Vulnerability of Cyber-Physical Systems to Optimal Stealthy Deterministic Attack Strategies

by

Ziyi Cheng

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Control Systems

Department of Electrical and Computer Engineering
University of Alberta

Abstract

Cyber-physical systems (CPSs) are an emerging technology with the potential to be transformational in the field of systems and control. They combine wireless and virtual components with physical infrastructure to create systems that are more adaptable, scalable, and resilient than their traditional counterparts. Unfortunately, the connections in these wireless networks may be vulnerable to attacks from hostile adversaries that seek to impair the system. This is why the security of CPSs has become a popular area of research in the past decade. Development of effective countermeasures requires a solid understanding of potential vulnerabilities, so it is necessary to study how attackers could successfully degrade system performance while remaining undetected.

Two deterministic attack models with different stealthiness conditions are considered. First of all, we study the properties and optimization of strictly stealthy attacks that cannot be detected by output and innovation-based detectors on CPSs. These attacks may target both actuator and sensor communication channels with the goal of impairing system performance. We provide a necessary and sufficient condition for a system to be susceptible to a strictly stealthy attack of any given time length. Furthermore, we analytically derive the optimal attack out of all possible strictly stealthy attacks with a particular length based on an energy constraint and a summation-based quadratic objective function.

Secondly, we examine optimal stealthy attacks that utilize a relaxed stealth-

iness condition. This condition ensures that the attacks are difficult for any innovation-based detectors to perceive. In order to determine the maximum performance degradation that the attacks may cause, a general optimization problem that can be solved numerically is formulated for a finite attack length. For non-divergent systems over an infinite horizon, the optimal constant and alternating attacks are derived analytically for any system configuration. Characteristics of a novel low-dimensional sinusoidal class of attacks are investigated and procedures for optimization are given. Furthermore, a condition is provided for constant and alternating attacks to be superior to most or all sinusoidal attacks. A mechanism to compare deterministic and stochastic attacks is also presented.

Finally, we illustrate the theoretical results using several numerical examples to demonstrate the effectiveness of the designed attacks.

Preface

- Chapter 2 has been published as: Donny Cheng, Jun Shang and Tongwen Chen,"Finite-Horizon Strictly Stealthy Deterministic Attacks on Cyber-Physical Systems," *IEEE Control Systems Letters*, vol. 6, pp. 1640–1645, 2022. The contents of the aforementioned paper were also selected by the American Control Conference 2022 Program Committee for presentation at the conference.

- Chapter 3 has been submitted for publication as: Donny Cheng, Jun Shang, and Tongwen Chen, "Optimal Stealthy Deterministic Attack Strategies on Cyber-Physical Systems", *IEEE Transactions on Control of Network Systems*.

Acknowledgements

I would like to express my deepest gratitude to all the people who have been a part of my postgraduate studies at the Univeristy of Alberta. First and foremost, I would like to thank my supervisor, Dr. Tongwen Chen. His counsel has been essential in finding my way in the world of academia. He is truly an advocate for his students and I can not thank him enough for his consistent support. This endeavour could not have been possible without him.

Furthermore, I am deeply indebted to Dr. Jun Shang, who first introduced me to the field of CPS security over a year ago and has since generously guided me throughout my academic work. I am constantly uplifted by his passion for research, keen technical ability, and warmhearted character. His mentorship has been critical in finding promising directions for investigation, growing my ideas to fruition, and recovering from setbacks gracefully.

Many thanks as well to all the members of the research group with whom I had the pleasure of working: Harikrishna Rao Mohan Rao, Mani Hemanth Dhullipalla, Dr. Junyi Yang, Jing Zhou, Dr. Mohammad Hossein Roohi, Dr. Hao Yu, Dr. Boyuan Zhou, Jinyuan Wei, Kota Akaike, Ziyi Guo, Nicola Tamascelli, Md Habibur Rahaman, Iman Amini, Dr. Shimin Wang, Md Rezwan Parvez, Haniyeh Seyed Alinezhad, and Li Deng. I will never forget their friendship, kindness, and encouragement.

I would like to acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC), Alberta Student Aid, and the University of Alberta Faculty of Graduate Studies and Research (FGSR) for their financial assistance. I would also like to recognize Drs. Mojgan Daneshmand, Zhan Shu, and Vien Van for their help in preparing applications to these agencies.

Finally, I am grateful to all my family members and friends who have fully supported me throughout the years. Mom, dad, and Wanlin, this thesis is dedicated to you.

Contents

1 Introduction **1**

1.1 Research Background 1

1.2 Literature Survey 2

1.2.1 Defensive Countermeasures 2

1.2.2 Types of Attacks on Cyber-Physical Systems 3

1.2.3 False Data Injection Attacks 4

1.2.4 Deterministic False Data Injection Attacks 4

1.3 Thesis Contributions 6

1.4 Thesis Outline 7

2 Finite-Horizon Strictly Stealthy Deterministic Attacks **8**

2.1 Problem Formulation 9

2.1.1 System Model 9

2.1.2 Attack Model 10

2.1.3 System and Attack Classification 12

2.1.4 Optimal Strictly Stealthy Attack Formulation 13

2.2 Main Results 14

2.2.1 Conditions for τ-Step Strict Vulnerability 14

2.2.2 Optimal $(\tau + 1)$-Step Strictly Stealthy Attack 18

2.3 Simulation 21

3 Optimal Stealthy Deterministic Attack Strategies **23**

3.1 Problem Formulation 24

3.1.1 Cyber-Physical System and Attack Model 24

3.1.2 System and Attack Classification 25
3.1.3 Optimal Stealthy Attack Formulation 25
3.2 Finite-Horizon Optimization Problem 26
3.2.1 Construction of the Optimization Problem 26
3.2.2 Receding Horizon Implementation of the Solution . . . 28
3.3 Alternating Constant Steady-State Attacks 29
3.3.1 Optimal Fully Constant and Alternating Attacks . . . 29
3.3.2 Optimality of Constant and Alternating Attacks 32
3.4 Low-Dimensional Sinusoidal Attacks 34
3.4.1 Equivalent Attack Model 34
3.4.2 Two-Dimensional Sinusoidal Attacks 35
3.4.3 Standard Three-Dimensional Sinusoidal Attacks 38
3.4.4 3D Sinusoidal Attacks with $w_1 = 0, \pi$ 41
3.4.5 3D Sinusoidal Attacks with $w_2 = 0$ 42
3.4.6 3D Sinusoidal Attacks with $w_2 = \pi$ 43
3.5 Comparison of Attack Strategies 45
3.6 Comparison with Stochastic Attacks 47
3.7 Simulation 49

4 Conclusions and Future Work **58**
4.1 Conclusions 58
4.2 Future Work 59

Bibliography **63**

List of Tables

3.1 Properties of optimal sinusoidal attacks on system (3.37). . . . 53

List of Figures

2.1 System architecture under attack. 9

2.2 The cumulative objective values at each time step for the optimal attack and a box plot for a set of random attacks. 22

3.1 Step-by-step objective values of optimal fully constant, fully alternating, and numerical attacks on a double integral system. 50

3.2 Evolution of $\|\Delta z_t\|$ under optimal fully constant, fully alternating, and numerical attacks on a double integral system. 51

3.3 Evolution of $\|\Delta e_t\|$ under attacks using the receding horizon and normal implementations. 52

3.4 2D sinusoidal attack objective values of varying frequencies compared with constant and alternating attacks. 53

3.5 Evolution of Δe_t under optimal sinusoidal and numerical attacks, represented by the solid and dashed lines, respectively. . 54

3.6 Objective values of 3D sinusoidal attacks with frequencies w_1 and w_2. 55

3.7 Objective values of 3D sinusoidal attacks at the edge cases. . . 56

3.8 Evolution of objective values for optimal deterministic and stochastic attacks. 57

List of Symbols

$\mathbb{R}$ ($\mathbb{N}$)	Set of Real (Natural) Numbers
$\mathbb{S}^n_{++}$ ($\mathbb{S}^n_{+}$)	Set of Positive (Semi-)Definite Matrices
$X^T, \mathrm{Tr}(X)$	Transpose and Trace of Matrix X
$\mathrm{im}(X), \ker(X)$	Image and Null Space of X
$\mathrm{Re}(X)$	Real Part of X
X^+	Moore–Penrose Pseudoinverse of X
$\rho(X)$	Spectral Radius of X
I (I_n)	(n by n) Identity Matrix
$[X;Y]$	Vertical Concatenation of X and Y: $[X^T, Y^T]^T$
$A \oplus B$	Minowski Sum of Sets A and B
$A - B$	Set Difference Between Sets A and B
$\mathcal{N}(\mu, \Sigma)$	Gaussian Distribution with Mean μ and Covariance Σ
$D_{KL}(P\|Q)$	Kullback–Leibler Divergence from Distributions Q to P

List of Acronyms

CPS	Cyber-Physical Systems
DoS	Denial-of-Service
FDI	False Data Injection
KLD	Kullback–Leibler Divergence
LQG	Linear Quadratic Gaussian
LQR	Linear Quadratic Regulator
QCQP	Quadratically Constrained Quadratic Program

Chapter 1

Introduction

In this chapter, the area of cyber-physical system (CPS) security in research is introduced. A survey of attack designs and countermeasures in recent literature is then provided. Furthermore, the primary contributions of this work are listed and an outline of the remainder of the thesis is given.

1.1 Research Background

In recent years, CPSs have been increasingly adopted for a variety of applications and industries. These systems utilize wireless communication networks to transmit data between components of the system, such as the sensors, state estimator, controller, and actuator. This growth has been driven by advancements in communication technologies that have the capability of creating systems that are less expensive and easier to maintain [35]. There are numerous budding control applications for CPSs, including intelligent transportation networks [40], smart medical devices [13], and more efficient manufacturing [18]. CPSs also have potential in the process control [15] and power [42, 43] industries as well.

However, the interconnections within CPSs also render them susceptible to attacks from malicious agents [35, 36]. In fact, sophisticated attacks have already been developed and launched, including multiple attacks on Ukraine's power grid in 2015 and 2016 [23] and the highly publicized Stuxnet attacks in 2010 [17]. This is why there has been a surge of publications on how to

protect CPSs from adversaries. However, developing suitable countermeasures requires knowledge of potential attack strategies to evaluate the existing risk and design effective mitigating actions. That is why an assortment of attack models has also been studied thoroughly. Many of these attacks feature minimal influence on signals available to the state estimator side of the CPS, such as output and innovation, to remain stealthy from any detectors that are equipped. This allows the attacker to cause more degradation over a longer period of time by delaying the initiation of emergency responses [3].

1.2 Literature Survey

The thesis examines the design of optimal false data injection (FDI) attacks on both actuator and output channels of closed-loop cyber-physical control systems under two distinct stealthiness constraints. This section discusses the classes of attacks and countermeasures that have been examined in recent publications. Then, a detailed literature review of FDI attacks is presented.

1.2.1 Defensive Countermeasures

A variety of countermeasures have been developed to protect closed-loop systems and remote state estimators from adversaries. Encryption of the data sent over communication channels is an effective way to hinder most attacks. For remote state estimation, optimal encryption strategies were developed using a game theoretical approach under different assumptions of attacker knowledge in [30]. Furthermore, a single-dimensional encryption scheme [33] was introduced to secure state estimators and features reduced computational complexity.

Advanced detection strategies have also been developed in order to identify stealthy attacks that may otherwise remain hidden. For example, a summation detector that extends the standard χ^2 detector to include historical innovations was introduced in [41]. Other detection mechanisms include causality-based detectors [34], sequential data verification [19], encode–decode schemes [24]

and watermark design for systems with unknown parameters [20]. Furthermore, resilient control and estimation policies have been proposed so that systems can operate effectively even while an attack is ongoing [38], [7]. Enhanced actuator saturation has also been introduced as a way to limit the potential damage an attack can cause [16]. Unfortunately, implementing some of these countermeasures can result in sub-optimal system performance in the absence of an attack [25].

1.2.2 Types of Attacks on Cyber-Physical Systems

A wide variety of attack formulations have been studied, and they can be grouped into a few general categories. First of all, there are denial-of-service (DoS) attacks that block communication channels to prevent data from being transmitted from one system component to another. For these attacks, a common problem that has been studied is optimal scheduling if the attacker can only block channels a limited number of times. This was investigated in [44] with the goal of maximizing an infinite-horizon linear quadratic Gaussian (LQG) cost function. In [46], DoS attacks against the control channels in a linear quadratic regulator (LQR) system were studied.

During replay attacks, the attacker saves a portion of transmitted data and then overrides the system's communications with this recording at a later time. Replay attacks are relatively simple to design and are effective at evading any innovation-based detectors [8]. Reset attacks are unique because they target a remote state estimator directly instead of the communication channels around it. They typically degrade the system by altering the stored state estimate, such as resetting it to the initial value at each time step [28]. Unlike the others, eavesdropping attacks do not seek to influence the system variables in any way. Rather, they attempt to breach a system's privacy by estimating the system state [39]. This allows adversaries to obtain proprietary information or initiate a more effective future attack.

1.2.3 False Data Injection Attacks

FDI attacks, sometimes also known as integrity or deception attacks, have attracted a lot of attention because they can be designed to severely impair CPS performance while retaining desirable stealthiness properties. In contrast to other attacks, FDI attacks directly corrupt the transmitted data by intercepting and modifying it appropriately. There are two primary models used for designing integrity attacks: the stochastic attack framework and the deterministic attack framework. They differ in how they handle the random noises inherent in the system.

Stochastic attacks address the probabilistic noises of the system directly and incorporate the influence of noise in their design. Stochastic attacks against state estimation have been studied extensively. The landmark paper is [10], in which it was found that the optimal stealthy linear attack that has no effect on the innovation distribution is to simply flip the sign on the transmitted residual. This was later extended in [11] with a relaxed stealthiness condition based on the Kullback–Leibler divergence (KLD) between the compromised and nominal innovation signals at each time step. Guo *et al.* also considered the optimal stealthy attack when the attacker is able to independently measure the system state, which is referred to as side information [12].

For closed loop systems, the optimal stealthy attacks were derived with a general attack form and the goal of maximizing the LQG control cost function in [31]. Bai *et al.* also studied optimal attacks on LQG systems but used a KLD stealthiness metric over an infinite sequence of outputs [1]. Attacks on general fixed-gain feedback closed-loop systems were discussed in [4].

1.2.4 Deterministic False Data Injection Attacks

While stochastic attacks may be more robustly resistant to detectors, they require the attacker to have consistent access to real-time data from the system. The deterministic attack model focuses its attention at the deviation

between system variables in attacked and healthy systems. Since these differences are all deterministic in nature, this allows attacks to be designed and computed offline without any need for real-time data as long as the system model is known. Furthermore, while the stochastic attack model typically assumes that the noises are zero-mean Gaussian and that the system uses fixed-gain feedback control, it can be shown that there is no need for any assumptions on the noise distributions or control scheme for deterministic attacks to be applicable.

The concept of stealthy deterministic attacks was first introduced in [26], which only considered attacks on the measurement channel. These results were later extended in [37] by incorporating attacks on the control channel as well. Necessary and sufficient conditions for an attack on both channels to be able to cause system states to diverge under strict and relaxed stealthiness constraints were derived. When divergence is not possible, a method to find a bound on the difference in the estimation error using the z-transform was proposed. The concepts of strict and relaxed stealthiness from [37] are used in this thesis.

In [45], a deterministic attack that completely eliminates its influence on residuals during steady state was designed to destabilize a CPS. Deterministic attacks against closed-loop control with network delays and state estimation were also studied in [29] and [14], respectively. However, these two papers assumed that the attacker had real-time knowledge of the data in the communication channels. Attempts to create attacks with highly limited information sets have also been made, such as in [9], in which it is assumed that the attacker has only the system state-space matrices available.

Almost all papers on deterministic attacks focus on destabilizing the target system. As far as we are aware, there is nothing in the literature on fully deterministic attacks that seek to maximize system performance degradation when they are unable to affect system stability. In this thesis, we try to fill this gap by studying the optimization of effective stealthy attack strategies in order

to determine the extent to which system performance may be undermined.

1.3 Thesis Contributions

To identify the vulnerability of CPSs to FDI attacks, this thesis studies optimal attacks under different stealthiness constraints. The major contributions are summarized as follows:

1. We extend the concept of strict vulnerability introduced in [37], which focused on the behaviour of the attack as $t \to \infty$, to τ-step strict vulnerability. This is a weaker condition, but strictly stealthy attacks still have the potential to cause substantial damage to τ-step strictly vulnerable systems in a short time span. For a τ-step strictly vulnerable system, we find the set of all possible $(\tau+1)$-step strictly stealthy attacks (strictly stealthy attacks from time step 0 to τ). Then, we study the optimization of deterministic strictly stealthy attacks by analytically deriving the optimal $(\tau+1)$-step strictly stealthy attack that maximizes its effect while utilizing limited attack energy.

2. The general optimization problem for the maximum performance degradation of a system under a finite-horizon attack with a relaxed stealthiness constraint is formulated. Over an infinite-horizon, we derive the optimal attack vectors for a stealthy fully constant or alternating attack and show that these attacks are superior to all other alternating attacks. Furthermore, we move the problem into the frequency domain by introducing stealthy sinusoidal attacks and discuss how to characterize these attacks to find the optimal attack parameters. We derive a sufficient condition for constant and alternating attacks to be optimal over all two-dimensional and most higher-dimensional sinusoidal attacks. Finally, we provide a way to directly compare the performance of the deterministic attacks in this thesis with optimal stochastic attacks in the literature.

1.4 Thesis Outline

The remainder of the thesis is organized as follows. In Chapter 2, we study the set of deterministic attacks on a CPS that do not affect the system innovation within a finite horizon and find the optimal attack out of all possibilities. In Chapter 3, we investigate the optimal attacks on CPSs under a relaxed stealthiness constraint in general over a finite horizon and using constant, alternating, and sinusoidal attack forms over an infinite horizon. Chapter 4 concludes this thesis and provides some potential ideas for future work.

Chapter 2

Finite-Horizon Strictly Stealthy Deterministic Attacks*

This chapter studies strictly stealthy FDI attacks on both channels of closed-loop CPSs. These attacks are designed to have zero influence on the output and innovation for the duration of the incursion. The necessary and sufficient conditions for a strictly stealthy attack of length τ to be possible against a system with known parameters is derived. It is also shown that an attack with length of $n+1$ is possible if and only if an attack of arbitrary length is also possible, where n is the order of the system. Furthermore, the set of all such attacks is provided as the null space of a matrix. Assuming the attacker also has a finite energy constraint, an optimization problem is formulated and solved to find the optimal attack over all feasible options.

