8-2018

ANALYSIS OF MULTIPLE SIGNAL ATTACKS ON CONTROL SYSTEMS

Nedas Jakstas
University of Tennessee, njakstas@vols.utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Nedas Jakstas entitled "ANALYSIS OF MULTIPLE SIGNAL ATTACKS ON CONTROL SYSTEMS." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.

Seddik M. Djouadi, Major Professor

We have read this thesis and recommend its acceptance:

Husheng Li, Kevin L. Tomsovic

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

ANALYSIS OF MULTIPLE SIGNAL ATTACKS ON CONTROL SYSTEMS

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Nedas Jakstas
August 2018

ACKNOWLEDGEMENTS

I would like to thank my professor and graduate advisor Dr. Seddik Djouadi for his guidance and advice throughout my thesis work. I extend my gratitude to Dr. Kevin Tomsovic and Dr. Husheng Li for willing to serve as my committee members. I would also like to thank Dr. Ricardo S. Sánchez-Peña for answering some questions I had about his MIMO Smith Predictor paper [2].

ABSTRACT

This thesis studies the effects of different signal injection attacks against a time-delayed networked cyber-physical system (CPS). A CPS is an industrial control system which integrates computer networks and physical processes. CPSs are used in critical areas such as transportation and manufacturing. A networked control system is one that allows the controller and plant to be geographically separated by sending the control and measurement signals over a communication network. The convenience of controlling a plant remotely comes at the cost of increased security risk. An adversary who gains access to the network may intercept the signals and corrupt them or simply prevent the transmission of the signals, which may cause considerable damage to the system. The four types of attacks simulated are i) covert misappropriation attack, ii) replay attack, iii) undetectable attack, and iv) worst-case signal attacks. In all of these cases, the attacker is assumed to have access to the communication network used to send the actuation and measurement signals. All of the attacks are implemented successfully. The covert misappropriation attack resulted in over percent error in the nominal output signal while remaining undetected. The replay attack resulted well above one-hundred percent error and is likely to cause considerable damage to the system. The undetectable actuator attack forced the controller to expend more energy than necessary for a brief period to achieve the nominal output. The worst-case attack caused the controller to expend significantly more energy during the entire simulation in order to achieve the nominal output.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Cyber-physical systems (CPS) connect physical processes through a network of connected elements, such as actuators, sensors, and controllers. They are used in many fields including transportation [10], medical devices [11], and industrial process control [12]. One example of a CPS is a natural gas pipeline network. Natural gas is collected from a well, sent through a gathering system where it is processed, and then sent through the transmission system [24]. The transmission system is composed of 272,000 miles of steel pipe and moves the natural gas thousands of miles to local distribution companies. The pressure inside the transmission system pipeline is controlled by compressor stations, located every 50 to 60 miles along the pipeline. Most of the compressor stations are completely automated, meaning the pressure inside the pipeline is regulated from a remote control station. The control station monitors and records operational data from each compressor station. The gas reaches a local distribution company, which also has a control center which monitors and controls the flow rate and pressure via various sensors and computer programs as the gas is sent to the customer [24]. Normal operation of these systems is often required to prevent considerable damage, whether it be to machines or employees. When control signals and sensor readings are able to be sent over a wireless communication network, this allows for controllers to be located off-site and used to control the plant remotely. The downside to this is that the network becomes vulnerable to attacks. Adversaries may gain access to the network, intercept the signals being sent, and corrupt them.

Therefore, these wireless control system networks must be secure to ensure safe operation. The type of attack described previously is called a “man-in-the-middle” attack and has been studied greatly on networked control systems [3],[4],[5],[6]. In [27], researchers successfully performed a replay attack on an implantable cardioverter defibrillator (ICD) introduced into the US market in 2003. This replay attack compromised the device’s integrity by changing stored information or therapy settings [27]. These ICDs are designed to administer an electrical shock to the patient when an irregular heartbeat is detected in order to restore a normal heartbeat rhythm. The device can then report that a shock was administered via wireless capabilities to a healthcare provider, who is able to modify the device’s settings without surgery [27]. Between 1990 and 2002, more than 2.6 million pacemakers and ICDs were implanted in patients in the United States [28]. The work in [27] showed that attacks on the privacy, integrity, and availability of the ICD data are possible. A reprogramming attack is demonstrated which changes the way the ICD operates, allowing a malicious entity to issue a commanded electrical shock. Also, the researchers demonstrated that the ICD disclosed sensitive information without encryption [27]. The demonstration of these security flaws in a device already implanted in patients is enough to cause concern for safety. Cyber-physical system attacks have successfully been carried out and are documented in [13],[14],[15],[16]. In [16], an attack is documented wherein an attacker used a laptop computer and a radio transmitter to take control of 150 sewage pumping stations and released one million liters of untreated sewage into a stormwater drain, where it flowed into local waterways. The reason for the attack is because Mr. Boden, the attacker, was

angry that he was not offered employment by the Maroochy Shire Council [16]. One of the most famous cyber-security attacks is the Stuxnet worm [17], which affected a uranium enrichment facility in Iran. The worm gained access to the supervisor control and data acquisition system (SCADA) used to operate the centrifuges and sent malicious control signals to cause them to malfunction with considerable damage to the system. The designers of the SCADA systems originally did not consider security concerns such as integrity checking, anti-replay checking, authentication, or anti-repudiation due to the assumption that SCADA systems would be isolated from other networks. SCADA communication protocols such as Modbus, Distributed Network Protocol 3 (DNP3), and Allen-Bradley Ethernet/Internet Protocol (IP) do not provide mechanisms to check for integrity or freshness of data received [25],[26]. Therefore, systems that employ these communication protocols are susceptible to denial-of-service attacks (DoS), man-in-the-middle attacks, and replay attacks [26]. The accomplishments of the previously referenced attacks and security flaws in the listed SCADA communication protocols raise concern for the security of critical cyber-physical systems operating today.

The purpose of this thesis is to analyze four different signal attacks on a networked control system with time delays to show how such attacks may affect the system. Chapter two introduces the nominal control system, describes the canal testbed to be used as a model for all attacks, and the controller used for the canal system. The chapter shows the system performance under normal operation. Chapter three shows the effect of a covert misappropriation attack, where control signals are summed with a

malicious signal before reaching the plant, and the effect of the malicious signal on the plant is removed before the output of the plant reaches the controller, thus remaining undetected. Chapter four presents the effect of a replay attack, in which sensor signals are monitored and recorded for a period of time and relayed to the controller at a future time while suppressing real-time sensor readings, thus rendering the attack undetectable. Chapter five shows the effect of an attack, which is undetectable by a dynamic monitor. A dynamic monitor is an algorithm which constantly checks if the system is under attack [8]. Chapter six shows the effect of a worst-case signal attack, in which the objective of the attack is not to remain undetected, but to maximize the cost function and perform the most damage to the system. Chapter seven contains concluding remarks.

CHAPTER TWO

INTRODUCTION OF THE NOMINAL SYSTEM

2.1 Introduction

The nominal system is described and illustrated in [1] and shown in Figure 1. The variables in Figure 1 represent the following parameters:

u_c – Output signal of the nominal controller – Dimension (2x1)

u – Input signal to the plant – Dimension (2x1)

w – Actuator disturbances – Dimension (2x1)

y – Output of the plant – Dimension (2x1)

n – Sensor noise – Dimension (2x1)

y_m – Output sensor data received by the controller (2x1)

y_{ref} – Output refence signal – Dimension (2x1)

In Figure 1, P represents the plant and C represents the nominal controller for the plant. Both the actuation and measurement signals are transmitted between the plant and the controller over a communication network uninterrupted and unmodified [1]. Communication delays and packet losses are not considered for simplicity. In the nominal case, $u = u_c$ and $y_m = y + n$.The plant is assumed to be linear time-invariant (LTI) and is driven by disturbances *w*, and a control signal $u = u_c$ such that

$$y = P_u u_c + P_w w.$$

Here, P_u and P_w are LTI operators that map signals u_c and w to the output y [1]. In other words, they are operators that represent the plant's effect on signals u_c and w. The output measured by the nominal controller is corrupted with noise such that

$$y_m = y + n.$$

The nominal controller outputs the control signal such that

$$u_c = C_y y_m + C_{ref} y_{ref}.$$

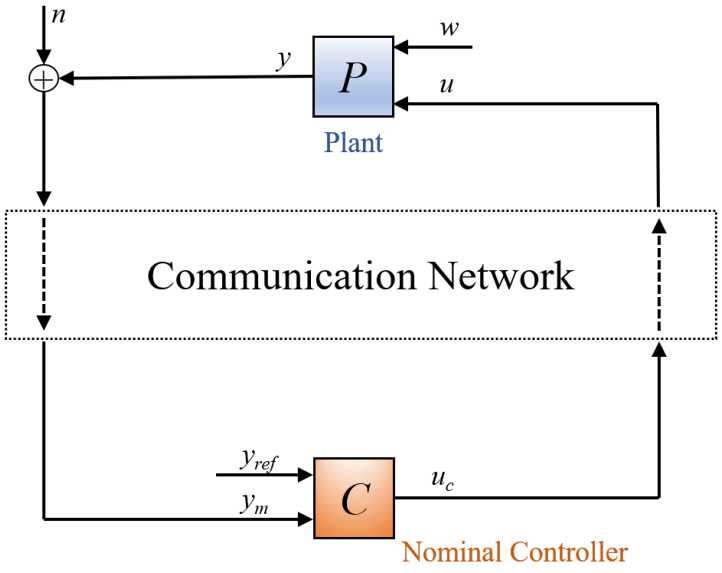

Figure 1. Nominal Control System Diagram [1].

2.2 MIMO Open Flow Canal

An open flow canal system described originally in [2] is used as a testbed to carry out the attacks. The system is composed of two pools with two sluice gates and a downstream spillway, illustrated in Figure 2 and Figure 3. A servomotor is used in each gate to drive the control gate positions u_1 and u_2. There are two level sensors located downstream of the first and second gates (h_1 and h_2, respectively). The reservoir is kept at a constant level of 3.5 m. The length of the first pool is 2 km and the length of the second pool is 4 km.

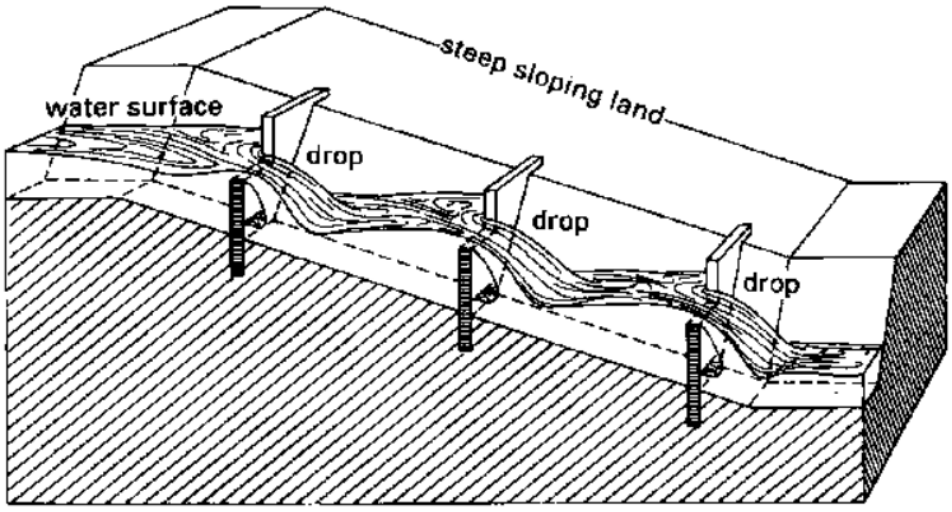

Figure 2. Open Flow Canal Illustration [33].

