SECURITY OF CYBER-PHYSICAL SYSTEMS: A CONTROL-THEORETIC PERSPECTIVE

BY

NABIL H. K. HIRZALLAH

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Petros G. Voulgaris, Chair
Professor Naira Hovakimyan
Professor Peter W. Sauer
Professor Srinivasa M. Salapaka
Associate Professor Mohamed Ali Belabbas

ABSTRACT

Motivated by the attacks on control systems through the cyber (digital) part, we study how signal attacks injected through actuators and/or sensors affect control system stability and performance. We ask the questions: What are the different types and scenarios of signal attacks? When are the attacks stealthy and unbounded? How to compute the worst stealthy bounded attacks? How to defend against such attacks through controller design? How to identify and estimate signal attacks before significant performance loss happens? We answer the above questions in this thesis using tools from control theory. We show that it is necessary to use a sampled-data framework to accurately assess the vulnerabilities of control systems. In addition, we show that the most lethal attacks are related to the structure of the system (location of zeros and poles, number of inputs and outputs). We show that dual rate control is a powerful tool to defend against these vulnerabilities, and we provide a related controller design. Furthermore, we show that the worst stealthy bounded attacks can be computed by an iterative linear program, and we show how to lessen their effects through iterative controller design. Finally, we study the trade-off between control and estimation of signal attacks and provide several controller designs utilizing the power of dual rate sampling.

To family and friends

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Prof. Petros Voulgaris. This thesis would not have seen light without his ideas, insights and direction. For him handling coprime factorization and visualizing controller structure is probably easier than drinking water. He was supportive and patient, and I learned a great deal from him. I am indebted to him for the rest of my life. My only regret is that I did not spend more personal time with him during my stay at Urbana.

I am also very thankful to my thesis committee (Prof. Naira, Prof. Srinivasa, Prof. Peter and Prof. Ali) for their valuable comments during the preliminary exam, and for shaping the person I am today through their courses and books. I learned a lot from Prof. Naira's course and her valuable course notes and book. She always emphasized the importance of intuition and architecture in control systems, and encouraged the audience to work on practical problems and find practical solutions. Reading her early papers that led to the development of L_1 adaptive control was a very interesting experience for me. I was very fortunate to take my first control course at UIUC with Prof. Srinivasa. The course was very important and fundamental to my research and he did a great job teaching it. I certainly will not forget the notes in advance and his unique teaching style, neither I will forget his friendly and humble personality. I was very lucky to take stochastic control with Prof. Ali. He included topics such as Ito calculus and the Fokker−Planck equation that I did not find in course notes of other instructors teaching the same course at UIUC and elsewhere. Unfortunately

I did not take a course with Prof. Peter although I have a huge interest in power systems. However, his book was very valuable for my coursework and for my master's thesis. Some of the theory in this dissertation was applied to models from his book. I am also thankful to him for changing his schedule to accommodate the final exam.

I would like to thank Dr. Mohammad Naghnaeian for the time we spent together in the lab, and for his contribution in Chapter 4. I am also very thankful to Dr. Juan Alvarez for providing teaching assistantship opportunities in his courses. I also would like to thank Prof. Daniel Liberzon and Prof. R. Srikant for offering a great learning experience in their courses. Special thanks also go to the staff of the ECE department and the Coordinated Science Lab for their help and support over the years.

I am very grateful to Abu Dhabi National Oil Company (ADNOC) for funding part of my stay here at the University of Illinois, and for funding my undergraduate studies. Without their generous support I would not be where I am today.

I am forever in debt to my parents and siblings for their infinite love and support. My decision to come to the US was not easy for them, and I was not close to them when they needed me the most. I am also very grateful to share my life with my wife. This thesis would not have been completed without her patience and support. I am also thankful to my in-laws for hosting my family for an extended period of time.

I met many of people during my time in Urbana who helped ease my stay and made it more interesting. I am thankful to all of them and I can't mention them all here because I will definitely miss a few names. Therefore I will only mention some of the older folks out of respect: Dr. Nahil Sobh, Dr. Irfan Ahmed, Dr. Emad Jassim and Dr. Abdulkareem Saadeh. Special thanks to the CIMIC community and to the soccer group. And special thanks to my friends in Abu Dhabi; they made my short rare visits joyful and hilarious.

I would like to dedicate this thesis to the memory of Prof. Bruce Francis

who passed away earlier this year. This dissertation is influenced heavily by his work, especially book *Optimal Sampled-Data Control Systems*. He had a remarkable writing style; I have benefited a lot from his books and posted online notes.

Finally, *"And the last of their calls will be: Praise to God, Lord of the worlds."*

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Motivation

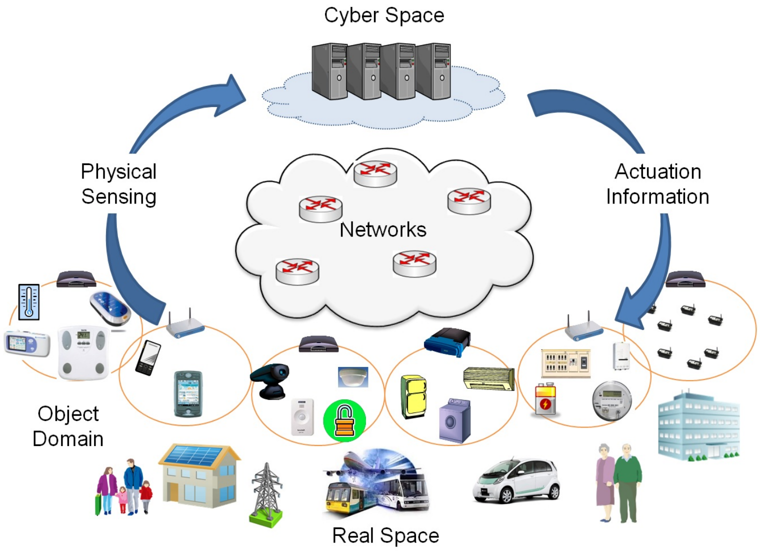


Figure 1.1: Cyber-physical systems.

Advancements in communication, sensing and computing technologies have led the control of physical systems or plants to be implemented over networks (cyberspace), leading to the creation of “cyber-physical” systems. Such systems are found in many applications including the smart grid and vehicle control units as depicted in Figure 1.1. However, the interaction between the cyber part and the physical part introduced security challenges that can be exploited by malicious agents. It has been shown through real world incidents and published research simulations that stealthy attacks can be carefully designed to cause significant damage in the control system [1, 2, 3]. The most famous is the Stuxnet attack in which a designed computer malware infected the Siemens programmable logic controllers (PLCs) of a nuclear enrichment plant in Iran. The malware spread through standard USB devices and it was estimated that it infected 100,000 computer systems. The attack started by recording centrifuge measurements for a period a time, and then intercepting the real measurements and replaying the recorded data indicating regular operating conditions, while at the same time injecting harmful actuation signals. It was estimated that around 1000 centrifuges were damaged by the attack.

Most work on securing cyber-physical systems focuses only on software security. While it is true that software security is the first line of defense, it is also clear that this line can be infiltrated as happened in the Stuxnet attack. Once attackers gain access to the cyber (digital) space (Figure 1.2), they can exploit their knowledge of the structural properties of the control system (e.g., pole-zero locations) to induce stealthy unbounded and bounded attacks that destroy the systems or at least affect their performance (even non-invasive intrusion is possible [4]). Moreover, even if the attack is not stealthy, it may be too late to react. Since these types of undetectable attacks are mainly related to system structure, it is very important to research the vulnerabilities of cyber-physical systems from a control-theoretic perspective, and find solutions that guarantee the security, stability and resiliency of control systems under different attack scenarios.

One might argue that control theory has already developed fault detection frameworks [5] that can handle various faults and disturbances, and that these frameworks are sufficient for handling attacks on control systems. However this argument fails to recognize that there are substantial differences between cyber-physical attacks and faults. These differences suggest that a fault-tolerant system may not be secure against carefully designed sophisticated attacks, and they stem from the fact that faults are considered random events that do not have malicious intents. In addition, the occur-

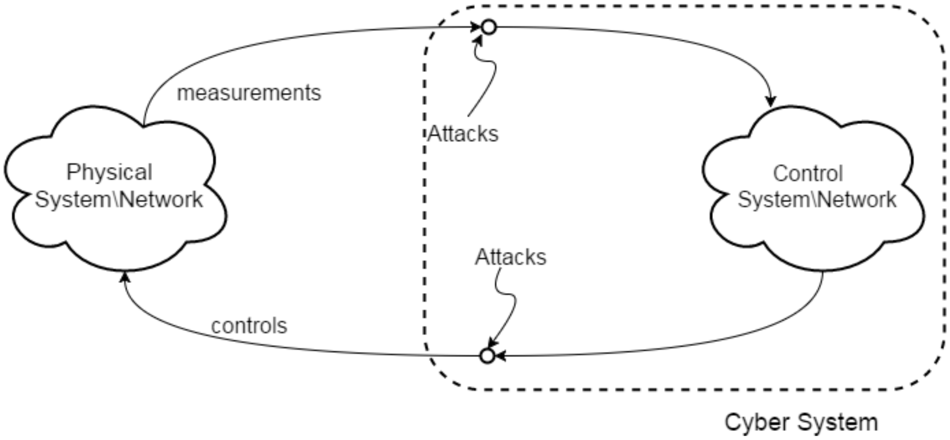


Figure 1.2: Cyber-physical systems with attacks on the cyber part.

rence of more than one fault at the same time is considered to have a low probability. On the other hand, malicious attacks can attack more than one point in the control system in a coordinated fashion. Moreover, they are carefully designed (intelligent) to exploit weaknesses in the system and may not be detectable (stealthy). Therefore, existing fault detection frameworks are not sufficient to design secure cyber-physical systems, and a new more thorough framework that takes into account the characteristics and signatures of attacks is needed to handle the various threats.

1.2 Related Work

Recent work on security of cyber-physical systems from a control-theoretic perspective has been focused on the characterization of feasible and optimal (for some cost function) attacks and proposing ways for detection and/or improving the resiliency of the control system subject to such attacks. The type of attacks studied can be generally split into two categories: static attacks (attacks that do not take into account the dynamics of the system and/or do not affect the states of the system directly) or dynamic attacks. Attacks under each category can be classified as stealthy or not stealthy depending on the assumptions and the detection methods used. Examples of static attacks include attacks on the power system state estimators [6, 7, 8, 9, 10, 11, 12], where a carefully designed bias can be added to the sensor measurements without being detected by the commonly used statistical detection methods. Another work on static attacks is by [13] and [14] where they showed that the states of the system cannot be accurately reconstructed if half of the sensors are attacked. Both papers propose computationally intensive methods to reconstruct the states when less than half of the sensors are attacked such that the system is observable from the remaining un-attacked sensors (solving an ℓ_0 optimization problem in [13] and constructing a bank of observers in [14]). In [15] the authors showed that to accurately estimate the states, it is sufficient for the system to be detectable from the un-attacked sensors (provided less than half are attacked). The bank of observers in [14] was substituted by a bank of Kalman filters in [16] to estimate the states under attack given that the measurements are corrupted by noise, by leveraging the noise statistics over a large enough time window, and by [17] using event-triggered observers. The work in [13] was also extended by [18] where the authors provide a framework to reconstruct the states that is robust to additive and multiplicative errors. On the other hand, research work related to dynamic attacks includes the study of replay attacks by [19, 20] in which the authors inject a designed random signal (watermarking signal unknown to the attacker) into the system to detect the attack at the expense of increasing the cost of the LQG controller. The authors provide and solve an optimization problem to design the watermarking signal to maximize the detection ability and minimize false alarms. In a similar context, [21] presented an information theoretic formulation of the problem, and showed that if the watermark is a Gaussian distributed random variable, then the maximal performance degradation for any given level of stealthiness for the attacker is achieved when the attacker replaces the control input with the realization of a Gaussian random variable. They also showed that the watermark signal that minimizes the stealthiness of a Gaussian attacker is also Gaussian. In [22] static and dynamic attack models on linear time-invariant systems are provided along with conditions for stealthiness. In addition they provide filter design methods to detect a class of (detectable) attacks under centralized and distributed fashion. In [23] the case for finding the worst constant bias (steady state) attack has been considered and a tractable procedure to compute it has been developed where the energy of the detection signal was considered as a measure of stealthiness. A similar framework was studied in [24] for the design of optimal attacks on automatic generation control (AGC) systems. In [25] coordinated actuator and sensor attacks are computed that create unbounded expectation of the estimation error while keeping the residual of the KF detector bounded. In [26] optimal attacks are computed on an LQG system that minimize the K-L divergence between the true and falsified state estimates such that the attack impact is above a specified a limit, showing that the optimal attacks are additive white noise. In [27] sufficient conditions of the existence of an optimal attack sequence that drives the states to a desired set are provided using dynamic programming, where the system is equipped with a Kalman filter for state estimation. In [28] optimal actuator attacks are designed using the minimum principle that maximizes a quadratic cost related to the error between the healthy (un-attacked) system and the attacked system while minimizing the attack cost, without including any stealthiness requirement. In [29], the authors investigate attacks on power systems that act by switching between loads in a coordinated manner inducing a sliding mode which drives the frequency to instability. These switching attacks are investigated again in [30] exciting the inter-area oscillation modes in a coordinated manner to drive groups of system generators out of step. This way only a small amount of the load is needed to be controlled by the attacker to induce instability. In [31] the authors presented a stochastic approach for optimal planning under malicious attack on sensors. The approach uses Markov decision processes theory (MDP) to obtain an optimal policy to drive a vehicle that has inconsistent measurements.. Resiliency against attacks is achieved by properly selecting the reward function of the MDP, thus avoiding actions that could hijack the vehicle to undesired regions of a state space. In [32] the authors studied denial of service (DoS) attacks and obtained sufficient conditions on the DoS duration, and frequency bounds for stabilization with finite data rates, which are characterized by the decay rates of the quantization range in the presence and absence of DoS attacks. In the same context of DOS attacks, [33] studies structural resilience of LTI systems under attacks. Specifically they provide conditions for the controllability of the systems under Dos and integrity attacks. In [34], the authors study controllability and stability properties of dynamical systems when actuator or sensor signals are under attack. The authors study the impact of these attacks and propose reactive countermeasures based on game theory. In [35] the authors study an attackers ability to control a maritime surface vessel by spoofing GPS signals. The authors formulate an optimization problem to find the attacker's control law, and provide a detection mechanism. In [36], the authors use tools from LMI theory to solve an optimization problem that finds the optimal artificial actuator saturation limits that maximize the reachable sets of the states, while guaranteeing that the dangerous states are not reachable.

1.3 Overview

This thesis addresses the problem of security of cyber-physical systems from a control theory perspective. In particular, the thesis contributes towards a comprehensive framework to analyze, identify, and evaluate the consequences of existing vulnerabilities in cyber-physical systems. This framework will help us propose defense and protection schemes.

The proposed framework is based on sampled-data (SD) systems since it captures the behavior of cyber-physical systems in the sense that the controllers are implemented digitally and the physical plants evolve in a continuous-time domain. Moreover, the SD approach allows us to assess the security of cyber-physical systems as it can help us visualize how additive attacks can be injected in various parts of the cyber-physical system and analyze their effects on stability and performance. In addition, SD approach is necessary because SD implementation generates additional vulnerability to stealthy attacks by introducing additional, so-called sampling zeros in the system as the sampling rate increases (e.g., [37]). Thus, while a system may be secure to stealthy attacks using a continuous-time analysis, it may not be secure using a more accurate, continuous- and discrete-time, SD framework.