This chapter is organized as follows. In Section 2.1, we describe the CPS and attack model. In Section 2.2, we present the main results. In Section 2.3, we use a numerical example to illustrate the theoretical results and proposed attack strategy.

*A version of this chapter has been published as: Donny Cheng, Jun Shang and Tongwen Chen, "Finite-Horizon Strictly Stealthy Deterministic Attacks on Cyber-Physical Systems," *IEEE Control Systems Letters*, vol. 6, pp. 1640–1645, 2022. The contents of this chapter were also selected by the American Control Conference 2022 Program Committee for presentation at the conference.

2.1 Problem Formulation

A diagram of a CPS under attack is provided in Fig. 2.1, which can described as follows.

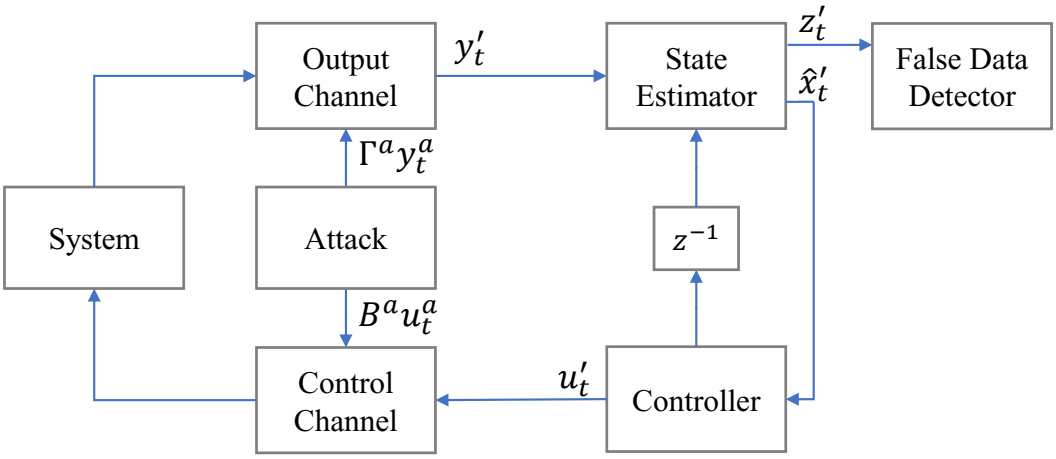

Figure 2.1: System architecture under attack.

2.1.1 System Model

The CPS is represented as a stochastic linear time-invariant discrete-time state-space model:

$$x_{t+1} = Ax_t + Bu_t + w_t \tag{2.1a}$$

$$y_t = Cx_t + v_t \tag{2.1b}$$

where $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^p$, $y_t \in \mathbb{R}^m$ are the state, control input, and measurement output vectors, respectively. $w_t \in \mathbb{R}^n$ and $v_t \in \mathbb{R}^m$ represent the stochastic process noise and measurement noise, respectively. $A \in \mathbb{R}^{n\times n}$ is the system matrix, $B \in \mathbb{R}^{n\times p}$ is the input matrix, and $C \in \mathbb{R}^{m\times n}$ is the output matrix. We assume that (A, B) is controllable and (A, C) is observable.

The system is equipped with an Luenberger observer that provides a state estimate, $\hat{x}_t$, with dynamics:

$$\hat{x}_{t+1} = A\hat{x}_t + Bu_t + K[y_{t+1} - C(A\hat{x}_t + Bu_t)]$$

where $K \in \mathbb{R}^{n \times m}$ is chosen such that $A - KCA$ is stable. Furthermore, the innovation and estimation error are defined by:

$$z_{t+1} = y_{t+1} - C(A\hat{x}_t + Bu_t),\ e_t = x_t - \hat{x}_t.$$

In similar papers, a fixed-gain feedback controller is typically assumed. In fact, we do not need to make any assumptions on the controller at all, and the input u_t can be arbitrarily designed from any available information.

2.1.2 Attack Model

In general, we assume the attacker can manipulate part of the control input and sensor output signals in an additive manner. Then, the system under attack can be described as:

$$\begin{aligned} x'_{t+1} &= Ax'_t + Bu'_t + B^a u^a_t + w_t \\ y'_t &= Cx'_t + \Gamma^a y^a_t + v_t \end{aligned}$$

where $(\cdot)'$ denotes the variable $(\cdot)$ under attack, $u^a_t \in \mathbb{R}^{p_a}$ and $y^a_t \in \mathbb{R}^{m_a}$ are the actuator and sensor attack signals, respectively, and $B^a \in \mathbb{R}^{n \times p_a}$ and $\Gamma^a \in \mathbb{R}^{m \times m_a}$ are the actuator and sensor attack matrices, respectively. Let $n_a = p_a + m_a$, which is the attack's total degrees of freedom.

We assume that $\Gamma^a = [e_{i_1}, \ldots, e_{i_{m_a}}]$, where e_i is the ith canonical basis vector of $\mathbb{R}^m$ and $\{i_1, \ldots, i_{m_a}\}$ is the set of the indices of the compromised outputs. Without loss of generality, B^a and Γ^a are assumed to have full column rank. Furthermore, we assume the attacker has full knowledge of the system parameters A, B, C, and K. In contrast to [37], the attacks on the input and output channels start at time steps 0 and 1, respectively, for notational simplicity.

The corresponding dynamics and definitions for the state estimate, control input, innovation, and estimation error are given by:

$$\begin{aligned} \hat{x}'_{t+1} &= A\hat{x}'_t + Bu'_t + K[y'_{t+1} - C(A\hat{x}'_t + Bu'_t)] \\ z'_{t+1} &= y'_{t+1} - C(A\hat{x}'_t + Bu'_t),\ e'_t = x'_t - \hat{x}'_t. \end{aligned}$$

We define the following variables to represent the difference between attacked and nominal systems:

$$\Delta x_t \coloneqq x'_t - x_t, \Delta \hat{x}_t \coloneqq \hat{x}'_t - \hat{x}_t,\ \Delta u_t \coloneqq u'_t - u_t$$
$$\Delta y_t \coloneqq y'_t - y_t,\ \Delta z_t \coloneqq z'_t - z_t,\ \Delta e_t \coloneqq e'_t - e_t.$$

Remark 2.1. These difference variables are deterministic because the stochastic parts of the attacked and nominal system are the same. Due to the linearity of system (2.1), the noises w_t and v_t cancel each other out.

Remark 2.2. The difference variables are used to quantify both the impact and stealthiness of the attack. In terms of impact, Δe_t gives a measure of how much the attack is impairing the system performance. For stealthiness, false data detectors typically monitor the innovation of the system to find potential anomalies. Thus, the magnitude of Δz_t can be used as a measure of the attack's stealthiness.

We can derive the following update equations:

$$\Delta e_{t+1} = (A - KCA)\Delta e_t + (B^a - KCB^a)u_t^a - K\Gamma^a y_{t+1}^a \tag{2.2}$$
$$\Delta z_{t+1} = CA\Delta e_t + CB^a u_t^a + \Gamma^a y_{t+1}^a. \tag{2.3}$$

Conveniently, Δe_t and Δz_t are both functions of only the attack inputs u_t^a and y_{t+1}^a. Based on our assumptions, we have initial conditions $\Delta e_0 = 0$ and $\Delta z_0 = 0$.

Remark 2.3. In this thesis, we consider general attacks on both channels of closed-loop control systems. However, it should be noted that this work can be easily applied to remote state estimation as well. Because the control input cancels itself out in the derivation of (2.2) and (2.3) for closed-loop control and is not present for state estimation, these two equations apply to both. The difference is that a control channel does not exist for remote state estimation, so the terms with B^a and u_t^a must be removed. Note that this is equivalent to an attack only on the output channel of a closed-loop control system.

2.1.3 System and Attack Classification

A false data detector that relies on innovation to detect faults will be unable to distinguish an attacked system from the nominal one if:

$$\Delta z_t = 0, \ \forall t \in \mathbb{N}. \tag{2.4}$$

Definition 2.1. As defined in [37], an attack sequence is said to be strictly stealthy if (2.4) holds. Additionally, the system in (2.1) is said to be strictly vulnerable if, for any $M_1 > 0$, there exists a strictly stealthy attack such that:

$$\lim_{t\to\infty} \sup \|\Delta e_t\| > M_1.$$

Otherwise, a system is strictly invulnerable if there exists $M_2 > 0$ such that:

$$\lim_{t\to\infty} \sup \|\Delta e_t\| \leq M_2$$

for any strictly stealthy attack.

This chapter extends the definition above. Some attacks may not be able to satisfy (2.4) $\forall t \in \mathbb{N}$, but they can achieve strict stealthiness in a finite horizon:

$$\Delta z_t = 0, \ \forall t = 1, 2, \ldots, \tau + 1. \tag{2.5}$$

Systems that permit such an attack may be strictly invulnerable, but these attacks can still deal significant damage in a short amount of time.

Definition 2.2. An attack sequence is said to be $(\tau + 1)$-step strictly stealthy if (2.5) holds. Additionally, system (2.1) is τ-step strictly vulnerable if, for any $M_1 > 0$, there exists a $(\tau + 1)$-step strictly stealthy attack such that:

$$\|\Delta e_{\tau+1}\| > M_1. \tag{2.6}$$

Remark 2.4. Due to the linearity of the system, any strictly stealthy attack can be scaled to become arbitrarily large while still keeping $\Delta z_t = 0$. Thus, as long as a $(\tau + 1)$-step strictly stealthy attack exists for a given system (2.1), we can scale it to achieve (2.6), and the system is τ-step strictly vulnerable.

Remark 2.5. While an attack remains strictly stealthy, it can be shown that $\Delta z_t = \Delta y_t = 0$. In other words, the output from a strictly stealthy attack is identical to the nominal one. Therefore, any detector that monitors the output of the system will fail to detect a strictly stealthy attack as well.

2.1.4 Optimal Strictly Stealthy Attack Formulation

We shall use the following formulation to find an optimal $(\tau + 1)$-step strictly stealthy attack for a τ-step strictly vulnerable system. First, we need to find an appropriate objective function and impose additional constraints on our problem. For notational simplicity, let $\zeta_t = [u_t^a; y_{t+1}^a]$, which combines the input and output attack vectors.

We can measure impact of the attack on the system using a summation of a quadratic function of Δe_t over $\tau + 1$ time steps:

$$J = \sum_{t=1}^{\tau+1} \Delta e_t^T W \Delta e_t \tag{2.7}$$

where $W \in \mathbb{S}_+^n$.

Note that if we do not create additional constraints on the attack input, ζ_t, then we can take $\|\zeta_t\| \to \infty$ to maximize J, which is a relatively uninteresting result. Instead, we will consider the case in which the attacker has limited attack energy over these $\tau + 1$ attacks. Let $\zeta = [\zeta_0; \zeta_1; \ldots; \zeta_\tau]$. Then, this can be represented as a quadratic constraint:

$$\zeta^T H \zeta \leq \varepsilon \tag{2.8}$$

where $H \in \mathbb{S}_{++}^{(\tau+1)n_a}$ and ε is some scalar constant. Putting this all together along with the strict stealthiness constraint in (2.5), we can form the optimization problem below to find the optimal attack:

$$\begin{aligned} \max_{\zeta} \quad & \sum_{t=1}^{\tau+1} \Delta e_t^T W \Delta e_t \\ \text{s.t.} \quad & \zeta^T H \zeta \leq \varepsilon \\ & \Delta z_t = 0, \ t = 1, 2, \ldots, \tau + 1. \end{aligned} \tag{2.9}$$

2.2 Main Results

2.2.1 Conditions for τ-Step Strict Vulnerability

During each step of a strictly stealthy attack, from (2.3), we have $KCA\Delta e_t + KCB^a u_t^a + K\Gamma^a y_{t+1}^a = 0$. Applying this to (2.2), the update equation for Δe_t during a strictly stealthy attack is:

$$\Delta e_{t+1} = A\Delta e_t + B^a u_t^a. \tag{2.10}$$

Let $\bar{\Gamma}$ be the matrix with columns being the canonical basis vectors of $\mathbb{R}^m$ that are not present in Γ^a. Moreover, let $B_z^a = \begin{bmatrix} CB^a & \Gamma^a \end{bmatrix}$ and $B_e^a = \begin{bmatrix} B^a & 0 \end{bmatrix}$, the matrix coefficients of ζ_t in (2.3) and (2.10), respectively. For notational simplicity, let $\bar{\mathcal{A}}_i = \bar{\Gamma}^T CA^i B^a$.

Lemma 2.1 below shows the equivalence between rank conditions and null spaces of M_i, introduced in [37, Th. 1], and N_i, which has a simpler form.

Lemma 2.1. *The following two conditions are equivalent:*

$$\operatorname{rank}(N_\tau) - \operatorname{rank}(N_{\tau-1}) < p_a \tag{2.11}$$

$$\operatorname{rank}(M_\tau) - \operatorname{rank}(M_{\tau-1}) < n_a \tag{2.12}$$

where:

$$N_i = \begin{bmatrix} \bar{\mathcal{A}}_0 & 0 & \dots & 0 \\ \bar{\mathcal{A}}_1 & \bar{\mathcal{A}}_0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathcal{A}}_i & \bar{\mathcal{A}}_{i-1} & \dots & \bar{\mathcal{A}}_0 \end{bmatrix}$$

$$M_i = \begin{bmatrix} B_z^a & 0 & \dots & 0 \\ CAB_e^a & B_z^a & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^i B_e^a & CA^{i-1} B_e^a & \dots & B_z^a \end{bmatrix}.$$

Moreover, the null space of M_i in the form $M_i\zeta = 0$ is equivalent to the null space of N_i in the form $N_i u^a = 0$, where $u^a = [u_0^a; u_1^a; \dots; u_\tau^a]$ and:

$$y_{i+1}^a = -(\Gamma^a)^T C \sum_{j=0}^{i} A^{i-j} B^a u_j^a,\ i = 0, 1, \dots, \tau. \tag{2.13}$$

Proof. Each block in M_i has m_a more rows and columns compared to N_i (corresponding to the addition of y^a_{t+1} to the attack vector) that are redundant. The added rows are associated with position of the non-zero elements of Γ^a in M_i. Since these rows include the only non-zero entry in a column (the 1 from Γ^a), they are linearly independent of the others. We can remove these rows by multiplying on the left of each block by $\bar{\Gamma}^T$, noting that $\bar{\Gamma}^T\Gamma^a = 0$. We then have the matrix $\mathcal{N}_i$:

$$\mathcal{N}_i = \begin{bmatrix} \bar{\mathcal{A}}_0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \bar{\mathcal{A}}_1 & 0 & \bar{\mathcal{A}}_0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \bar{\mathcal{A}}_i & 0 & \bar{\mathcal{A}}_{i-1} & 0 & \dots & \bar{\mathcal{A}}_0 & 0 \end{bmatrix}.$$

Note that the only difference between $\mathcal{N}_i$ and N_i are several zero columns, so they have the same rank. Additionally, since $\mathcal{N}_i$ is M_i with $(i+1)m_a$ linearly independent rows removed, then:

$$\text{rank}(\mathcal{N}_i) = \text{rank}(N_i) = \text{rank}(M_i) - (i+1)m_a. \tag{2.14}$$

It can then be easily shown that (2.11) implies (2.12) and vice versa.

Furthermore, solving for the rows of $M_i\zeta = 0$ that are not included in N_i, we can also recover the design of the attack on the output channel:

$$\Gamma^a y^a_{i+1} + \sum_{j=0}^{i} CA^{i-j}B^a u^a_j = 0,\ i = 0, 1, \dots, \tau. \tag{2.15}$$

This is equivalent to (2.13). The solution for u^a_i remains the same. To see this, note that the solution for the remaining rows (represented by $\mathcal{N}_i\zeta = 0$) and $N_i u^a = 0$ are equivalent in terms of u^a. ■

Theorem 2.1. *A system is τ-step strictly vulnerable if and only if (2.12) is satisfied.*

Proof. The solution to the dynamics under attack in (2.10) is given by:

$$\Delta e_{t+1} = \sum_{i=0}^{t} A^{t-i}B^a u^a_i. \tag{2.16}$$

From (2.3) and (2.5), we have the following condition for the attack to be strictly stealthy at a given time step:

$$CA\Delta e_t + CB^a u_t^a + \Gamma^a y_{t+1}^a = 0. \tag{2.17}$$

Combining (2.16) and (2.17), we have:

$$\Gamma^a y_{t+1}^a = -C \sum_{i=0}^{t} A^{t-i} B^a u_i^a. \tag{2.18}$$

Solving (2.18) for y_{t+1}^a, we have (2.13) since $(\Gamma^a)^T \Gamma^a = I_{m_a}$. In (2.13), there is no freedom to design y_{t+1}^a since the equation is a fixed function of the sequence of u_t^a. Effectively, this formulation puts all of the degrees of freedom for the attacker in u_t^a, which is desirable as it is the part of the attack that influences Δe_t.

To meet the stealthiness condition in (2.18), we require it to have a solution. This can be represented as:

$$\bar{\Gamma}^T C \sum_{i=0}^{t} A^{t-i} B^a u_i^a = 0. \tag{2.19}$$

As long as this condition is met for some $t = 0, 1, \ldots, \tau$, then a strictly stealthy attack exists up until time step τ (a sequence of $\tau+1$ attacks). A combination of these conditions up to and including the one for $t = \tau$ can be written in matrix form as:

$$N_\tau u^a = 0. \tag{2.20}$$

However, an attack sequence in the null space of N_τ is not sufficient to guarantee the existence of a $(\tau + 1)$-step strictly stealthy attack. This is because we assume the non-trivial attack starts from $t = 0$ on the input channel. An attack vector with $u_0^a = 0$ will contradict this assumption.

