, 2015.

Gyorgy Dan and Henrik Sandberg. Stealth attacks and protection schemes for state estimators in power systems. In *IEEE International Conference on Smart Grid Communications*, pages 214–219, 2010.

Mohammad Esmalifalak, Huy Nguyen, Rong Zheng, and Zhu Han. Stealth false data injection using independent component analysis in smart grid. In *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 244–248, 2011.

Mohammad Ashiqur Rahman, Ehab Al-Shaer, and Md Ashfaqur Rahman. A formal model for verifying stealthy attacks on state estimation in power grids. In *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 414–419, 2013.

Aditya Ashok, Manimaran Govindarasu, and Venkataramana Ajjarapu. Online detection of stealthy false data injection attacks in power system state estimation. *IEEE Transactions on Smart Grid*, 9(3):1636–1646, 2016.

Mehmet Necip Kurt, Yasin Yılmaz, and Xiaodong Wang. Real-time detection of hybrid and stealthy cyber-attacks in smart grid. *IEEE Transactions on Information Forensics and Security*, 14(2):498–513, 2018.

André Teixeira, Saurabh Amin, Henrik Sandberg, Karl H Johansson, and Shankar S Sastry. Cyber security analysis of state estimators in electric power systems. In *IEEE conference on decision and control (CDC)*, pages 5991–5998, 2010.

Yacine Chakhchoukh and Hideaki Ishii. Enhancing robustness to cyber-attacks in power systems through multiple least trimmed squares state estimations. *IEEE Transactions on Power Systems*, 31(6):4395–4405, 2016.

Ognjen Vuković, Kin Cheong Sou, György Dán, and Henrik Sandberg. Network-layer protection schemes against stealth attacks on state estimators in power systems. In *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 184–189, 2011.

Mohammad Esmalifalak, Zhu Han, and Lingyang Song. Effect of stealthy bad data injection on network congestion in market based power system. In *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 2468–2472. IEEE, 2012.

Yifa Liu, Wenchao Xue, Shuping He, and Long Cheng. Stealthy false data injection attacks against extended kalman filter detection in power grids. In *2021 8th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)*, pages 459–464. IEEE, 2021.

Jiwei Tian, Buhong Wang, Zhen Wang, Kunrui Cao, Jing Li, and Mete Ozay. Joint adversarial example and false data injection attacks for state estimation in power systems. *IEEE Transactions on Cybernetics*, 2021.

Chensheng Liu, Hao Liang, and Tongwen Chen. Network parameter coordinated false data injection attacks against power system ac state estimation. *IEEE Transactions on Smart Grid*, 12(2):1626–1639, 2020.

Siddhartha Deb Roy and Sanjoy Debbarma. A novel oc-svm based ensemble learning framework for attack detection in agc loop of power systems. *Electric Power Systems Research*, 202:107625, 2022.

Shan Liu, Bo Chen, Takis Zourntos, Deepa Kundur, and Karen Butler-Purry. A coordinated multi-switch attack for cascading failures in smart grid. *IEEE Transactions on Smart Grid*, 5(3):1183–1195, 2014.

Shan Liu, Salman Mashayekh, Deepa Kundur, Takis Zourntos, and Karen Butler-Purry. A framework for modeling cyber-physical switching attacks in smart grid. *IEEE Transactions on Emerging Topics in Computing*, 1(2):273–285, 2013.

Daranith Choeum and Dae-Hyun Choi. Oltc-induced false data injection attack on volt/var optimization in distribution systems. *IEEE Access*, 7:34508–34520, 2019.

Brian DO Anderson. Output-nulling invariant and controllability subspaces. *IFAC Proceedings Volumes*, 8(1):337–345, 1975.

Karl Henrik Johansson. The quadruple-tank process: A multivariable laboratory process with an adjustable zero. *IEEE Transactions on control systems technology*, 8(3):456–465, 2000.