In this thesis, we employ the SD framework to characterize the necessary and sufficient conditions for the existence of stealthy unbounded actuator and/or sensor attacks. We define the attack detection or monitoring mechanism and then we show that stealthy unbounded attacks are related to the unstable zeros or poles of the system. We also show that unbounded coordinated sensor and actuator attacks are always feasible regardless of the structure of the system. The analysis is done in an input-output fashion, which provides a framework for future analysis and study of different types of attacks on cyber-physical systems as it can provide mappings between an attack and its effect on various points of the feedback loop. We then provide a defense scheme based on dual sampling to detect stealthy unbounded actuator attacks, and we investigate trade-offs in the controller design.

After considering stealthy unbounded attacks, we shift our focus to stealthy bounded attacks. While bounded attacks may not induce catastrophic events, their effect on cyber-physical systems can still be severe. This is because since bounded attacks cannot cause instability in linear time-invariant (LTI) stable systems, they can be injected repeatedly into the system without being detected, degrading the performance and efficiency of the system and inducing stress on the physical part. An example would be injecting a signal that would increase the voltage across a machine's terminal or increasing the speed of uranium centrifuges. Moreover, after a loss in performance is observed, the system operator will have a hard time deciding whether the loss of performance is due to a random failure or a carefully designed attack since the attack is stealthy, bounded and with bounded effects. In addition, bounded attacks are more practical to inject since usually actuators have saturation limits. In this part, we consider the problem of characterizing the worst bounded and stealthy attacks under different attack resource and stealthiness constraints. We employ discrete input-output maps describing the effect of the attack signal on the performance variable and the monitoring variables. The objective is to find a traceable computation procedure to find the worst stealthy attack signal. We define the "worst" attack as the attack that induces the maximum damage on the performance variable in a ℓ_∞ sense. Using this computation, we will be able to assess the vulnerability and resiliency of the system with respect to the considered attack. We consider different attack resource constraints and stealthiness intervals, and provide an iterative controller synthesis procedure that alternates between computing worst attacks and designing optimal controllers that enhance performance and minimize the impact of worst attacks.

Next we consider the problem of estimating signal attacks on the actuators and/or sensors of control systems using the available measurements. The estimated attack signal will help the operator decide whether it is a persistent intelligent attack or just a nominal disturbance. We show that the design of the controller for estimation and controller for rejection are coupled, and that a trade-off exists between their individual performances. The quality of the estimate depends on the performance of the attack rejection controller. Then we provide controller design methods to estimate the signal attack. To guarantee that all unbounded attacks are detectable, we use a faster sampling loop for the estimation controller so that unstable zeros in the map from the attack signal to the measurements are removed. Furthermore, dual rate estimation allows for the construction of unknown input observers.

1.4 Outline and Contribution

In the following we present the outline of this thesis.

Chapter 2: Preliminary Results
In this chapter we present some of the standard results in the literature that are used to build the main results of this thesis.

Chapter 3: Conditions for Existence of Unbounded Actuator and/or Sensor Attacks
In this chapter we introduce the SD framework that we will be using to investigate the security of cyber-physical systems. We introduce attacks on actuators and sensors, represented as additive and unbounded disturbances on the digital (i.e., cyber) part of the controlled system. We examine from an input-output perspective the exact conditions under which such attacks can be stealthy, which brings up the role of unstable zeros, poles and structure of the open loop, continuous-time, physical plant, regardless of the specific controller and/or detection (e.g., Kalman) filter in use.

Chapter 4: Dual Rate Control for Detecting Unbounded Actuator Attacks
In this chapter we introduce multirate sampled-data (MRSD) control as a solution to detect unbounded actuator attacks on cyber-physical systems. We show that, if there is a single sensor that is guaranteed to be secure and the plant is observable from that sensor, then there exists a class of multirate sampled data controllers that ensure that all attacks remain detectable. These dual rate controllers are sampling the output faster than the zero order hold rate that operates on the control input, and as such, they can even provide better nominal performance than single rate, at the price of higher sampling of the continuous output.

Chapter 5: On the Computation of Worst Attacks: An LP Framework

In this chapter we consider the problem of characterizing the worst bounded and stealthy attacks. This problem involves a maximization of a convex function subject to convex constraints, and hence, in principle, it is not easy to solve. However, by employing an ℓ_∞ framework, we show how tractable linear programming (LP) methods can be used to obtain the worst attack design for different attack scenarios. Moreover, we provide a controller synthesis iterative method to minimize the worst impact of such attacks and test its efficacy in a power system component.

Chapter 6: On the Estimation of Signal Attacks

In this chapter we consider the problem of estimating signal attacks injected into the actuators or sensors of control systems, assuming the attack is detectable (can be seen at the output). We show that there exists a trade-off between attack rejection and control, and that the estimator design depends on the controller used. We use dual rate sampling to enhance detectability of the attacks and we provide different methods to design the estimator. The first method is by solving a model matching problem subject to causality constraints. The second method exploits dual rate sampling to accurately reconstruct the unknown input. The third method is using a dual rate unknown input observer. We provide conditions on the existence of these estimators, and show that dual rate unknown input observers always exist if the multirate system does not have a zero at 1.

Chapter 7: Summary

In this chapter we summarize the work done in this thesis..

CHAPTER 2

STANDARD RESULTS

In this chapter we present some of the standard results in the literature ([37, 38, 39, 40, 41, 42, 43, 44]) that are used to build the main results of this thesis.

Some standard notation we use is as follows: $\mathbb{Z}_+$, $\mathbb{R}^n$, $\mathbb{C}^n$ and $\mathbb{R}^{n\times m}$ denote the sets of non-negative integers, n-dimensional real vectors, n-dimensional complex vectors and $n \times m$ dimensional real matrices, respectively. For any $\mathbb{R}^n$ or $\mathbb{C}^n$ vector x we denote x' its transpose and $|x| := \max_i \sqrt{x_i^2}$ where $x' = [x_1, x_2, ..., x_n]$; for a sequence of real n-dimensional vectors, $x = \{x(k)\}_{k\in\mathbb{Z}_+}$ we denote $||x||_\infty := \sup_k |x(k)|$; for a sequence of real $n \times m$ dimensional real matrices $G = \{G_k\}_{k\in\mathbb{Z}_+}$ we denote its z-transform $G(z) := \sum_{k=0}^{\infty} G_k z^{-k}$; and if viewed as the pulse response of the LTI system G then $||G||_1 = \sup_{||x||_\infty\leq 1} ||Gx||_\infty$.

2.1 Discrete-Time Systems

2.1.1 Basic Concepts

The discrete-time set is taken to be the integers $\{0, 1, 2, \dots\}$. A discrete-time signal is a sequence $\{v(0), v(1), v(2), \dots\}$, where each $v(k)$ is a real number.

The λ-transform of a v is defined to be

$$\hat{v}(\lambda) := v(0) + v(1)\lambda + v(2)\lambda^2 + \cdots = \sum_{k=0}^{\infty} v(k)\lambda^k.$$

Any causal LTI system G has a matrix of the form

$$\left[G\right] = \begin{bmatrix} g(0) & 0 & 0 & \dots \\ g(1) & g(0) & 0 & \dots \\ g(2) & g(1) & g(0) & \dots \\ \vdots & \vdots & \vdots & \end{bmatrix}.$$

A system G is causal if and only if $\left[G\right]$ is block-lower triangular; is time invariant if and only if $\left[G\right]$ is constant along block-diagonal, i.e., Toeplitz.

The impulse response is the sequence represented by the first column of $\left[G\right]$, and the transfer function is the λ-transform of the impulse response:

$$\hat{g}(\lambda) = g(0) + g(1)\lambda + g(2)\lambda^2 + \cdots = \sum_{k=0}^{\infty} g(k)\lambda^k.$$

Let $\psi = Gv$; then $\psi(k)$ is can be found using the convolution equation:

$$\psi(k) = \sum_{l=0}^{k} g(k-l)v(l).$$

2.1.2 Lifting Discrete-Time Signals and Systems

Let h be the period of a base clock, and let $v(k)$ be discrete-time signal with period h/n where n is some positive integer. That is, $v(0)$ occurs at time $t = 0$, $v(1)$ at $t = h/n$, $v(2)$ at $t = 2h/n$, etc. The lifted signal $\underline{v}$, is defined as follows: If

$$v = \{v(0), v(1), v(2), \dots\},$$

then

$$\underline{v} = \left\{ \begin{bmatrix} v(0) \\ v(1) \\ \vdots \\ v(n-1) \end{bmatrix}, \begin{bmatrix} v(0) \\ v(1) \\ \vdots \\ v(n-1) \end{bmatrix}, \ldots \right\}.$$

The dimension of $\underline{v}(k)$ equals n times that of $v(k)$, and $\underline{v}$ is regarded as referred to the base period; that is, $\underline{v}(k)$ occurs at time $t = kh$. The lifting operator L is defined to be the map $v \to \underline{v}$. Te vector representation of the equation $\underline{v} = Lv$ when $n = 2$ is

$$\begin{bmatrix} \underline{v}(0) \\ \underline{v}(1) \\ \underline{v}(2) \\ \vdots \end{bmatrix} = \left[\begin{array}{c|c|c|c|c|c} I & 0 & 0 & 0 & 0 & \ldots \\ 0 & I & 0 & 0 & 0 & \ldots \\ \hline 0 & 0 & I & 0 & 0 & \ldots \\ 0 & 0 & 0 & I & 0 & \ldots \\ \hline 0 & 0 & 0 & 0 & I & \ldots \\ 0 & 0 & 0 & 0 & 0 & \ldots \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \end{array}\right] \begin{bmatrix} v(0) \\ v(1) \\ v(2) \\ \vdots \end{bmatrix}.$$

For the partition shown, $\begin{bmatrix} L \end{bmatrix}$ is neither lower-triangular nor Toeplitz; therefore, as a system L is non-causal and time-varying. For $n = 2$, L^{-1} is represented as

$$\begin{bmatrix} L^{-1} \end{bmatrix} \left[\begin{array}{cc|cc|cc|c} I & 0 & 0 & 0 & 0 & 0 & \ldots \\ \hline 0 & I & 0 & 0 & 0 & 0 & \ldots \\ \hline 0 & 0 & I & 0 & 0 & 0 & \ldots \\ \hline 0 & 0 & 0 & I & 0 & 0 & \ldots \\ \hline 0 & 0 & 0 & 0 & I & 0 & \ldots \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{array}\right].$$

It can be shown that L is norm preserving, we will not present this result here, but it can be found in [37].

For a discrete-time FDLTI system G with underlying period h/n, lifting the input and output signals so that the lifted signals correspond to the base period h results in a lifted system $\underline{G} = LGL^{-1}$. Given $\hat{g}$ in terms of state-space data, $\hat{g}(\lambda) = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$, then

$$\underline{\hat{g}}(\lambda) = \left[\begin{array}{c|cccc} A^n & A^{n-1}B & A^{n-2}B & \dots & B \\ \hline C & D & 0 & \dots & 0 \\ CA & CB & D & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ CA^{n-1} & CA^{n-2}B & CA^{n-3}B & \dots & D \end{array}\right], \tag{2.1}$$

which is easy to prove.

Next we present how to find the state space description of dual rate systems, which is used many times in this thesis. Consider $P = L\mathcal{S}_f P_c \mathcal{H}$, where $\mathcal{H}$ is the hold function operating at a rate of h, and $\mathcal{S}_f$ is the sampling function operating at a faster rate h/n. Let P_c be given by

$$P_c = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right].$$

Let A_f and B_f be the fast discretization of A and B, and A_s and B_s be the slow discretization of A and B, i.e.,

$$\begin{aligned} A_s &:= e^{Ah}, & A_f &:= e^{Ah/n}, \\ B_s &:= \int_0^h e^{A\tau} B \mathrm{d}\tau, & B_f &:= \int_0^{h/n} e^{A\tau} B \mathrm{d}\tau. \end{aligned}$$

Working on P, we get $P = L\mathcal{S}_f P_c \mathcal{H} = L(\mathcal{S}_f P_c \mathcal{H}_f)\mathcal{S}_f \mathcal{H}$; this is because

$\mathcal{H}_f\mathcal{S}_f\mathcal{H} = \mathcal{H}$. The matrix representation of $\mathcal{S}_f\mathcal{H}$ is

$$\left[\mathcal{S}_f\mathcal{H}\right] = \begin{bmatrix} \left.\begin{matrix} I & 0 \\ \vdots & \vdots \\ I & 0 \end{matrix}\right\} n & & \\ \left.\begin{matrix} 0 & I \\ \vdots & \vdots \\ 0 & I \end{matrix}\right\} n & & \\ & & \ddots \end{bmatrix}.$$

From this and $\left[L\right]$ it can be inferred that

$$L\mathcal{S}_f\mathcal{H} = \begin{bmatrix} I \\ \vdots \\ I \end{bmatrix}, (n \text{ blocks}),$$

that is

$$\mathcal{S}_f\mathcal{H} = L^{-1}\begin{bmatrix} I \\ \vdots \\ I \end{bmatrix}.$$

Therefore, $P = L(\mathcal{S}_f P_c \mathcal{H}_f)L^{-1}\begin{bmatrix} I \\ \vdots \\ I \end{bmatrix}$, using the results from (2.1), we get the following:

$$
\begin{aligned}
P &= \left[\begin{array}{c|cccc} A_s & A_f^{n-1}B_f & A_f^{n-2}B_f & \dots & B_f \\ \hline C & D & 0 & \dots & 0 \\ CA_f & CB_f & D & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ CA_f^{n-1} & CA_f^{n-2}B_f & CA_f^{n-3}B & \dots & D \end{array}\right] \begin{bmatrix} I \\ I \\ \vdots \\ I \end{bmatrix} \\
&= \left[\begin{array}{c|c} A_s & B_s \\ \hline C & D \\ CA_f & CB_f + D \\ \vdots & \vdots \\ CA_f^{n-1} & CA_f^{n-2}B_f + \cdots + CB_f + D \end{array}\right].
\end{aligned}
$$

2.2 System Zeros

Definition 1. *z_0 is a zero of $G(z) = C(zI - A)^{-1}B + D$ if the rank of $G(z_0)$ is less than the normal rank of $G(z)$.*

2.2.1 Computing Zeros

The state-space equations of a system may be written as

$$
P(z)\begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}, \quad P(z) = \begin{bmatrix} zI - A & -B \\ C & D \end{bmatrix}.
$$

The zeros are the values $z = z_0$ for which $P(z)$ loses rank, resulting in a zero output for some non-zero input. Numerically, the zeros are found as non-trivial solutions (with $u_z \neq 0$ and $x_z \neq 0$) to the following problem:

$$
(zI_g - M)\begin{bmatrix} x_z \\ u_z \end{bmatrix} = 0, \; M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \; I_g = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix},
$$

where u_z is the zero input direction, and x_z is the associated state initial condition.

2.2.2 Remarks on Zeros

- In the time domain, the presence of zeros implies blocking of certain input signals. If z_0 is a zero $G(z)$, then there exists an input signal of the form $u_z z_0^{-k}$ where u_z is the zero input direction, and a set of initial conditions x_z such that $y(k) = 0$ for $k > 0$.

- There are no zeros if the outputs y contain direct information about all the states (example $y = x$). More generally, there are no zeros if rank $C = n$ (n is the states dimension) and $D = 0$.

- Zeros usually appear when there are fewer inputs or outputs than states, or when $D \neq 0$. Consider $m \times m$ plant $G(z) = C(zI - A)^{-1}B + D$ with n states. We then have for the number of (finite) zeros of $G(z)$

  - $D \neq 0:$ At most $n - m + \text{rank}(D)$ zeros.
  - $D = 0:$ At most $n - 2m + \text{rank}(CB)$ zeros.
  - $D = 0$ and $\text{rank}(CB) = m:$ Exactly $n - m$ zeros.

- Zeros in the input-output map of connected network systems exist under certain conditions even if the dynamics of the single nodes have no zeros [45].

- Discretization introduces unstable zeros under certain conditions (fast sampling and relative degree of the continuous-time plant is greater than two) even if the continuous-time plant is minimum phase [46].