In general, note that $\dim \ker(N_{t+1}) \geq \dim \ker(N_t)$ because we can always take $u_0^a = 0$ in $\ker(N_{t+1})$ to recover $\ker(N_t)$. Assuming $\dim \ker(N_0) > 0$, we require $\dim \ker(N_1) > \dim \ker(N_0)$ for a 1-step strictly stealthy attack to exist. We further require $\dim \ker(N_2) > \dim \ker(N_1)$ for a 2-step strictly

stealthy attack to exist, and so on. In other words, a necessary condition for a $(\tau+1)$-step strictly stealthy attack to exist is:

$$\dim\ker(N_i) > \dim\ker(N_{i-1}),\ \forall i = 0, 1, \ldots, \tau \tag{2.21}$$

where we can define $\dim\ker(N_{-1}) = \operatorname{rank}(N_{-1}) = 0$. From the rank-nullity theorem, this is equivalent to:

$$\operatorname{rank}(N_i) - \operatorname{rank}(N_{i-1}) < p_a,\ \forall i = 0, 1, \ldots, \tau. \tag{2.22}$$

Now, we show by contradiction that if the condition in (2.22) holds for $i = \tau$, then it also holds for $i = 0, 1, \ldots, \tau-1$. Suppose $\operatorname{rank}(N_{\tau-1}) - \operatorname{rank}(N_{\tau-2}) = p_a$ and $\operatorname{rank}(N_\tau) - \operatorname{rank}(N_{\tau-1}) < p_a$. Then, the last block row of $N_{\tau-1}$ has p_a rows that are linearly independent of the other rows above it. However, this also implies that the last block row of N_τ will have p_a rows that are linearly independent of the rows above such that $\operatorname{rank}(N_\tau) - \operatorname{rank}(N_{\tau-1}) = p_a$, regardless of $\bar{\mathcal{A}}_\tau$. This is a contradiction. Thus, if $\operatorname{rank}(N_\tau) - \operatorname{rank}(N_{\tau-1}) < p_a$, then $\operatorname{rank}(N_{\tau-1}) - \operatorname{rank}(N_{\tau-2}) < p_a$ since the difference in rank cannot exceed p_a.

Along with (2.22) and Lemma 2.1, this proves the necessity. The condition is also sufficient because a strictly stealthy attack can be designed against these systems in accordance with Corollary 2.2. ■

Remark 2.6. Note that the proof of Theorem 2.1 also implies that if a system is τ-step strictly vulnerable, then it is also i-step strictly vulnerable $\forall i = 0, 1, \ldots, \tau-1$.

Corollary 2.1. *A system is n-step strictly vulnerable if and only if it is also ∞-step strictly vulnerable.*

Proof. From the Cayley–Hamilton theorem, we can write A^i with $i \geq n$ as a linear combination of the next n smaller powers of A:

$$A^i = c_{n-1}A^{i-1} + \cdots + c_1 A^{i-n+1} + c_0 A^{i-n}. \tag{2.23}$$

For N_n, this means we can express $\bar{\Gamma}^T C A^n B^a$ as a linear combination of the n elements above it. Thus, we can use elementary row operations to obtain the following matrix:

$$\bar{N}_n = \begin{bmatrix} \bar{\mathcal{A}}_0 & 0 & \dots & 0 & 0 \\ \bar{\mathcal{A}}_1 & \bar{\mathcal{A}}_0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \bar{\mathcal{A}}_{n-1} & \bar{\mathcal{A}}_{n-2} & \dots & \bar{\mathcal{A}}_0 & 0 \\ 0 & \hat{A}_1 & \dots & \hat{A}_{n-1} & \bar{\mathcal{A}}_0 \end{bmatrix} \tag{2.24}$$

where $\hat{A}_j = \bar{\Gamma}^T C(A^{n-j} - \sum_{i=j}^{n-1} c_i A^{i-j}) B^a$. If the system is n-step strictly vulnerable, then the last block row has less than p_a rows that are linearly independent of all the other rows. Now assume the system is $(\tau-1)$-step strictly vulnerable ($\tau > n$). In N_τ, we can remove the first $\tau - n + 1$ block elements by subtracting the n block rows above it with appropriate coefficients. Then, comparing with (2.24), it can be shown that the last block row has less than p_a rows that are linearly independent of all the other rows such that $\text{rank}(N_\tau) - \text{rank}(N_{\tau-1}) < p_a$. ■

Remark 2.7. Note that ∞-step strictly vulnerability is equivalent to the strict vulnerability defined in [37]. Condition (2.12) reduces to the one in [37, Th. 1] when $\tau = n$.

Corollary 2.2. *All $(\tau+1)$-step strictly stealthy attack sequences for a τ-step strictly vulnerable system must satisfy:*

$$\zeta \in \ker(M_\tau). \tag{2.25}$$

Proof. From Theorem 2.1, for a $(\tau+1)$-step attack to be strictly stealthy, the input and output attack vector sequences must satisfy (2.20) and (2.13), respectively. From Lemma 2.1, this is equivalent to (2.25). ■

2.2.2 Optimal $(\tau+1)$-Step Strictly Stealthy Attack

Theorem 2.2. *The solution to the optimization problem in (2.9) is:*

$$\zeta^* = \pm\sqrt{\varepsilon} Z Q \begin{bmatrix} 0 \\ v_{max} \end{bmatrix} \tag{2.26}$$

where $Z = H^{-\frac{1}{2}}$, Q *is from the* QR *factorization of* $(M_\tau Z)^T$, *and* v_{max} *is the normalized eigenvector corresponding to the dominant eigenvalue of* W_{22}, *defined from the following partition:*

$$Q^T Z^T \sum_{i=1}^{\tau+1} W^{(i)} Z Q = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}. \tag{2.27}$$

Here, $W_{11} \in \mathbb{R}^{r\times r}$, $W_{12} \in \mathbb{R}^{r\times(\tau+1)n_a - r}$, $W_{21} \in \mathbb{R}^{(\tau+1)n_a - r\times r}$, *and* $W_{22} \in \mathbb{R}^{(\tau+1)n_a - r\times(\tau+1)n_a - r}$ *and the* $n_a \times n_a$ *submatrices in the* j*th block row and* k*th block column of* $W^{(i)}$ *is given by:*

$$W_{jk}^{(i)} = \begin{cases} (B_e^a)^T (A^{i-j})^T W A^{i-k} B_e^a & j,k = 1,2,\ldots,i \\ 0 & \textit{otherwise.} \end{cases} \tag{2.28}$$

Proof. From Corollary 2.2, to meet the stealthiness constraint in (2.5), the attack must satisfy (2.25). Thus, we can replace the stealthiness condition with the linear equality constraint:

$$M_\tau \zeta = 0. \tag{2.29}$$

Using the solution for Δe_t in (2.16), we can write (2.7) as a quadratic function of ζ only: $J = \zeta^T \check{W} \zeta$, where:

$$\check{W} = \sum_{i=1}^{\tau+1} W^{(i)} \tag{2.30}$$

and $W^{(i)}$ represents the contribution at the ith time step. We can partition $W^{(i)}$ into a block matrix with $\tau+1$ block rows and columns made up of $n_a \times n_a$ submatrices. From direct computation, the element in the jth block row and kth block column of $W^{(i)}$ is given by (2.28).

Now we will convert the energy constraint in (2.8) to a norm inequality. Since $H \in \mathbb{S}_{++}^{(\tau+1)n_a}$, then $Z = H^{-\frac{1}{2}}$ exists. Let $x = Z^{-1}\zeta \Rightarrow \zeta = Zx$. Then, problem (2.9) becomes:

$$\begin{aligned} \max_x \quad & x^T Z^T \check{W} Z x \\ \text{s.t.} \quad & x^T x \le \varepsilon \\ & M_\tau Z x = 0. \end{aligned} \tag{2.31}$$

Without the linear equality constraint, the solution to this optimization problem is equal to the eigenvector corresponding to the dominant eigenvalue of the quadratic matrix in the objective function scaled such that $\|x\|^2 = \varepsilon$. Thus, our goal is to incorporate the linear equality constraint into the other parts of the optimization problem. A procedure to accomplish this is given in [6]. Let $\bar{M}_\tau = (M_\tau Z)^T \in \mathbb{R}^{(\tau+1)n_a \times (\tau+1)m}$ with rank r. There exists a QR decomposition of $\bar{M}_\tau$:

$$\bar{M}_\tau = Q \begin{bmatrix} R_1 & S \\ 0 & 0 \end{bmatrix} \Pi \tag{2.32}$$

where $Q \in \mathbb{R}^{(\tau+1)n_a \times (\tau+1)n_a}$ is an orthogonal matrix, $R_1 \in \mathbb{R}^{r \times r}$ is a full rank upper triangular matrix, S is some arbitrary matrix, and Π is a permutation matrix. Now, let us split x into two separate arguments: $x = Q[y; z]$, where $y \in \mathbb{R}^r$ and $z \in \mathbb{R}^{(\tau+1)n_a - r}$. The equality constraint becomes: $\bar{M}_\tau^T Q[y; z] = 0$. Applying the QR decomposition:

$$\Pi^T \begin{bmatrix} R_1^T & 0 \\ S^T & 0 \end{bmatrix} Q^T Q \begin{bmatrix} y \\ z \end{bmatrix} = 0 \Rightarrow \Pi^T \begin{bmatrix} R_1^T \\ S^T \end{bmatrix} y = 0. \tag{2.33}$$

Since Π is invertible and R_1 has rank r, it is clear that the unique solution is $y = 0$. The optimization problem in (2.31) simplifies to:

$$\begin{aligned} \max_z \quad & \begin{bmatrix} 0 & z^T \end{bmatrix} Q^T Z^T \breve{W} Z Q \begin{bmatrix} 0 \\ z \end{bmatrix} \\ \text{s.t.} \quad & z^T z \leq \varepsilon. \end{aligned} \tag{2.34}$$

We can partition the matrix in the objective function as in (2.27). Then, we have the standard optimization problem:

$$\begin{aligned} \max_z \quad & z^T W_{22} z \\ \text{s.t.} \quad & z^T z \leq \varepsilon. \end{aligned} \tag{2.35}$$

This has a clear optimal solution: $z^* = \pm\sqrt{\varepsilon} v_{max}$, where v_{max} is the eigenvector corresponding to the largest eigenvalue of W_{22}. After reverting the two transformations we made, the optimal attack sequence is given in (2.26). ■

2.3 Simulation

We shall use the following system to numerically illustrate the results. Suppose we have a fourth order system with parameters:

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad B^a = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \Gamma^a = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

and $C = I_4$. Computing M_i, we have $\text{rank}(M_0) = 3$, $\text{rank}(M_1) = 6$, $\text{rank}(M_2) = 9$, and $\text{rank}(M_3) = 13$. Using Theorem 2.1, we can see that this system is 2-step strictly vulnerable. The permissible 3-step attack sequences satisfy $M_2\zeta = 0$, where $\zeta = [\zeta_0; \zeta_1; \zeta_2]$.

Let $\varepsilon = 1$. Take the quadratic weighting matrices to be:

$$W = \begin{bmatrix} 3 & 1 & 0 & 1.5 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 5 & 1 \\ 1.5 & 0 & 1 & 2 \end{bmatrix} \quad H = I_{12}. \tag{2.36}$$

From here, we can compute $\check{W}$ using (2.28) and (2.30). Here, we can see that $Z = I_{12}$. Q, R_1, and Π can be computed such that $Q^TQ = I_{12}$ and $\text{rank}(R_1) = r = 9$. Then, computing $Q^TZ^T\check{W}ZQ$ and partitioning, we can find:

$$W_{22} = \begin{bmatrix} 1.5000 & 0.3959 & -0.0995 \\ 0.3959 & 1.8528 & 0.4514 \\ -0.0995 & 0.4514 & 2.3139 \end{bmatrix}. \tag{2.37}$$

The largest magnitude eigenvalue of W_{22} is $\lambda_{max} = 2.6957$ and the corresponding normalized eigenvector is $v_{max} = [0.1259; 0.5580; 0.8202]$. From (2.26), we can solve for one solution of ζ^* as:

$$\zeta_0^* = \begin{bmatrix} 0.4658 \\ -0.4658 \\ 0 \\ 0 \end{bmatrix}, \ \zeta_1^* = \begin{bmatrix} 0.1969 \\ -0.1969 \\ -0.4658 \\ 0 \end{bmatrix}, \ \zeta_2^* = \begin{bmatrix} 0.0891 \\ -0.0891 \\ -0.1969 \\ -0.4658 \end{bmatrix}.$$

The other solution is obtained by flipping the sign on all attack vectors.

Applying this attack, the cumulative objective value at each time step is shown in Fig. 2.2. Within the figure, the box represents values within

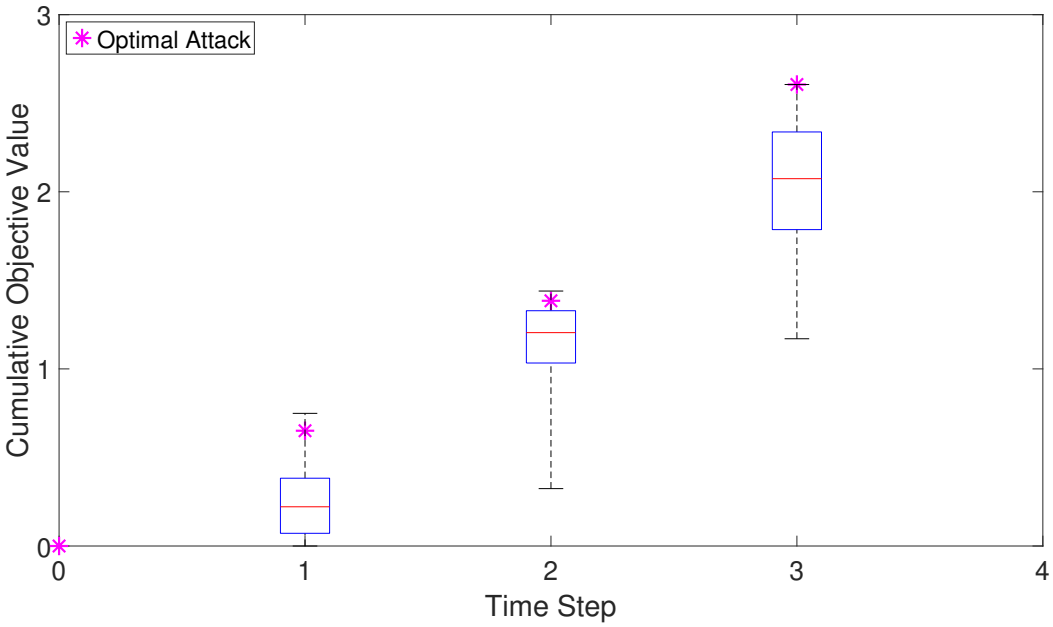

Figure 2.2: The cumulative objective values at each time step for the optimal attack and a box plot for a set of random attacks.

the 25th and 75th percentile, the line within the box indicates the median, and the whiskers mark the minimum and maximum values. This figure also includes some statistics for the objective values from random attacks generated from 10,000 Monte Carlo simulations to show the optimality of our solution. Although random attacks may be able to achieve higher objective values for time steps 1 and 2, the attack from Theorem 2.2 has the largest impact on the system over the three time steps as designed.

Chapter 3

Optimal Stealthy Deterministic Attack Strategies*

This chapter also studies the problem of FDI attacks against both communication channels of a closed-loop CPS. However, relaxed definitions of stealthiness that differ from Chapter 2 are considered instead. It is still difficult for innovation-based detectors to detect these attacks, although they are no longer undetectable. We formulate and simplify an optimization problem for the attack over both finite and infinite horizons. Using the intuition from the numerical solutions, we identify effective analytical attack forms for the infinite-horizon, including constant, alternating, and sinusoidal attacks, and optimize them. Algorithms are created to characterize the objective value of two and three-dimensional sinusoidal attacks of any frequency, including edge cases. Because the proposed analysis of sinusoidal attacks is computationally intensive, a useful condition that is relatively easy to check is provided for a constant or alternating attack to be superior to most sinusoidal attacks. We also propose a method to compare these optimal infinite-horizon deterministic attacks with their stochastic counterparts. Finally, simulations show the effectiveness of the proposed strategies to degrade system performance.

This chapter is organized as follows. Section 3.1 introduces some additional definitions and formulations required for this chapter. In section 3.2, the

*A version of this chapter has been submitted for publication as: Donny Cheng, Jun Shang, and Tongwen Chen, "Optimal Stealthy Deterministic Attack Strategies on Cyber-Physical Systems", *IEEE Transactions on Control of Networked Systems.*

optimization problem for an optimal attack over a finite horizon is derived. Section 3.3 discusses optimal constant and alternating attacks. The analysis and optimization of sinusoidal attacks are studied in Section 3.4. Section 3.5 provides a condition for the optimal constant and alternating attacks to outperform their sinusoidal counterparts. Section 3.6 discusses how we can compare the performance of deterministic attacks with stochastic ones. We then illustrate the effect of the proposed attack strategies using simulations in Section 3.7.

3.1 Problem Formulation

3.1.1 Cyber-Physical System and Attack Model

This chapter is an extension of Chapter 2 to attacks that satisfy a relaxed stealthiness condition. Therefore, the system and attack model for this chapter is the same as in Chapter 2. Although a short summary is provided here, please refer to Section 2.1 for details.

We consider a the state-space realization of a stochastic linear time-invariant discrete-time system:

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + w_t \\ y_t &= Cx_t + v_t \end{aligned} \tag{3.1}$$

The system is equipped with a general controller and fixed-gain observer on the remote side that generates a control input and state estimate, respectively, at each time step. The attacker has the capability to intercept the signals being sent over both communication channels and modify a component of these transmissions. Additionally, the attacker has the system parameters A, B, C, and K available. The analysis makes use of variables that represent the difference between attacked and nominal systems, which are all deterministic in nature. The dynamics for the two most important difference variables are derived and reiterated here:

$$\Delta e_{t+1} = (A - KCA)\Delta e_t + (B^a - KCB^a)u_t^a - K\Gamma^a y_{t+1}^a \tag{3.2}$$

$$\Delta z_{t+1} = CA\Delta e_t + CB^a u_t^a + \Gamma^a y_{t+1}^a. \tag{3.3}$$

3.1.2 System and Attack Classification

For this chapter, some additional definitions for stealthiness are required. An innovation-based false data detector will have difficulty distinguishing an attacked system from the nominal one if:

$$\|\Delta z_t\| \leq \delta, \ \forall t \in \mathbb{N} \tag{3.4}$$

for some small constant $\delta > 0$.

Definition 3.1. As defined in [37], an attack sequence is said to be stealthy if (3.4) holds. Additionally, the system in (3.1) is said to be vulnerable if, for any $M_1 > 0$, there exists a stealthy attack such that:

$$\limsup_{t\to\infty} \|\Delta e_t\| > M_1.$$

Otherwise, a system is invulnerable if there exists $M_2 > 0$ such that:

$$\limsup_{t\to\infty} \|\Delta e_t\| \leq M_2$$

for any stealthy attack.

When the focus is on infinite-horizon attacks, the stealthiness condition can be further relaxed by only considering the stealthiness condition when $t \to \infty$. This idea is similar to the notion of complete stealthiness introduced in [45], for which the focus is the norm of the innovation as $t \to \infty$.