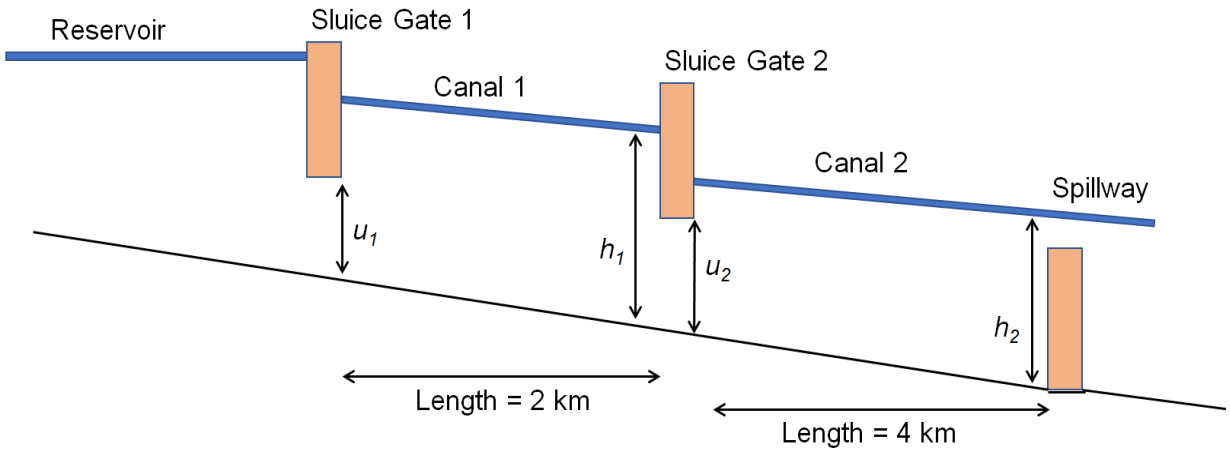

Figure 3. Simplified Open Flow Canal Diagram [2].

This system can be modeled by two Saint-Venant equations. The simplified model is derived in [2] and shown below. The system takes the form

$$y = \begin{bmatrix} h_1(s) \\ h_2(s) \end{bmatrix} = G(s) \begin{bmatrix} u_1(s) \\ u_2(s) \end{bmatrix}$$

where

$$G(s) = \begin{bmatrix} \frac{k_{11}}{T_{11}s+1} e^{-\tau_{11}s} & \frac{k_{12}}{T_{12}s+1} e^{-\tau_{12}s} \\ \frac{k_{21}}{T_{21}s+1} e^{-\tau_{21}s} & \frac{k_{22}}{T_{22}s+1} e^{-\tau_{22}s} \end{bmatrix}.$$

Each delay τ_{ij} is associated with the travelling time of water between each input and output [2]. By defining t_1 to be the travelling time of water to cover the first pool and t_2 to be the travelling time of water to cover the second pool as in Lemma 2.2 in [2], the following relation can be shown:

$$\tau_{11} = t_1, \qquad \tau_{12} = 0, \qquad \tau_{21} = t_1 + t_2, \qquad \tau_{22} = t_2.$$

By using this relation and separating the rational and irrational part of the plant, the following final plant is derived [2]:

$$y = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & e^{-t_2 s} \end{bmatrix} \begin{bmatrix} \frac{k_{11}}{T_{11}s+1} & \frac{k_{12}}{T_{12}s+1} \\ \frac{k_{21}}{T_{21}s+1} & \frac{k_{22}}{T_{22}s+1} \end{bmatrix} \begin{bmatrix} e^{-t_1 s} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Where $k_{11} = 4.87$, $k_{12} = -4.35$, $k_{21} = 1.2$, $k_{22} = 1.4$, $T_{11} = 1800$ seconds, $T_{12} = 2100$ seconds, $T_{21} = 1900$ seconds, $T_{22} = 1500$ seconds, $t_1 = 420$ seconds and $t_2 = 900$ seconds.

2.3 Uncontrolled System Response

The uncontrolled system response to a reference step change in h_1 from 0.64 meters to 0.94 meters at 200 minutes and h_2 held at a constant 0.96 meters is shown in Figure 4.

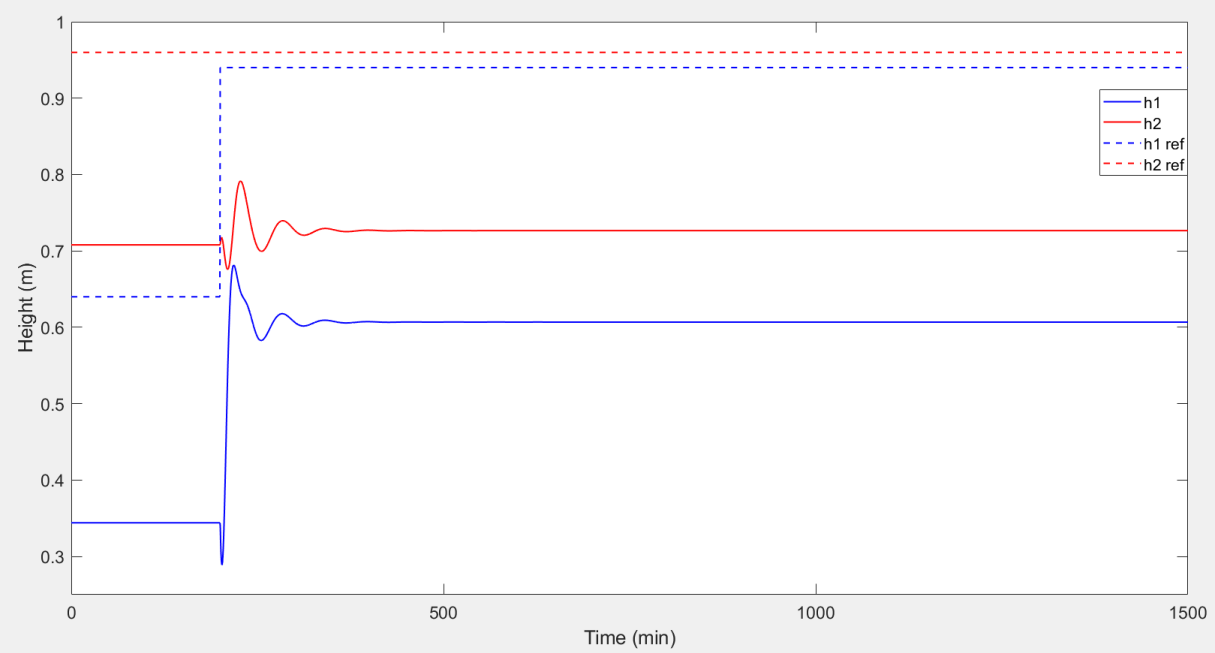

Figure 4. Uncontrolled Canal System Response.

From observation, it is clear that the outputs h_1 and h_2 do not follow the setpoints. Instead of h_1 rising from 0.64 meters to 0.94 meters, it only rises from 0.35 meters to 0.61 meters. The setpoint of h_2 is held at 0.96 meters, but the output only settles at 0.71 meters. Also, the coupling of the two outputs is clear; the reference step change in h_1 produces a large disturbance in the output h_2, which deviates from its setpoint 0.08 meters at the peak of the disturbance. Another problem is the large overshoot evident in the system response. For these reasons, a controller must be used to compensate for the tracking error between the setpoint and the output, to decouple the two outputs, and to reduce the overshoot present in the system response.

2.4 Controller Description

The controller for this system is originally derived in [2]. The rational part of the plant model is:

$$G_m(s) = \begin{bmatrix} \frac{4.87}{1800s+1} & \frac{-4.35}{2100s+1} \\ \frac{1.20}{1900s+1} & \frac{1.40}{1500s+1} \end{bmatrix}$$

To guarantee nominal internal stability, any stabilizing controller, its sensitivity, its complement, and the transfer function from the reference to the control signal are respectively:

$$K(s) = Q(s)[I - G_m(s)Q(s)]^{-1} \quad (1)$$

$$S_m(s) = I - G_m(s)Q(s)$$

$$T_m(s) = G_m(s)Q(s)$$

$$T_{ur}(s) = K(s)S_m(s) = Q(s)$$

for any stable and proper matrix *Q(s)* [2]. To achieve complete diagonal sensitivities and achieve robust stability, *Q(s)*, $T_m(s)$, and $S_m(s)$ are chosen to be:

$$Q(s) = [W_\delta(s)G_m(s)]^{-1}\begin{bmatrix} q(s) & 0 \\ 0 & q(s) \end{bmatrix} \quad (2)$$

$$T_m(s) = G_m(s)Q(s) = [W_\delta(s)]^{-1}\begin{bmatrix} q(s) & 0 \\ 0 & q(s) \end{bmatrix}$$

$$S_m(s) = I - T_m(s)$$

where $W_\delta(s)$ is a weight used to counteract the delay uncertainties and is taken to be

$$W_\delta(s) = \frac{k_\delta s}{s+p_\delta} I_{2x2}$$

where I_{2x2} is the identity matrix, $k_\delta = 4.4$, and $p_\delta = 2.4 * 10^{-3}$ [2]. Next, a low-order performance weight is selected to reject constant and low frequency disturbances as follows:

$$W_e(s) = \frac{s+z_e}{2s} I_{2x2}$$

where $z_e = 0.4 * 10^{-3}$ [2]. To achieve robustness, $q(s)$ is chosen as:

$$q(s) = \frac{q_0 s}{(s+p_q)(\tau s+1)} \quad (3)$$

where $q_0 = 0.52$, $\tau = 10$, and $p_q = q_0 * p_\delta / k_\delta$ in order to cancel the pole in $s = 0$ of $W_e(s)$. Using equations (1), (2), and (3), the following proper controller is derived:

$$K(s) = [W_\delta(s) G_m(s)]^{-1} \begin{bmatrix} q(s) & 0 \\ 0 & q(s) \end{bmatrix} [I - G_m(s) Q(s)]^{-1}$$

which simplifies to [2]

$$K(s) = G_m^{-1}(s)\frac{q_0(s+p_\delta)}{s[sk_\delta\tau+k_\delta(\tau p_q+1)-q_0]}$$

Finally, $K(s)$ is replaced as a function of $Q(s)$ using equation (1) which is easier to implement, resulting in the final controller used for this system.

2.5 Nominal System Performance

The nominal system response is simulated in Figure 5. A noise signal is not applied in the simulations in this thesis in order to make comparisons clearer. The setpoint of the output h_1 is stepped up from 0.64 meters to 0.94 meters at 200 minutes, while the setpoint of h_2 is kept at a constant 0.96 meters. From 800 to 1200 minutes, a disturbance pulse signal is produced equivalent to raising the sluice gate u_1 by 0.01 meters is applied. Figure 5 5 shows the nominal canal system water levels h_1 and h_2 response to a reference point change in gate u_1. From Figure 5, it can be observed that the tracking error of h_1 goes to zero due to the integral action of the controller. Also, the controller successfully rejects the applied disturbance due to the selection of parameter z_e in the performance weight W_e. The decoupling characteristic of the controller is apparent; the amplitude of output h_1 is raised 0.3 meters, but the output h_2 only moves 0.01 meters from its setpoint during the step. In the next chapter, the covert misappropriation attack is introduced and performed on the system.