2.3 Strong Observability, Strong Detectability and Unknown Input Observers

Consider the following discrete linear time-invariant system:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned} \tag{2.2}$$

Definition 2. *(Strong Observability). A linear system of the form* (2.2) *is said to be strongly observable if, for any initial state $x(0)$ and any unknown sequence of inputs $u(0), u(1), \ldots,$ there is a positive integer L such that $x(0)$ can be recovered from the outputs $y(0), y(1), \ldots, y(L)$.*

To relate the concept of strong observability to the system matrices, if we simply iterate the output equation in (2.2) for $L+1$ time-steps, we get:

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \\ y(L) \end{bmatrix} = \underbrace{\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{L-1} \end{bmatrix}}_{\mathcal{O}_L} x(0) + \underbrace{\begin{bmatrix} D & 0 & 0 & \ldots & 0 \\ CB & D & 0 & \ldots & 0 \\ CAB & CB & D & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{L-1}B & CA^{L-2}B & CA^{L-3}B & \vdots & D \end{bmatrix}}_{\mathcal{J}_L} \begin{bmatrix} u(0) \\ u(1) \\ u(2) \\ \vdots \\ u(L) \end{bmatrix}.$$

Theorem 3. *Consider the system* (2.2) *with $x(k) \in \mathbb{R}^n$. The system is strongly observable if and only if*

$$rank([\mathcal{O}_L \quad \mathcal{J}_L]) = n + rank(\mathcal{J}_L)$$

for some $L \leq n$. [39].

Theorem 4. *Consider the system* (2.2)*. The system is strongly observable if and only if the system has no invariant zeros [40, 39].*

Definition 5. *(Strong Detectability) The linear system* (2.2) *is strongly de-*

tectable if $y(k) = 0$ for all k implies that $x(k) \to 0$.

Theorem 6. *Consider the system* (2.2)*. The system is strongly detectable if and only if the system is strictly minimum phase [40, 39].*

We consider an observer of the form

$$\hat{x}(k+1) = E\hat{x}(k) + Fy(k : k+L). \tag{2.3}$$

Definition 7. *The system* (2.3) *is said to be an unknown input observer of the states in* (2.2) *with delay L if $\hat{x}(k) - x(k) \to 0$ as $k \to \infty$, regardless of the $u(k)$.*

Theorem 8. *The system in* (2.2) *has an unknown input observer (possibly with delay) if and only if* (2.2) *is strongly detectable [39].*

2.4 Controller Parameterization

Definition 9. *(Coprime Factorization): A doubly coprime factorization of P is a set of maps $N, M, \tilde{N}, \tilde{M}$, with $P = NM^{-1} = \tilde{M}^{-1}\tilde{N}$ satisfying*

$$\begin{pmatrix} \tilde{X} & -\tilde{Y} \\ -\tilde{N} & \tilde{M} \end{pmatrix} \begin{pmatrix} M & Y \\ N & X \end{pmatrix} = I$$

for some stable $X, Y, \tilde{X}, \tilde{Y}$. Further, M and N are referred to as right coprime factors while $\tilde{M}$ and $\tilde{N}$ are referred to as left coprime factors of P.

Lemma 10. *Let a doubly-coprime factorization of P be given as in Definition 9. Controller K stabilizes P if and only if K has a right coprime factorization*

$K = Y_1 X_1^{-1}$ such that the map

$$\begin{pmatrix} M & Y_1 \\ N & X_1 \end{pmatrix}$$

is stable and stably invertible.

Theorem 11. *Let a doubly-coprime factorization of P be given as in Definition 9. All stabilizing controllers are given by*

$$K = (Y - MQ)(X - NQ)^{-1} = (\tilde{X} - A\tilde{N})^{-1}(\tilde{Y} - Q\tilde{M}),$$

where Q is stable.

The parameterization for when P is stable is straightforward. A doubly-coprime factorization of P is then given by

$$\begin{pmatrix} I & 0 \\ -P & I \end{pmatrix} \begin{pmatrix} I & 0 \\ P & I \end{pmatrix} = I.$$

The controller parametrization is then given by

$$K = -Q(I - PQ)^{-1}.$$

CHAPTER 3

ON THE EXISTENCE OF UNBOUNDED ACTUATOR AND/OR SENSOR ATTACKS

3.1 Introduction

In this chapter we examine the conditions for the existence of stealthy additive signal attacks on the actuators and sensors of feedback control systems. We define the notion of stealthiness and associate the existence of stealthy unbounded attacks to structural properties of the control system.

3.2 System Model

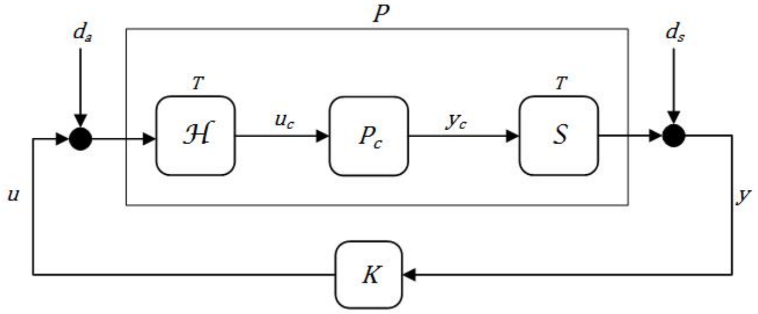


Figure 3.1: The standard SD system.

We consider the physical, continuous-time, LTI plant $P_c = [A_c, B_c, C_c, D_c]$ of Figure 3.1 that is controlled by a digital controller K using the standard zero order hold and sampling devices $\mathcal{H}$ and $\mathcal{S}$ respectively. In particular, in the absence of any disturbances d_a and d_s, the digital controller input

$u = \{u(k)\}$ converts to the continuous-time input

$$u_c(t) = (\mathcal{H}u)(t) = u(k) \;\text{ for }\; kT \le t < (k+1)T,$$

where T is the hold period, and the digital output $y = \{y(k)\}$ sequence is obtained by sampling the continuous-time output y_c with the same period T, i.e.,

$$y(k) = (\mathcal{S}y_c)(k) = y_c(kT).$$

The corresponding discrete-time LTI plant P is defined by the relation $y = Pu$, i.e., $P = \mathcal{S}P_c\mathcal{H}$, and has a description $P = [A_d, B_d, C_d, D_d]$ where the state space matrices are obtained from the corresponding continuous-time as

$$\begin{aligned} &A_d := e^{A_cT} \in \mathbb{R}^{n\times n}, \quad B_d := \int_0^T e^{A_c\tau}B_c\mathrm{d}\tau \in \mathbb{R}^{n\times n_u}, \\ &C_d := C_c \in \mathbb{R}^{n_y\times n}, \quad\; D_d := D_c \in \mathbb{R}^{n_y\times n_u}. \end{aligned} \tag{3.1}$$

We assume that the employed realization of the continuous plant P_c is minimal, which implies that the same holds true for the discrete plant P in the absence of pathological sampling (e.g., [37]), i.e., for almost all periods T.

Also in this figure, we consider the possibility of attacks in terms of additive disturbances d_a and d_s respectively at the digital input u and at the output y of P. These attacks on the digital part of the system can be on actuators only ($d_s = 0$), sensors only ($d_a = 0$), or on both, coordinated or not. As they act on the cyber part of the system we allow them to be unbounded sequences.

We assume that there is an attack detection mechanism in place that monitors u and y and can detect an attack only if the effect of d_a and/or d_s on these signals is beyond a given noise level threshold $\theta > 0$, i.e., only if

$$\left|\begin{bmatrix} y \\ u \end{bmatrix}(k)\right| > \theta$$

for some k. Note that we implicitly assume that there are other inputs such as noise, not shown in Fig 3.1, that have some effect on u and y which is what relates to the nonzero noise level θ. Accordingly, a stealthy attack of interest will be the case when the attack inputs d_a and/or d_s can grow unbounded while maintaining their effect on u and y below the detection limit; i.e., their effect cannot be distinguished from that of the normal noise inputs. Specifically, if d represents any of d_a or d_s, then the attack will be stealthy if

$$\limsup_{k\to\infty} |d(k)| = \infty$$

while

$$\left| \begin{bmatrix} y \\ u \end{bmatrix} (k) \right| \leq \theta$$

for all $k = 0, 1, 2, \ldots$. In the sequel we consider various (unbounded) attack scenarios and analyze the conditions of their detectability, i.e., when such attacks can or cannot be stealthy.

Remark 12. *In the following it is assumed that an attacker has knowledge of the description of P, e.g., the transfer function $P(\lambda)$ or its state space realization. In fact, only knowledge of the unstable zero and pole locations (and directions in case of MIMO P) is necessary for our analysis to hold. The attacker generates d_a or/and d_s based only on this knowledge.*

3.3 Actuator Attacks

We start with the case when only actuator attacks d_a are present ($d_s = 0$) and proceed in characterizing their effect on the monitoring vector $\begin{bmatrix} y \\ u \end{bmatrix}$. Towards this end, let P be factored (e.g., [47, 43, 42]) as

$$P = \tilde{M}^{-1}\tilde{N} = NM^{-1},$$

where the stable systems $\tilde{N}, \tilde{M}$ and N, M are left and right coprime respectively, and consider the controller K with a similar coprime factorization as

$$K = \tilde{X}^{-1}\tilde{Y} = YX^{-1}.$$

The mappings from d_a to y and u are given respectively as $(I - PK)^{-1}P$ and $K(I - PK)^{-1}P$. Given that K stabilizes P, it holds that

$$\tilde{M}X - \tilde{N}Y =: W$$

is a stable and stably invertible map (unit). Moreover, it can be easily checked that

$$\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} (I - PK)^{-1}P \\ K(I - PK)^{-1}P \end{bmatrix} d_a = \begin{bmatrix} X \\ Y \end{bmatrix} W^{-1}\tilde{N}d_a. \tag{3.2}$$

As X and Y are right coprime and W is a unit, it follows that a stealthy attack is possible if and only if $\tilde{N}d_a$ is bounded for an unbounded d_a. That is, when

$$\limsup_{k\to\infty} |d_a(k)| = \infty$$

it holds that

$$\left\| \begin{bmatrix} y \\ u \end{bmatrix} \right\| < \infty$$

if and only if

$$\left\| \tilde{N}d_a \right\| < \infty.$$

The following proposition is a direct consequence of the previous analysis.

Proposition 13. *Let P be a "tall" system, i.e., the number of outputs is greater than or equal to the number of inputs. Assume further that $P(\lambda)$ has no zero on the unit circle $|\lambda| = 1$. Then, an (unbounded) actuator stealthy attack is possible if and only if $P(\lambda)$ has a non-minimum phase zero other* *than at $\lambda = 0$, i.e., a zero for $0 < |\lambda| < 1$.*

Proof. Note that the unstable zeros of P are zeros of $\tilde{N}$. Assume that

$$P(z_0)d_0 = 0,$$

where $0 < |z_0| < 1$ and $d_0 \neq 0$ is the zero direction of z_0 which can be chosen with $|d_0| = 1$. So we have that

$$\tilde{N}(z_0)d_0 = 0$$

and consequently any input

$$d_a(k) = d_0 \epsilon {z_0}^{-k}$$

will lead via Equation (3.2) to

$$\left\| \begin{bmatrix} y \\ u \end{bmatrix} \right\| < \epsilon C_0$$

where the constant $C_0 > 0$, depends on the closed loop maps. For example, C_0 could be taken as

$$C_0 = \left\| \begin{bmatrix} (I - P(\lambda)K(\lambda))^{-1}P(\lambda)d_0\frac{1}{1-(\lambda/z_0)} \\ K(\lambda)(I - P(\lambda)K(\lambda))^{-1}P(\lambda)d_0\frac{1}{1-(\lambda/z_0)} \end{bmatrix} \right\|.$$

Thus, if ϵ is small enough, e.g., $0 < \epsilon < \frac{\theta}{C_0}$, the input remains undetected.

To prove the reverse, note that if P has no unstable zeros, then the same holds for $\tilde{N}$ and thus $\left\| \tilde{N} d_a \right\| < \infty$ implies that $\|d_a\| < \infty$, so no unbounded stealthy attacks are possible. □

Remark 14. *We remark here that if P has zeros on the boundary $|\lambda| = 1$ with multiplicity one but no other unstable zeros (other than at $\lambda = 0$), then* *unbounded stealthy attacks are not possible. Indeed, if z_0 is a simple zero with $|z_0| = 1$, then the corresponding input that can be masked ("zeroed out") is of the form $d_a(k) = \epsilon d_0 {z_0}^{-k}$ which is bounded with $|d_a(k)| < \epsilon$, and becomes undetected for small enough ϵ. But this case is uninteresting, as the disturbance has a level of noise (which can be taken care by any reasonably robust controller). On the other hand, if there are more than one multiplicities, unbounded stealthy attacks are possible. For example, if P is SISO and $z_0 = 1$ is a zero with multiplicity 2, then an unbounded input of the form $d_a(k) = \epsilon k$, $k = 0, 1, \ldots$ remains undetected for small enough ϵ. More generally, in the MIMO case when a zero at the boundary has multiplicity greater than one, one has to check the Smith-McMillan form of $P(\lambda)$ for invariant factors with multiplicity greater than one corresponding to these zeros: unbounded stealthy attacks are possible if and only if there are such factors.*

Remark 15. *When there is a zero of P at $\lambda = 0$ there is no corresponding (causal) input signal to be "zeroed out."*

The case when P is "fat", i.e. when the number of outputs in y is less than the number of inputs in u, is always conducive to stealthy attacks as one input can mask the effect of the other. Indeed, consider a two-input one-output $P = [P_1 \ P_2]$; the effect of attacks at the individual control channels d_{a1} and d_{a2} on the output y is

$$y = P_1 d_{a1} + P_2 d_{a2} + [P_1 \ P_2]u,$$

and thus, picking for example,

$$d_{a2} = -P_2^{-1} P_1 d_{a1}$$

with d_{a1} arbitrary and unbounded leads to $y = [P_1 \ P_2]u$, i.e., complete masking of the attacks. [1]

[1] Strictly speaking, P_2^{-1} may not exist if P_2 is strictly proper , i.e., P_2 has a zero at $\lambda = 0$;

3.4 Sensor Attacks

The case of sensor-only attack $d_s \neq 0, d_a = 0$ can be viewed in a similar spirit. In particular, by considering coprime factorizations for P and K as before, the effect of d_s on the monitor vector is as

$$\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} (I - PK)^{-1} \\ K(I - PK)^{-1} \end{bmatrix} d_s = \begin{bmatrix} X \\ Y \end{bmatrix} W^{-1} \tilde{M} d_s. \tag{3.3}$$

Therefore, using the same rationale as in the previous case, we can claim that an attack is detectable if and only if there are no d_s with $\|d_s\| = \infty$ and $\left\|\tilde{M} d_s\right\| < \infty$. This in turn means that attacks are detectable if and only if $\tilde{M}$ has no unstable zeros, which is equivalent to P being a stable system. More specifically, we have the following which can be proved as in the Proposition 13.

Proposition 16. *Assume that $P(\lambda)$ has no pole on the unit circle $|\lambda| = 1$. Then, a sensor stealthy attack is possible if and only if $P(\lambda)$ has a pole with $|\lambda| < 1$.*

Regarding poles of $P(\lambda)$ on the boundary ($|\lambda| = 1$), similar remarks hold as in the actuator attack case. Namely, if these poles are simple then there is no stealthy attack. If they have multiplicities, then their multiplicities in the corresponding invariant factors in the Smith-McMillan form determine whether stealthy attacks are possible.