Definition 3.2. An attack sequence is said to be steady-state stealthy if the following condition holds:

$$\lim_{t\to\infty} \|\Delta z_t\| \leq \delta. \tag{3.5}$$

3.1.3 Optimal Stealthy Attack Formulation

In a finite horizon, to quantify the impact of the attack on the system, let us consider a summation objective function to maximize over τ time steps:

$$J = \sum_{t=1}^{\tau} \Delta e_t^T Q_0 \Delta e_t \tag{3.6}$$

for some weighting matrix $Q_0 \in \mathbb{S}_+^n$. Then, combining this with the stealthiness condition, we can form the optimization problem:

$$\begin{aligned} \max_{\zeta} \quad & \sum_{t=1}^{\tau} \Delta e_t^T Q_0 \Delta e_t \\ \text{s.t.} \quad & \|\Delta z_{t+1}\| \leq \delta,\ t = 0, 1, \ldots, \tau - 1. \end{aligned} \tag{3.7}$$

Now, moving to an infinite horizon, we consider a similar objective function. However, we are only interested in the step-by-step objective value at the steady state, since the costs in a finite interval become negligible. Thus, we have:

$$J = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \Delta e_t^T Q_0 \Delta e_t \tag{3.8}$$

and the corresponding optimization problem, assuming that the attack must also be steady-state stealthy, is:

$$\begin{aligned} \max_{\zeta} \quad & \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \Delta e_t^T Q_0 \Delta e_t \\ \text{s.t.} \quad & \lim_{t \to \infty} \|\Delta z_{t+1}\| \leq \delta. \end{aligned} \tag{3.9}$$

Only invulnerable systems will be examined for the infinite-horizon case because attacks that are designed to create a divergent step-by-step cost as $t \to \infty$ are possible for vulnerable systems, as discussed in [37].

3.2 Finite-Horizon Optimization Problem

3.2.1 Construction of the Optimization Problem

First, let us consider joint actuator and sensor stealthy attacks in a finite horizon. Let $\bar{A} = A - KCA$, $\bar{B} = [B^a - KCB^a, -K\Gamma^a]$, $\bar{C} = CA$, and $\bar{D} = [CB^a, \Gamma^a]$. Equations (3.2) and (3.4) can then be represented by the following equation and constraint:

$$\begin{aligned} \Delta e_{t+1} &= \bar{A}\Delta e_t + \bar{B}\zeta_t \\ \|\Delta z_{t+1}\| &= \|\bar{C}\Delta e_t + \bar{D}\zeta_t\| \leq \delta. \end{aligned} \tag{3.10}$$

Although at the beginning of an attack, $\Delta e_0 = 0$, we do not make this assumption below to allow us to chain many of these optimization problems together in order to obtain an attack of arbitrary length.

Proposition 3.1. *The optimization problem in* (3.7) *is equivalent to a concave objective, convex constraint quadratically constrained quadratic program (QCQP) in terms of the attack vector* $\zeta = [\zeta_0; \zeta_1; \ldots; \zeta_{\tau-1}]$*:*

$$\begin{aligned} \max_{\zeta} \quad & \frac{1}{2}\zeta^T Q \zeta + f^T \zeta \\ s.t. \quad & \frac{1}{2}\zeta^T H^{(i)} \zeta + {g^{(i)}}^T \zeta + d^{(i)} \leq 0,\ i = 1, 2, \ldots, \tau \end{aligned} \tag{3.11}$$

where $Q = \sum_{i=1}^{\tau} Q^{(i)}$*,* $f = \sum_{i=1}^{\tau} f^{(i)}$*, and:*

$$\begin{aligned} Q_{jk}^{(i)} &= \begin{cases} -2\bar{B}^T {\bar{A}^{(i-j)}}^T Q_0 \bar{A}^{i-k} \bar{B} & j, k = 1, 2, \ldots, i \\ 0 & otherwise \end{cases} \\ f_j^{(i)} &= \begin{cases} -2\bar{B}^T {\bar{A}^{(i-j)}}^T \bar{A}^i \Delta e_0 & j = 1, 2, \ldots, i \\ 0 & otherwise \end{cases} \\ \Xi_i &= [\bar{C}\bar{A}^{i-2}\bar{B}, \bar{C}\bar{A}^{i-3}\bar{B}, \ldots, \bar{C}\bar{B}, \bar{D}, 0, \ldots, 0] \\ H^{(i)} &= \Xi_i^T \Xi_i \\ g^{(i)} &= \Xi_i^T \bar{C}\bar{A}^{i-1}\Delta e_0 \\ d^{(i)} &= \Delta e_0^T {\bar{A}^{(i-1)}}^T \bar{C}^T \bar{C}\bar{A}^{i-1}\Delta e_0 - \delta. \end{aligned}$$

Proof. For some initial condition, Δe_0, the solution of (3.10) is solely dependent on the input, ζ_t:

$$\Delta e_{t+1} = \sum_{i=0}^{t} \bar{A}^{t-i}\bar{B}\zeta_i + \bar{A}^{t+1}\Delta e_0. \tag{3.12}$$

For the objective function, we can write (3.6) as a sum of a linear and quadratic function of ζ using (3.12) by direct computation. The terms of (3.6) are taken into account separately and then summed together; $Q^{(i)}$ and $f^{(i)}$ represent the contribution of the ith time step. Similarly, the stealthiness constraint at each time step can also be written as a sum of a linear and quadratic function of ζ as well as a constant by using (3.12) again. $H^{(i)}$, $g^{(i)}$, and $d^{(i)}$ represent this constraint at the ith time step, associated with the innovation Δz_i.

Given some matrix $M \in \mathbb{S}_+^n$ and any compatible matrix A, $A^T M A \in \mathbb{S}_+^n$. Since $Q_0 \in \mathbb{S}_+^n$, it is clear that $-Q^{(i)} \in \mathbb{S}_+^n\ \forall i$. Q is the sum of τ negative semi-definite matrices, so it is also negative semi-definite and the objective function

is concave. Similarly, $H^{(i)} \in \mathbb{S}_+^n \ \forall i$ because it is the product of a matrix and its transpose. Then, the search space is the intersection of τ convex sets, so it is itself convex. ■

Remark 3.1. Unfortunately, due to the non-convexity of the objective function, it is non-trivial to find the global optimum of the optimization problem in (3.11) numerically. Nonetheless, using a global non-convex solver, such as MATLAB's fmincon, yields a solution relatively efficiently and reliably. Alternatively, we can attempt to obtain an approximate solution using the semi-definite relaxation [21].

Remark 3.2. Numerically solving this optimization problem can give us an optimal attack of length τ. However, if we make τ too large, it could make the problem computationally intractable. Instead, we can solve a series of these optimization problems by first solving one problem with a reasonably small τ and then using the final state of the last problem, Δe_τ, as the initial state, Δe_0, for the next problem. This allows us to create an attack sequence of arbitrary length more effectively.

Remark 3.3. Although (3.11) is a difficult optimization problem to solve, numerical solutions provide us with some insight into the most effective attack strategies. The optimal attack tends to converge to either a steady-state constant, alternating, or sinusoidal-like signal after a short transient. This phenomenon is demonstrated using some of the examples in Section 3.7 in Figs. 3.1, 3.2, and 3.5. Therefore, the remainder of this chapter will discuss these attack strategies over an infinite horizon in detail.

3.2.2 Receding Horizon Implementation of the Solution

Particularly if the system is open-loop unstable, the solution to the optimization problem in (2.9) may be greedy and generate a final state that causes the next optimization problem to become infeasible if applied as Δe_0. That is, for some systems and at some states that can be reached with stealthy attacks, there may not exist any stealthy attack of an arbitrary length. Inspired by

model predictive control, one possible solution to avoid this issue is to use a receding-horizon implementation, which is summarized in Algorithm 1.

Algorithm 1: Receding Horizon Implementation for Attack Design.

Input: τ, t_{util}
begin
  $\Delta e_0 \leftarrow 0$;
  Compute matrices Q and $H^{(i)}$;
  Loop
    Compute f, $g^{(i)}$, and $d^{(i)}$ from Δe_0;
    $\zeta_t^* \leftarrow$ solution of optimization problem (3.11) over τ time steps;
    Find $\Delta e_{t_{util}}$ from (3.2) and the first t_{util} elements of ζ_t^*;
    $\Delta e_0 \leftarrow \Delta e_{t_{util}}$;

For the receding horizon implementation, we solve the optimization problem in Proposition 3.1 for some finite horizon τ but only apply a part of the solution specified by a number of time steps t_{util}. Then, we solve the optimization problem for the next τ time steps again. This prevents the attack from approaching an infeasible point because it must ensure that a stealthy attack is still possible past t_{util}, rather than only ensuring feasibility before the current optimization problem ends. The effectiveness of this algorithm is demonstrated using Fig. 3.3 from Section 3.7.

3.3 Alternating Constant Steady-State Attacks

First, we discuss attacks that are either constant or alternate over time. These attacks take the form:

$$\zeta_t = \zeta_{ss} + (-1)^t \zeta_{al} \tag{3.13}$$

where ζ_{ss} and ζ_{al} are the constant and alternating components, respectively.

3.3.1 Optimal Fully Constant and Alternating Attacks

We first consider a fully constant or alternating attack, $\zeta_t = (\pm 1)^t \zeta_c$. In all subsequent equations, if there are two signs, the top and bottom signs

correspond to a fully constant and alternating attack, respectively, to avoid redundancy. We can solve (3.2) to obtain the steady-state solution. For alternating attacks, the solution depends on whether the latest attack vector was positive or negative:

$$\lim_{t\to\infty} \Delta e_t = \sum_{i=0}^{\infty} \bar{A}^i \bar{B} (\pm 1)^{t+1+i} \zeta_c.$$

We can rewrite this solution as:

$$\lim_{t\to\infty} \Delta e_t = (\pm 1)^{t+1} \sum_{i=0}^{\infty} (\pm \bar{A})^i \bar{B} \zeta_c.$$

We see that the alternating attack causes $\lim_{t\to\infty} \Delta e_t$ to switch between two states and effectively flips the sign for the system matrix $\bar{A}$. Since we know that $\bar{A}$, and therefore also $-\bar{A}$, are Schur, the infinite sum converges:

$$\lim_{t\to\infty} \Delta e_t = (\pm 1)^{t+1} (I \mp \bar{A})^{-1} \bar{B} \zeta_c. \tag{3.14}$$

Then, the infinite-horizon objective function (3.8) becomes $J = \zeta_c^T M \zeta_c$, where:

$$M = \bar{B}^T [(I \mp \bar{A})^{-1}]^T Q_0 (I \mp \bar{A})^{-1} \bar{B}. \tag{3.15}$$

At steady state, the constraint in (3.10) becomes the condition:

$$\| [(\pm 1)^{t+1} \bar{C} (I \mp \bar{A})^{-1} \bar{B} + (\pm 1)^t \bar{D}] \zeta_c \| \leq \delta. \tag{3.16}$$

Let:

$$N = [\bar{C}(I \mp \bar{A})^{-1} \bar{B} \pm \bar{D}]^T [\bar{C}(I \mp \bar{A})^{-1} \bar{B} \pm \bar{D}]. \tag{3.17}$$

Since $N \in \mathbb{S}_+^{n_a}$, its square root, $\sqrt{N}$, exists. Let $\text{rank}(N) = r$. We can then define the QR decomposition of $\sqrt{N}$ as:

$$\sqrt{N} = Q \begin{bmatrix} R \\ 0 \end{bmatrix} \Pi$$

where Q is a unitary matrix, Π is a permutation matrix, and $R \in \mathbb{R}^{r \times n_a}$ is full row rank. Now, we try to find the optimal value of ζ_c in the attack vector for fully constant and alternating attacks with respect to our objective function.

Theorem 3.1. *Assume that* $\ker(N) \subseteq \ker(M)$. *Let* $M' = Q^T M Q$, $N' = RR^T$, *and define the partition:*

$$M' = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

where $M_{11} \in \mathbb{R}^{r \times r}$. *Then, the optimal fully constant or alternating steady-state stealthy attack is given by:*

$$\zeta_c = \delta Q \begin{bmatrix} N'^{-\frac{1}{2}} z \\ 0 \end{bmatrix} \tag{3.18}$$

where z *is the normalized eigenvector corresponding to the dominant eigenvalue of the matrix* $N'^{-\frac{1}{2}} M_{11} N'^{-\frac{1}{2}}$.

Proof. From (3.15) and (3.16), the infinite-horizon optimization problem in (3.9) can be written as:

$$\begin{aligned} \max_{\zeta_c} \quad & \zeta_c^T M \zeta_c \\ \text{s.t.} \quad & \zeta_c^T N \zeta_c \leq \delta^2. \end{aligned} \tag{3.19}$$

A problem of this form can typically be solved as an eigenvalue problem simply by making a transformation, $\xi = \sqrt{N}\zeta_c$. However, this requires that $N \in \mathbb{S}^{n_a}_{++}$, which is not true in general, so this transformation may not be bijective. However, note that attacks in $\ker(N)$ do not affect the objective function because $\ker(N) \subseteq \ker(M)$. Thus, we can use the QR factorization to remove this subspace entirely from the optimization problem. Let:

$$\zeta_c = Q \begin{bmatrix} x \\ y \end{bmatrix}$$

where $x \in \mathbb{R}^r$ and $y \in \mathbb{R}^{n_a - r}$. Plugging this in:

$$\sqrt{N}\zeta_c = \Pi \begin{bmatrix} R^T & 0 \end{bmatrix} Q^T Q \begin{bmatrix} x \\ y \end{bmatrix} = \Pi R^T x.$$

We can see that y represents the component of the attack vector in $\ker(N)$. Since we have already established that these attacks are redundant because they will not affect the objective function, we can set $y = 0$ for simplicity. The objective then becomes:

$$J = \begin{bmatrix} x^T & 0 \end{bmatrix} Q^T M Q \begin{bmatrix} x \\ 0 \end{bmatrix}.$$

Then, we can optimize with respect to x only, noting that Π is an orthogonal matrix:

$$\begin{aligned} \max_{x} \quad & x^T M_{11} x \\ \text{s.t.} \quad & x^T N' x \leq \delta^2. \end{aligned}$$

Now, $N' \in \mathbb{S}^n_{++}$ so we can define $z = N'^{\frac{1}{2}} x$ such that the optimization problem becomes:

$$\begin{aligned} \max_{z} \quad & z^T N'^{-\frac{1}{2}} M_{11} N'^{-\frac{1}{2}} z \\ \text{s.t.} \quad & z^T z \leq \delta^2. \end{aligned}$$

The solution of this optimization problem is simply the eigenvector associated with the dominant eigenvalue in the objective matrix scaled such that $\|z\| = \delta$. Then, reverting all transformations, we have the optimal attack in (3.18). ■

Corollary 3.1. *If* $\ker(N) \nsubseteq \ker(M)$*, then there exists a constant or alternating steady-state stealthy attack with* $\zeta_c \in \ker(N) - \ker(M)$ *that can cause arbitrarily large damage to the system.*

Proof. An attack $\zeta_c \in \ker(N) - \ker(M)$ has no effect on the innovation but does affect the objective value. Thus, by the linearity of the system, we can scale ζ_c by an arbitrarily large constant, which would result in an arbitrarily large amount of damage to the system. However, the attack would still have zero impact on the innovation at steady state. ■

3.3.2 Optimality of Constant and Alternating Attacks

In fact, we do not have to consider any other form of the attack in (3.13) other than the ones in Section 3.3.1. The following theorem explains why this is so.

Theorem 3.2. *The optimal combined constant and alternating steady-state attack* (3.13) *is one that is fully constant,* $\zeta_t = \zeta_{ss}$*, or fully alternating,* $\zeta_t = (-1)^t \zeta_{al}$*.*

Proof. Applying the general attack in (3.13), from (3.14), we have at steady state:

$$\lim_{t \to \infty} \Delta e_t = (I - \bar{A})^{-1} \bar{B} \zeta_{ss} + (-1)^{t+1} (I + \bar{A})^{-1} \bar{B} \zeta_{al}.$$

Therefore, the objective function (3.8) becomes:

$$J = \zeta_{ss}^T M_+ \zeta_{ss} + \zeta_{al}^T M_- \zeta_{al}$$

where M_+ and M_- are the M matrices in (3.15) for the fully constant and alternating cases, respectively. The cross term is removed because every two time steps, it will cancel itself out due to the sign change of the alternating component. The constraint in (3.10) changes every time step between:

$$\|[\bar{C}(I-\bar{A})^{-1}\bar{B}+\bar{D}]\zeta_{ss} \pm [\bar{C}(I+\bar{A})^{-1}\bar{B}-\bar{D}]\zeta_{al}\|^2 \leq \delta^2.$$

Let $N_+^T N_+$ and $N_-^T N_-$ be the N matrices in (3.17) for the fully constant and alternating cases, respectively. Expanding the norm, this is equivalent to:

$$\zeta_{ss}^T N_+^T N_+ \zeta_{ss} + \zeta_{al}^T N_-^T N_- \zeta_{al} \leq \delta^2 - 2|\zeta_{ss}^T N_+^T N_- \zeta_{al}|. \tag{3.20}$$

Relax the problem by removing the cross terms on the right hand side of the constraint:

$$\zeta_{ss}^T N_+^T N_+ \zeta_{ss} + \zeta_{al}^T N_-^T N_- \zeta_{al} \leq \delta^2.$$

The relaxed optimization problem becomes:

$$\begin{aligned} \max_{\bar{\zeta}_{ss},\bar{\zeta}_{al}} \quad & \zeta_{ss}^T M_+ \zeta_{ss} + \zeta_{al}^T M_- \zeta_{al} \\ \text{s.t.} \quad & \zeta_{ss}^T N_+^T N_+ \zeta_{ss} + \zeta_{al}^T N_-^T N_- \zeta_{al} \leq \delta^2. \end{aligned}$$

Notice that the two optimization variables are completely decoupled in the objection function and the constraint. Suppose we have a candidate optimal solution:

$$\zeta_{ss} = a\zeta_{ss}^* \neq 0,\ \zeta_{al} = b\zeta_{al}^* \neq 0$$

for some $a, b > 0 \in \mathbb{R}$. Then, let $M_+^* = \zeta_{ss}^T M_+ \zeta_{ss}$, $M_-^* = \zeta_{al}^T M_- \zeta_{al}$, $N_+^* = \zeta_{ss}^T N_+^T N_+ \zeta_{ss}$, and $N_-^* = \zeta_{al}^T N_-^T N_- \zeta_{al}$. Then, the objective and constraint are:

$$J = a^2 M_+^* + b^2 M_-^*$$

$$a^2 N_+^* + b^2 N_-^* \leq \delta^2.$$

Since all parameters in the equation above are positive, we can always obtain a more or equally optimal solution by decreasing a and increasing b if $M_+^* \leq M_-^*$

or vice versa if $M_+^* \geq M_-^*$. Therefore, the optimal attack will always have $a = 0$ or $b = 0$, corresponding to a fully alternating and full constant attack, respectively.