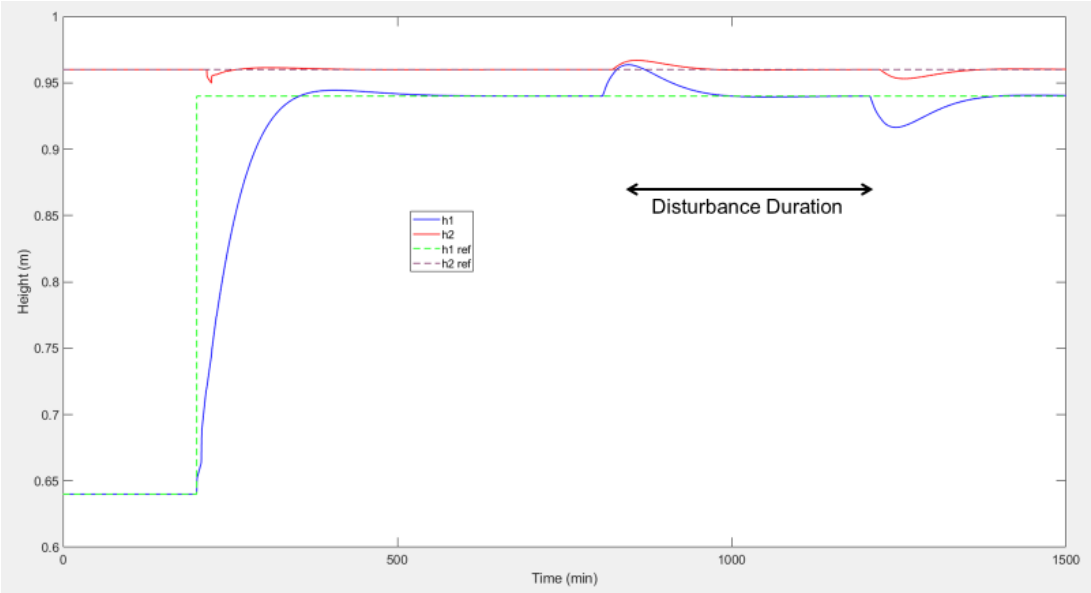

Figure 5. Nominal Canal System Response.

CHAPTER THREE

COVERT MISAPPROPRIATION ATTACK

3.1 Introduction

The covert misappropriation attack is a type of man-in-the-middle attack and is described in [1]. The covert agent is assumed to have infiltrated the communication network in a way that allows the control signal u_c and sensor measurement signal y_m to be read and modified. This attack allows the adversary to intercept the actuation signal, add a malicious component to it designed to force the plant to output a signal other than the setpoint specified by the nominal controller, and then remove the effect of the malicious signal from the plant's output before it is fed back to the controller. This attack allows the attacker to send the plant's output to a desired value without the nominal controller's knowledge. The system under a covert misappropriation attack is illustrated in Figure 6 6 [1]. The attacker calculates the nominal plant's response γ to its injected signal μ and subtracts it from the measured plant's output. The two components of the covert agent structure are the model of the nominal plant Π_u and the covert controller Θ. It can be deduced that if the attacker knows the exact model of the plant, then the injected signal's effect on the nominal plant can be calculated exactly and the attack is completely undetectable [1]. This case has already been studied in [7]. For this example, Π_u will be assumed to contain model errors in order to show that the attack can still be carried out undetected with modeling errors. The covert plant can be represented by

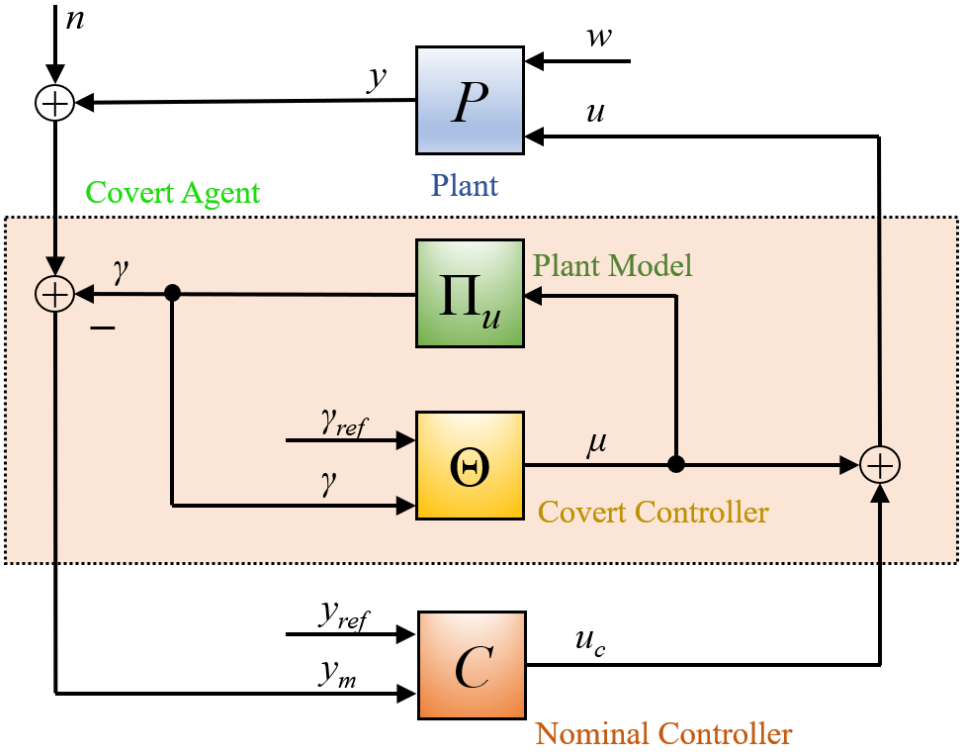

Figure 6. Covert Misappropriation Attack Diagram [1].

$$\Pi_u = P_u + \Delta$$

where Δ represents the additive model error of the covert plant [1]. For this attack, the attacker does not need any knowledge of the nominal controller in order to perform the attack undetected. In this case, the actuation signal that is received by the plant is now

$$u = u_c + \mu.$$

The measured output signal received by the nominal controller is now

$$y_m = y + n - \gamma.$$

The signals γ and μ are calculated in the feedback loop

$$\gamma = \Pi_u \mu$$

$$\mu = \Theta_\gamma \gamma + \Theta_{ref} \gamma_{ref}.$$

3.2 Canal System Under Covert Misappropriation Attack

The same open-flow canal testbed is used to demonstrate the covert misappropriation attack. The covert agent's plant model is given by [1]

$$\Pi_u = \begin{bmatrix} 1 & 0 \\ 0 & e^{-\alpha_2 t_2 s} \end{bmatrix} \begin{bmatrix} \frac{\alpha_2 4.87}{\alpha_2 1800s+1} & \frac{-\alpha_1 4.35}{\alpha_1 2100s+1} \\ \frac{\alpha_2 1.20}{\alpha_1 1900s+1} & \frac{\alpha_1 1.40}{\alpha_2 1500s+1} \end{bmatrix} \begin{bmatrix} e^{-\alpha_1 t_1 s} & 0 \\ 0 & 1 \end{bmatrix}$$

where $\alpha_1 = 2.0$ and $\alpha_2 = 0.5$ and are used to represent the model errors of the covert plant. These model errors contribute to increasing the time delays, gains, and time constants by up to a factor of 2 [1]. The covert controller Θ is calculated using the same approach derived in Section 2.2, with the rational part of the plant taken to be

$$G_{m_\Theta}(s) = \begin{bmatrix} \frac{\alpha_2 4.87}{\alpha_2 1800s+1} & \frac{-\alpha_1 4.35}{\alpha_1 2100s+1} \\ \frac{\alpha_2 1.20}{\alpha_1 1900s+1} & \frac{\alpha_1 1.40}{\alpha_2 1500s+1} \end{bmatrix}.$$

The covert misappropriation attack is simulated with the setpoint γ_{ref} equal to zero before 200 minutes and equal to 0.1 meters after 200 minutes. The results are shown in Figure 7. The effect of the attack is clear. The measured output of h_1 seen by the nominal controller is equal to 0.94 meters, while the actual level of h_1 is 0.84 meters. This results in a 10.64 percent error from the nominal reference setpoint of h_1, and without any mitigation techniques, this attack is undetectable by the controller. In the next chapter, the replay attack is introduced and performed on the system.

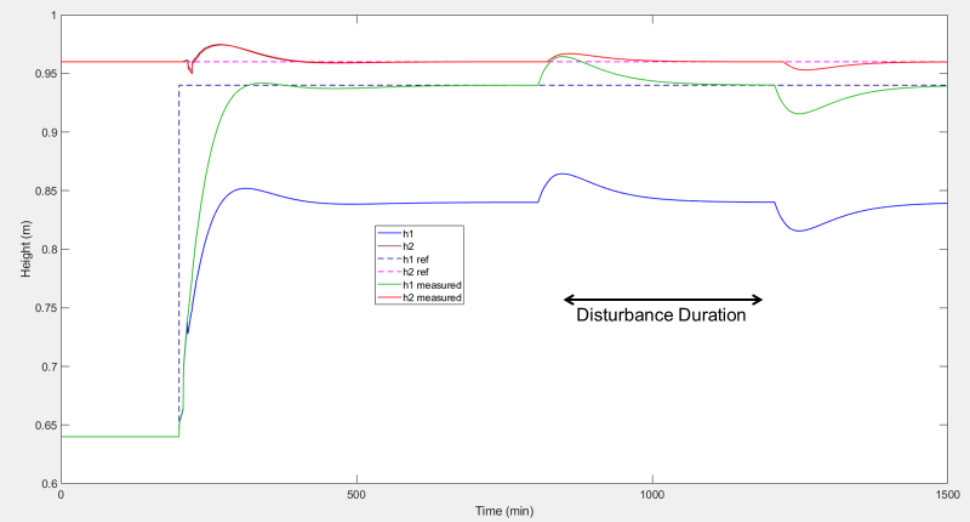

Figure 7. System Response Under Covert Misappropriation Attack.

CHAPTER FOUR

REPLAY ATTACK

4.1 Introduction

In this chapter, the nominal system will be subjected to a replay attack. A replay attack is a type of network attack “in which an attacker detects a data transmission and fraudulently has it delayed or repeated” [21]. The adversary intercepts and records the data transmission and transmits the same data at a later time while suppressing the real-time data. Without proper mitigation, a network subject to a replay attack would interpret the repeated transmission as legitimate [21]. Even if the data is encrypted and the adversary cannot decrypt it, a repeated transmission of legitimate data across a network would still be received as legitimate.

4.2 Mathematical Explanation

Let the nominal control signal be $u_c(t)$, the control signal received by the plant be $u(t)$, and the malicious control signal under a replay attack be $u_a(t)$. Let the nominal output signal of the plant be $y(t)$. Let the nominal output measurement signal be $y_m(t)$ and the output measurement signal under a replay attack be $y_m^*(t)$. Let the sensor noise at time t be $n(t)$. Let the time of system operation begin at t_0 and end at t_f. Let the time length of recording $y_m(t)$ span from t_{r1} to t_{r2} and the time of the replay attack span from t_{a1} to t_{a2}. Therefore, the signals can be represented as:

$$y_m(t) = y(t) + n(t), \qquad t_0 \le t < t_{a1}, \qquad t_{a2} < t \le t_f$$

$$y_m^*(t) = y_m(t), \qquad t_{r1} \le t \le t_{r2}$$

$$y_m(t) = y_m^*(t), \qquad t_{a1} \le t \le t_{a2}$$

$$u(t) = u_c(t), \qquad t_0 \le t < t_{a1}, \qquad t_{a2} < t \le t_f$$

$$u(t) = u_a(t), \qquad t_{a1} \le t \le t_{a2}.$$

The diagram of the system under a replay attack during time $t_{a1} \le t \le t_{a2}$ is shown in Figure 8.