3.5 Coordinated Actuator Sensor Attacks

In the case when a coordination of actuator and sensor attack is possible, unbounded stealthy attacks are always possible even in the case where P

but one can always pick $d_{a1}(\lambda) = \lambda \bar{d}_{a1}(\lambda)$ with $\bar{d}_{a1}$ unbounded and make $(P_2^{-1} P_1 d_{a1})(\lambda)$ meaningful, i.e., corresponding to a sequence $\{d_{a1}(k)\}$ defined for nonnegative integers k.

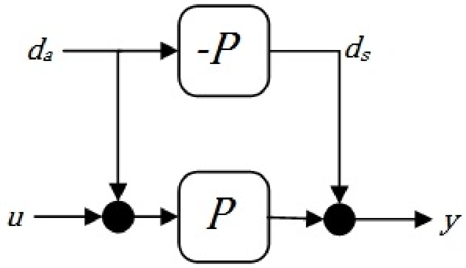


Figure 3.2: Coordinated actuator and sensor attacks.

is stable and minimum phase. Indeed, in this case the effect of d_a can be completely masked by canceling its effect at the output via d_s: just pick

$$d_s = -Pd_a$$

with d_a arbitrary and unbounded, then $y = Pu$, as depicted in Figure 3.2. Therefore, unless there are outputs that are not attacked, this situation is not of interest as there is no hope to detect the attack. If there are such attack-free outputs, then the problem reverts to the actuator-only attack case, with these outputs used for analysis and design.

3.6 Conclusion

In this chapter we introduced a sampled-data framework to study the effect of attacks on cyber-physical systems. We defined the attack detection mechanism and derived the input-output maps for actuator and sensor attacks on the monitoring signals. We showed that unbounded stealthy actuator attacks are related to the open-loop discrete plant unstable zeros, while unbounded stealthy sensor attacks are related to the open-loop discrete plant unstable poles. We also showed that coordinated actuator and sensor attacks can always be designed to be stealthy and unbounded regardless of the locations of poles and zeros of the plant.

CHAPTER 4

DUAL RATE CONTROL FOR DETECTING UNBOUNDED ACTUATOR ATTACKS

4.1 Introduction

Multirate sampling has been studied extensively in the context of sampled-data control in the past and many relevant analysis and synthesis results were obtained (e.g., [48, 49, 50, 51, 52, 53, 54, 55]). An interesting property of multirate sampling is its ability to remove certain unstable zeros of the discrete-time system when viewed in the lifted LTI domain, which in turn allows for fulfilling certain potential design requirements such as gain margin levels, or strong stabilization, that are not possible to satisfy with single rate. We plan to utilize this property and study in detail in the context of stealthy attack detection. We show that the proposed dual rate control structure removes all the vulnerabilities to unbounded stealthy actuator attacks. This is shown to hold also when the plant has more controls than measurements (i.e., a “fat” plant). We show that if a single measurement output remains secure, and if the modes of the system are observable from this output, then dual rate systems always provide the ability to detect actuator as well as combined sensor-actuator attacks.

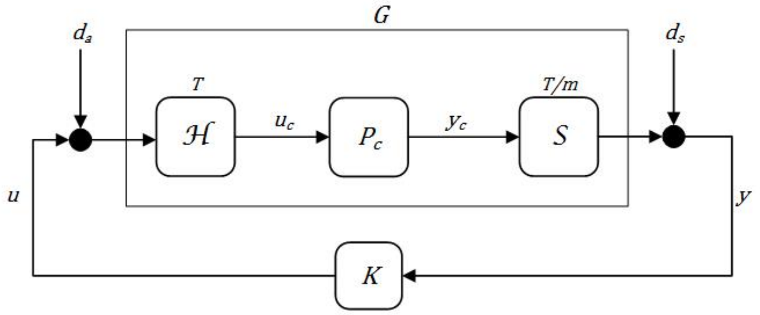


Figure 4.1: A dual rate SD system.

4.2 Analysis of Dual Rate Control Systems with Respect to Detecting Stealthy Actuator Unbounded Attacks

We consider the SD scheme of Figure 4.1 (temporarily without any disturbances) where the output is sampled with period T/m, where m is a sufficiently large integer, i.e., $y(k) = (\mathcal{S}_m y_c)(t) := y_c(kT/m)$. To this end, let the corresponding discrete-time system mapping u to y be

$$G = \mathcal{S}_m P_c \mathcal{H}.$$

For this MR discrete system we have that

$$\Lambda^m G = G\Lambda,$$

where Λ is the 1-step right shift operator on discrete sequences $\{x(k)\}$, i.e., $(\Lambda x)(k+1) = x(k)$ with $(\Lambda x)(0) = 0$. Using standard lifting techniques (e.g., [37]) one can obtain a shift invariant (LTI) description $\tilde{G}$ of the discrete dynamics by grouping the plant input and output signals as $\tilde{u}(k) = u(k)$ and $\tilde{y}(k) = [y_c'(kT/m)\ y_c'((k+1)T/m) \ldots y_c'((k+m-1)T/m)]'$ (similarly for $\tilde{d}_a$ and $\tilde{d}_s$). A state space description for $\tilde{G}$ can be obtained from the original system.

Define state space matrices

$$\begin{aligned} A &:= e^{A_c T/m} \in \mathbb{R}^{n\times n}, \quad B := \int_0^{T/m} e^{A_c \tau} B_c \mathrm{d}\tau \in \mathbb{R}^{n\times n_u}, \\ C &:= C_c \in \mathbb{R}^{n_y \times n}, \qquad D := D_c \in \mathbb{R}^{n_y\times n_u}. \end{aligned}$$

Then

$$\tilde{G} = \left[\begin{array}{c|c} \tilde{A} & \tilde{B} \\ \hline \tilde{C} & \tilde{D} \end{array}\right], \tag{4.1}$$

where

$$\begin{aligned} \tilde{A} &= A^m \in \mathbb{R}^{n\times n}, \tilde{B} = \sum_{k=0}^{m-1} A^k B \in \mathbb{R}^{n\times n_u}, \\ \tilde{C} &= \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{m-1} \end{bmatrix} \in \mathbb{R}^{mn_y\times n}, \\ \tilde{D} &= \begin{bmatrix} D \\ CB + D \\ \vdots \\ C\sum_{k=0}^{m-2} A^k B + D \end{bmatrix} \in \mathbb{R}^{mn_y\times n_u}. \end{aligned}$$

Also, it becomes useful to define a discrete-time system

$$P_m := \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right].$$

This system corresponds to the single-rate sampling and hold scheme of the original plant P_c with a period of T/m, i.e., $P_m = \mathcal{S}_m P_c \mathcal{H}_m$ where $\mathcal{H}_m$ is accordingly generating a continuous signal u_c from the discrete u as $u_c(t) = (\mathcal{H}_m u)(t) = u(k)$ for $kT/m \le t < (k+1)T/m$. It is clear that P_m has the same dimension as P_c; i.e., it maps n_u inputs to n_y outputs. Moreover, given

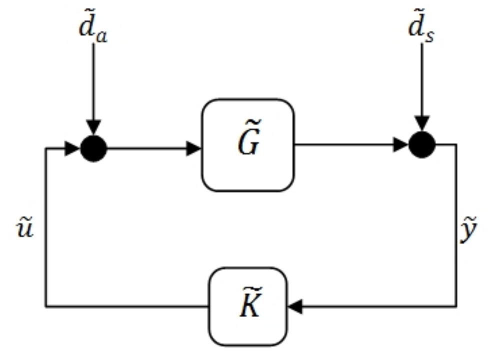


Figure 4.2: The lifted system.

that P_c holds a controllable and observable realization, and the sampling is not pathological, it follows that the inherited realization of P_m is also controllable and observable. Based on our assumptions on the sampling, it is also easily verified that the realization of $\tilde{G}$ as above is controllable and observable. Let $\tilde{M}_{\tilde{G}}$ and $\tilde{N}_{\tilde{G}}$ be the left coprime factors of $\tilde{G}$. We will use the state-space realization of $\tilde{N}_{\tilde{G}}$ as

$$\tilde{N}_{\tilde{G}} = \left[\begin{array}{c|c} \tilde{A} + H\tilde{C} & \tilde{B} + H\tilde{D} \\ \hline \tilde{C} & \tilde{D} \end{array}\right], \tag{4.2}$$

where H is chosen such that $\tilde{A}+H\tilde{C}$ is Schur stable. It is easy to show that $\tilde{G}$ and $\tilde{N}_{\tilde{G}}$ have the same non-minimum phase zeros. We consider now the closed loop in the lifted domain in Figure 4.2 where the controller is $\tilde{K}$ and proceed to argue that the lifted loop is not susceptible to stealthy actuator attacks $\tilde{d}_a$, and thus the original MR loop of Figure 4.1 is not susceptible either. To this end, the integer m is chosen such that the following assumptions are satisfied.

Assumption 17. *The matrix B is full column rank.*

Assumption 18. *The matrix* $\mathcal{O} := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{m-2} \end{bmatrix}$ *is full column rank.*

The first assumption is standard and holds generically if B_c is full column rank in the continuous system. The second assumption holds for large enough m, in particular $m = n + 1$, if the pair (A, C) is observable, which is true as P_m is minimal. It can also hold, however, even with a small m generically. Also, if Assumption 18 holds, $\tilde{G}$ is a tall system. Then the following lemma characterizes the zeros of $\tilde{G}$.

Lemma 19. *Consider the lifted system $\tilde{G}$ as in (4.1) together with Assumptions (17) and (18). Then $\tilde{G}$ has at most one non-minimum phase zero and is located at $\lambda = 1$.*

Proof. Since $\tilde{N}_{\tilde{G}}$ and $\tilde{G}$ have the same non-minimum phase zeros, we will prove this lemma for $\tilde{N}_{\tilde{G}}$. Notice that since $\tilde{N}_{\tilde{G}}$ is tall, $|\lambda_0| \leq 1$ is a zero if and only if there exists a non-zero vector $\nu \in \mathbb{R}^{n_u}$ such that

$$\begin{aligned}
&\tilde{N}_{\tilde{G}}(\lambda_0)\,\nu = \\
&\left[\lambda_0 \tilde{C}\left[I - \lambda_0\left(\tilde{A} + H\tilde{C}\right)\right]^{-1}\left(\tilde{B} + H\tilde{D}\right) + \tilde{D}\right]\nu = 0.
\end{aligned}$$

Notice that

$$\left[I - \lambda_0\left(\tilde{A} + H\tilde{C}\right)\right]^{-1}$$

is well-defined as all the eigenvalues of $\tilde{A} + H\tilde{C}$ are inside the unit circle. Now, let

$$\xi = \left[I - \lambda_0\left(\tilde{A} + H\tilde{C}\right)\right]^{-1}\left(\tilde{B} + H\tilde{D}\right)\nu.$$

Then, pre-multiplying by

$$\left[I - \lambda_0\left(\tilde{A} + H\tilde{C}\right)\right]$$

and using

$$\lambda_0 \tilde{C}\xi + \tilde{D}\nu = 0,$$

we get

$$\begin{aligned} \lambda_0 \tilde{C} \xi + \tilde{D} \nu &= 0, && (4.3) \\ \left(I - \lambda_0 \tilde{A} \right) \xi - \tilde{B} \nu &= 0. && (4.4) \end{aligned}$$

Pre-multiplying (4.3) by X, where X is a matrix in $\mathbb{R}^{(m-1)n_y \times mn_y}$ given as

$$X = \begin{bmatrix} I & -I & 0 & \cdots & 0 \\ 0 & I & -I & & \\ \vdots & & \ddots & \ddots & \\ 0 & \cdots & 0 & I & -I \end{bmatrix}. \tag{4.5}$$

we get

$$\lambda_0 X \tilde{C} \xi + X \tilde{D} \nu = \mathcal{O} \left[\lambda_0 \left(I - A \right) \xi - B \nu \right] = 0.$$

Since $\mathcal{O}$ is full column rank by Assumption 18, it holds true that

$$\lambda_0 \left(A - I \right) \xi + B \nu = 0,$$

which together with (4.4) gives

$$\left[\left(I - \lambda_0 \tilde{A} \right) B + \lambda_0 \left(A - I \right) \tilde{B} \right] \nu = 0.$$

Simplifying further yields

$$\left(1 - \lambda_0 \right) B \nu = 0.$$

Therefore, if ν is nonzero then $\lambda_0 = 1$ since, by Assumption 17, B is full column rank. □

According to Lemma 19, the lifted system, $\tilde{G}$, has no zeros inside the unit circle. However, it may have a zero at $\lambda = 1$. Based on Proposition 13 and Remark 14, an (unbounded) actuator stealthy attack will not be possible if

$\lambda = 1$ is zero of $\tilde{G}$ with multiplicity of at most one. Indeed, this is the case as it is proved in the following theorem:

Theorem 20. *Consider the dual rate SD scheme as in Figure 4.2. Then, there does not exist any (unbounded) actuator stealthy attack if Assumptions 17 and 18 are met.*

Proof. As discussed before, we need to show that $\lambda = 1$ is a zero of $\tilde{G}$ or equivalently $\tilde{N}_{\tilde{G}}$ with the multiplicity of at most one. It can be argued that ([42]-Section 6.5) $\lambda = 1$ is a zero of algebraic multiplicity greater than one if and only if the matrix

$$T := \begin{bmatrix} \tilde{N}_{\tilde{G}}(1) & 0 \\ \frac{\mathrm{d}}{\mathrm{d}\lambda}\tilde{N}_{\tilde{G}}(\lambda)\,|_{\lambda=1} & \tilde{N}_{\tilde{G}}(1) \end{bmatrix}$$

has a right null chain; that is, there exists a vector

$$\nu = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix},$$

with $\nu_1 \neq 0$, such that $T\nu = 0$. By the way of contradiction, we will show that if $T\nu = 0$ then $\nu_1 = 0$. Direct calculations show that if $T\nu = 0$ then

$$\left[\tilde{C}\left[I - \left(\tilde{A} + H\tilde{C}\right)\right]^{-1}\left(\tilde{B} + H\tilde{D}\right) + \tilde{D}\right]\nu_1 = 0, \tag{4.6}$$

$$\begin{aligned}&\left[\tilde{C}\left[I - \left(\tilde{A} + H\tilde{C}\right)\right]^{-2}\left(\tilde{B} + H\tilde{D}\right)\right]\nu_1 \\ &+ \left[\tilde{C}\left[I - \left(\tilde{A} + H\tilde{C}\right)\right]^{-1}\left(\tilde{B} + H\tilde{D}\right) + \tilde{D}\right]\nu_2 = 0.\end{aligned} \tag{4.7}$$

Define

$$\xi_1 \;=\; \left[I - \left(\tilde{A} + H\tilde{C}\right)\right]^{-1}\left(\tilde{B} + H\tilde{D}\right)\nu_1,$$

$$\xi_2 = \left[I - \left(\tilde{A} + H\tilde{C}\right)\right]^{-1}\left[\xi_1 + \left(\tilde{B} + H\tilde{D}\right)\nu_2\right].$$

Pre-multiplying ξ_1 and ξ_2 by $\left[I - \left(\tilde{A} + H\tilde{C}\right)\right]$ and grouping terms we get

$$\begin{aligned}\left(I - \tilde{A}\right)\xi_1 - \tilde{B}\nu_1 &= H\left(\tilde{C}\xi_1 + \tilde{D}\nu_1\right), & (4.8)\\ -\xi_1 + \left(I - \tilde{A}\right)\xi_2 - \tilde{B}\nu_2 &= H\left(\tilde{C}\xi_2 + \tilde{D}\nu_2\right). & (4.9)\end{aligned}$$

From (4.6)-(4.9),

$$\begin{aligned}\tilde{C}\xi_1 + \tilde{D}\nu_1 &= 0, & (4.10)\\ \tilde{C}\xi_2 + \tilde{D}\nu_2 &= 0, & (4.11)\\ \left(I - \tilde{A}\right)\xi_1 - \tilde{B}\nu_1 &= 0, & (4.12)\\ -\xi_1 + \left(I - \tilde{A}\right)\xi_2 - \tilde{B}\nu_2 &= 0. & (4.13)\end{aligned}$$

Furthermore, pre-multiplying (4.10) and (4.11) gives

$$\begin{aligned}X\tilde{C}\xi_1 + X\tilde{D}\nu_1 &= \mathcal{O}\left[(I - A)\xi_1 - B\nu_1\right] = 0,\\ X\tilde{C}\xi_2 + X\tilde{D}\nu_2 &= \mathcal{O}\left[(I - A)\xi_2 - B\nu_2\right] = 0,\end{aligned}$$

where X is as in (4.5), which in turn imply

$$\begin{aligned}(I - A)\xi_1 - B\nu_1 &= 0, & (4.14)\\ (I - A)\xi_2 - B\nu_2 &= 0. & (4.15)\end{aligned}$$

Eliminating ξ_2 between (4.13) and (4.15), we get

$$-(I - A)\xi_1 - \left[(I - A)\tilde{B} - \left(I - \tilde{A}\right)B\right]\nu_2 = 0.$$

Notice that $(I - A)\tilde{B} - \left(I - \tilde{A}\right)B = 0$ and hence the last equation implies

$$(I - A)\xi_1 = 0,$$

which in turn, together with (4.14), implies $B\nu_1 = 0$. By Assumption 17, $B\nu_1 = 0$ implies $\nu_1 = 0$ and this completes the proof. □

As a final comment from the previous analysis, we offer conditions when $\tilde{G}$ has a zero $\lambda = 1$. We note that, as proved in the previous theorem, these zeros are not a problem since they cannot generate unbounded stealthy attacks.