Since one of the decision variables is 0, the original constraint in (3.20) is always satisfied as well since $\zeta_{ss}^T N_+^T N_- \zeta_{al} = 0$. Since this is the optimal solution of a more relaxed problem, it is also the solution of the original one. ■

3.4 Low-Dimensional Sinusoidal Attacks

3.4.1 Equivalent Attack Model

Note that in the stealthiness constraint introduced in (3.3) and (3.4), the attack vector, ζ_t, is constrained to an offset hyper-ellipsoid in general. The following lemma makes a transformation to introduce a new equivalent attack vector, $\bar{\zeta}_t$, that is instead constrained to a centered hyper-ellipsoid with a time-varying size. Let $P = \bar{D}\bar{D}^+$, the orthogonal projection matrix onto $\text{im}(\bar{D})$.

Lemma 3.1. *An equivalent system to* (3.2) *and* (3.3) *with stealthiness constraint* (3.4) *is:*

$$\begin{aligned} \Delta e_{t+1} &= (\bar{A} - \bar{B}\bar{D}^+\bar{C})\Delta e_t + \bar{B}\bar{\zeta}_t \\ \|\bar{D}\bar{\zeta}_t\|^2 &\leq \delta^2 - \|(I-P)\bar{C}\Delta e_t\|^2 \end{aligned} \tag{3.21}$$

where:

$$\zeta_t = \bar{\zeta}_t - \bar{D}^+\bar{C}\Delta e_t. \tag{3.22}$$

Proof. From (3.3) with stealthiness constraint (3.4), we remove all of the components of $CA\Delta e_t$ in $\text{im}(\bar{D})$ by shifting the attack vector as in (3.22). Then, splitting $CA\Delta e_t$ into the components within and orthogonal to $\text{im}(\bar{D})$, the innovation constraint becomes:

$$\|P\bar{C}\Delta e_t + (I-P)\bar{C}\Delta e_t + \bar{D}\bar{\zeta}_t - \bar{D}\bar{D}^+\bar{C}\Delta e_t\| \leq \delta$$

which, due to orthogonality, simplifies to:

$$\|(I-P)\bar{C}\Delta e_t\|^2 + \|\bar{D}\bar{\zeta}_t\|^2 \leq \delta^2$$

and thus we obtain the system in (3.21). ■

For the design of sinusoidal attacks, we need to make the key technical assumption that $\bar{D}$ has full row rank. Note that $\bar{D}$ must then be square or the system becomes strictly vulnerable [5]. In this case, the constraint becomes a hyper-sphere of constant radius δ, which means we can freely use spherical coordinates of a fixed, centered $(n_a - 1)$-sphere.

Corollary 3.2. *If $\bar{D} \in \mathbb{R}^{m \times m}$ is invertible, then an equivalent system to* (3.21) *is:*

$$\begin{aligned} \Delta e_{t+1} &= (\bar{A} - \bar{B}\bar{D}^{-1}\bar{C})\Delta e_t + \bar{B}\bar{D}^{-1}\zeta_t' \\ \|\zeta_t'\|^2 &\leq \delta^2 \end{aligned} \tag{3.23}$$

where:

$$\zeta_t = \bar{D}^{-1}\zeta_t' - \bar{D}^{-1}\bar{C}\Delta e_t. \tag{3.24}$$

Proof. If $\bar{D}$ is invertible, then $P = I$. The result is then a direct outcome of defining $\zeta' = \bar{D}\bar{\zeta}$. ■

Remark 3.4. One of the simplest ways for the attacker to achieve an invertible $\bar{D}$ is by compromising the entire signal in the output channel such that $\bar{D} = \Gamma^a = I$.

To simplify the design of sinusoidal attacks in the frequency domain, we create a new system that represents $\zeta_t' \to \sqrt{Q_0}\Delta e_t$. Then, the norm squared of this system's output is directly the average cost of the objective in (3.8). This system has parameters denoted with $\check{\cdot}$ as follows:

$$\check{A} = \bar{A} - \bar{B}\bar{D}^{-1}\bar{C},\ \check{B} = \bar{B}\bar{D}^{-1},\ \check{C} = \sqrt{Q_0}. \tag{3.25}$$

As a further technical condition, we require $\check{A}$ to be Schur in order for the frequency response to have a valid interpretation. The dimension of the attack refers to the length of the attack vector ζ_t'.

3.4.2 Two-Dimensional Sinusoidal Attacks

If $\zeta_t' \in \mathbb{R}^2$, to fully utilize the constraint such that $\|\zeta_t'\| = \delta$, the sinusoidal attack must satisfy the Pythagorean theorem, taking the form of polar coordinates. There are two distinct cases, depending on whether the first element

lags or leads relative to the second, respectively:

$$\zeta_t' = \delta[\sin(wt); \cos(wt)]$$
$$\zeta_t' = \delta[\cos(wt); \sin(wt)].$$

Remark 3.5. Without loss of generality, we can set the initial phases to zero because we are looking at the steady-state behaviour. Furthermore, we will only consider $w \in [0, \pi]$ for these sinusoidal attacks to avoid aliasing, since the Nyquist frequency is π.

Remark 3.6. Note that if $w = 0$ or $w = \pi$, then we have a constant or alternating attack, respectively. However, there is a discontinuity in the objective value at these limits. The reason is twofold:

1. The mean squared of a sinusoid $A\sin(wt + \phi)$ when $w \to 0, \pi$ is $A^2/2$, but when $w = 0$ and $w = \pi$, the mean squared jumps to A^2.

2. The initial phase of the sinusoids does not matter when $w \to 0, \pi$, but it does when $w = 0, \pi$ as it determines where the attack vector is located on the unit circle.

We can find the objective associated with each frequency, w, by representing each element i of the attack as a phasor with respect to w that has a magnitude δ and a relative phase ϕ_i. This allows us to compute the effect of interference on the output side. If we set $\phi_1 = 0$ as our reference, then $\phi_2 = \pm\frac{\pi}{2}$. Passing the phasors into system (3.25) scales the amplitude and shifts the phase. The output phasors are then added together elementwise. The magnitudes in this final phasor vector represent the amplitudes of the output sinusoids at steady state, so the average steady-state step-by-step cost of the attack is given by half of the norm squared of the output vector amplitudes.

Denote the transfer function of (3.25) as a function of frequency as $T(w)$. The i-th column and element on the i-th row and j-th column of T are denoted t_i and t_{ij}, respectively. We can then characterize all 2D attacks using Algorithm 2. For the discretization of the frequencies, we can use a logarithmic-

based discretization scheme in order to gain a higher resolution for smaller frequencies.

Algorithm 2: Characterization of 2D Sinusoidal Attacks.

begin
- $W \leftarrow$ Discretization of frequencies on the interval $(0, \pi)$;
- Compute the discrete-time frequency domain response or Bode plot, $T(w)$, over W;
- **for** $w \in W$ **do**
  - **for** $i = 1, 2$ **do**
    - $j \leftarrow (i \bmod 2) + 1$;
    - $o \leftarrow \delta t_i(w)$;
    - $o \leftarrow o + \sqrt{-1}\delta t_j(w)$;
    - $J_w^{(i)} \leftarrow \|o\|^2/2$;
  - $J_w \leftarrow \max_i J_w^{(i)}$;
- $w^*, J^* \leftarrow \arg\max_w J_w, \max_w J_w$;

Remark 3.7. For any dimensional attack, the edge cases for which $w = 0, \pi$ correspond to constant and alternating attacks. This gives us a simpler, alternative method for finding the optimal fully constant and alternating attacks when the technical conditions for sinusoidal attacks are met. For a constant attack, associated with a frequency of zero, let ζ'_{ss} be the attack vector. The objective value is

$$J_0 = (\zeta'_{ss})^T T^T(0) T(0) \zeta'_{ss}.$$

We then have the following optimization problem:

$$\begin{aligned} \max_{\zeta'_{ss}} \quad & (\zeta'_{ss})^T T^T(0) T(0) \zeta'_{ss} \\ \text{s.t.} \quad & (\zeta'_{ss})^T \zeta'_{ss} \leq \delta^2. \end{aligned}$$

This is an eigenvalue problem once again, so the optimal attack is the eigenvector associated with the dominant eigenvalue of $T^T(0)T(0)$, λ_0^*, scaled to have a norm of δ. The associated cost is $\delta^2 \lambda_0^*$. Similarly, for $w = \pi$, which represents a fully alternating attack, the optimal attack is the eigenvector associated with the dominant eigenvalue of $T^T(\pi)T(\pi)$, λ_π^*, scaled to have a norm of δ. The associated cost is $\delta^2 \lambda_\pi^*$.

3.4.3 Standard Three-Dimensional Sinusoidal Attacks

In the three-dimensional case, the sinusoidal attack takes the form of spherical coordinates. Without loss of generality, there are six cases in total:

$$\zeta'_t = \delta[\sin(w_1 t); \cos(w_1 t)\sin(w_2 t); \cos(w_1 t)\cos(w_2 t)]. \tag{3.26}$$

and the same attack vectors with their elements permuted. We first consider when $w_1, w_2 \in (0, \pi)$. For attack elements with the product of two sinusoids, the product-to-sum trigonometric identities can be used to separate the attack into 3 distinct frequencies: w_1, $w_1 + w_2$, and $w_1 - w_2$. Conveniently, sinusoids of different frequencies are mutually orthogonal, so in most cases, we can study each frequency involved independently for convenience.

We can rewrite all elements of the attack vector (3.26) in terms of sine to find the relative phase between them:

$$\zeta'_t = \delta \begin{bmatrix} \sin(w_1 t) \\ \frac{1}{2}\left[\sin((w_1 + w_2)t) + \sin((w_1 - w_2)t + \pi)\right] \\ \frac{1}{2}\left[\sin((w_1 + w_2)t + \frac{\pi}{2}) + \sin((w_1 - w_2)t + \frac{\pi}{2})\right] \end{bmatrix}.$$

The main difference compared to the two-dimensional attacks is that we have to consider 3 different phasors at each output, associated with the three different frequencies, and then add their associated objective values together. Furthermore, we have to handle two special cases.

The first case is when one of the frequencies is outside of the range $[0, \pi]$, it may become equivalent to another one. The reason for this is that in discrete time, given some frequency $w \in [0, \pi]$, the following frequencies are equivalent: w, $-w$, and $2\pi - w$, perhaps with a phase shift of π. When this occurs, we must merge the outputs of the same frequency together before taking the norm as they are no longer orthogonal. This is only possible for two cases:

- For $w_1 + w_2$ to be equivalent to w_1, we require $w_1 + w_2 = 2\pi - w_1$ such that $2w_1 + w_2 = 2\pi$.
- For $w_1 - w_2$ to be equivalent to w_1, we require $w_1 - w_2 = -w_1$ such that $2w_1 - w_2 = 0$.

- It is impossible for $w_1 + w_2$ to be equivalent to $w_1 - w_2$ because this would either require $w_1 - w_2 = -(w_1 + w_2)$ or $w_1 + w_2 = 2\pi - (w_1 - w_2)$ such that $w_1 = 0$ or $w_1 = \pi$.

The second special case is if $w_1 + w_2 = \pi$ and $w_1 = w_2$. Then, one of the three frequencies becomes constant or alternating. In this case, we now need to double the objective value that this signal contributes and also find the optimal initial phase. There are two degrees of freedom for initial phase: ϕ_1 and ϕ_2, associated with sinusoids of frequencies w_1 and w_2, respectively. From the product-to-sum identities, the initial phases of sinusoids associated with sinusoids of frequencies $w_1 + w_2$ and $w_1 - w_2$ are then $\phi_1 + \phi_2$ and $\phi_1 - \phi_2$, respectively. Since these phases are linearly independent with respect to ϕ_1 and ϕ_2, we can arbitrarily choose the position of these constant and alternating components on the unit circle. Say the indexes of the attack's constant components are α and β, which correspond to the second and third elements of attack vector (3.26), respectively. The output of the constant component is given by $t_\alpha(0)y + t_\beta(0)x$, where (x, y) is a point on the unit circle. Then, we can optimize by solving:

$$\begin{aligned} \max_{y,x} \quad & \begin{bmatrix} y & x \end{bmatrix} \begin{bmatrix} t_\alpha(0)^T t_\alpha(0) & t_\alpha(0)^T t_\beta(0) \\ t_\beta(0)^T t_\alpha(0) & t_\beta(0)^T t_\beta(0) \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} \\ \text{s.t.} \quad & y^2 + x^2 \leq \delta^2. \end{aligned} \tag{3.27}$$

The solution is the eigenvector corresponding to the dominant eigenvalue of the quadratic weighting matrix above. Then, the initial phases have to satisfy $\phi_1 - \phi_2 = \text{arctan2}(-y, x)$. The same can be done for an alternating component by replacing all frequencies arguments with π. Furthermore, the initial phase condition changes to $\phi_1 + \phi_2 = \text{arctan2}(y, x)$. The characterization of 3D steady-state sinusoidal attacks in the interior of our search space can be computed using Algorithm 3. When $w_1, w_2 = 0, \pi$, we also have discontinuities that are handled separately below.

Algorithm 3: Characterization of 3D Sinusoidal Attacks.

begin

$W_1, W_2 \leftarrow$ Discretization of the frequencies on the interval $(0, \pi)$;

Compute the discrete-time frequency domain response or Bode plot, $T(w)$, over $W = W_1 \oplus W_2 \cup W_1 \oplus (-W_2)$;

for $w_1 \in W_1$, $w_2 \in W_2$, *and each permutation* p *of* (3.26), *with indexes of the elements in order being* i, j, *and* k **do**

$o^{(w_1)} \leftarrow \delta t_i(w_1)$;

if $w_1 = w_2$ **then**

`// A constant attack component exists`

Solve problem (3.27), with dominant eigenvalue λ_0^*;

$o^{(-)} \leftarrow \sqrt{2\lambda_0^*}/2$;

else

$o_1^{(-)} \leftarrow -\delta t_j(w_1 - w_2)/2$;

$o_2^{(-)} \leftarrow \sqrt{-1}\delta t_k(w_1 - w_2)/2$;

if $2w_1 - w_2 = 0$ **then**

`// Frequencies` $w_1 - w_2$ `and` w_1 `are equivalent`

$o^{(w_1)} \leftarrow o^{(w_1)} - o_1^{(-)} + o_2^{(-)}$;

$o^{(-)} \leftarrow 0$;

else

$o^{(-)} \leftarrow o_1^{(-)} + o_2^{(-)}$;

if $w_1 + w_2 = \pi$ **then**

`// An alternating attack component exists`

Solve problem (3.27), with dominant eigenvalue λ_π^*;

$o^{(+)} \leftarrow \sqrt{2\lambda_\pi^*}/2$;

else

$o_1^{(+)} \leftarrow \delta t_j(w_1 + w_2)/2$;

$o_2^{(+)} \leftarrow \sqrt{-1}\delta t_k(w_1 + w_2)/2$;

if $2w_1 + w_2 = 2\pi$ **then**

`// Frequencies` $w_1 + w_2$ `and` w_1 `are equivalent`

$o^{(w_1)} \leftarrow o^{(w_1)} - o_1^{(+)} + o_2^{(+)}$;

$o^{(+)} \leftarrow 0$;

else

$o^{(+)} \leftarrow o_1^{(+)} + o_2^{(+)}$;

$J_{w_1,w_2}^{(p)} \leftarrow (\|o^{(w_1)}\|^2 + \|o^{(+)}\|^2 + \|o^{(-)}\|^2)/2$;

$w_1^*, w_2^*, p^*, J^* \leftarrow \arg\max_{w_1,w_2,p} J_{w_1,w_2}^{(p)}, \max_{w_1,w_2,p} J_{w_1,w_2}^{(p)}$;

3.4.4 3D Sinusoidal Attacks with $w_1 = 0, \pi$

If $w_1 = 0$, we have the following attack vectors:

$$\zeta_t' = \delta[a; b\sin(w_2 t); b\cos(w_2 t)] \tag{3.28}$$

where $a, b \in \mathbb{R}$ such that $a^2 + b^2 = 1$, and the same attack vectors with their elements permuted.

Lemma 3.2. *For a fixed w_2 in (3.28), the optimal cost is achieved with either $a = 0$ or $b = 0$.*

Proof. Let t_a, t_{bs}, and t_{bc} be the columns of T associated with the position of the elements with a, $b\sin(w_2 t)$, and $b\cos(w_2 t)$ in the attack vector in (3.28), respectively. Then, our objective function is:

$$J = \delta^2 \left[a^2 \left\| t_a(0) \right\|^2 + \frac{b^2}{2} \left\| t_{bs}(w_2) + t_{bc}(w_2) e^{\sqrt{-1}\frac{\pi}{2}} \right\|^2 \right].$$

Since the terms with a and b are completely decoupled in the objective function and this is a positive semi-definite function with respect to a and b, we select $a = \pm 1$ if $\|t_1(0)\| \geq \frac{1}{2} \left\| t_{bs}(w_2) + t_{bc}(w_2) e^{\sqrt{-1}\frac{\pi}{2}} \right\|$ and $b = \pm 1$ otherwise for the maximum. ■

Remark 3.8. Note that $a = \pm 1$ is a subset of the constant attack. As for $b = \pm 1$, this is effectively a two-dimensional sinusoidal attack, leaving out one of the elements in the attack vector. Thus, the problem can be directly converted to the 2D equivalent by removing the element in ζ_t' and the column of the $\bar{B}\bar{D}^{-1}$ matrix in (3.23) associated with the position of a.

If $w_1 = \pi$ instead, then the analysis is equivalent to $w_1 = 0$. The attack vector in (3.28) becomes:

$$z_t' = \delta[(-1)^t a; -b\sin((\pi - w_2)t); b\cos((\pi - w_2)t)].$$

We can show that $a = 0$ or $b = 0$ is optimal in a similar manner. The only difference in this case is that the sine element leads the cosine element by $\pi/2$ rather than lagging. However, this is already addressed by swapping the sine

and cosine terms in the attack vector of (3.28) anyways. Since we study all frequencies $w_2 \in (0, \pi)$, then we are effectively surveying the same attacks and frequencies. Thus, we can ignore the case when $w_1 = \pi$.