Appendix A

LQG Derivation

Define the Bellman equation for the LQR:

$$V_t(z) = (z - x_d)^T Q(z - x_d) + \min_{\omega}(\omega^T R\omega + V_{t+1}(Az + B\omega)), \tag{A.1}$$

where $t = 0, \ldots, N$ and x_d is the desired state of the system. An ω that minimizes the above equation will give the optimal control input. We will assume that

$$V_{t+1}(z) = z^T G_{t+1} z - H_{t+1} z - z^T T_{t+1} + C_{t+1}, \tag{A.2}$$

we will show that $V_t(z)$ has the same form.

$$\begin{aligned} V_t(z) = (z - x_d)^T Q(z - x_d) + \min_{\omega}(\omega^T R\omega + (Az + B\omega)^T G_{t+1}(Az + B\omega) \\ - H_{t+1}(Az + B\omega) - (Az + B\omega)^T T_{t+1} + C_{t+1}) \end{aligned} \tag{A.3}$$

In order to find the optimal control input, we must take the derivative of (A.3) with respect to ω

$$\begin{aligned} 0 &= 2\omega^T R + 2(Az + B\omega)^T G_{t+1} B - H_{t+1} B - T_{t+1}^T B \\ \omega &= \frac{1}{2}(R + B^T G_{t+1} B)^{-1}(-2B^T G_{t+1} Az + B^T H_{t+1} + B^T T_{t+1}). \end{aligned} \tag{A.4}$$

Now, we can substitute the optimal control signal into (A.3):

$$V_t(z) = (z - x_d)^T Q(z - x_d) + \frac{1}{2}(-2z^T A^T G_{t+1} B + H_{t+1}^T B + T_{t+1}^T B) \tag{A.5}$$
$$(R + B^T G_{t+1} B)^{-1} R \frac{1}{2}(R + B^T G_{t+1} B)^{-1}(-2B^T G_{t+1} Az + B^T H_{t+1} + B^T T_{t+1})$$
$$+ (z^T A^T + \frac{1}{2}(-2z^T A^T G_{t+1} B + H_{t+1}^T B + T_{t+1}^T B)(R + B^T G_{t+1} B)^{-1} B^T) G_{t+1}(Az$$
$$+ \frac{1}{2} B(R + B^T G_{t+1} B)^{-1}(-2B^T G_{t+1} Az + B^T H_{t+1} + B^T T_{t+1})) - H_{t+1}(Az + \frac{1}{2} B(R + B^T G_{t+1} B)^{-1}$$
$$(-2B^T G_{t+1} Az + B^T H_{t+1} + B^T T_{t+1})) - (z^T A^T + \frac{1}{2}(-2z^T A^T G_{t+1} B + H_{t+1}^T B + T_{t+1}^T B)$$
$$(R + B^T G_{t+1} B)^{-1} B^T) T_{t+1} + C_{t+1}.$$

Since (A.4) does not contain C_{t+1}, we will only consider terms in (A.5) that contain z in order to simplify the calculations. After come simplification, we can obtain

$$V_t(z) = z^t(Q + A^T G_{t+1} A - A^T G_{t+1} B(R + B^T G_{t+1} B)^{-1} B^T G_{t+1} A) z \tag{A.6}$$
$$- (x_d^T Q + H_{t+1} A - H_{t+1} B(R + B^T G_{t+1} B)^{-1} B^T G_{t+1} A) z - z^T (Q x_d +$$
$$A^T T_{t+1} - A^T G_{t+1} B(R + B^T G_{t+1} B)^{-1} T_{t+1}),$$

where G_t, H_t, and T_t are defined using backwards recursion by the equations below:

$$H_k = H_N + H_{k+1} A - H_{k+1} B(B^T G_{k+1} B + R)^{-1} B^T G_{k+1} A, \tag{A.7}$$

$$G_k = A^T(G_{k+1} - G_{k+1} B(B^T G_{k+1} B + R) B^T G_{k+1}) A + Q, \tag{A.8}$$

$$T_k = T_N + A^T T_{k+1} - A^T G_{k+1} B(B^T G_{k+1} B + R)^{-1} B^T G_{k+1}. \tag{A.9}$$

Appendix B

PEDG State Space Equations

The system configuration is calculated using the IEEE Standard 399-1997 with some modifications. The states are transferred to the d-q frame with respect to a global d-q frame. The state space matrices A, B, and C are calculated using the dynamic equations of the aggregated inverter model.