Proposition 21. *Let P_c be "tall." Then $\tilde{G}$ has a zero at $\lambda = 1$ if and only if P_m does.*

Proof. Suppose $\tilde{G}$ has a zero at $\lambda = 1$. Then, there exist vectors ξ and ν, at least one of them nonzero, such that (4.3) and (4.4) hold for $\lambda_0 = 1$. In particular, from (4.3) we get

$$C\xi + D\nu = 0. \tag{4.16}$$

Furthermore, pre-multiplying (4.3) by X results in

$$\mathcal{O}\left[(I - A)\,\xi - B\nu\right] = 0$$

which in turn implies

$$(I - A)\,\xi - B\nu = 0. \tag{4.17}$$

(4.16) and (4.17) imply that P_m has a zero at $\lambda = 1$.

Conversely, if P_m has a zero at $\lambda = 1$,

$$\begin{bmatrix} I - A & -B \\ C & D \end{bmatrix} \begin{bmatrix} \xi \\ \nu \end{bmatrix} = 0, \tag{4.18}$$

for some ξ and ν. Pre-multiplying it by

$$\begin{bmatrix} \sum_{k=1}^{m-1} A^k & 0 \\ 0 & I \\ -C & I \\ -C - CA & I \\ \vdots & \\ -C\sum_{k=0}^{m-2} A^k & I \end{bmatrix} \tag{4.19}$$

gives

$$\begin{bmatrix} I - \tilde{A} & -\tilde{B} \\ \tilde{C} & \tilde{D} \end{bmatrix} \begin{bmatrix} \xi \\ \nu \end{bmatrix} = 0.$$

That is, $\tilde{G}$ has a zero at $\lambda = 1$. □

Proposition 22. *Let P_c be "fat." Then $\tilde{G}$ has always a zero at $\lambda = 1$.*

Proof. The proof relies on the fact that since P_c or equivalently P_m is fat, there always exist two vectors ξ and ν with at least one of them nonzero such that (4.18) holds. Then, the rest of the proof follows similarly to that of the converse part of Proposition 21. □

Remark 23. *We would like to point out that an equivalent way of obtaining the same results, i.e., ability to detect zero attacks, is to hold the control input longer rather than sampling the output faster. That is, if we consider a dual rate system where the hold operates with a period of mT while the output is sampled with T, then the corresponding lifted system will enjoy the same properties as before in terms of unstable zeros. Obviously, the (nominal) controller performance will be reduced as the control is slower. On the other hand, there is a potential benefit of lower cost of actuation in this case. An example offered in the next session illustrates this point.*

4.3 Examples

In this section we provide examples of systems that are vulnerable to zero dynamics attacks, and then we apply the techniques presented in the previous sections to remove these vulnerabilities. Using one of the examples, we also present an approach to perform controller design trade-offs.

4.3.1 Quadruple-Tank

This example is a Quadruple-Tank Process (QTP) [56] that was used for system security analysis in [57]. The continuous-time nonlinear plant model is given by

$$\begin{aligned}
\dot{h}_1(t) &= -\frac{a_1}{A_1}\sqrt{2gh_1(t)} + \frac{a_3}{A_1}\sqrt{2gh_3(t)} + \frac{\gamma_1 k_1}{A_1}u_1(t) \\
\dot{h}_2(t) &= -\frac{a_2}{A_2}\sqrt{2gh_2(t)} + \frac{a_4}{A_1}\sqrt{2gh_4(t)} + \frac{\gamma_2 k_2}{A_2}u_2(t) \\
\dot{h}_3(t) &= -\frac{a_3}{A_3}\sqrt{2gh_3(t)} + \frac{(1-\gamma_2)k_2}{A_3}u_2(t) \\
\dot{h}_4(t) &= -\frac{a_4}{A_4}\sqrt{2gh_4(t)} + \frac{(1-\gamma_1)k_1}{A_4}u_1(t),
\end{aligned}$$

where A_i is the cross-section area tank i, a_i the cross-section area of the outlet hole, h_i the height of water in tank i, k_i is pump constants, γ_i the flow ratios and g the gravity acceleration. We regard the outputs as the water levels of tanks 1 and 2, i.e. h_1 and h_2. The voltage applied to pump i is u_i, and the corresponding flow is $k_i u_i$. At a certain operating condition, the system is linearized and sampled at $T = 0.5$ sec to get the following discrete-time open loop system [57]:

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

$$B_d = \begin{bmatrix} 0.00515 & 0.0016 \\ 0.0019 & 0.00447 \\ 0 & 0.0737 \\ 0.0850 & 0 \end{bmatrix}, C_d = \begin{bmatrix} 0.2 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \end{bmatrix}, D_d = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The discrete system has an unstable zero at $\lambda = 0.97$ with direction $\nu = \begin{bmatrix} 0 & 0 & 1 & -0.96 \end{bmatrix}'$ which indicates that stealthy actuator attacks of the form $d_a(k) = \epsilon\nu(.97)^{-k}$ are possible. Next, we apply multirate control to move the unstable zero outside the unit circle. We sample faster at rate $T = 0.5/2 = 0.25$ sec while keeping the hold at rate $T = 0.5$ sec. The resulting open loop state space representation after lifting is

$$\tilde{A} = A_d, \ \tilde{B} = B_d,$$
$$\tilde{C} = \begin{bmatrix} 0.2 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0.1975 & 0 & 0.0043 & 0 \\ 0 & 0.1977 & 0 & 0.0045 \end{bmatrix},$$
$$\tilde{D} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0.005183 & 8.095e-005 \\ 9.437e-005 & 0.004496 \end{bmatrix}.$$

The new open loop system has no unstable zeros, which indicates that it is not susceptible to stealthy actuator attacks. We note that only a small m is enough to accomplish our goal, i.e., $m = 2$. In fact, Assumption 18, a sufficient condition for Theorem 13 to hold, is not even satisfied in this case and yet the unstable zeros are removed. We also note that the unstable zero at the single rate system was due to the unstable zero of the continuous time dynamics. On top of physical unstable zeros, sampling can create additional ones as indicated in the following power system example, where a simulation of a stealthy attack is shown.

4.3.2 Automatic Voltage Regulator

The automatic voltage regulator (AVR), or the generator excitation control, specifies the terminal voltage magnitude of a synchronous generator by controlling the reactive power. A simplified block diagram of a linearized AVR is shown in Figure 4.3 [58]. An increase in the reactive power load of the generator results in a drop in the voltage magnitude across its terminals. The voltage drop is sensed by a potential transformer and then is rectified and compared to the reference voltage magnitude. The error signal is then amplified and raises the generator terminal voltage by controlling the excitation field. For a set of typical system parameters $K_A = 10, \tau_A = 0.1, K_E = 1, \tau_e = 0.4, K_g = 1, \tau_g = 1, K_r = 1, \tau_r = 0.05$ as in Figure 4.3, the open loop state space representation of the single rate system after discretization at a sample rate $T = 0.5$ sec is

$$A_d = \begin{bmatrix} 0.0105 & 0.3949 & 3.86 & 2.869 \\ -0.0057 & -0.1817 & -1.369 & -0.587 \\ 0.00117 & 0.03359 & 0.1793 & -0.4597 \\ 0.00092 & 0.03197 & 0.3163 & 0.8918 \end{bmatrix},$$

$$B_d = \begin{bmatrix} -0.005738 \\ 0.001174 \\ 0.0009193 \\ 0.0002165 \end{bmatrix}, C_d = \begin{bmatrix} 0 & 0 & 0 & 5000 \end{bmatrix}, D_d = \begin{bmatrix} 0 \end{bmatrix},$$

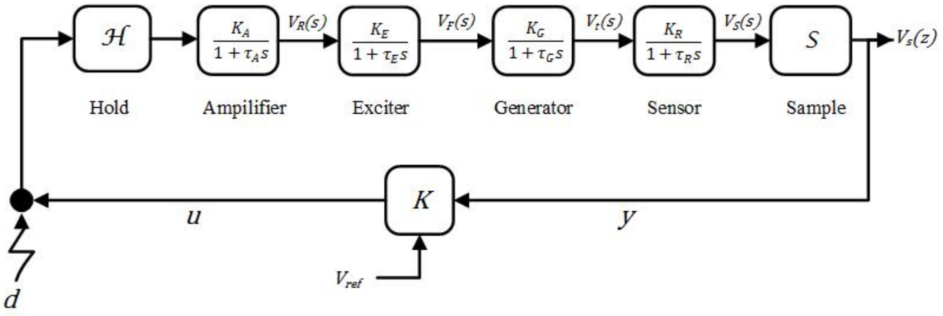
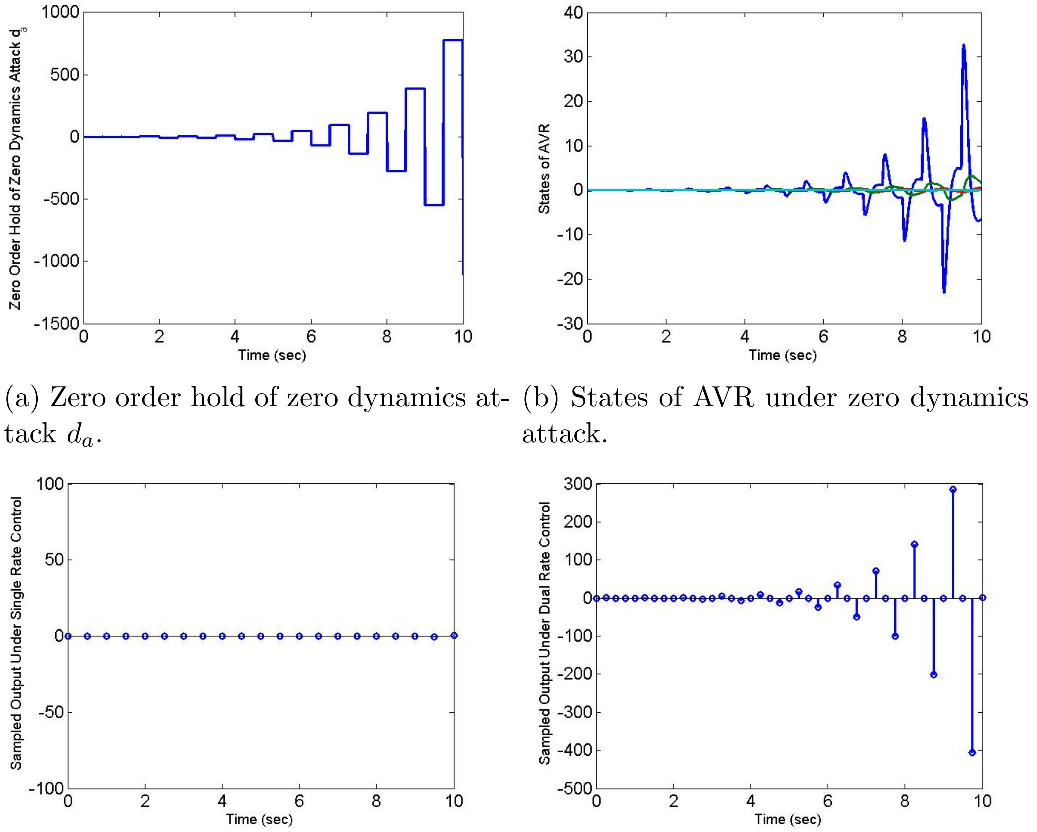


Figure 4.3: A simplified automatic voltage regulator block diagram.

which has an unstable zero at $\lambda = -0.7045$. We note that although the continuous system has no unstable zeros, sampling at the relatively slow rate of 0.5 sec per sample created an unstable zero. Next, we consider an attack input of the form $d_a(k) = \epsilon {z_0}^{-k}$, where ϵ is a small number and z_0 is the zero of the system. Figures 4.4a–4.4c show a plot of the attack held at $T = 0.5$ sec along with the states and the sampled output of the system. We can notice that while the states of the system are exploding, the sampled output remains zero and no attack is detected. Next, we change the single rate block diagram to a multirate architecture to move the unstable zero to the safe region. We sample faster at rate $T/m = 0.5/2 = 0.25$ sec per sample while keeping the hold at rate $T = 0.5$ sec. The resulting open loop state space representation after lifting is

$$\tilde{A} = A_d,\ \tilde{B} = B_d, \tilde{C} = \begin{bmatrix} 0 & 0 & 0 & 5000 \\ 2.185 & 86.13 & 1092 & 4902 \end{bmatrix}, \tilde{D} = \begin{bmatrix} 0 \\ 0.196 \end{bmatrix}.$$

The resulting open loop system has no unstable zeros. We note that only a small m is enough to accomplish our goal, i.e., $m = 2$. Again, Assumption 18, a sufficient condition for Theorem 13 to hold, is not even satisfied in this case and yet the unstable zeros are removed. We consider the same attack input as above and simulate the system. The sampled output at rate $T = 0.25$ sec is shown in Figure 4.4d. It is obvious that the multirate scheme detects the



(a) Zero order hold of zero dynamics attack d_a.

(b) States of AVR under zero dynamics attack.

(c) Sampled output of AVR using single rate control.

(d) Sampled output of AVR using dual rate control.

Figure 4.4: (a)-(c) show simulation of zero dynamics attack on a sampled-data AVR system under single rate control. (d) Shows the sampled output under dual rate control.

attack on the system.