3.4.5 3D Sinusoidal Attacks with $w_2 = 0$

If $w_2 = 0$, we have the attack vectors:

$$\zeta_t' = \delta[\sin(w_1 t); c\cos(w_1 t); d\cos(w_1 t)] \tag{3.29a}$$

$$\zeta_t' = \delta[\cos(w_1 t); c\sin(w_1 t); d\sin(w_1 t)] \tag{3.29b}$$

where $c, d \in \mathbb{R}$ such that $c^2 + d^2 = 1$, and the same attack vectors with their elements permuted.

In this case, we have to consider the initial phase of the element with amplitude 1 in (3.29) since all elements have the same frequency; it will therefore be involved in interference at the output. That is why we also include the case in which we have cosine for this case only. Algorithm 4 can be used to find the optimal attack when $w_2 = 0$. The algorithm makes use of the following lemma.

Let t_1, t_c, and t_d be the columns of T associated with the position of the elements with amplitude 1, c, and d in the attack vector (3.29), respectively, and $\phi_0 = \{\pm\frac{\pi}{2}\}$ be the relative phase of the elements with amplitude c and d relative to the one with amplitude 1.

Lemma 3.3. *Let $x = [c; d]$. The optimal values of c and d in (3.29) can be obtained by solving the semidefinite programming problem:*

$$\begin{aligned} \max_{X,x} \quad & \mathrm{Tr}(XQ) + fx + \gamma \\ s.t. \quad & \mathrm{Tr}(X) \leq 1 \\ & \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0 \end{aligned} \tag{3.30}$$

where

$$Q = \frac{\delta^2}{2}\begin{bmatrix} t_c^*(w_1)t_c(w_1) & t_c^*(w_1)t_d(w_1) \\ t_d^*(w_1)t_c(w_1) & t_d^*(w_1)t_d(w_1) \end{bmatrix}$$
$$f = \delta^2 \operatorname{Re}\left(e^{\sqrt{-1}\phi_0} t_1^*(w_1)\begin{bmatrix} t_c(w_1) & t_d((w_1)) \end{bmatrix}\right)$$
$$\gamma = \frac{\delta^2}{2} t_1^*(w_1)t_1(w_1).$$

Proof. The cost from finding the output as a phasor of a fixed frequency w_2 would be:

$$J = \frac{\delta^2}{2}\left\| t_1(w_1) + e^{\sqrt{-1}\phi_0}\begin{bmatrix} t_c(w_1) & t_d((w_1)) \end{bmatrix}\begin{bmatrix} c \\ d \end{bmatrix}\right\|^2.$$

By expanding the norm, we can formulate the following QCQP to optimize this objective value with respect c and d:

$$\begin{aligned} \max_x \quad & x^T Q x + f x + \gamma \\ \text{s.t.} \quad & x^T x \leq 1. \end{aligned} \tag{3.31}$$

Since this is a QCQP with one constraint, we can solve this with the semidefinite programming relaxation in (3.30) quickly and precisely [2]. ■

3.4.6 3D Sinusoidal Attacks with $w_2 = \pi$

If $w_2 = \pi$, we only have three unique attack vectors:

$$\zeta_t' = \delta[\sin(w_1 t); (-1)^t c\cos(w_1 t); (-1)^t d\cos(w_1 t)] \tag{3.32}$$

and the same attack vectors with the sine element permuted. This is equivalent to

$$\zeta_t' = \delta[\sin(w_1 t); c\cos(w_1' t); d\cos(w_1' t)]$$

where $w_1' = \pi - w_1$. We no longer have to consider the case in which the sine and cosine elements are swapped in each attack vector because this would lead to the same relative phase between elements with the same frequency. Considering that $c, d \in \mathbb{R}$, this greatly simplifies the optimization problem in (3.31) to have:

$$\begin{aligned} Q &= \frac{\delta^2}{2}\begin{bmatrix} t_c^*(w_1')t_c(w_1') & t_c^*(w_1')t_d(w_1') \\ t_d^*(w_1')t_c(w_1') & t_d^*(w_1')t_d(w_1') \end{bmatrix} \\ f &= 0 \end{aligned} \tag{3.33}$$

Algorithm 4: Characterization of 3D Steady State Sinusoidal Attacks when $w_2 = 0$.

begin

$W_1 \leftarrow$ Discretization of the frequencies on the interval $(0, \pi)$;

Compute the discrete-time frequency domain response or Bode plot, $T(w)$, over W_1;

for $w_1 \in W_1$, $\phi \in \{\pm\pi/2\}$, *and each valid permutation* p *of* (3.29)*, with indexes of the elements in order being* i*,* j*, and* k **do**

$o_1 \leftarrow t_i(w_1)\delta$;

$o_c \leftarrow t_j(w_1)\delta e^{\sqrt{-1}\phi}$;

$o_d \leftarrow t_k(w_1)\delta e^{\sqrt{-1}\phi}$;

$o_{cd} \leftarrow [o_c, o_d]$;

$Q \leftarrow o_{cd}^* o_{cd}/2$;

$f \leftarrow \mathrm{Re}(o_1^* o_{cd})$;

$\gamma \leftarrow o_1^* o_1/2$;

Solve optimization problem (3.31) and store optimal objective value in $J_{w_1}^{(p,\phi)}$ and optimal arguments c and d in $c_{w_1}^{(p,\phi)}$ and $d_{w_1}^{(p,\phi)}$, respectively.;

$w_1^*, p^*, \phi^*, J^* \leftarrow \arg\max_{w_1,p,\phi} J_{w_1}^{(p,\phi)}, \max_{w_1,p,\phi} J_{w_1}^{(p,\phi)}$;

$c^*, d^* \leftarrow c_{w_1^*}^{(p^*,\phi^*)}, d_{w_1^*}^{(p^*,\phi^*)}$;

which can be solved as an eigenvalue problem because there is no linear term. Then, $[c; d]$ is the eigenvector corresponding to the largest eigenvalue of Q scaled to have a norm of 1. Algorithm 5 can be used to find the optimal attack when $w_2 = \pi$.

Algorithm 5: Characterization of 3D Steady State Sinusoidal Attacks when $w_2 = \pi$.

begin

$W_1 \leftarrow$ Discretization of the frequencies on the interval $(0, \pi)$;

$W_1' \leftarrow \{\pi - w_1 \mid w_1 \in W_1\}$;

Compute the discrete-time frequency domain response or Bode plot, $T(w)$, over $W_1 \cup W_1'$;

for $w_1 \in W_1$ *and each valid permutation* p *of* (3.32)*, with indexes of the elements in order being* i*,* j*, and* k **do**

$o_{cd} \leftarrow [t_j(w_1'), t_k(w_1')]$;

$Q \leftarrow o_{cd}^* o_{cd}$;

$\gamma \leftarrow t_i^*(w_1) t_i(w_1)$;

$\{\lambda_{max}, v_{max}\} \leftarrow$ dominant eigenvalue and corresponding eigenvector of Q;

$[c_{w_1}^{(p)}, d_{w_1}^{(p)}]^T \leftarrow v_{max}$;

$J_{w_1}^{(p)} \leftarrow \delta^2 \left[\lambda_{max} + \gamma\right]/2$;

$w_1^*, p^*, J^* \leftarrow \arg\max_{w_1, p} J_{w_1}^{(p)}, \max_{w_1, p} J_{w_1}^{(p)}$;

$c^*, d^* \leftarrow c_{w_1^*}^{(p^*)}, d_{w_1^*}^{(p^*)}$;

3.5 Comparison of Attack Strategies

As shown in Section 3.4, the design of sinusoidal attacks is quite an involved process. Thus, it is desirable to find a fast and easy way to check if constant or alternating attacks are optimal so that we can avoid performing unnecessary analysis. Here, we provide a condition for pure sinusoidal attacks to be sub-optimal.

Theorem 3.3. *Let* $\bar{t}_{ij}$ *be the largest magnitude value of* $t_{ij}(w)$ *over all* $w \in [0, \pi]$*.* $J_{\sin}$*, the average step-by-step cost from sinusoidal attacks without a*

constant or alternating component, is upper bounded by:

$$J_{\sin} \leq \frac{\delta^2}{2} \sum_{i=1}^{n} \sum_{j=1}^{n_a} \sum_{k=1}^{n_a} \bar{t}_{ij} \bar{t}_{ik}. \tag{3.34}$$

Proof. Without loss of generality, the sinusoidal attack along each element of the attack vector is taken from the set of spherical coordinates from a $(n_a - 1)$-sphere of radius δ. That is, the set:

$$\{\delta \cos(\theta_1), \delta \sin(\theta_1) \cos(\theta_2), \delta \sin(\theta_1) \cdots \cos(\theta_{n_a-1}), \delta \sin(\theta_1) \cdots \sin(\theta_{n_a-1})\}$$

where $\theta_i = w_i t$. From the product-to-sum trigonometric identities, the product of n sinusoidal functions can be written as a linear combination of 2^n sinusoidal functions, each with magnitude 2^{-n} and a frequency in the set of values that can be represented with $\eta_1 w_1 + \eta_2 w_2 + \cdots + \eta_n w_n$, where $\eta_i \in \{1, -1\}$. Let $\mathcal{W}$ be the set of all frequencies that are included in the attack from the any element of the attack vector. Then, the cost from a single output element, i, is:

$$J_{\sin,i} = \frac{1}{2} \sum_{w \in \mathcal{W}} \left| [t_{i1}(w), t_{i2}(w), \ldots, t_{in_a}(w)] \Phi_0^{(w)} v^{(w)} \right|^2$$

where $v^{(w)}$ is a vector that represents the magnitude of the sinusoid with frequency w for each element of the attack and $\Phi_0^{(w)}$ is a diagonal matrix with elements that contain the relative phase between them. To meet the stealthiness condition, the sum of all $v^{(w)}$ is a vector of δ. Because we are looking for an upper bound for the sinusoidal attacks, we assume all interference is perfectly constructive to maximize the output amplitude. This can be represented as:

$$J_{\sin,i} \leq \frac{1}{2} \sum_{w \in \mathcal{W}} \left| [|t_{i1}(w)|, |t_{i2}(w)|, \ldots, |t_{in_a}(w)|] v^{(w)} \right|^2.$$

For further simplicity, we remove the dependency on w by using the maximum magnitude of each element of the transfer function:

$$J_{\sin,i} \leq \frac{1}{2} \sum_{w \in \mathcal{W}} \left([\bar{t}_{i1}, \bar{t}_{i2}, \ldots, \bar{t}_{in_a}] v^{(w)} \right)^2.$$

By expanding the vector multiplication and the square, this simplifies to:

$$J_{\sin,i} \leq \frac{1}{2} \sum_{w \in \mathcal{W}} \sum_{j=1}^{n_a} \sum_{k=1}^{n_a} \bar{t}_{ij} \bar{t}_{ik} v_j^{(w)} v_k^{(w)}.$$

Since $v_j^{(w)} v_k^{(w)} \leq \delta v_j^{(w)}$ and $\sum_{w \in \mathcal{W}} v_j^{(w)} = \delta$, then:

$$J_{\sin,i} \leq \frac{\delta^2}{2} \sum_{j=1}^{n_a} \sum_{k=1}^{n_a} \bar{t}_{ij} \bar{t}_{ik}.$$

Summing up all of the contributions from all n outputs, we have (3.34). ■

Corollary 3.3. *Let* $\lambda^* = \max(\rho[T^T(0)T(0)], \rho[T^T(\pi)T(\pi)])$. *Then, a sufficient condition for the constant or alternating attack to be optimal over any two-dimensional sinusoidal attack is:*

$$\lambda^* \geq \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n_a} \sum_{k=1}^{n_a} \bar{t}_{ij} \bar{t}_{ik}.$$

It is also a sufficient condition for constant or alternating attacks to be optimal over any three-dimensional sinusoidal attacks except those that satisfy $w_1 = w_2$ *or* $w_1 + w_2 = \pi$.

Proof. From Remark 3.7, the optimal cost of a constant or alternating attack is $\delta^2 \lambda^*$. Two-dimensional sinusoidal attacks may not have any constant or alternating components whereas three-dimensional attack only have constant or alternating components if $w_1 = w_2$ or $w_1 + w_2 = \pi$. The result then directly follows from Theorem 3.3. ■

3.6 Comparison with Stochastic Attacks

In this section, we will provide a framework in which deterministic attacks may be compared to stochastic attacks in order to compare their impact on the system. From a stochastic attack perspective, the work that is most analogous to this chapter is [11], which utilized a norm bound on the KLD from the original innovation to the compromised innovation at each time step as their stealthiness condition. Although Guo *et al.* studied attacks on remote state

estimation using a Kalman filter rather than closed-loop control, our attack methods can still be used as explained in Remark 2.3.

To facilitate the comparison for this section, we match the assumption in [11] that the process and measurement noises follow zero mean Gaussian distributions: $w_k \sim \mathcal{N}(0, Q)$ and $v_k \sim \mathcal{N}(0, R)$, where $Q \in \mathbb{S}_+^n$ and $R \in \mathbb{S}_{++}^m$. Furthermore, we assume that the state estimator uses a Kalman filter that has reached steady state. Then, we can make a connection between the stealthiness condition in (3.4) and the KLD:

Proposition 3.2. *The KLD from the distribution of the original innovation to the compromised innovation at each time step of a stealthy attack is upper bounded by:*

$$D_{KL}(z_t' \| z_k) \leq \frac{\lambda_{max}\delta^2}{2} \tag{3.35}$$

where λ_{max} is the dominant eigenvalue of $(CPC^T + R)^{-1}$.

Proof. It can be shown that the KLD from one k-variate normal distribution, $\mathcal{N}_1(\mu_1, \Sigma_1)$, to another, $\mathcal{N}_0(\mu_0, \Sigma_0)$, is:

$$D_{KL}(\mathcal{N}_0 \| \mathcal{N}_1) = \frac{1}{2}\left[\mathrm{tr}(\Sigma_1^{-1}\Sigma_0) + (\mu_1 - \mu_0)^T\Sigma_1^{-1}(\mu_1 - \mu_0) - k + \ln\left(\frac{|\Sigma_1|}{|\Sigma_0|}\right)\right] \tag{3.36}$$

Given our assumption on the noises, the nominal innovation at each step also follows a zero mean Gaussian distribution: $z_t \sim \mathcal{N}(0, CPC^T + R)$, where P is the steady-state error covariance. During a deterministic attack, the compromised innovation at each time step is simply the nominal one shifted by Δz_t, so it follows the Gaussian distribution: $z_t' \sim \mathcal{N}(\Delta z_t, CPC^T + R)$. Then, applying this to (3.36) and simplifying, we obtain:

$$D_{KL}(z_t' \| z_t) = \frac{1}{2}\Delta z_t^T(CPC^T + R)^{-1}\Delta z_t$$

From the Rayleigh quotient and the stealthiness constraint, since $\Delta z_t^T \Delta z_t \leq \delta^2$, then $\Delta z_t^T(CPC^T + R)^{-1}\Delta z_t \leq \lambda_{max}\delta^2$. ■

This can be considered a tight upper bound in some sense because the bound is achievable.

3.7 Simulation

For simplicity, we will assume that $\delta = 1$ and $Q_0 = I$ for all numerical examples unless otherwise stated. First, we will demonstrate the results of Section 3.3 with an invulnerable system for which the attacker does not have many degrees of freedom. For comparability, we analyze the double integrator system used in [37] and [27]:

$$x_{t+1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} x_t + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_t + w_t$$
$$y_t = x_t + \begin{bmatrix} 1 \\ 0 \end{bmatrix} y_t^a + v_t$$

and estimator gain $K = \begin{bmatrix} 0.6 & 0 \\ -1.4 & 1.6 \end{bmatrix}$. If we compute and apply the fully constant attack, we have:

$$\zeta_t = 1.6,\ \lim_{t\to\infty} \|\Delta e_{t+1}\| = 1.7088,\ J = 2.92.$$

For this system configuration, this is optimal over the fully alternating attack:

$$\zeta_t = 0.2392(-1)^t,\ \lim_{t\to\infty} \|\Delta e_{t+1}\| = 1.0473,\ J = 1.0969.$$

Figs. 3.1 and 3.2 show the objective value and $\|\Delta z_t\|$, respectively, of the attack at each time step. We can see in these figures that the constant attack is in agreement with solving (3.11) using numerical methods as well. For most systems, the attack converges to a steady state very quickly. Because we designed the attacks to be steady-state stealthy, it is guaranteed that $\lim_{t\to\infty} \|\Delta z_t\| = \delta$. However, it is not necessarily a stealthy attack because the stealthiness constraint could be violated during the transient.

We will also use this system to demonstrate the effectiveness of the receding horizon implementation of the solution to optimization problem (2.9), as described in Section 3.2.2. Simulations of the attacks computed using the receding horizon and normal implementations are presented together in Fig. 3.3. At the end of the horizon, the attack that uses the nominal solution causes the Δe_t to move away from the origin. This may be beneficial because

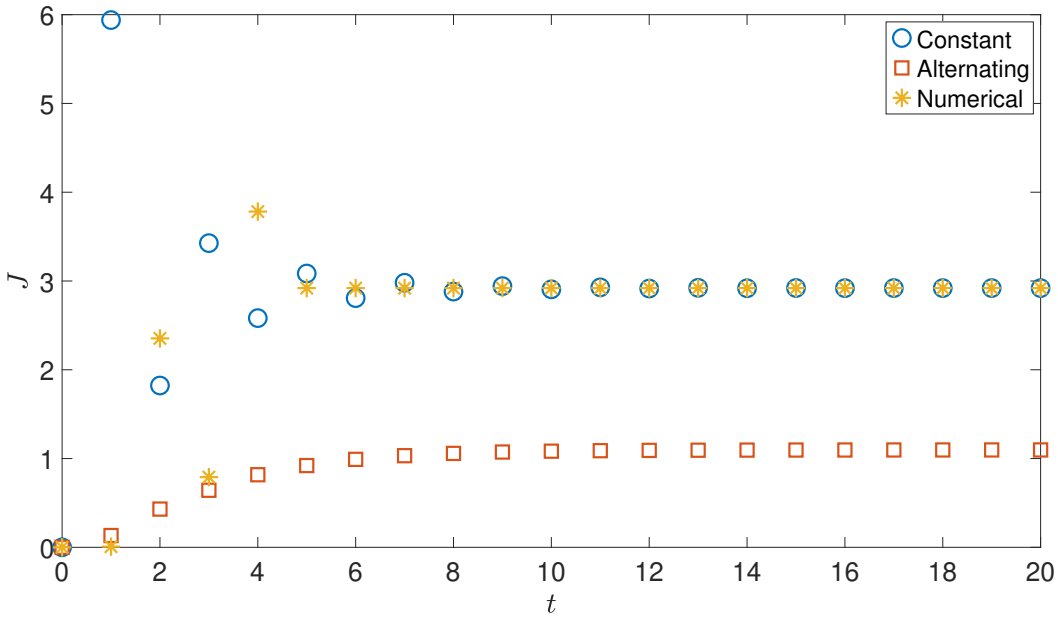

Figure 3.1: Step-by-step objective values of optimal fully constant, fully alternating, and numerical attacks on a double integral system.

the objective value is maximized during these time steps, but if this final state is used as Δe_0 in the next optimization problem, the problem becomes infeasible. In contrast, the receding horizon implementation, which uses $\tau = 40$ and $t_{util} = 20$, avoids this issue by keeping Δe_t constant throughout.