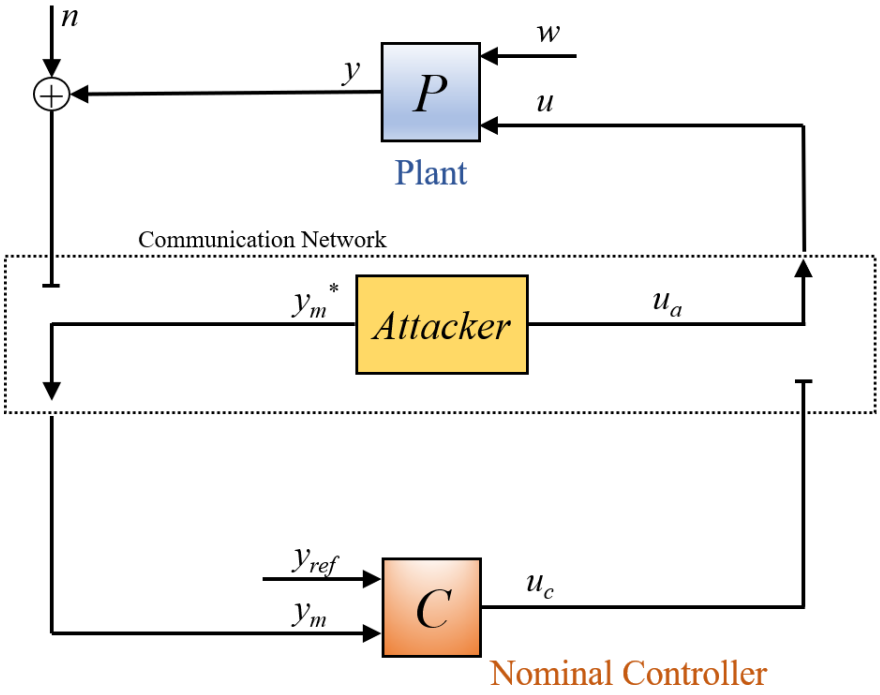

Figure 8. System Diagram Under a Replay Attack.

4.3 Canal System Under Replay Attack

For this section, the measured sensor output signal y_m will be recorded for a period of time, and then relayed to the controller at a future time period while preventing the transmission of y_m during that period. To create the attacked signal for simulation purposes, the nominal system's output is recorded from 33.33 minutes to 366.67 minutes. Then from 1533.3 minutes to 1866.7 minutes, the nominal system's output signal is suppressed and replaced with the previously recorded output signal. The nominal signal and the attacked signal are illustrated in Figure 9.

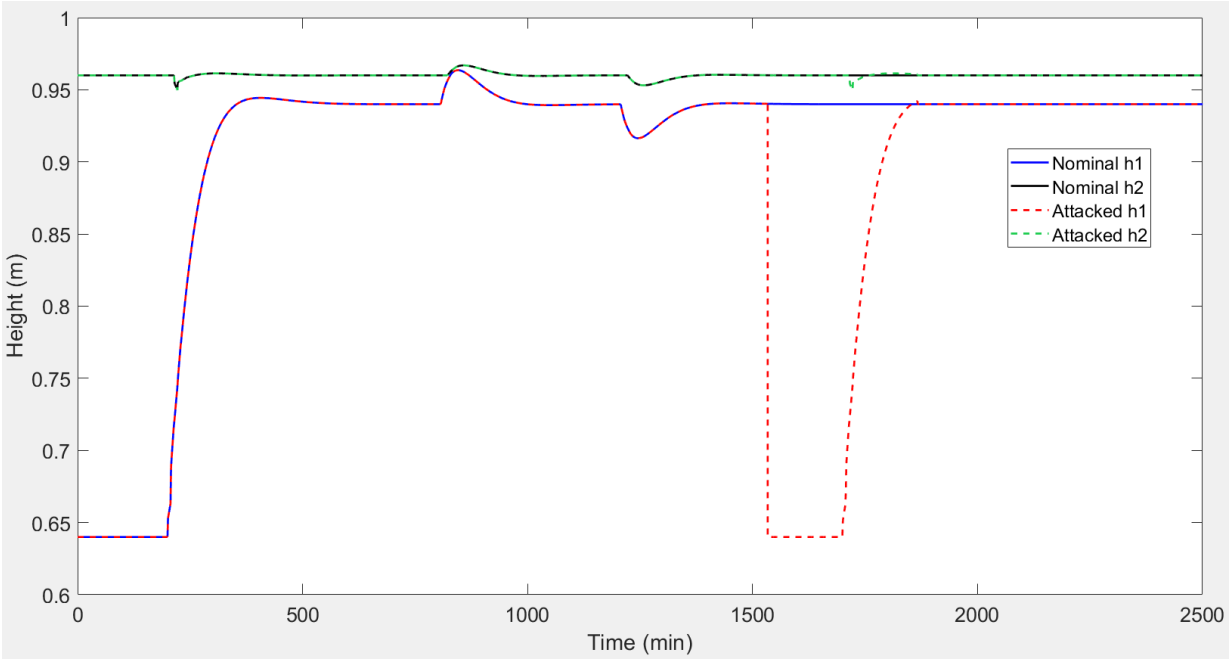

Figure 9. Nominal System Output and Replay Attack Signal.

The canal simulation is performed again, but this time the feedback signal to the controller is disconnected and replaced with the created attacked signals. The system response to the replay attack is shown in Figure 10. The replay attack's effect on the canal system is evident. When the step change from 0.94 meters to 0.64 meters of the attacked signal of h_1 at 1533 minutes occurs, the controller response causes the nominal plant to send the height of h_1 to 2.45 meters. This is due to the reference point of h_1 being set to 0.94 meters and the false h_1 output signal indicating that the plant's output has fallen to 0.64 meters. Therefore, the controller sees this false output and tries to correct it by sending actuator signals to quickly raise the output of h_1 back to 0.94 meters, when in reality, the output was already at 0.94 meters and the controller was fooled into raising the output of h_1 even more. This attack led to a 160.6 percent error in the h_1 response when compared to the nominal output, and without any attack mitigation techniques, this attack would be undetected by the controller and cause considerable damage to the system. The effect of this attack is much greater than that of the covert misappropriation attack, where the percent error in the h_1 response was 10.64 percent. On the other hand, the replay attack is likely to be noticed by employees due to the extreme change in the output of h_1, but the covert misappropriation attack can likely be executed for a longer period of time due to the smaller effect of the attack on h_1. The replay attack could be used to cause damage to the system, where the covert misappropriation attack may be better used in order to deceive the controller for a longer period of time. In the next chapter, the undetectable attack is introduced and performed on the system.

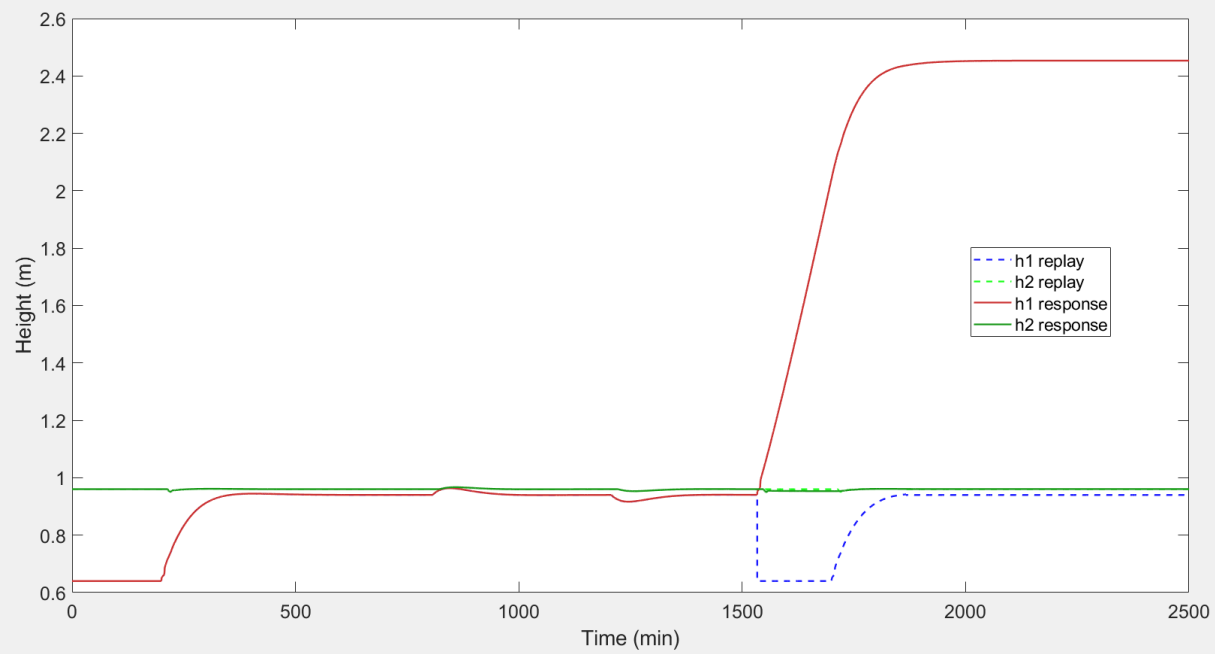

Figure 10. System Response to a Replay Attack.

CHAPTER FIVE

UNDETECTABLE ATTACK

5.1 Introduction

In this chapter, the system will be subjected to an undetectable sensor attack and an undetectable actuator attack. These attacks are described and defined in [9],[32] and solved explicitly in [8]. In this reference, an undetectable attack is defined as one "that is undetectable by a dynamic monitor." A dynamic monitor is "an algorithm which has access to the system dynamics and outputs $(A, B, C, y(t), t \geq 0)$, and checks for the presence of attacks at all times" [9],[32] where A, B, C are the state-space matrices of the system and represent the system dynamics. The reference states that an attack (Δ_u, Δ_y) is undetectable if and only if for some initial state $x_0 \in \mathbb{R}^n$:

$$y\big(0, u, (\Delta_u, \Delta_y), t\big) = y(x_0, u, 0, t), \qquad \forall t \geq 0$$

where $y(x_0, u, (\Delta_u, \Delta_y), t)$ is the output of the system due to the initial state x_0, the control signal u, the under-attack actuator signal Δ_u, and the under-attack sensor signal Δ_y [9],[32]. The undetectable sensor attack signal is derived in [8] and results in the signal:

$$\Delta_y(t) = Ce^{At}x_0, \qquad t \geq 0.$$

Therefore, the sensor attack is always possible if the attacker knows the system's state-space matrices A, C and the initial state x_0. The undetectable actuator attack is also derived in [8] and results in the signal:

$$\Delta_u(\tau) = B^* e^{A^*(t-\tau)} W_c^{-1}(t) e^{At} x_0, \qquad 0 \leq \tau \leq t$$

where

$$W_c(t) := \int_0^t e^{At} B B^* e^{A^*\tau} d\tau$$

is the controllability gramian, and A^*, B^* represent the conjugate transpose of system matrices A and B, respectively.