4.3.3 Automatic Generation Control

The main objectives of any AGC system are to maintain the frequency of the grid and to maintain the power interchanges between neighboring areas at their scheduled values. This is achieved by controlling the units participating in AGC to follow the load profile and correct for errors in the load forecast. Figure 4.5 shows a load frequency control (LFC) block diagram of a single machine [59], [58]. A change in frequency ($\Delta\omega$) is sensed by the governor, which in turn orders the turbine to raise or lower the generation of electric power until the frequency is stabilized. The figure shows what is called primary frequency control augmented with a secondary controller K to make sure the error in frequency settles to zero. The open loop state space representation of the single rate AGC system after discretization at a sample rate $T = 0.5$ sec is

$$\frac{d}{dt}\begin{bmatrix}\Delta\omega\\ \Delta P_{mech}\\ \Delta P_{valve}\end{bmatrix} = \begin{bmatrix}0.78 & 0.03 & 0.008\\ -8.34 & 0.23 & 0.163\\ -16.17 & -0.42 & -0.019\end{bmatrix}\begin{bmatrix}\Delta\omega\\ \Delta P_{mech}\\ \Delta P_{valve}\end{bmatrix} + \begin{bmatrix}0.009\\ 0.424\\ 0.832\end{bmatrix}\Delta P_{ref}$$

$$y = \begin{bmatrix}1 & 0 & 0\end{bmatrix}\begin{bmatrix}\Delta\omega\\ \Delta P_{mech}\\ \Delta P_{valve}\end{bmatrix},$$

which has an unstable zero at $\lambda = -0.5721$. We consider an attack input of the form $d_a(k) = \epsilon {z_0}^{-k}$ where ϵ is a small number and z_0 is the zero of the system. Figures 4.6a–4.6c show plots of the attack held at 0.5 sec along with the states and the continuous and sampled output of the system. Next, we change the single rate block diagram to a multirate architecture to move the unstable zero to the safe region. We sample faster at rate $T/m = 0.5/2 = 0.25$ sec per sample while keeping the hold at rate $T = 0.5$ sec. The resulting open

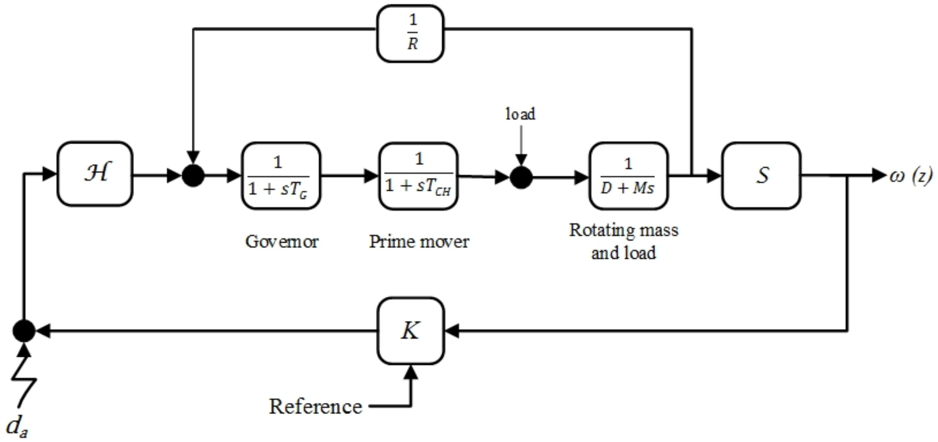


Figure 4.5: A simplified automatic generation control block diagram.

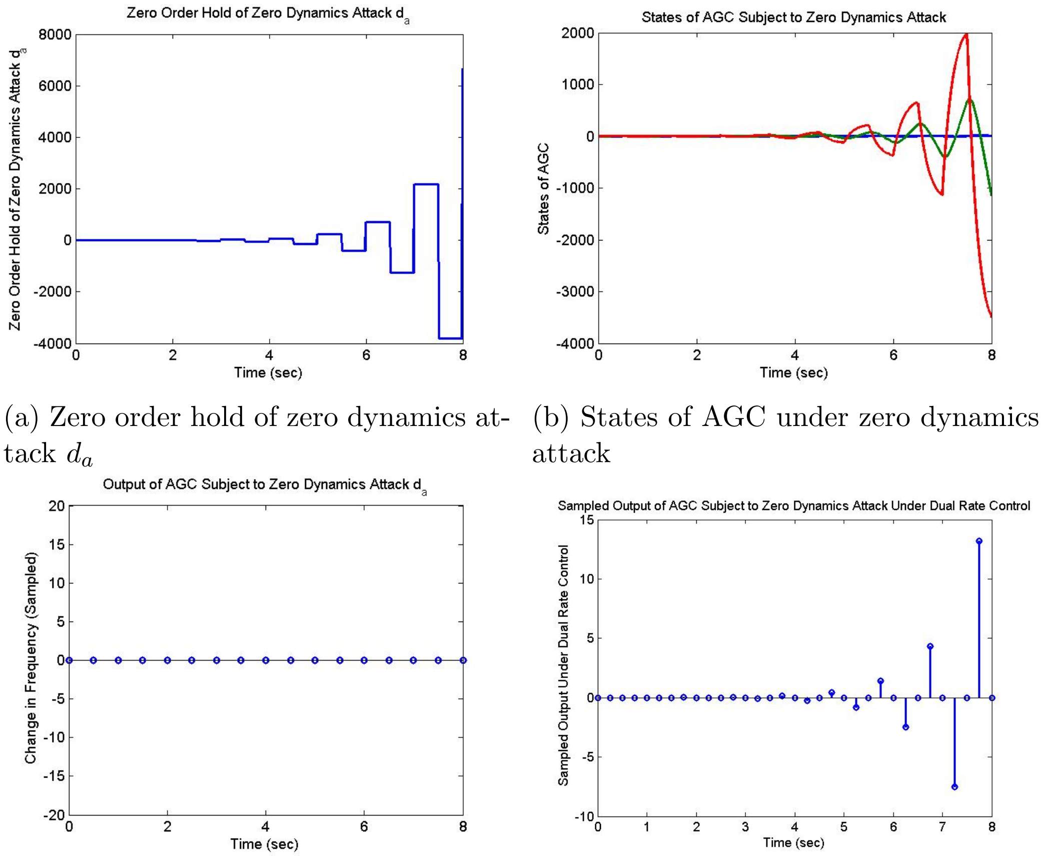
loop state space representation after lifting is

$$y = \begin{bmatrix} 1 & 0 & 0 \\ 0.9462 & 0.01925 & 0.003557 \end{bmatrix} \begin{bmatrix} \Delta\omega \\ \Delta P_{mech} \\ \Delta P_{valve} \end{bmatrix} + \begin{bmatrix} 0 \\ 0.001711 \end{bmatrix} \Delta P_{ref},$$

where A and B stay the same. For the multirate scheme, we consider the same attack input as above and simulate the system. The sampled output at rate $T = 0.25$ sec is shown in Figure 4.6d. The simulation shows that the multirate scheme detects the attack on the system.

4.3.4 Controller Trade-Offs

In this section we consider the automatic voltage regulation system previously discussed in order to investigate trade-offs in the controller design. In particular, we will close the control loop by designing linear quadratic Gaussian (LQG) controllers for the dual rate system and compare the cost with that of single rate LQG controllers. We set up a baseline LQG formulation for the dynamics of the open loop AVR with sampled measurements with a



(a) Zero order hold of zero dynamics attack d_a

(b) States of AGC under zero dynamics attack

(c) Sampled output of AGC using single rate control

(d) Sampled output of AGC using dual rate control

Figure 4.6: (a)-(c) show simulation of zero dynamics attack on a sampled-data AGC system under single rate control. (d) shows the sampled output under dual rate control.

period s as

$$\mathrm{d}x(t) = A_c x(t)\mathrm{d}t + \omega_c(t)\mathrm{d}t + B_c u_c(t)\mathrm{d}t,$$
$$y(k) = C_c x(ks) \ + v(k), \ k = 0, 1, \ldots$$

We assume that the process noise $\{\omega_c(t), t \geq 0\}$ is a Brownian motion with $\mathcal{E}\{\mathrm{d}\omega_c(t)\mathrm{d}\omega_c(t)'\} = \Xi_c$, the observation noise $\{v(k), k = 0, 1...\}$ is a zero mean white Gaussian sequence with covariance $\Theta = \mathcal{E}\{v(k)v(k)'\}$, and $x(0)$ is zero mean Gaussian with covariance $S_0 = \mathcal{E}\{x(0)x(0)'\}$. Moreover, it is assumed that the random variables $x(0), v(k), \omega_c(t)$ are independent. We assume that the hold period is h. The objective is to minimize the following cost:

$$J = \mathcal{E}\left\{ \limsup_{k\to\infty}(1/kh) \int_0^{kh} (x'Q_c x + u_c R_c u_c')\mathrm{d}t \right\}$$

with the usual positive definiteness conditions $Q_c = Q_c' \geq 0$ and $R_c = R_c' > 0$, which transforms to

$$J = \mathcal{E}\left\{ \limsup_{k\to\infty}(1/k) \sum_{k=0}^{\infty} (x_k'Qx_k + 2x_k'Su_k + u_k'Ru_k) \right\}$$

with $x_k := x(kh)$, $u_k := u_c(kh)$ and

$$\begin{aligned} Q &= \int_0^h e^{A_c'\tau} Q_c e^{A_c\tau}\mathrm{d}\tau \\ S &= \int_0^h e^{A_c't} Q_c \big(\int_0^t e^{A_c(t-\tau)} B_c \mathrm{d}\tau \big)\mathrm{d}t \\ \Xi &= \int_0^s e^{A_c(s-\tau)} B_c \Xi_c B_c' e^{A_c'(s-\tau)}\mathrm{d}\tau \end{aligned}$$

$$R = \textstyle\int_0^h \big[\big(\int_0^t B_c' e^{A_c'(t-\tau)}\mathrm{d}\tau \big) Q_c \big(\int_0^t e^{A_c(t-\tau)} B_c \mathrm{d}\tau \big) + R_c \big]\mathrm{d}t.$$

The hold and sample periods h and s are assumed to be integer related and in particular $h = ms$ with $m = 1, 2, \ldots$. In this synchronous dual rate case, rather than using lifting techniques to solve the problem, we take the separation principle approach which applies also to asynchronous sampling

Table 4.1: LQG Cost

$s\backslash h$	0.5	1
0.25	0.6704	0.6705
0.5	0.6868	0.6877
1	0.7033	0.7047

(e.g., [60]) to find the optimal cost by computing

$$J_o = \mathrm{trace}\big[PF'(R + B_h'X)B_hF + X\Xi\big],$$

where X and P are the unique positive semidefinite symmetric solutions of the algebraic Riccati equations

$$\begin{aligned} X &= A_h'XA_h - (S + A_h'B_hX)(R + B_h'XB_h)^{-1}(XB_h'A_h + S') + Q \\ P &= A_sPA_s' - A_sPC_s'(C_sPC_s' + \Theta)^{-1}C_sPA_s' + \Xi \end{aligned}$$

and

$$F = (B_h'XB_h + R)^{-1}(B_h'XA_h + S'),$$

where the various A, B, C matrices above are corresponding to the matrices in Equation (3.1) for $T = h$ and $T = s$, i.e., $A_h = e^{A_ch}$, $A_s = e^{A_cs}$, etc. Table 4.1 summarizes the LQG cost for different single rate and dual rate sample and hold for the case $\Xi_c = 10^3$, $\Theta = 10$, $Q_c = I_4$ and $R_c = I$. The entries (s, h) for which stealthy attacks are not possible are (0.25, 0.5), (0.25, 1), and (0.5, 1). We notice that, because of faster sampling, we get better performance in dual rate control than single rate control. Also, as expected, we get better performance between dual rate controllers when we increase the rate of sampling rather than decreasing the rate of the zero order hold. Faster sampling, however, may require more expensive devices and so a trade-off is present.

4.4 Conclusion

We presented a simple dual rate sampled data scheme which guarantees detectability of unbounded actuator and/or sensor attacks, if a secure output that maintains observability of the open loop modes is available. The main observation is that the sampled data nature in the implementation of the cyber-physical system cannot be ignored as sampling can generate additional vulnerabilities due to the extra unstable zeros it may introduce, particularly if high rates are necessary to achieve certain performance level. The proposed method solves this issue by the use of multirate sampling that ensures that zeros exist only in harmless locations in the lifted domain. A few examples were also presented that show how the multirate scheme detects the unbounded actuator attacks. In addition, the examples incorporated a dual rate controller cost comparison based on LQG control.

Several other possibilities can be studied in this context. The use of asynchronous sampling (e.g., [49, 61]) can provide alternative ways to detect stealthy attacks; or even the network's random delays can be helpful in that respect. The speed of detecting, however, needs to be taken into consideration, even if the attack is detectable. The methods of generalized holds [62] are also relevant as they move zeros, and with careful analysis of their robustness properties (e.g., [63, 64]) can provide acceptable and simple solutions as well.

CHAPTER 5

ON THE COMPUTATION OF WORST ATTACKS: AN LP FRAMEWORK

5.1 Introduction

In this work, we consider signal attacks where the general problem from the attacker's perspective is to find the attack input $d = \{d(k)\}$ so that it is stealthy while inflicting the maximum damage on the performance variable $z = \{z(k)\}$. We showed in the previous chapters that unbounded attacks for LTI systems are related to the unstable zeros and/or poles of the open loop system. However, in this chapter we consider the problem of characterizing the worst bounded and stealthy attacks. This problem involves a maximization of a convex function subject to convex constraints. We propose different attack resource constraints to make the problem more practical. More specifically, we assume that the attacker has a finite time window $\{0, 1, \ldots, t_a\}$ to attack the system and inflict the maximum damage before the attack is over, and we attempt to solve the following three attack scenarios:

- Scenario 1: Attacker can attack in a finite time window up to $t = t_a$; his goal is to inflict the maximum damage anywhere (before or after t_a) while remaining stealthy for all t.
- Scenario 2: Attacker can attack in a finite interval up to $t = t_a$; his goal is to inflict the maximum damage anywhere (before or after t_a) while remaining stealthy for $t \leq t_a$ (does not care if detected after the attack is over).
- Scenario 3: Attacker can attack in a finite interval up to $t = t_a$; his goal

is to inflict the maximum damage at $t \leq t_a$ while remaining stealthy for $t \leq t_a$.

We show that by employing a ℓ_∞ framework, tractable linear programming (LP) methods can be used to compute the worst attack for the above three scenarios.

Our work is closely related to [23], [25], [26], [28], [65], [66], [67] and [24]. However, we do not assume a constant d such as in [23] where they assume the system is in steady state. In addition, the work in the mentioned references does not address attack impact and stealthiness after the attack is concluded, nor does it relate to either a specific detection method (e.g. residual detectors) which assumes certain thresholding mechanisms that may be stochastic, or to a specific controller in use. We study these problems in a more general input-output fashion that does not depend on the particular controller used, and in a totally deterministic worst case scenario. In other words, the assumed noise thresholds are based on the existence of a worst case magnitude bounded noise. In this sense, the noise is allowed to "conspire" with the attacker to keep the detection signals within what is assumed normal operation.

In the second part of this chapter, we build on the worst attack design problem and provide a novel K-d controller synthesis iterative method to minimize the performance cost without increasing the impact of the worst attack. Each iteration is an LP and alternates between finding the worst attack d for a given controller K, and finding the next K that minimizes the performance cost while keeping a non-increasing upper bound on the worst case impact inflicted by d.

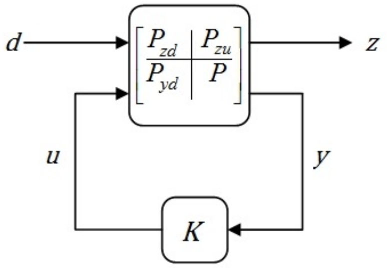


Figure 5.1: General setup of input-output maps.

5.2 Problem Setup

We consider the case of a general signal attack d on a closed loop system of Figure 5.1. Let $\Phi(K)$ describe the effect of d on the performance variable z and on the monitoring signal ψ, i.e. let

$$\Phi = \begin{bmatrix} \Phi_{zd} \\ \Phi_{\psi d} \end{bmatrix} =: d \mapsto \begin{bmatrix} z \\ \psi \end{bmatrix}.$$

The monitoring signal ψ consists of the measured output y and the control signal u; it can however contain any other information that is recorded and measured, e.g., reference inputs. In this setup, we assume that there may be other external disturbances and noise inputs which are normal, i.e., not malicious attackers, which are not shown in the figure. Also, the entire formulation deals with discrete-time systems and signals.