Now, consider a similar system with the same K matrix in which the attacker has more degrees of freedom:

$$x_{t+1} = \begin{bmatrix} 0.6 & 0 \\ 1 & 0.5 \end{bmatrix} x_t + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_t + w_t$$
$$y_t = x_t + y_t^a + v_t$$

such that $y_t^a \in \mathbb{R}^2$. We calculate the upper bound for sinusoidal attacks, $\bar{J}_{\text{sin}}$, in Theorem 3.3 and the constant and alternating attack objective value, λ^*, in Corollary 3.3:

$$\bar{J}_{\text{sin}} = 15.2818, \ \lambda^* = 10.2912.$$

From Corollary 3.3, we can see that there is no guarantee that the constant and alternating attacks are optimal over all sinusoidal attacks. Thus, it

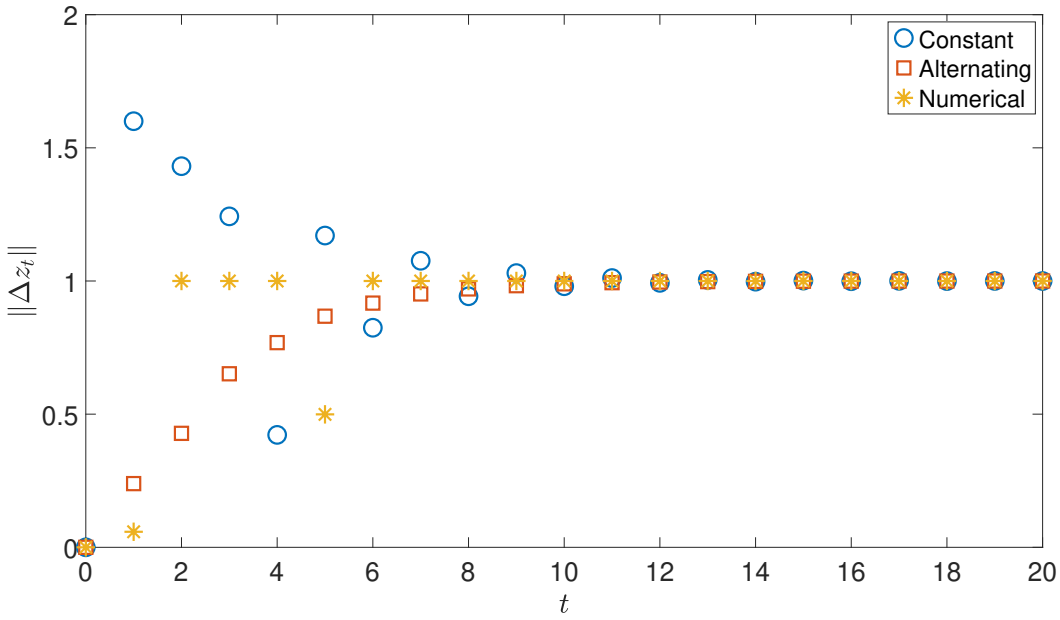

Figure 3.2: Evolution of $\|\Delta z_t\|$ under optimal fully constant, fully alternating, and numerical attacks on a double integral system.

is necessary to perform some analysis of two-dimensional sinusoidal attacks. Applying Algorithm 2, we can compute and plot the impact of sinusoidal attacks as a function of frequency, as shown in Fig. 3.4.

The optimal sinusoidal attack achieves $J = 11.1701$ when the first element leads at $w = 0.3271$, surpassing the constant attack by a sizable margin. In Fig. 3.5, Δe_t is shown for the system under this optimal sinusoidal attack along with the numerically optimal attack obtained by solving (3.11). An initial phase of 0.65π was added to the sinusoidal attack in order to align the two attacks. We can see that although the optimal numerical attack has a complex structure, it is well approximated by the optimal sinusoidal attack. Correspondingly, the difference between the objective values of the two attacks is relatively small. Over these 60 time steps, the costs of the sinusoidal and numerical attacks are 645 and 694, respectively. A significant part of this discrepancy is due to the difference in the transient during the first few time steps.

We now consider a third-order system, which has a sinusoidal attack analy-

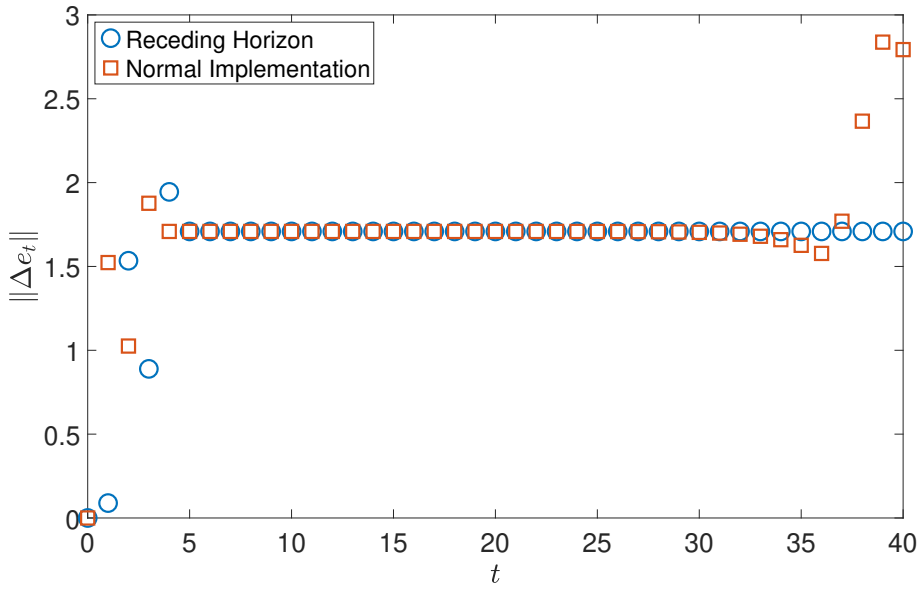

Figure 3.3: Evolution of $\|\Delta e_t\|$ under attacks using the receding horizon and normal implementations.

sis that is substantially more complex than the two-dimensional case. Consider the system:

$$\begin{aligned} x_{t+1} &= \begin{bmatrix} 0.83 & -0.04 & 1.05 \\ 0.26 & 0.60 & 0.33 \\ 0.27 & -1.12 & -0.33 \end{bmatrix} x_t + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u_t + w_t \\ y_t &= \begin{bmatrix} -1.34 & -1.05 & 2.00 \\ 1.02 & 1.36 & -0.85 \\ -1.05 & -1.97 & 0.11 \end{bmatrix} x_t + y_t^a + v_t \end{aligned} \tag{3.37}$$

and an estimator gain of:

$$K = \begin{bmatrix} 1.86 & -1.77 & -1.41 \\ -0.47 & 1.63 & 1.39 \\ 1.71 & 1.25 & -1.04 \end{bmatrix}.$$

We can then characterize all the three-dimensional sinusoidal attacks and the edge cases using Algorithm 3. The results of this characterization are shown in Figs. 3.6 and 3.7. For brevity, note that these plots show the maximum value over all permutations of the attack vector at each given frequency. Furthermore, a comparison between the optimal attacks in each case is provided in Table 3.1. The "format" column shows how the elements of the attack

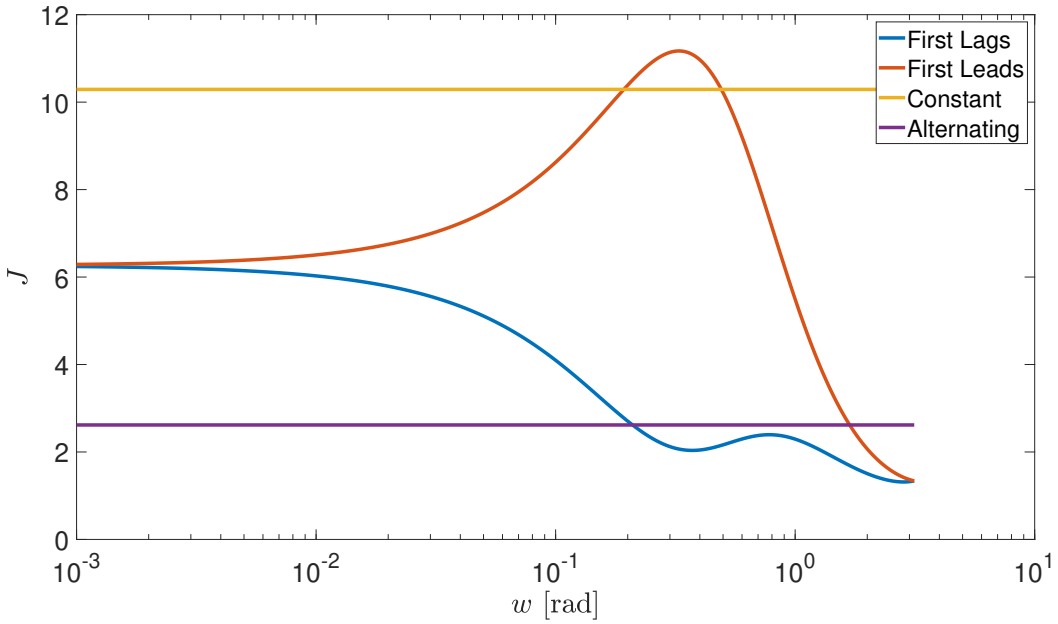

Figure 3.4: 2D sinusoidal attack objective values of varying frequencies compared with constant and alternating attacks.

vector is constructed using the indexes from (3.26), (3.28), (3.29), and (3.32), respectively.

Table 3.1: Properties of optimal sinusoidal attacks on system (3.37).

Attack	Objective	Frequency	Format	Other
Standard	537	$w_1 = 0.5248$ $w_2 = 0.0023$	3,2,1	
$w_1 = 0$	778	$w_2 = 0.5219$	3,2,1	
$w_2 = 0$	1114	$w_1 = 0.5206$	1,2,3 (3.29b)	$c = 0.6983$ $d = 0.7158$
$w_2 = \pi$	558	$w_1 = 2.6282$	3,1,2	$c = -0.7295$ $d = 0.6840$

For this system, the value of sinusoidal attacks is clear as they are capable of performing over four times better than the optimal constant and alternating attacks, which have objective values of only 248 and 34, respectively.

Finally, we examine the second order system parameters and noise covari-

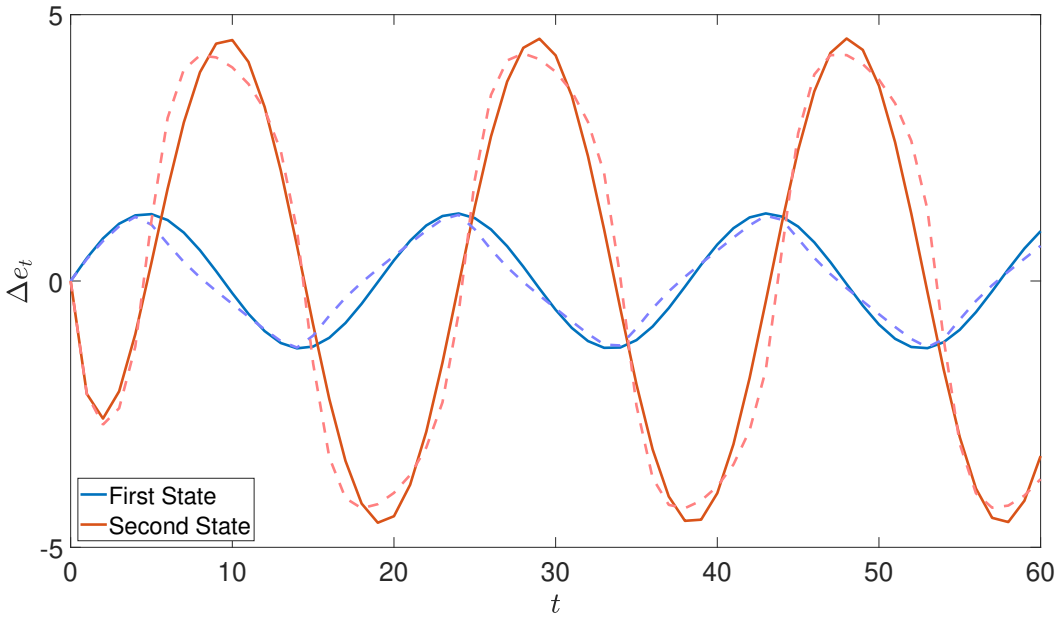

Figure 3.5: Evolution of Δe_t under optimal sinusoidal and numerical attacks, represented by the solid and dashed lines, respectively.

ance matrices from [11], which is used below to compare the performance of the optimal deterministic attacks discussed in this chapter with their stochastic counterparts:

$$A = \begin{bmatrix} 0.7 & 0.2 \\ 0.05 & 0.64 \end{bmatrix} \quad C = \begin{bmatrix} 0.5 & -0.8 \\ 0 & 0.7 \end{bmatrix}$$
$$Q = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.7 \end{bmatrix} \quad R = \begin{bmatrix} 1 & 0 \\ 0 & 0.8 \end{bmatrix}.$$

The Kalman filter then has a steady state gain of:

$$K = \begin{bmatrix} 0.2647 & 0.2064 \\ -0.2650 & 0.4188 \end{bmatrix}.$$

As the stealthiness constraint, we assume that the KLD from nominal to compromised innovation at each time step has to be less than 1. Then, from Proposition 3.2, we can compute that $\delta = 1.3949$. For both types of attacks, we shall use the step-by-step objective function in (3.8). We do not use the estimation error covariance to compare because the compromised covariance is the same as the nominal one for deterministic attacks studied in this thesis. However, note that the second moment with respect to the origin of the estimation error distribution, which changes during a deterministic attack, may

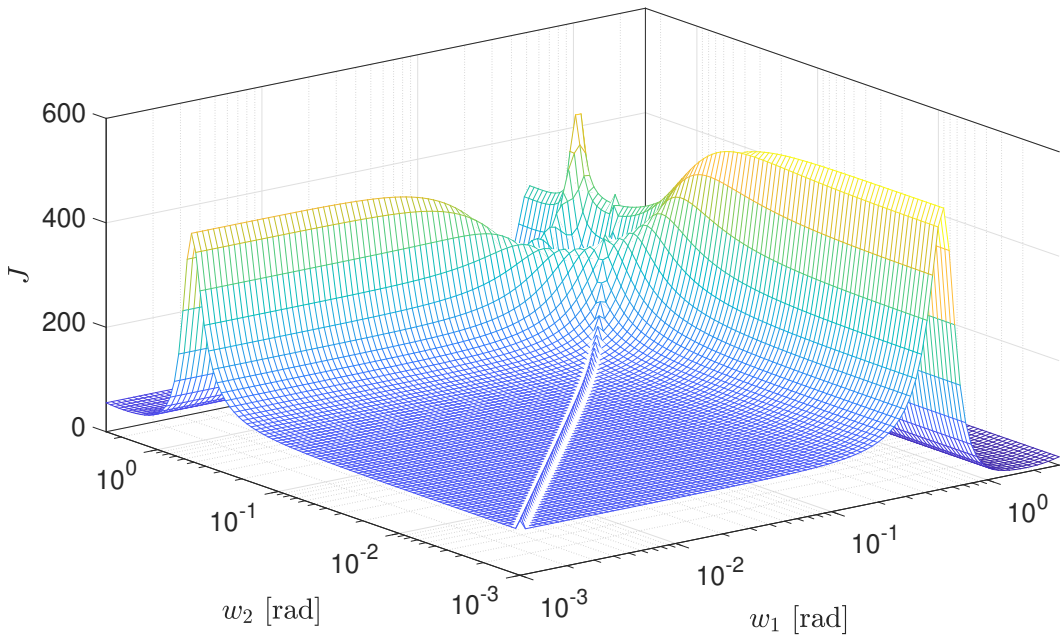

Figure 3.6: Objective values of 3D sinusoidal attacks with frequencies w_1 and w_2.

also be used for comparison to some optimal stochastic attacks [32] but is not considered here.

Applying Theorem 3.3, we can see that the sinusoidal upper bound is 8.4702, which is lower than the optimal objective value for the constant attack of 8.8354. Since this is a two dimensional system, we can then guarantee that the constant attack is better than a sinusoidal attack of any frequency. For reference, the alternating attack only achieves an objective value of 0.1798, so it is disregarded.

We can compare this attack performance with the optimal solution from [11]. For the stochastic attack, we first let the system run for 20 time steps in order to ensure that it reaches a steady state; then we apply the attack starting at the 21st time step. Furthermore, we perform 10,000 Monte Carlo simulations to obtain an average of the objective value at each time step. The objective values over the first 40 time steps of each attack is provided in Fig. 3.8. We can see that the attack performance is comparable, although the stochastic attack is superior by a small margin. This is reasonable because the

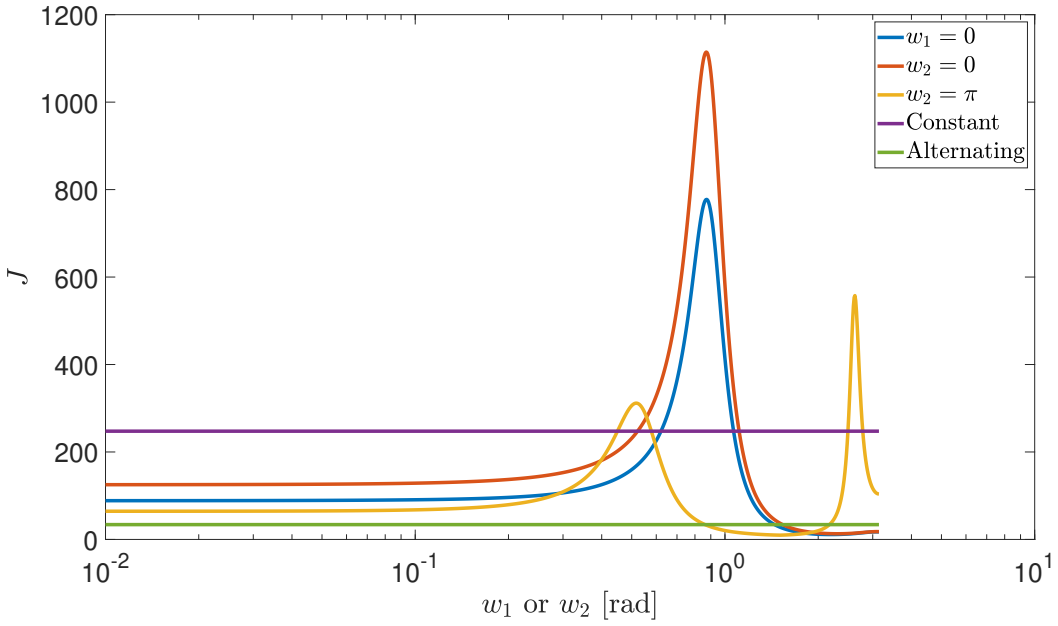

Figure 3.7: Objective values of 3D sinusoidal attacks at the edge cases.

stochastic attack assumes the attacks have a significantly larger information set available than the deterministic one. In other words, the attacker requires real-time knowledge about how the noises affect the system. Although this comparison is informative, it should be noted that it is not definitive. Neither attack can be considered strictly optimal. On one hand, the deterministic attack does not make full use of the stealthiness constraint since we utilize the upper bound in (3.35). On the other hand, the goal of the stochastic attack is to maximize the compromised estimation error. Although they are related, the objective measure presented in this thesis is different, resulting in some suboptimality.