5.2 Canal System Under Undetectable Attack

The inherent time delays of the canal system are not accounted for in the undetectable sensor and actuator attacks derived in [8]. Therefore, the Padé approximation [18] is used to estimate the delays in the irrational plant to obtain a rational approximation. This estimation was computed using the $pade()$ command in MATLAB [19]. The minimum realization of the resulting system was then obtained using the MATLAB command $minreal()$ [20]. The resulting transfer function matrix of the system is:

$$G_p = \begin{bmatrix} \frac{-0.002706s+1.288*10^{-5}}{s^2+0.005317s+2.646*10^{-6}} & \frac{-0.002071}{s+0.0004762} \\ \frac{0.0006316s^2-4.411*10^{-6}s+6.683*10^{-9}}{s^3+0.00751s^2+1.426*10^{-5}s+5.569*10^{-9}} & \frac{-0.0009333s+2.074*10^{-6}}{s^2+0.002889s+1.481*10^{-6}} \end{bmatrix}$$

The Padé-approximated system's nominal system response and the original system's response are plotted in Figure 11. The approximation resulted in the system containing uncontrollable states. Because the actuator attack relies on the controllability gramian of the system and requires no uncontrollable states, the undetectable attacks were only performed on the part of the system relating input 1 to output 1, which contains no uncontrollable states.

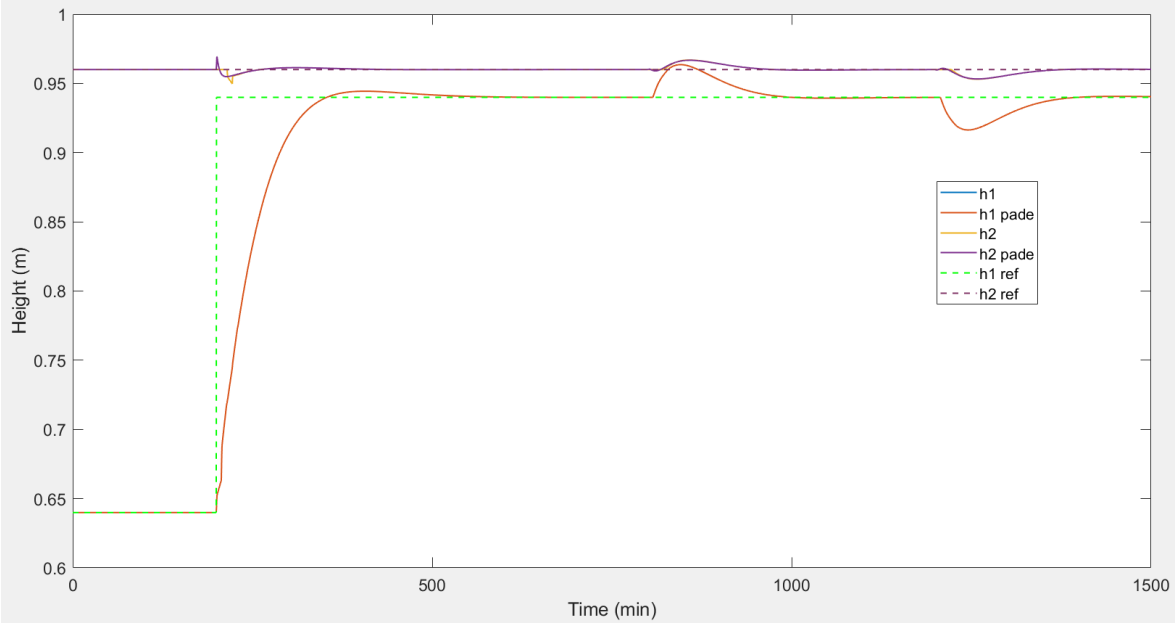

Figure 11. Nominal System Output with Original Time Delays and Nominal System Output with Padé Approximated Time Delays.

This corresponds to the transfer function in row one, column one of G_p, which will be referred to as G_{p11}. The state-space representation of G_{p11} is obtained and the undetectable sensor and actuator attacks are simulated and shown in Figures 12-19. It can be observed from Figure 12Figure 12 that the amplitude of the sensor attack signal Δ_y is relatively small, only reaching a peak of $2.12 * 10^{-4}$ meters. The sensor attack signal also quickly converges to zero. This is due to the system dynamics of the approximated canal system. It can also be observed from Figure 13 and Figure 14 that the effect of the sensor attack on the control input u_1 and the output of G_{p11} is negligible due to the dynamics of the system. Figure 16Figure 16 and Figure 17 show that the actuator attack signal Δ_u has a negligible effect on the canal system due to the system dynamics. Figure 18 and Figure 19 show that even when both attacks are applied simultaneously, their effect is still negligible. In the next chapter, the worst-case bounded sensor signal attack is introduced and applied to the system.

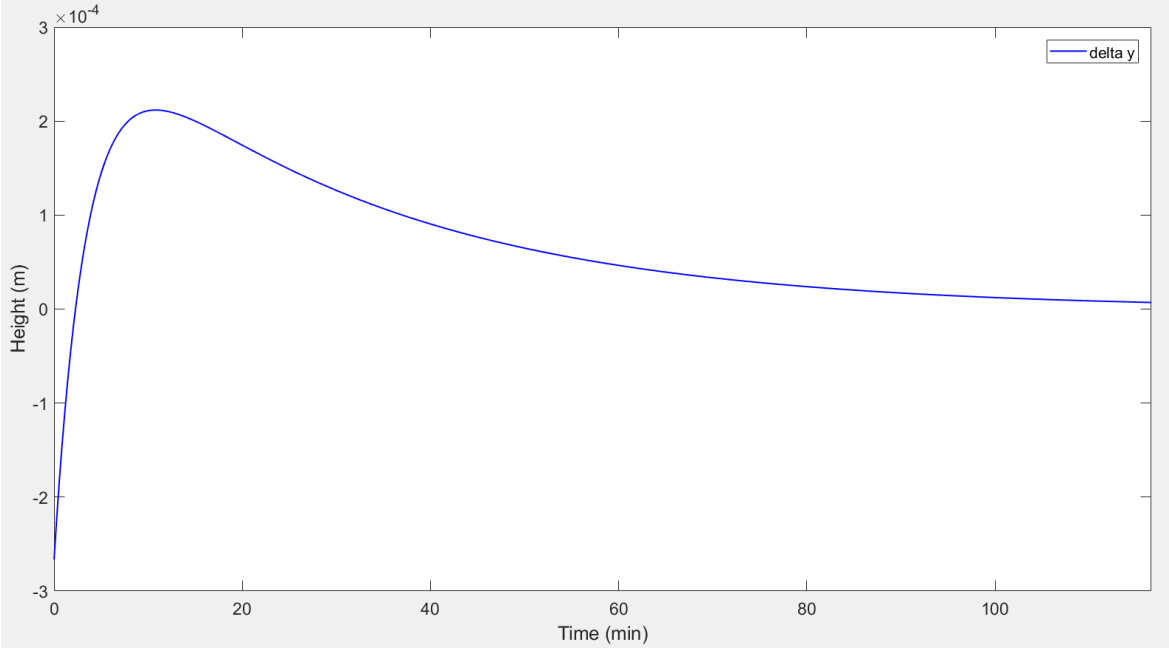

Figure 12. Undetectable Sensor Attack Signal Δ_y. The signal only reaches a height of $2 * 10^{-4}$ meters.

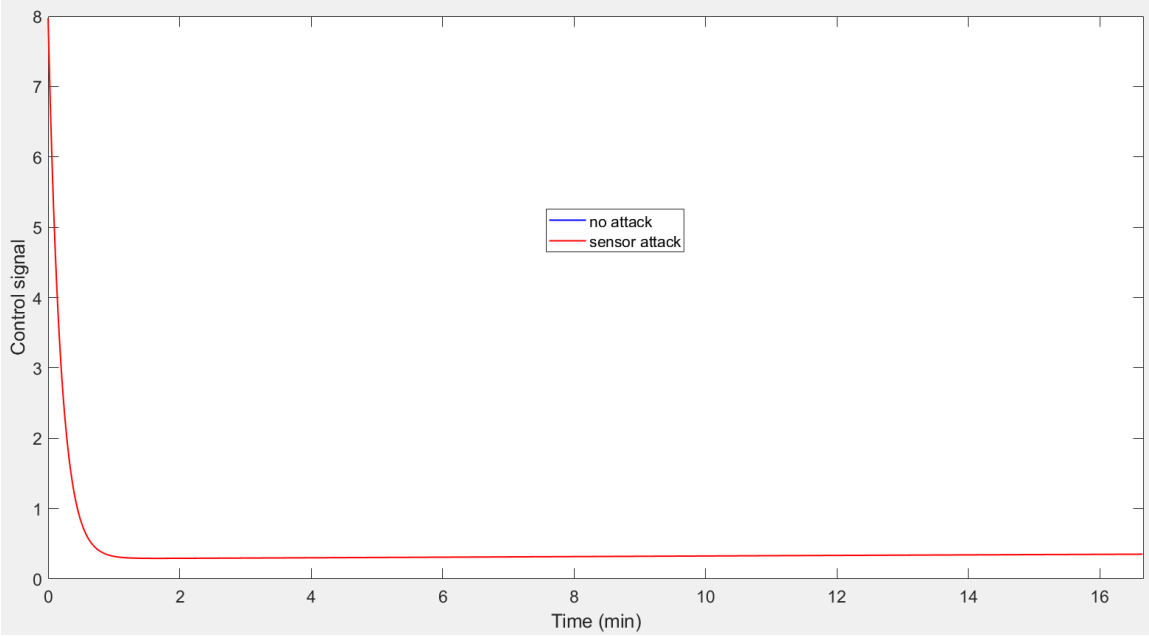

Figure 13. Undetectable Sensor Attack Effect on Control Signal u_1. The effect of the attack is negligible.

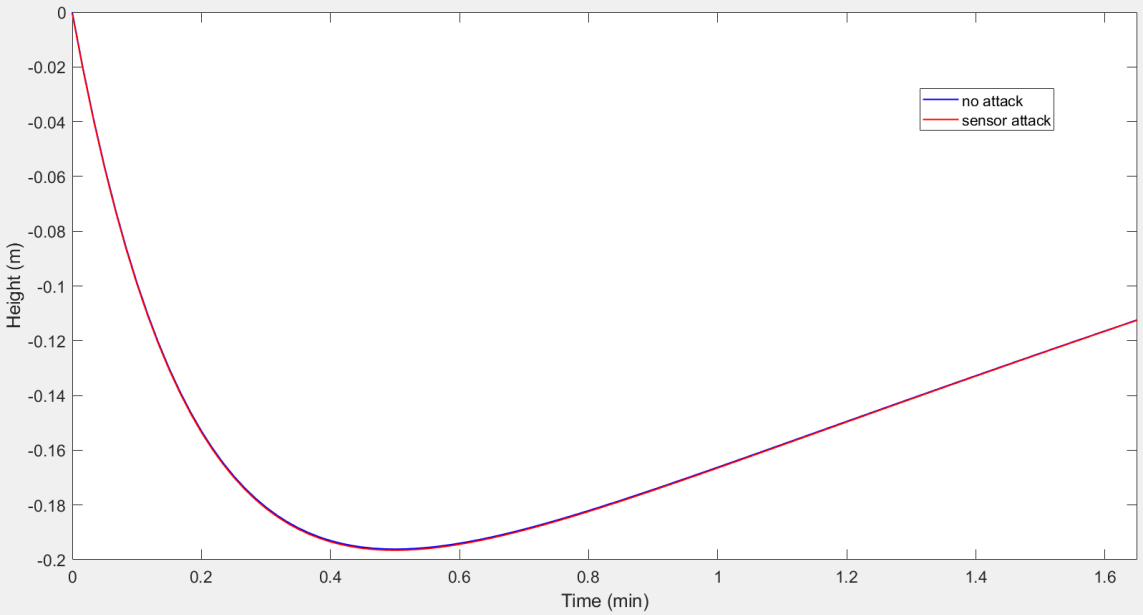

Figure 14. Undetectable Sensor Attack Effect on the Output of G_{p11}. The output stays the same under the attack.