The attacker's goal can be stated in general as

$$\begin{aligned} &\max_{d} \|z\|_\infty \\ &\text{s.t. } \|\psi\|_\infty \leq \theta, \end{aligned} \tag{5.1}$$

where θ is an alarm threshold, associated with the afore mentioned normal set of disturbances. In Chapter 2, we established exact conditions for stealthiness of unbounded actuator and sensor attacks which can totally destroy the system. These attacks are ultimately related to the open loop plant P, and for LTI systems in particular, to the non-minimum phase (unstable) zeros and unstable poles of P. We note, as pointed out in Chapter 1, that unstable zeros can also be due to the sampled data implementation of controllers.

In this general setup of Figure 5.1, we elaborate on the existence of stealthy unbounded attacks using an input-output approach. In particular, considering a left coprime factorization ([47, 43, 42]) for the part of the generalized system that connects inputs to the measured output

$$y = P_{yd}d + Pu$$

in the open loop, we have

$$[P_{yd}\ P] = \tilde{M}^{-1}[\tilde{N}_{yd}\ \tilde{N}].$$

Using a left coprime factorization for the stabilizing controller $K = YX^{-1}$, we can express

$$\psi = \begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix} W^{-1}\tilde{M}P_{yd}d = \begin{bmatrix} X \\ Y \end{bmatrix} W^{-1}\tilde{N}_{yd}d,$$

where

$$W = \tilde{M}X - \tilde{N}Y.$$

Since W is stable and, by stability of the closed loop, has a stable inverse W^{-1} we have that the detectability of d depends on the unstable zeros of $\tilde{N}_{yd}$; i.e., unbounded stealthy attacks d are possible if and only if $\tilde{N}_{yd}$ has unstable zeros.

For actuator-only attacks

$$P_{yd} = P = \tilde{M}^{-1}\tilde{N} \Longrightarrow \tilde{N}_{yd} = \tilde{N}$$

while for sensor-only attacks

$$P_{yd} = I = \tilde{M}^{-1}\tilde{M} \Longrightarrow \tilde{N}_{yd} = \tilde{M}.$$

Hence, this shows how the unstable zeros of P (which are the unstable zeros of $\tilde{N}$) and the unstable poles of P (which are the unstable zeros of $\tilde{M}$) relate to the actuator and sensor attacks considered in Chapter 2. Multirate sampling can potentially remove unstable zeros of $\tilde{N}_{yd}$ as was shown in Chapter 3 for unbounded actuator attacks, but it cannot work for total sensor unbounded attacks.

In the following we consider the case of bounded in magnitude (and time) attacks with various levels of stealth. The question we want to address is how to compute the worst possible bounded attacks and how to defend against such attacks by a suitable controller design.

5.3 Computation of Worst Attack

We consider the problem of computing the worst case attack in (5.1) when the attacker has a finite time window $\{0, 1, \ldots, t_a\}$ to attack the system. In addition, we require the attack to remain stealthy after the attack is over. This allows for repeatedly attacking the system without triggering the monitoring signal alarm.

Specifically, consider the optimization problem in (5.1); we are interested in finding the worst, stealthy, bounded (in magnitude and time) attack. Assume the LTI closed loop system $\Phi(K)$ is stable and let t_{zd} and $t_{\psi d}$ be design parameters related to the decay rate of the pulse responses of of Φ_{zd} and $\Phi_{\psi d}$ respectively. These parameters determine the time windows that the attacker cares about for impact and stealthiness respectively. Suppose the intruder can only attack the system during a finite interval $\{0, 1, \ldots, t_a\}$, with attack magnitude less than or equal to α. Then, a corresponding problem of interest can be formulated as

$$
\begin{aligned}
&\max_{d} \|z\|_{\infty}^{[0,t_a+t_{zd}]} \\
&\text{s.t. } \|\psi\|_{\infty}^{[0,t_a+t_{\psi d}]} \leq \theta, \\
&\quad |d(k)| \leq \alpha, \;\; k = 0, 1, \dots, t_a, \\
&\quad d(k) = 0, \quad k = t_a + 1, \dots,
\end{aligned}
\tag{5.2}
$$

where

$$
\|z\|_{\infty}^{[0,t_a+t_{zd}]} = \max_{0 \leq k \leq t_a + t_{zd}} = |z(k)|\,,
$$

and similarly

$$
\|\psi\|_{\infty}^{[0,t_a+t_{\psi d}]} = \max_{0 \leq k \leq t_a + t_{\psi d}} = |\psi(k)|\,.
$$

The system of equations governing the output z when subjected to the attack input d for each instance of time is given by

$$
\begin{bmatrix} z(0) \\ z(1) \\ \vdots \\ z(t_a) \\ z(t_a+1) \\ \vdots \\ z(t_a+t_{zd}) \end{bmatrix}
=
\begin{bmatrix}
\Phi_{zd}(0) & 0 & 0 & \cdots \\
\Phi_{zd}(1) & \Phi_{zd}(0) & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots \\
\Phi_{zd}(t_a) & \Phi_{zd}(t_a-1) & \Phi_{zd}(t_a-2) & \cdots \\
\Phi_{zd}(t_a+1) & \Phi_{zd}(t_a) & \Phi_{zd}(t_a-1) & \cdots \\
\vdots & \vdots & \vdots & \vdots \\
\Phi_{zd}(t_a+t_{zd}) & \Phi_{zd}(t_a+t_{zd}-1) & \Phi_{zd}(t_a+t_{zd}-2) & \cdots
\end{bmatrix}
\begin{bmatrix} d(0) \\ d(1) \\ \vdots \\ d(t_a) \\ 0 \\ \vdots \\ 0 \end{bmatrix},
\tag{5.3}
$$

where $d(k) = 0$ for $t > t_a$. The objective is to find the sequence $\{d(k)\}$, $k = \{0, \dots, t_a\}$ that maximizes $\|z\|_{\infty}^{[0,t_a+t_{zd}]}$ such that $\|\psi\|_{\infty}^{[0,t_a+t_{\psi d}]} \leq \theta$. This corresponds to selecting the optimal row in (5.3) to be maximized and finding the optimal d that would maximize this row. In view of the above, the following proposition is obvious.

Proposition 24. *Problem* (5.2) *can be formulated as the following optimiza-*

tion problem:

$$\begin{aligned}
&\max_{d,n\in\{0,1,\ldots,t_a+t_{zd}\}} \sum_{k=0}^{n} \Phi_{zd}(n-k)d(k) \\
&s.t. \left|\sum_{k=0}^{\tau} \Phi_{\psi d}(\tau-k)d(k)\right| \leq \theta,\ \tau = 0,1,\ldots,t_a+t_{\psi d}, \\
&\quad |d(k)| \leq \alpha,\ k = 0,1,\ldots,t_a, \\
&\quad d(k) = 0, \quad k = t_a+1,\ldots\ .
\end{aligned} \tag{5.4}$$

After finding the worst case attack $\hat{d}$, the worst case impact can be obtained by computing $\left\|\Phi_{zd}\hat{d}\right\|_\infty$.

Note also that an optimal $\hat{d}$ can always be selected so that

$$\left|\sum_{k=0}^{n} \Phi_{zd}(n-k)\hat{d}(k)\right| = \sum_{k=0}^{n} \Phi_{zd}(n-k)\hat{d}(k),$$

thus the expression for the cost in (5.4).

Remark 25. *The objective function looks for the optimal row in the set $\{0,\ldots,t_a+t_{zd}\}$. We can always choose a sufficiently long t_{zd}, determined by the decay rate of Φ_{zd} and the bound α on d, to ensure that we capture the worst case $\|z\|_\infty = \sup_{t_{zd}} \|z\|_\infty^{[0,t_a+t_{zd}]}$.*

Remark 26. *Note that the first set of constraints ensures the monitoring signal ψ is below a threshold level ($\|\psi\|_\infty \leq \theta$) during and after the attack interval. Since we assume that $\Phi_{\psi d}$ is stable and that $d(k) = 0$ for $t > t_a$, if $t_{\psi d}$ is chosen long enough, depending on the decay rate of $\Phi_{\psi d}$ and the bound α, one can guarantee that d is undetectable for all t. Therefore, to guarantee stealthiness for all t it is sufficient to enforce the monitoring constraints up to $t_a+t_{\psi d}$. The last set of constraints ensures the attack is bounded and decays to zero at the end of the attack interval.*

Remark 27. *Remarks 25 and 26 basically state that for a priori computable*

t_{zd} and $t_{\psi d}$, problems (5.2) *and* (5.4) *solve the following problem:*

$$\begin{aligned} &\max_{d} \|z\|_{\infty} \\ &s.t.\ \ \|\psi\|_{\infty} \leq \theta, \\ &\qquad |d(k)| \leq \alpha,\ \ k = 0, 1, \ldots, t_a, \\ &\qquad d(k) = 0, \quad k = t_a + 1, \ldots\ . \end{aligned} \tag{5.5}$$

Remark 28. *Problem* (5.4) *is LP for a fixed n (fixed row) which can be solved efficiently. Fixing n transforms the objective function to a linear function under linear (polytopic) constraints. However, one has to solve (in principle) $t_a + t_{zd}$ LPs.*

Following is a simple search algorithm (Algorithm 1) to solve Problem 5.4:

Algorithm 1 Compute worst attack $\hat{d}$

```
Input $\Phi_{zd}$, $\Phi_{\psi d}$, $t_a$, $t_{zd}$, $t_{\psi d}$, $t_d$, $\theta$ and $\alpha$.
for $i = 1 : t_a + t_{zd}$ do
    Solve
        $\max_{d_i} \sum_{k=0}^{i} \Phi_{zd}(i-k) d_i(k)$
        s.t. $\left| \sum_{k=0}^{\tau} \Phi_{\psi d}(\tau - k) d_i(k) \right| \leq \theta,\ \tau = 0, 1, \ldots, t_a + t_{\psi d},$
             $|d_i(k)| \leq \alpha,\ k = 0, 1, \ldots, t_a,$
             $d_i(k) = 0, \quad k = t_a + 1, \ldots, t_a + t_d.$
    Compute and store $\|\Phi_{zd} d_i\|_{\infty}$
end
Compare $\|\Phi_{zd} d_i\|_{\infty}$ and determine the worst attack $\hat{d}$.
```

In the sequel, we consider certain cases which further simplify the computations. Specifically, we consider the problem of computing the worst case attack when the attacker has a finite time window $k = \{0, \ldots, t_a\}$ to attack the system such as in Proposition 24. However, in this case we assume that the intruder does not mind being detected after the attack is over, i.e., stealthiness constraints are checked up to $t = t_a$ only. The following corollary describes how to construct the optimal d.

Corollary 29. *Consider the optimization Problem in* (5.2) *with* $t_{\psi d} = 0$ *(finite stealthiness interval). Then, its solution can be obtained by solving*

$$\begin{aligned} &\max_{d, n \in \{t_a, \dots, t_a + t_{zd}\}} \sum_{k=0}^{n} \Phi_{zd}(n-k)d(k) \\ &s.t. \Big| \sum_{k=0}^{\tau} \Phi_{\psi d}(\tau - k) d(k) \Big| \leq \theta, \ \tau = 0, 1, \dots, t_a, \\ &|d(k)| \leq \alpha, \ k = 0, 1, \dots, t_a, \\ &d(k) = 0, \quad k = t_a + 1, \dots \ . \end{aligned} \tag{5.6}$$

Proof. We will prove that the optimal row to be maximized is in the set $\{z(t_a), \dots, z(t_a + t_{zd})\}$. Let $\hat{d}$ be the worst attack that maximizes $\mu =: \|z\|_\infty^{[0, t_a + t_{zd}]}$ found by solving for the maximum impact over all the rows of (5.3) where the stealthiness constraints are enforced up to $t = t_a$. Assume that $\hat{d}$ was found by maximizing any row before $z(t_a)$ calling it row i. Since the stealthiness constraints are imposed only up to t_a and $\Phi(K)$ is LTI, we can delay $\hat{d}$ by $t_a - i$ steps (shift $\hat{d}$ to the right) so that $\|z\|_\infty^{[0, t_a + t_{zd}]} = \mu$ is achieved by maximizing the row $z(t_a)$ without violating the stealthiness constraints. In addition, we cannot shift the attack beyond $z(t_a)$ since $\hat{d}(k) = 0$ for $t > t_a$. As a result, maximizing $\|z\|_\infty^{[0, t_a + t_{zd}]}$ is equivalent to maximizing $\|z\|_\infty^{[t_a, t_a + t_{zd}]}$ for $t_{\psi d} = 0$.

□

Remark 30. *The optimization problem in* (5.6) *differs from the problem in* (5.4) *in two ways: First, the stealthiness constraints set in* (5.6) *is a subset of the set in* (5.4)*, since in* (5.6) *the objective is to remain stealthy only during the attack interval, where in* (5.4) *the stealthiness condition is enforced at all times. Therefore, the attack designed using Corollary 29 yields worse impact in the* ℓ_∞ *sense than the attack designed using Proposition 24. The second* *difference is in the objective function where in* (5.4) *we have to maximize each row in* (5.3) *to find the worst attack (i.e.,* $t_a + t_{zd}$ *LPs), while in* (5.6) *we only need to maximize the last rows associated with* $[z(t_a), \dots, z(t_a + t_{zd})]'$ *(i.e.,* $t_{zd} + 1$ *LPs). An immediate corollary is as follows.*

Corollary 31. *Let* $t_{zd} = t_{\psi d} = 0$*, i.e., the attacker cares to inflict maximum damage in the window up to* t_a *while he does not care for stealthiness after* t_a*. Then, the optimal* d *is obtained by solving the following single LP:*

$$\begin{aligned} &\max_{d} \sum_{k=0}^{t_a} \Phi_{zd}(t_a - k) d(k) \\ &s.t. \left| \sum_{k=0}^{\tau} \Phi_{\psi d}(\tau - k) d(k) \right| \leq \theta, \ \tau = 0, 1, \dots, t_a, \\ &\quad |d(k)| \leq \alpha, \ k = 0, 1, \dots, t_a. \end{aligned} \tag{5.7}$$

The above corollary states that computing the worst attack when the attack impact and stealthiness constraints are desired to be inside the attack interval only is equivalent to solving (5.6) for $n = t_a$.

Remark 32. *If* $\Phi_{\psi d}$ *has an unstable zero that is not found in* Φ_{zd}*, and* α *is not specified, then the optimization problems in Corollary 29 and Corollary 31 will yield unbounded zero dynamics attacks (Chapter 4).*

5.4 Example - Worst Attack Computation

In this section we work on an example of a real power system component and compute the worst attack for different scenarios.