The system matrix is given by

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix}$$

where, A_{11}, ..., A_{44} are defined as follows:

$$\begin{aligned} A_{11} &= \frac{-1}{C_1 R_s}, \\ A_{12} &= \begin{bmatrix} \frac{-\sqrt{3} m \cos\phi}{2C_1} & \frac{-\sqrt{3} m \sin\phi}{2C_1} \end{bmatrix}, \\ A_{13} &= \begin{bmatrix} 0 & 0 \end{bmatrix}, \\ A_{14} &= \begin{bmatrix} 0 & 0 \end{bmatrix}, \end{aligned}$$

$$A_{21} = \begin{bmatrix} \dfrac{\sqrt{3}m\cos(\phi)}{3L_1} & \dfrac{\sqrt{3}m\sin(\phi)}{3L_1} \end{bmatrix}^T,$$

$$A_{22} = \begin{bmatrix} \dfrac{-3R_1 - R_f}{3L_1} & -\omega \\ \omega & \dfrac{-3R_1 - R_f}{3L_1} \end{bmatrix},$$

$$A_{23} = \begin{bmatrix} \dfrac{-1}{2L_1} & \dfrac{\sqrt{3}}{6L_1} \\ \dfrac{-\sqrt{3}}{6L_1} & \dfrac{-1}{2L_1} \end{bmatrix},$$

$$A_{24} = \begin{bmatrix} \dfrac{R_f}{3L_1} & 0 \\ 0 & \dfrac{R_f}{3L_1} \end{bmatrix},$$

$$A_{31} = \begin{bmatrix} 0 & 0 \end{bmatrix}^T,$$

$$A_{32} = \begin{bmatrix} \dfrac{1}{2C_f} & \dfrac{\sqrt{3}}{6C_f} \\ \dfrac{-\sqrt{3}}{6C_f} & \dfrac{1}{2C_f} \end{bmatrix},$$

$$A_{33} = \begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix},$$

$$A_{34} = \begin{bmatrix} \dfrac{-1}{2C_f} & \dfrac{-\sqrt{3}}{6C_f} \\ \dfrac{\sqrt{3}}{6C_f} & \dfrac{-1}{2C_f} \end{bmatrix},$$

$$A_{41} = \begin{bmatrix} 0 & 0 \end{bmatrix}^T,$$

$$A_{42} = \begin{bmatrix} \dfrac{R_f}{3L_2} & 0 \\ 0 & \dfrac{R_f}{3L_2} \end{bmatrix},$$

$$A_{43} = \begin{bmatrix} \dfrac{1}{2L_2} & \dfrac{-\sqrt{3}}{6L_2} \\ \dfrac{\sqrt{3}}{6L_2} & \dfrac{1}{2L_2} \end{bmatrix},$$

$$A_{44} = \begin{bmatrix} \dfrac{-3R_2 - R_f}{3L_2} & -\omega \\ \omega & \dfrac{-3R_2 - R_f}{3L_2} \end{bmatrix}.$$

The input matrix B is given by

$$B = \begin{bmatrix} B_{11} & B_{12} & B_{14} & B_{15} & B_{16} & B_{17} \end{bmatrix}^T$$

where, B_{11}, B_{12}, B_{13}, B_{14}, B_{15}, B_{16}, and B_{17} are defined as follows:

$$\begin{aligned}
B_{11} &= \begin{bmatrix} \frac{1}{C_1 R_s} & 0 & 0 \end{bmatrix}, \\
B_{12} &= \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \\
B_{13} &= \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \\
B_{14} &= \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \\
B_{15} &= \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \\
B_{16} &= \begin{bmatrix} 0 & \frac{-1}{L_2} & 0 \end{bmatrix}, \\
B_{17} &= \begin{bmatrix} 0 & 0 & \frac{-1}{L_2} \end{bmatrix}.
\end{aligned}$$

Values for the constants are provided in Table B.1.

R_s	0.1 Ω	m	0.849 pu
C_1	$1800e^{-6}$ F	ϕ	1.37 rad
L_1	$1e^{-3}$ H	ω	120π rad/s
R_1	0.15 Ω	R_2	0.8 Ω
R_f	0.7 Ω	L_2	$0.5e^{-3}$ H
C_f	$30e^{-6}$ F		

Table B.1: Values for the constants in the state space representation