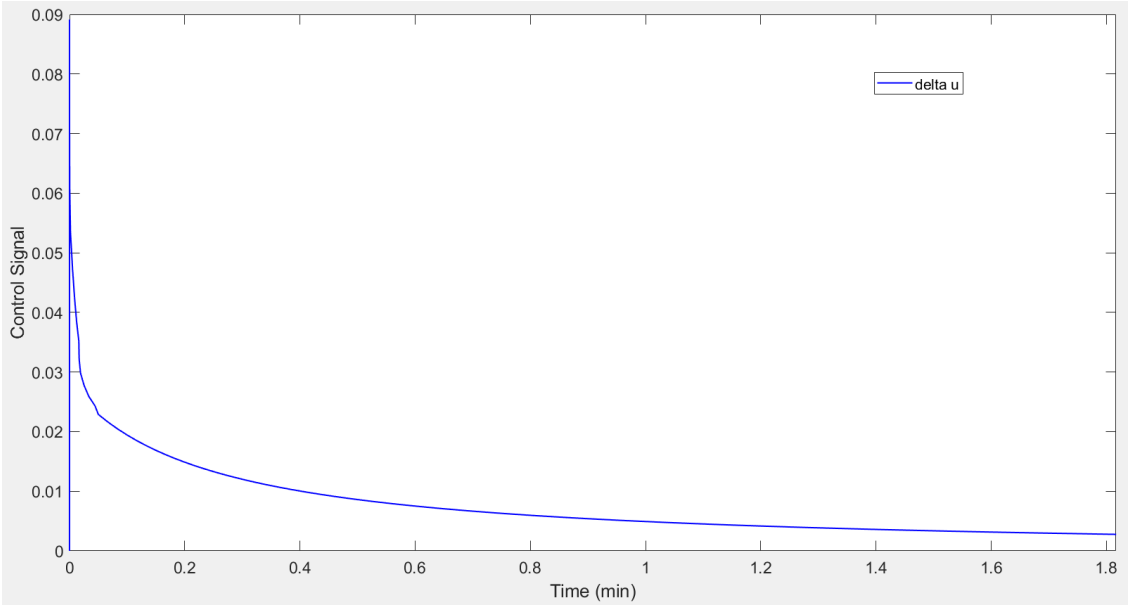

Figure 15. Undetectable Actuator Attack Signal Δ_u. The signal quickly converges to zero due to the system dynamics.

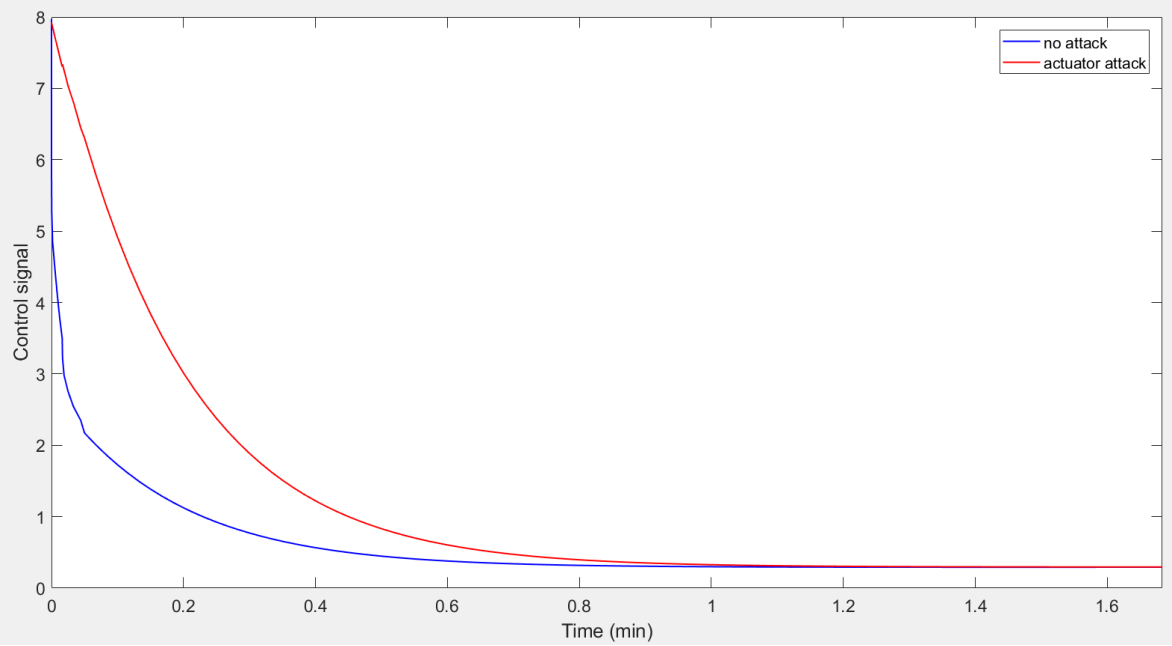

Figure 16. Undetectable Actuator Attack Effect on Control Signal u_1. The controller s forced to exert more energy in the attacked case.

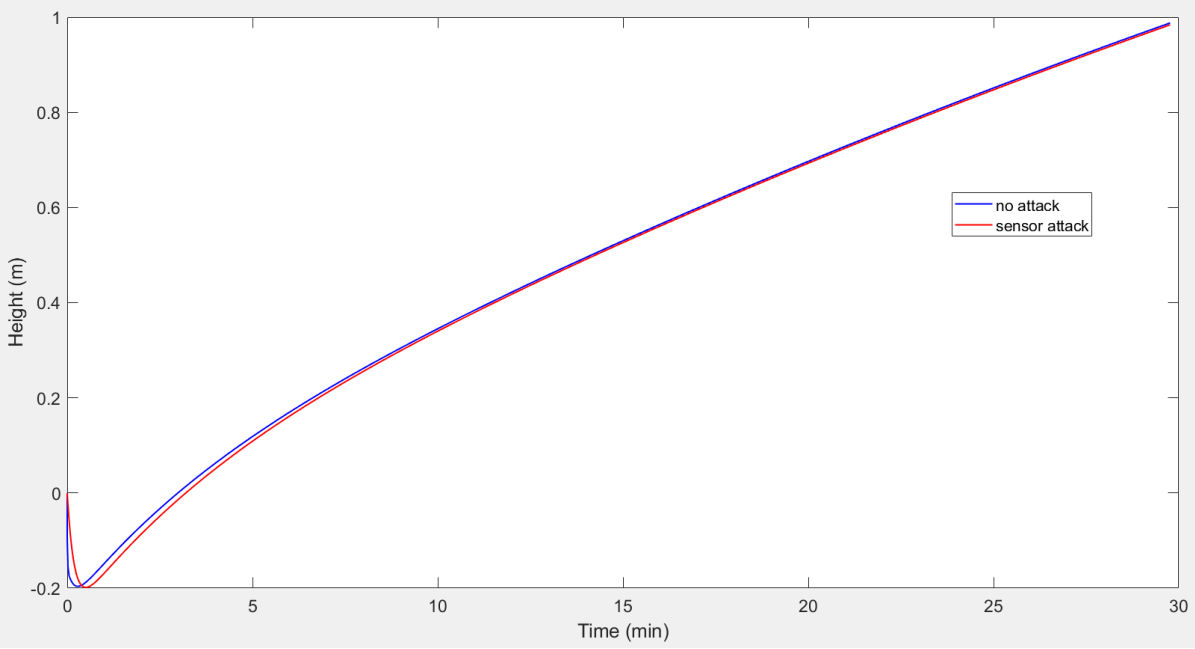

Figure 17. Undetectable Actuator Attack Effect on the Output of G_{p11}. The output stays relatively the same.

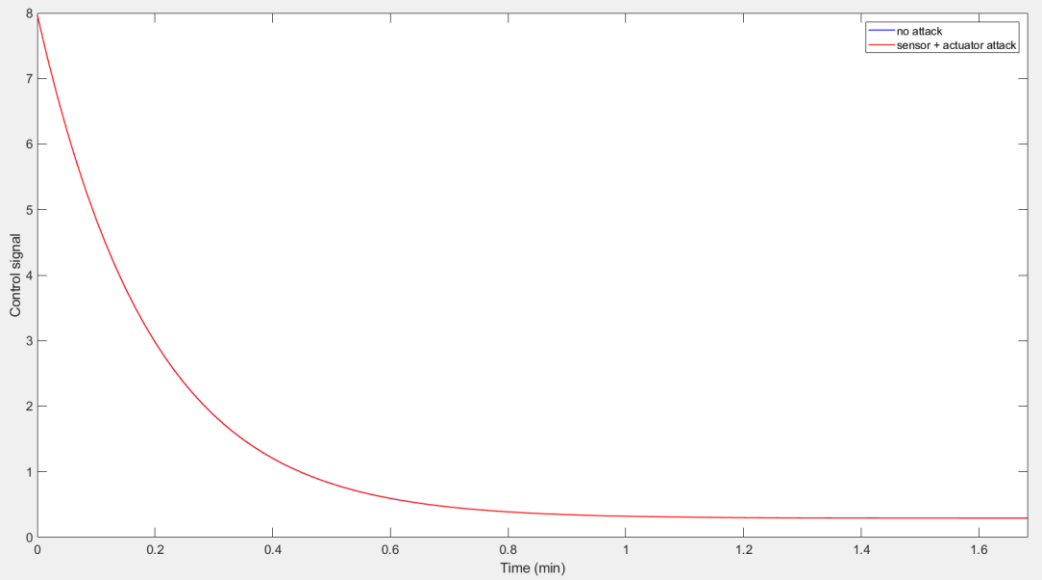

Figure 18. Undetectable Sensor and Actuator Attack Effect on Control Signal u_1. The effect of the attack is negligible.

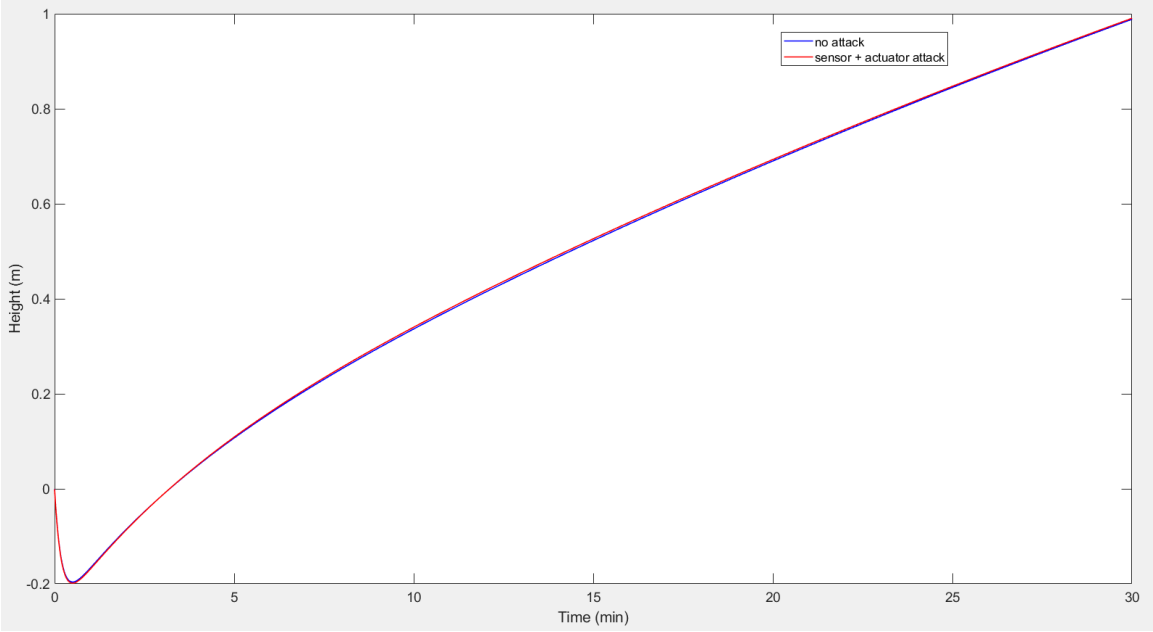

Figure 19. Undetectable Sensor and Actuator Attack Effect on the Output of G_{p11}. The output stays the same.