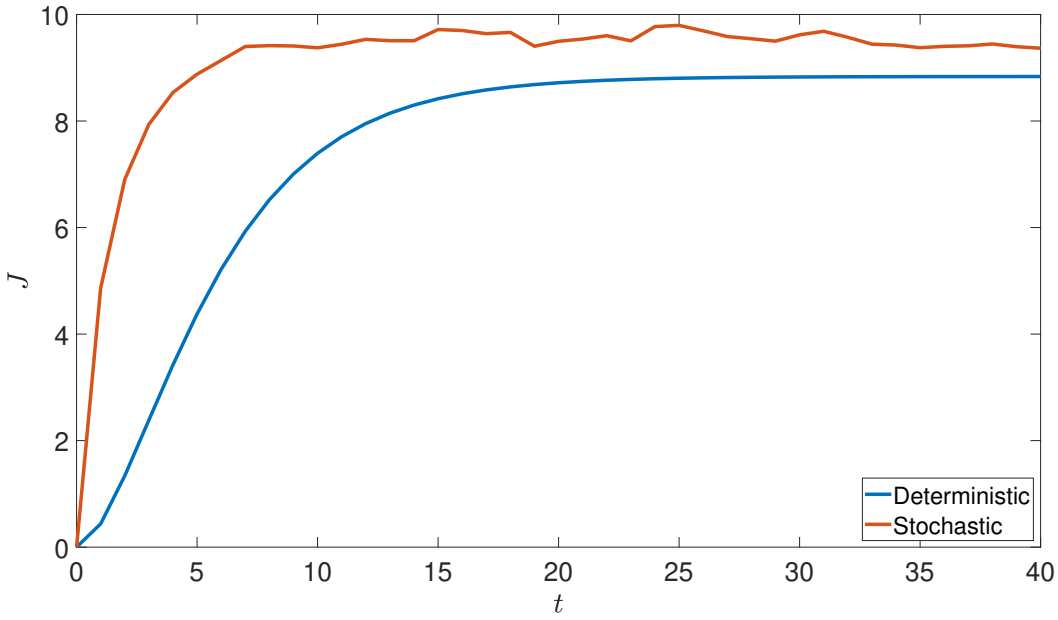

Figure 3.8: Evolution of objective values for optimal deterministic and stochastic attacks.

Chapter 4

Conclusions and Future Work

In this chapter, remarks are provided to conclude this thesis, and then some potential research directions are suggested for future work.

4.1 Conclusions

This thesis studies the design of optimal stealthy attacks on the communication channels of CPSs. The goal of this research is to identify system vulnerabilities and determine the risk that is present. This will facilitate the design of defensive countermeasures and other mitigating actions. The main results of this thesis are summarized as follows:

1. CPSs were studied with respect to their vulnerability to strictly stealthy deterministic attacks. A necessary and sufficient condition was presented for a system and associated attack parameters to permit a strictly stealthy attack of a particular length. It was shown that strict vulnerability in [37] is a special case of τ-step strict vulnerability introduced in this thesis. Then, a method for finding all possible strictly stealthy attacks of a certain length was given. Out of these possibilities, the optimal attack was obtained analytically given an energy constraint and an objective function. Finally, a numerical example was used to illustrate the proposed concepts and attack strategy.

2. We investigated optimal deterministic FDI attacks on CPSs with a re-

laxed stealthiness constraint. Given some basic model parameters, we derived the optimization problem that can be solved to obtain the most effective attack over a finite horizon. For invulnerable systems, these attacks tend to quickly converge to a constant, alternating, or a sinusoidal-like signal. Thus, we determined the optimal constant and alternating attacks analytically. Furthermore, two and three-dimensional sinusoidal attacks were characterized in detail, including a number of edge cases. In order to reduce unnecessary computations, a sufficient condition was obtained for the optimal constant and alternating attack to outperform any two-dimensional and most three-dimensional sinusoidal attacks. Finally, we introduce an approach to directly compare the optimal stealthy deterministic and stochastic attacks.

4.2 Future Work

A few directions for future research on CPS security are provided below:

1. Investigate the exact detection rate of deterministic attacks at each time step with respect to a specific detector. This information would be useful for both attackers and defenders alike for a more direct and intuitive measure of stealthiness than a bound on Δz_t or the KLD. This may be difficult to solve analytically but should be possible using numerical methods. For example, assume there are zero-mean Gaussian noises, a Kalman filter is used as the state estimator, and that the system has a χ^2 detector equipped. Then, z_t' must remain within an ellipsoid centered on the origin in order to remain undetected. For a deterministic attack, the compromised innovation follows a multivariate Gaussian distribution with a non-zero mean. The detection rate would then be one minus the integration of the probability distribution function of z_t' over the centered ellipsoid. For example, this computation can be performed using the *pmvnEll* function in the R programming language by treating the Gaussian distribution as centered at the origin and the ellipsoid

as offset. Studying the detection rate and its properties for a more relaxed assumptions or a broader set of detectors would be an interesting research topic.

2. Develop countermeasures that are catered towards detecting or mitigating deterministic FDI attacks. Although many countermeasures have been studied in the literature, there has been little research on how to exploit the properties of deterministic attacks to determine appropriate defensive actions. In general, deterministic attacks are invulnerable against additive watermarks in linear systems because the attacks themselves are additive in nature. Furthermore, detectors would no longer be effective if a system is strictly vulnerable. Instead, two avenues of investigation are suggested:

   (a) A robust countermeasure that should be effective is encryption, as discussed in [30]. It would be interesting to examine the effect of encrypted communication channels on a system under FDI attacks and evaluate its effectiveness under a deterministic framework.

   (b) A major weakness of deterministic attacks is that they cause the compromised innovation distribution to no longer be zero-mean. Thus, an obvious countermeasure that can be deployed is one that monitors the average of the transmitted innovation. In this context, the technical design and theoretical properties of such a detector, such as detection and false alarm rates, should be a relatively simple but compelling research problem.

   Fortuitously, the suggested two defense mechanisms would have no negative impact on the nominal system.

3. As shown in Chapter 3, for some systems, there are some values of Δe_t that are reachable under stealthy attacks but do not permit a future stealthy attack of an arbitrary length. For such systems, it would be

interesting to find a general method to analytically or numerically determining the "sustainable stealthy attack set", or values of Δe_t from which stealthy attacks are always possible. One idea to determine such a set is as follows. Let P be projection matrix in Section 3.4.1. First, a stealthy attack on the current time step is only possible if Δe_t falls within the set:

$$\mathcal{A}_0 = \{\Delta e_t \mid \|(I - P)CA\Delta e_t\| \leq \delta\}$$

From here, a stealthy attack from $\Delta e_t \in \mathcal{A}_0$ on the next time step after is only possible if there exists a stealthy attack vector that can take an element of $\mathcal{A}_0$ to Δe_t. Otherwise, it should be removed from the set, forming $\mathcal{A}_1$. This process can be repeated to obtain progressively smaller sets $\mathcal{A}_i$ until a sufficiently accurate estimate of the sustainable stealthy attack set is obtained. Ideally, we would compute $\lim_{i\to\infty} \mathcal{A}_i = \mathcal{A}$, which is the exact sustainable stealthy attack set. Note that if $\Delta e_t \in \mathcal{A}_i$, then there exists an attack sequence such that a feasible stealthy attack will be possible at Δe_{t+i}. This method is similar to the techniques used to compute control invariant sets, also known as viability kernels [22]. Viewing the design of stealthy attacks as a system with input constraints, the reachable set of Δe_t could also be investigated to obtain a holistic understanding of attacker capabilities.

4. Extend some of the work on comparing the performance between stochastic and deterministic attack designs that was presented in Section 3.6. This may include applying the comparison to more general system models, such as for closed-loop control systems or an arbitrary stabilizing fixed-gain estimator, rather than just Kalman filtering at steady state. Even more interesting results could be obtained by proposing and studying some standard unified attack model that combines or incorporates elements of both stochastic and deterministic attacks.

5. Study the design of optimal attacks when there is a greatly limited infor-

mation set or there is some uncertainty in the system model available to the attacker. In almost all studies on FDI attacks on CPS, it is assumed that the attacker knows most system parameters exactly. However, this information is likely proprietary and would be difficult to obtain in practice. Data-driven and model-less control methods could be investigated and extended in order to design effective and stealthy attacks using only the transmitted data available to the attacker.

Bibliography

[1] C. Z. Bai, F. Pasqualetti, and V. Gupta. Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs. *Automatica*, 82:251–260, 2017.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] L. Cao, X. Jiang, Y. Zhao, S. Wang, D. You, and X. Xu. A survey of network attacks on cyber-physical systems. *IEEE Access*, 8:44219–44227, 2020.

[4] Y. Chen, S. Kar, and J. M. F. Moura. Optimal attack strategies subject to detection constraints against cyber-physical systems. *IEEE Transactions on Control of Network Systems*, 5(3):1157–1168, 2018.

[5] D. Cheng, J. Shang, and T. Chen. Finite-horizon strictly stealthy deterministic attacks on cyber-physical systems. *IEEE Control Systems Letters*, 6:1640–1645, 2022.

[6] G. H. Golub. Some modified matrix eigenvalue problems. *SIAM Review*, 15(2):318–334, 1973.

[7] Y. Guan and X. Ge. Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks. *IEEE Transactions on Signal and Information Processing over Networks*, 4(1):48–59, 2018.

[8] H. Guo, Z. Pang, J. Sun, and J. Li. An output-coding-based detection scheme against replay attacks in cyber-physical systems. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 68(10):3306–3310, 2021.

[9] H. Guo, J. Sun, and Z. Pang. Stealthy FDI attacks against networked control systems using two filters with an arbitrary gain. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 69(7):3219–3223, 2022.

[10] Z. Guo, D. Shi, K. H. Johansson, and L. Shi. Optimal linear cyber-attack on remote state estimation. *IEEE Transactions on Control of Network Systems*, 4(1):4–13, 2017.

[11] Z. Guo, D. Shi, K. H. Johansson, and L. Shi. Worst-case stealthy innovation-based linear attack on remote state estimation. *Automatica*, 89:117–124, 2018.

[12] Z. Guo, D. Shi, K. H. Johansson, and L. Shi. Worst-case innovation-based integrity attacks with side information on remote state estimation. *IEEE Transactions on Control of Network Systems*, 6(1):48–59, 2019.

[13] S. A. Haque, S. M. Aziz, and M. Rahman. Review of cyber-physical system in healthcare. *International Journal of Distributed Sensor Networks*, 10(4), 2014. Art. ID. 217415.

[14] L. Hu, Z. Wang, Q. L. Han, and X. Liu. State estimation under false data injection attacks: Security analysis and system protection. *Automatica*, 87:176–183, 2018.

[15] X. Ji, G. He, J. Xu, and Y. Guo. Study on the mode of intelligent chemical industry based on cyber-physical system and its implementation. *Advances in Engineering Software*, 99:18–26, 2016.

[16] S. H. Kafash, J. Giraldo, C. Murguia, A. A. Cardenas, and J. Ruths.

Constraining attacker capabilities through actuator saturation. In *2018 American Control Conference*, pages 986–991, 2018.

[17] R. Langner. Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security & Privacy*, 9(3):49–51, 2011.

[18] J. Lee, B. Bagheri, and H. A. Kao. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing Letters*, 3:18–23, 2015.

[19] Y. Li, L. Shi, and T. Chen. Detection against linear deception attacks on multi-sensor remote state estimation. *IEEE Transactions on Control of Network Systems*, 5(3):846–856, 2018.

[20] H. Liu, Y. Mo, J. Yan, L. Xie, and K. H. Johansson. An online approach to physical watermark design. *IEEE Transactions on Automatic Control*, 65(9):3895–3902, 2020.

[21] Z. Q. Luo, W. K. Ma, A. M. C. So, Y. Ye, and S. Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine*, 27(3):20–34, 2010.

[22] J. N. Maidens, S. Kaynama, I. M. Mitchell, M. M. Oishi, and G. A. Dumont. Lagrangian methods for approximating the viability kernel in high-dimensional systems. *Automatica*, 49(7):2017–2029, 2013.

[23] G. M. Makrakis, C. Kolias, G. Kambourakis, C. Rieger, and J. Benjamin. Industrial and critical infrastructure security: Technical analysis of real-life security incidents. *IEEE Access*, 9:165295–165325, 2021.

[24] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas. Coding schemes for securing cyber-physical systems against stealthy data injection attacks. *IEEE Transactions on Control of Network Systems*, 4(1):106–117, 2017.

[25] Y. Mo, R. Chabukswar, and B. Sinopoli. Detecting integrity attacks on SCADA systems. *IEEE Transactions on Control Systems Technology*, 22(4):1396–1407, 2014.

[26] Y. Mo and B. Sinopoli. False data injection attacks in control systems. In *Proceedings of the 1st Workshop on Secure Control Systems*, pages 56–62, 2010.

[27] Y. Mo and B. Sinopoli. On the performance degradation of cyber-physical systems under stealthy integrity attacks. *IEEE Transactions on Automatic Control*, 61(9):2618–2624, 2016.

[28] Y. Ni, Z. Guo, Y. Mo, and L. Shi. On the performance analysis of reset attack in cyber-physical systems. *IEEE Transactions on Automatic Control*, 65(1):419–425, 2020.

[29] Z. Pang, G. Liu, D. Zhou, F. Hou, and D. Sun. Two-channel false data injection attacks against output tracking control of networked systems. *IEEE Transactions on Industrial Electronics*, 63(5):3242–3251, 2016.

[30] J. Shang, M. Chen, and T. Chen. Optimal linear encryption against stealthy attacks on remote state estimation. *IEEE Transactions on Automatic Control*, 66(8):3592–3607, 2021.

[31] J. Shang, D. Cheng, J. Zhou, and T. Chen. Asymmetric vulnerability of measurement and control channels in closed-loop systems. *IEEE Transactions on Control of Network Systems*, 2022. DOI: 10.1109/TCNS.2022.3165086.

[32] J. Shang, H. Yu, and T. Chen. Worst-case stealthy innovation-based linear attacks on remote state estimation under Kullback–Leibler divergence. *IEEE Transactions on Automatic Control*, 2021. DOI: 10.1109/TAC.2021.3125430.

[33] J. Shang, J. Zhou, and T. Chen. Single-dimensional encryption against innovation-based stealthy attacks on remote state estimation. *Automatica*, 136, 2022. Art. no. 110015.

[34] D. Shi, Z. Guo, K. H. Johansson, and L. Shi. Causality countermeasures for anomaly detection in cyber-physical systems. *IEEE Transactions on Automatic Control*, 63(2):386–401, 2018.

[35] H. Song, D. B. Rawat, S. Jeschke, and C. Brecher, editors. *Cyber-Physical Systems: Foundations, Principles and Applications*. Academic Press, 2017.

[36] S. Sridhar, A. Hahn, and M. Govindarasu. Cyber–physical system security for the electric power grid. *Proceedings of the IEEE*, 100(1):210–224, 2012.

[37] T. Sui, Y. Mo, D. Marelli, X. Sun, and M. Fu. The vulnerability of cyber-physical system under stealthy attacks. *IEEE Transactions on Automatic Control*, 66(2):637–650, 2021.

[38] K. Wang, E. Tian, J. Liu, L. Wei, and D. Yue. Resilient control of networked control systems under deception attacks: A memory-event-triggered communication scheme. *International Journal of Robust and Nonlinear Control*, 30(4):1534–1548, 2020.

[39] L. Wang, X. Cao, H. Zhang, C. Sun, and W. X. Zheng. Transmission scheduling for privacy-optimal encryption against eavesdropping attacks on remote state estimation. *Automatica*, 137, 2022. Art. no. 110145.

[40] G. Xiong, F. Zhu, X. Liu, X. Dong, W. Huang, S. Chen, and K. Zhao. Cyber-physical-social system in intelligent transportation. *IEEE/CAA Journal of Automatica Sinica*, 2(3):320–333, 2015.

[41] D. Ye and T. Zhang. Summation detector for false data-injection attack in

cyber-physical systems. *IEEE Transactions on Cybernetics*, 50(6):2338–2345, 2020.

[42] R. V. Yohanandhan, R. M. Elavarasan, P. Manoharan, and L. Mihet-Popa. Cyber-physical power system (CPPS): A review on modeling, simulation, and analysis with cyber security applications. *IEEE Access*, 8:151019–151064, 2020.

[43] X. Yu and Y. Xue. Smart grids: A cyber–physical systems perspective. *Proceedings of the IEEE*, 104(5):1058–1070, 2016.

[44] H. Zhang, P. Cheng, L. Shi, and J. Chen. Optimal denial-of-service attack scheduling against linear quadratic gaussian control. In *2014 American Control Conference*, pages 3996–4001, 2014.

[45] T. Zhang and D. Ye. False data injection attacks with complete stealthiness in cyber–physical systems: A self-generated approach. *Automatica*, 120, 2020. Art. no. 109117.

[46] J. Zhou, J. Shang, Y. Li, and T. Chen. Optimal DoS attack against LQR control channels. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 68(4):1348–1352, 2021.