CHAPTER SIX

WORST-CASE BOUNDED SIGNAL ATTACK

6.1 Introduction

In this Chapter, the canal system will be subject to a worst-case sensor attack. The attack in this Chapter is originally described in [8]. The goal of this attack is not to remain undetected, but to inflict maximal damage by maximizing the control cost function of the system. For this attack, a linear-quadratic controller is used for the system instead of the controller previously derived. The control law minimizes the cost function shown below while the worst-case attack signal maximizes it [22]:

$$\max_{\substack{\|\Delta_x\|_\infty \le M \\ \Delta_x \in L^\infty([0,\infty),\mathbb{R}^n)}} \min_{u \in L^2} \left\{ J(u) = \int_0^\infty \{z^T Q z + u^T R u + 2 z^T N u\}\, dt \right\}.$$

The MATLAB function $lqi()$ is used to compute state-feedback control law [22]. An LQI controller was used instead of the LQR controller due to steady-state errors occurring when using the LQR controller. The integral term in the LQI controller forces the steady-state error to converge to zero. The state-feedback control of the LQI control scheme is of the form $u = -Kz = -K[x; x_i]$. The new state-space representation of the system with the LQI control implemented is derived as:

$$\dot{x} = A\dot{x} + Bu$$
$$\text{define } \tilde{x} = \int x$$
$$\dot{\tilde{x}} = x = Ix + 0u$$
$$\tilde{z} = \begin{bmatrix} \dot{x} \\ \dot{\tilde{x}} \end{bmatrix} = \begin{bmatrix} Ax + Bu \\ Ix + 0u \end{bmatrix}$$
$$\dot{\tilde{z}} = \begin{bmatrix} A & 0 \\ I & 0 \end{bmatrix} \tilde{z} + \begin{bmatrix} B \\ 0 \end{bmatrix} u$$

The control law diagram is shown below in Figure 20 [22].

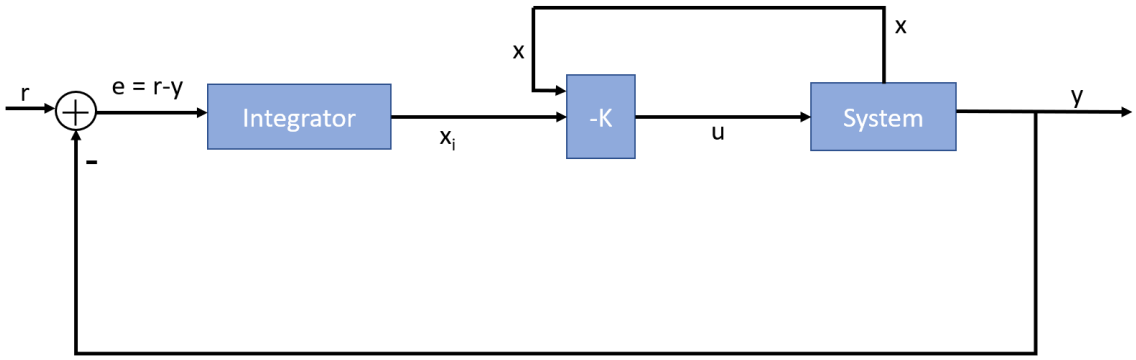

Figure 20. Diagram of the LQI Control Scheme [22].

The function $[K, S, e] = lqi(SYS, Q, R, N)$ calculates the optimal gain matrix K, given a state-space model SYS for the plant and weighting matricies Q, R, N [22]. The Padé-approximated system previously computed is used for SYS. Q, R, N are chosen to be $5 * 10^{-5} * I_{10x10}$, $900 * I_{2x2}$, and 0, respectively.

The performance of the nominal controller is shown below in Figure 21.

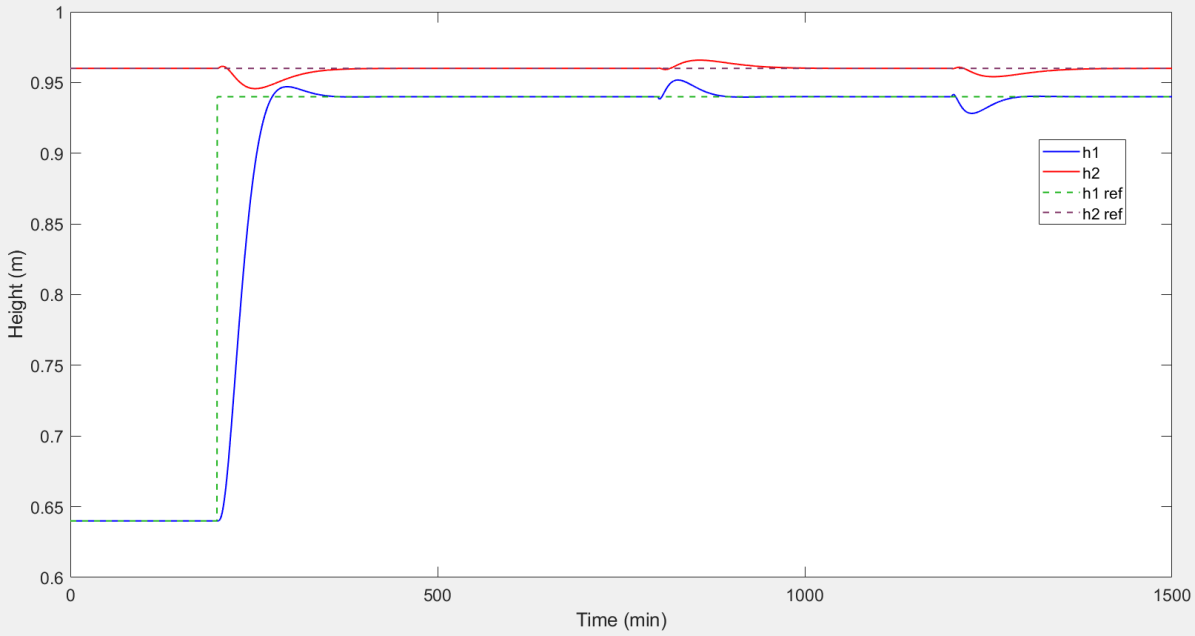

Figure 21. Nominal System Performance with LQI Control.

It should be noted that this type of controller cannot be used on this system in reality, as the state variables of this system measure by the controller are not tied to physical parameters but can be related to physical states using a similarity transformation. The purpose of this Chapter is simply to show the effect of such an attack on a canal system with these properties. The optimal sensor attack for infinite horizon LQI control is derived in [8] and [30], and also used in [29]. The cost function for this problem is derived as

$$J(u, x_0) \coloneqq \lim_{h\to\infty} \left\{ \int_0^h (x^T C^T C x + u^T u) d\tau + x^T(h) Q x(h) \right\}, Q \geq 0$$

where Q is a positive semi-definite matrix, x is the vector of state variables, u is the vector of control variables, and C is the output matrix in the state-space representation of the plant [8],[29],[30]. The minimizing control input is given by $u = -B^T P x$ where P is the solution of the algebraic Riccati equation [23]:

$$PA + A^T P - PBB^T P + C^T C = 0.$$

In the infinite horizon case for bounded sensor attack signals, $\Delta_x \in L^\infty([0,\infty), \mathbb{R}^n)$, meaning $ess\sup_{t\in[0,\infty)} \|\Delta_x(t)\| \leq \delta$, where *ess sup* stands for essential supremum, for some $\delta \geq 0$, and with $\|\Delta\|_\infty \leq M$, the optimal sensor attack signal is

$$\Delta_x(t) = M v_1, \qquad t > 0$$

where $L^\infty([0,\infty), \mathbb{R}^n)$ is the space M is a scalar representing the maximum bound threshold an attacker can expend, and v_1 is the normalized right singular vector corresponding to the maximal singular value of $B^T P$ [8],[29],[30]. The optimal worst-case actuator attack is not used in this thesis, but it is derived in [8],and [31], and used in [29].

6.2 Numerical Results

The nominal system is subjected to this sensor attack with $M = 10$ and the results are shown in the Figures 22 and 23.

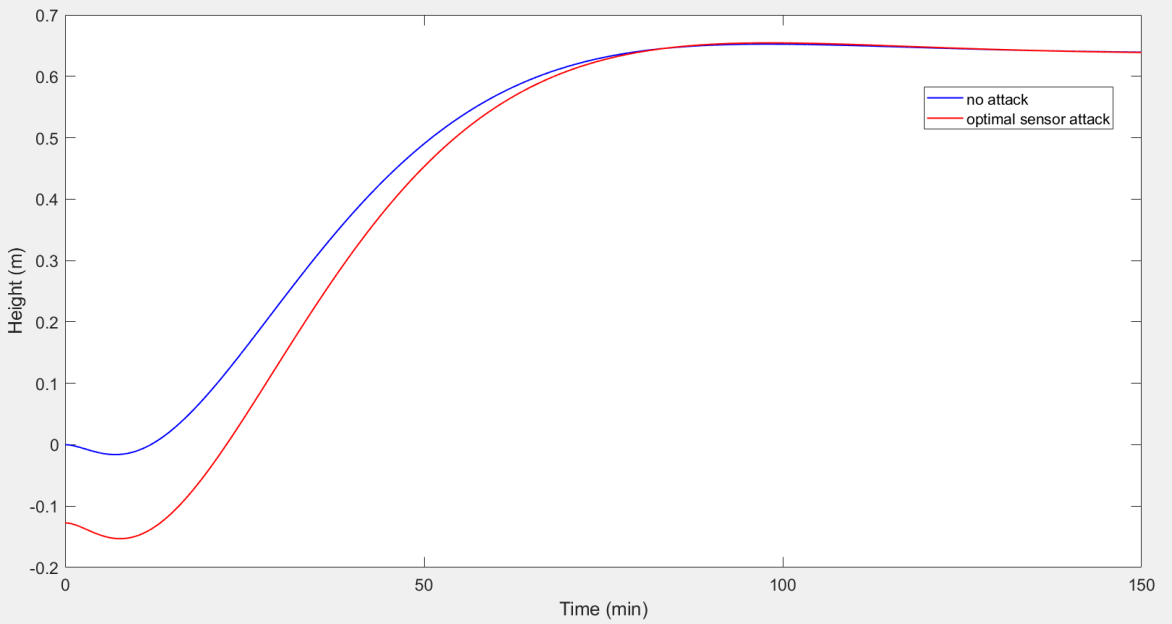

Figure 22. Output h_1 with and without Optimal Sensor Attack.

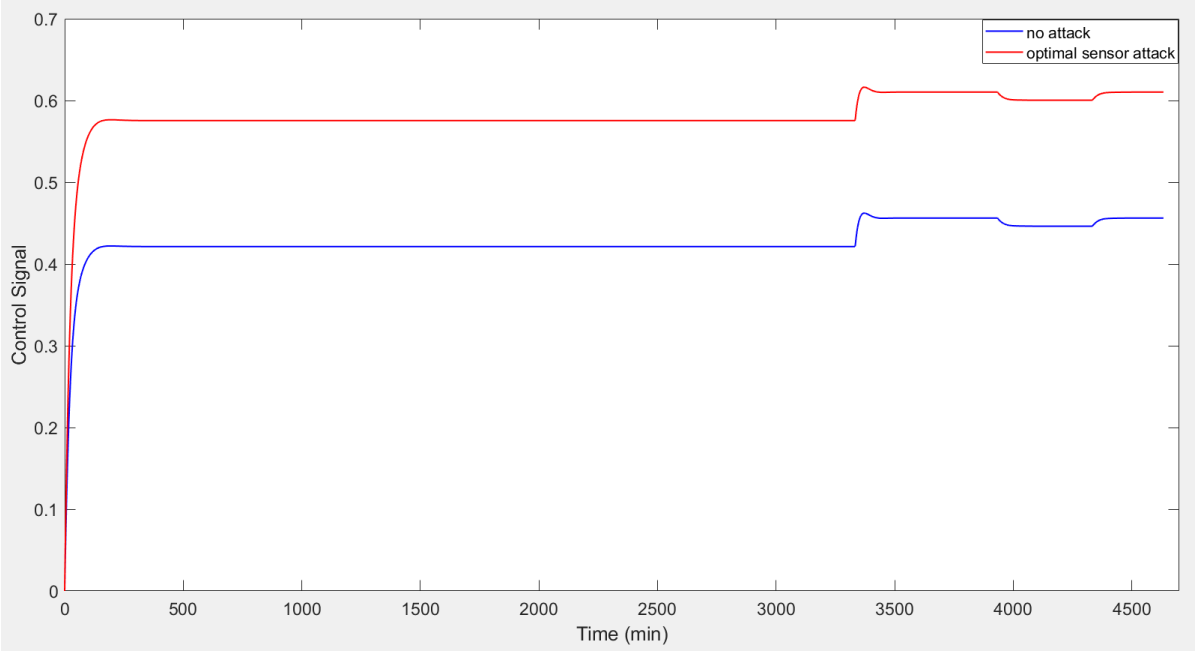

Figure 23. Control Input u_1 with and without Optimal Sensor Attack.

From Figure 22 it can be observed that the controller eventually corrects the output h_1 to converge to the nominal value after 150 minutes. On the other hand, it can be observed from Figure 23 that the control signal energy expended is 0.16 units higher than in the nominal case for the entire duration of the simulation. Therefore, this attack succeeded in maximizing the cost function of the system. Compared to the covert misappropriation attack, the effect of this attack does not affect the output h_1 as much but forces the controller to expend more energy. The replay attack on this system has a higher impact on the output of system due to the 160.6 percent error in h_1. The worst-case bounded sensor attack forces the controller to exert much more energy than the undetectable attack. The next chapter contains concluding remarks.

CHAPTER SEVEN

CONCLUSION

Four types of malicious cyber-physical system signal attacks have been described, and their effects on a time-delayed system have been analyzed. All of the attacks were implemented successfully. The covert misappropriation and replay attacks caused significant error between the nominal output and the under-attack output. The effects of the undetectable sensor attack were negligible due to the system dynamics and only one part of the system being attacked. The effect of the undetectable actuator attack was also negligible but resulted in a larger control signal and therefore expended more energy than necessary. The covert misappropriation attack resulted in a 10.64 percent error in the nominal output signal while remaining undetected. The replay attack resulted in a 160.6 percent error and is likely to cause considerable damage to the system. The undetectable actuator attack forced the controller to expend more energy than necessary for a brief period to achieve the nominal output. The worst-case attack caused the output h_1 to deviate around 0.12 meters from the nominal response, but after 150 minutes, the output signal converged the nominal value. However, the controller was still affected throughout the whole simulation and the attack caused it to excerpt significantly more energy than required in the nominal case. Future work includes performing the undetectable sensor and actuator attacks on a different networked control system, in which all states of the plant are controllable with system dynamics such that the effect of the attack is more pronounced. Another area that could

be studied in the future is developing mitigation strategies and detection algorithms for the covert misappropriation attack, undetectable attacks, and worst-case sensor attack.

REFERENCES

[1] R. S. Smith, "Covert Misappropriation of Networked Control Systems: Presenting a Feedback Structure," in *IEEE Control Systems*, vol. 35, no. 1, pp. 82-92, Feb. 2015. doi: 10.1109/MCS.2014.2364723

[2] Ricardo S. Sánchez-Peña, Yolanda Bolea, Vicenç Puig, MIMO Smith predictor: Global and structured robust performance analysis, Journal of Process Control, Volume 19, Issue 1, 2009, Pages 163-177

[3] Amin S., Cárdenas A.A., Sastry S.S. (2009) Safe and Secure Networked Control Systems under Denial-of-Service Attacks. In: Majumdar R., Tabuada P. (eds) Hybrid Systems: Computation and Control. HSCC 2009. Lecture Notes in Computer Science, vol 5469. Springer, Berlin, Heidelberg

[4] Z.H. Pang, G.P. Liu, Z. Dong, Secure Networked Control Systems under Denial of Service Attacks*, IFAC Proceedings Volumes, Volume 44, Issue 1, 2011, Pages 8908-8913, ISSN 1474-6670, ISBN 9783902661937

[5] A. Teixeira, H. Sandberg and K. H. Johansson, "Networked control systems under cyber attacks with applications to power networks," *Proceedings of the 2010 American Control Conference*, Baltimore, MD, 2010, pp. 3690-3696. doi: 10.1109/ACC.2010.5530638

[6] Y. Yuan, H. Yuan, L. Guo, H. Yang and S. Sun, "Resilient Control of Networked Control System Under DoS Attacks: A Unified Game Approach," in *IEEE Transactions on Industrial Informatics*, vol. 12, no. 5, pp. 1786-1794, Oct. 2016. doi: 10.1109/TII.2016.2542208

[7] R. S. Smith, “A decoupled feedback structure for covertly appropriating networked control systems,” in *Proc. Int. Federation Automatic Control World Congr.*, Aug. 2011, pp. 90–95.

[8] S. M. Djouadi, "Finite Energy and Bounded Cyber-Attacks on Closed-Loop Control Systems," ORNL Report, 2013.

[9] F. Pasqualetti, F. Dorfler, and F. Bullo, Attack Detection and Identification in Cyber-Physical Systems Part I: Models and Fundamental Limitations, arXiv:1202.6144v2 [math.OC] 10 Mar 2012.

[10] J. K.-S. Lau, C.-K. Tham, and T. Luo, “Participatory cyber physical system in public transport application,” in Utility and Cloud Computing (UCC), 2011 Fourth IEEE International Conference on. IEEE, 2011, pp. 355–360.

[11] I. Lee and O. Sokolsky, “Medical cyber physical systems,” in Proceedings of the 47th Design Automation Conference. ACM, 2010, pp. 743–748.

[12] Y. Wang, M. C. Vuran, and S. Goddard, “Cyber-physical systems in industrial process control,” ACM Sigbed Review, vol. 5, no. 1, p. 12, 2008.

[13] T. Greene, “Experts hack power grid in no time.” NetworkWorld. https://www.networkworld.com/article/2277908/lan-wan/experts-hack-power-grid-in-no-time.html. April 2008.

[14] J. Leyden, Teen derails tram after hacking train network. The Register, January 11th, 2008.

[15] A. Greenberg, A. Hackers cut cities’ power. In Forbes, January 2008.

[16] Slay, J., Miller, M.: ‘Lessons learned from the Maroochy water breach’ (Springer, 2007).

[17] N. Falliere, L. O. Murchu, and E. Chien. (2010, Feb.). W32.Stuxnet dossier. Symantec Security Response, Tech. Rep. [Online]. Available: http://www. symantec.com/content/en/us/enterprise/media/security_response/ whitepapers/w32_stuxnet_dossier.pdf

[18] Golub, G. H. and C. F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, 1989, pp. 557-558.

[19] MathWorks ®, (2017). Control System Toolbox ™: pade (R2018a). Retrieved June 26, 2018 from https://www.mathworks.com/help/control/ref/pade.html.

[20] MathWorks ®, (2017). Control System Toolbox ™: minreal (R2018a). Retrieved June 26, 2018 from https://www.mathworks.com/help/control/ref/minreal.html.

[21] “What Is a Replay Attack?” Techopedia.com, www.techopedia.com/definition/21695/replay-attack.

[22] MathWorks ®, (2017). Control System Toolbox ™: lqi (R2018a). Retrieved June 26, 2018 from https://www.mathworks.com/help/control/ref/lqi.html.

[23] M. Green and D.J.N. Limebeer, Linear Robust Control, Prentice Hall, 1995.

[24] “How Does the Natural Gas Delivery System Work?” American Gas Association, www.aga.org/natural-gas/delivery/how-does-the-natural-gas-delivery-system-work-/.

[25] Gao, W., Morris, T., Reaves, B., and Richey, D. On SCADA control system command and response injection and intrusion detection. In Proceedings of the eCrime Researchers Summit (eCrime), IEEE, 2010, 1–9.

[26] Van Long Do, Lionel Fillatre, Igor Nikiforov, Peter Willet. Security of SCADA Systems Against Cyber-Physical Attacks. IEEE Aerospace and Electronic Systems Magazine, Institute of Electrical and Electronics Engineers, 2017. 〈hal-01544580〉

[27] D. Halperin *et al*., "Pacemakers and Implantable Cardiac Defibrillators: Software Radio Attacks and Zero-Power Defenses," *2008 IEEE Symposium on Security and Privacy (sp 2008)*, Oakland, CA, 2008, pp. 129-142. doi: 10.1109/SP.2008.31

[28] W. H. Maisel, M. Moynahan, B. D. Zuckerman, T. P. Gross, O. H. Tovar, D.-B. Tillman, and D. B. Schultz. Pacemaker and ICD generator malfunctions: Analysis of Food and Drug Administration annual reports. Journal of the American Medical Association, 295(16):1901-1906, April 2006.

[29] Drira, Anis, "Characterization of Optimal Cyber Attacks on Control Systems. " PhD diss., University of Tennessee, 2015. http://trace.tennessee.edu/utk_graddiss/3575

[30] S. M. Djouadi, A. M. Melin, E. M. Ferragut, J. A. Laska and J. Dong, "Finite energy and bounded attacks on control system sensor signals," *2014 American Control Conference*, Portland, OR, 2014, pp. 1716-1722. doi: 10.1109/ACC.2014.6859001

[31] S. M. Djouadi, A. M. Melin, E. M. Ferragut, J. A. Laska, J. Dong and A. Drira, "Finite energy and bounded actuator attacks on cyber-physical systems," *2015 European Control Conference (ECC)*, Linz, 2015, pp. 3659-3664. doi: 10.1109/ECC.2015.7331099

[32] F. Pasqualetti, F. Dorfler, and F. Bullo, Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design, IEEE Conf. on Decision and Control and European Control Conference, Orlando, FL, pp. 2195-2201, Dec. 2011.

[33] “CHAPTER 5 - IRRIGATION SYSTEM.” *Fao.org*, FAO of the UN, www.fao.org/docrep/R4082E/r4082e06.htm.

VITA

Nedas Jakstas was born in Memphis, TN in 1995. He began studying at the University of Tennessee, Knoxville in 2013. He graduated with a B.S. in Electrical Engineering in 2017 and continued studying Control Systems as a graduate student.