5.4.1 Automatic Voltage Regulator

The automatic voltage regulator (AVR) or the generator excitation control, specifies the terminal voltage magnitude of a synchronous generator

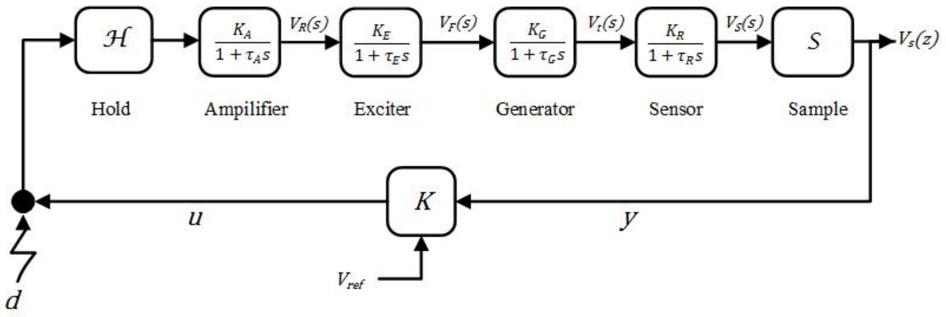


Figure 5.2: A simplified automatic voltage regulator block diagram.

by controlling the reactive power. A simplified block diagram of a linearized AVR is shown in Figure 5.2 [58]. For a set of typical system parameters $K_A = 10, \tau_A = 0.1, K_E = 1, \tau_e = 0.4, K_G = 1, \tau_G = 1, K_R = 1, \tau_R = 0.05$ as in Figure 4.3. We consider actuator attacks as depicted in Figure 4.3 and seek to find the attack with the worst impact on V_F (excitation voltage) while keeping the monitoring vector

$$\psi = \begin{bmatrix} y \\ u \end{bmatrix}$$

below a noise level threshold θ. Let K be a suitable controller for the system and let

$$\begin{aligned} P &= \mathcal{S} \frac{K_A}{1+\tau_A s} \frac{K_E}{1+\tau_E s} \frac{K_G}{1+\tau_G s} \frac{K_R}{1+\tau_R s} \mathcal{H} \quad \text{and} \\ P_F &= \mathcal{S} \frac{K_A}{1+\tau_A s} \frac{K_E}{1+\tau_E s} \mathcal{H}. \end{aligned} \tag{5.8}$$

Then closed loop system $\Phi(K)$ describing the effect of d on $z = V_F$ and the monitoring vector ψ is given by

$$\Phi(K) = \begin{bmatrix} \Phi_{zd} \\ \Phi_{\psi d} \end{bmatrix} =: d \mapsto \begin{bmatrix} z \\ y \\ u \end{bmatrix} = \begin{bmatrix} \frac{P_F}{1+PK} \\ \frac{P}{1+PK} \\ \frac{PK}{1+PK} \end{bmatrix}.$$

Given

$$K = \frac{0.1z - 0.09}{z - 1},$$

then $\Phi(K)$ becomes

$$\begin{bmatrix} \Phi_{zd} \\ \\ \Phi_{\psi d} \end{bmatrix} = \begin{bmatrix} \frac{0.8423z^4 - 1.162z^3 - 0.1551z^2 + 0.5433z - 0.06808}{z^5 - 3.186z^4 + 3.794z^3 - 2.043z^2 + 0.4705z - 0.03522} \\ \\ \frac{0.01114z^4 + 0.05639z^3 - 0.03266z^2 - 0.03337z - 0.001502}{z^5 - 3.186z^4 + 3.794z^3 - 2.043z^2 + 0.4705z - 0.03522} \\ \\ \frac{(1.11z^4 + 5.75z^3 - 2.59z^2 - 2.99z - 0.135) \times 10^{-3}}{z^5 - 3.186z^4 + 3.794z^3 - 2.043z^2 + 0.4705z - 0.03522} \end{bmatrix}$$

sampled at $T = 0.1$ sec. We note that the has $\Phi_{\psi d}$ an unstable zero at $z = 1.42$. We compute the attack for 3 cases:

- In the first case we employ Proposition 24 to compute the worst attack for an attack interval $\{0, \dots, t_a\}$ that is stealthy for all t.
- In the second case, we compute the worst attack for an attack interval $\{0, \dots, t_a\}$ using Proposition 29, i.e., stealthiness requirement for $t \leq t_a$ only.
- In the third case, we compute the worst attack using Corollary 31, i.e., $\max\limits_{d} z(t_a)$ where the stealthiness requirement holds for $t \leq t_a$ only.

For all cases, we fix $t_a = 500$ (corresponding to 5 seconds), $\theta = 0.1$, $\alpha = 100$. Figures 5.3, 5.4 and 5.5 show the computed worst attack signals for cases 1, 2 and 3 with their impact on the performance variable z and and monitoring signal ψ. Case 1 was obtained by maximizing $z(260)$ (corresponding to 2.6 seconds), Case 2 was obtained by maximizing $z(520)$ (corresponding to 5.2 seconds) and case was obtained by maximizing $z(500)$ (corresponding to 2.6 seconds). We note that the maximum impact on z in case 2 is larger than in case 3 which in turn is larger than in case 1, confirming remark 30.

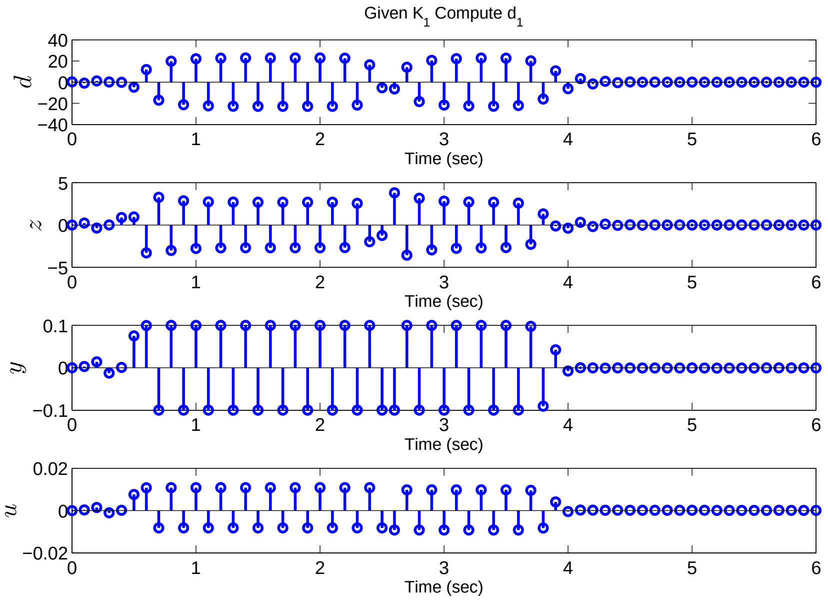


Figure 5.3: Case 1 - Stealthy for all t. Worst attack computation with effect on z, y and u.

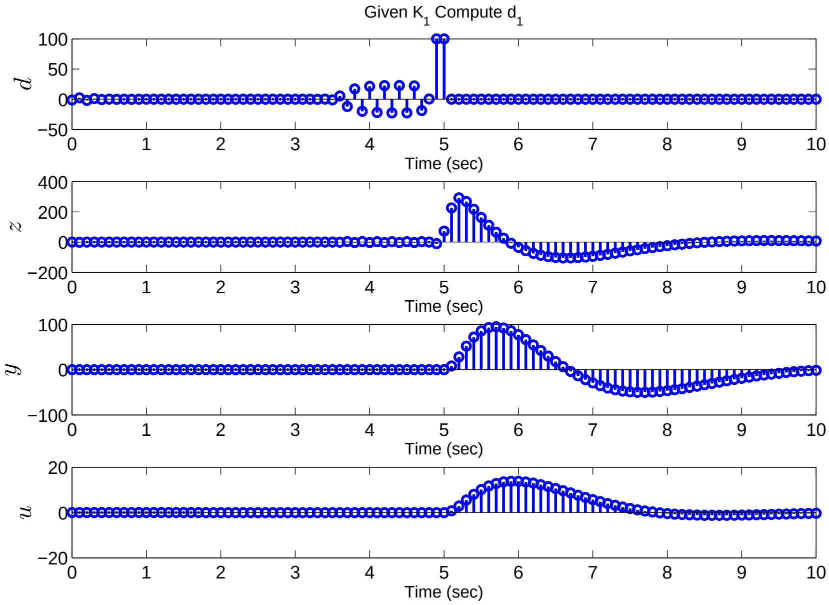


Figure 5.4: Case 2 - Stealthy for $t \leq t_a$ and $t_{zd} \neq 0$. Worst attack computation with effect on z, y and u.

We also show in Figure 5.6a a plot for the maximum impact on z for all t_a for case 1. This is obtained by iterating $t_a \geq 0$ and solving (5.4) until

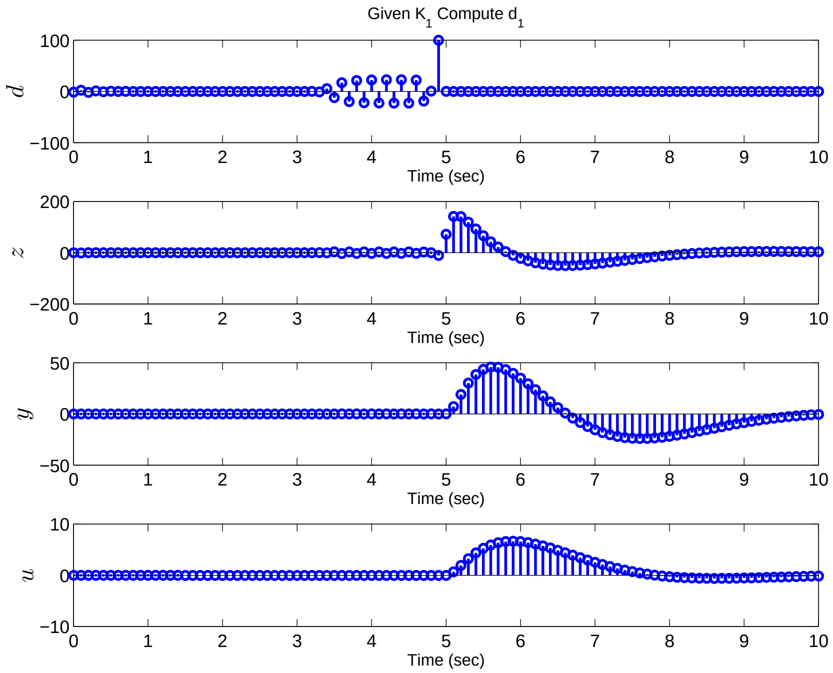
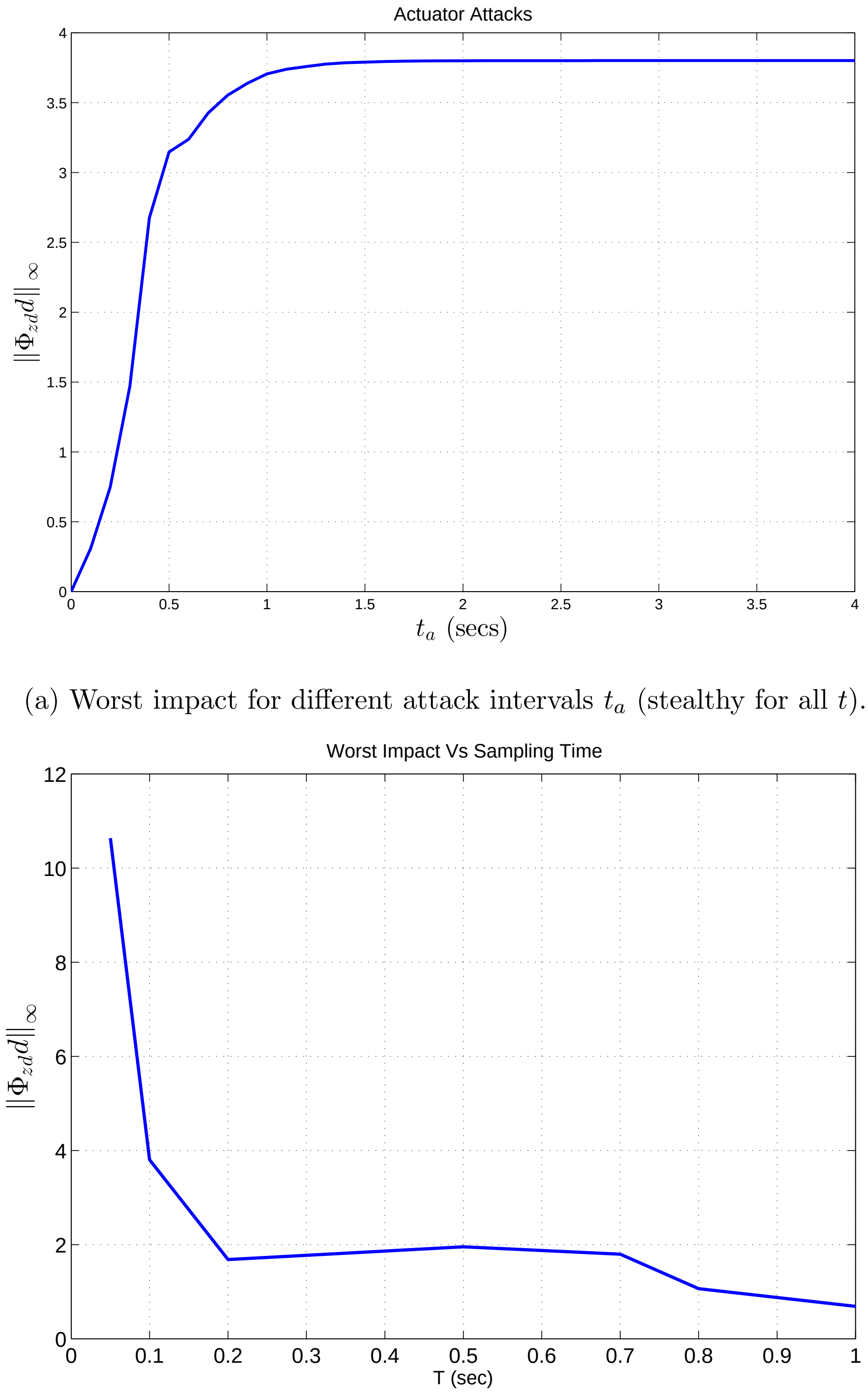


Figure 5.5: Case 3 - Stealthy for $t \leq t_a$ and $t_{zd} = 0$. Worst attack computation with effect on z, y and u.

$\|z\|_\infty$ stops increasing. We note from Figure 5.6a that $\|z\|_\infty$ stops increasing after $t_a = 200$ (corresponding to 2 sec). As a result, for this example solving (5.1) is equivalent to solving (5.4) for $t_a \geq 200$. Furthermore, we show in Figure 5.6b how of the worst impact yielded by the optimization problem in 5.4 changes with changing the sampling and hold time (T). It is not clear from the figure if a direct relationship between $\|z\|_\infty$ and T exists. This is because although for a faster rate the cardinality of the attack sequence for a fixed time interval increases allowing for extra optimization variables, the number of stealthiness constraints also increases, reducing the set of feasible solutions.



(a) Worst impact for different attack intervals t_a (stealthy for all t).

(b) Worst impact for different sample and hold rate (stealthy for all t).

Figure 5.6: (a): Worst impact for or different attack intervals t_a.
(b): Worst impact for different sample and hold rate.

5.5 Controller Design for Resiliency - K-d iteration

In view of the previous discussion, a controller design procedure can be formulated based on LP. In particular, given a desired ℓ_1 performance level γ for attacks d, find K such that $\|\Phi_{zd}(K)\|_1 \leq \gamma$, and to ensure that for a given attack level characterized by $\|d\|_\infty \leq \alpha$, where α is an attack resource parameter, the "undetected loss" of the closed loop given by

$$\begin{aligned} \mu_\alpha := \max_d \; & \|\Phi_{zd}(K)d\|_\infty \\ \text{s.t.} \; & \|\Phi_{\psi d}(K)d\|_\infty \leq \theta, \\ & \|d\|_\infty \leq \alpha \end{aligned}$$

remains below a desired level μ. Computing μ_α for a given K corresponds to the problem of computing the worst d of the previous section. A synthesis procedure can be developed by a "K-d" type of iteration:

- Given K_i with $\|\Phi_{zd}(K_i)\|_1 = \gamma_i$ find d_i from:

$$\begin{aligned} \mu_i := \max_d \; & \|\Phi_{zd}(K_i)d\|_\infty \\ \text{s.t.} \; & \|\Phi_{\psi d}(K_i)d\|_\infty \leq \theta, \\ & \|d\|_\infty \leq \alpha. \end{aligned}$$

- Given d_i find K_{i+1} from:

$$\begin{aligned} \gamma_{i+1} := \min_K \; & \|\Phi_{zd}(K)\|_1 \\ \text{s.t.} \; & \|\Phi_{zd}(K)d_i\|_\infty \leq \mu_i. \end{aligned}$$

- At each iteration i the problem is an LP with

$$\gamma_i \leq \gamma_{i-1} \leq \gamma_0,$$

$$\mu_i \leq \gamma_i \|d_i\|_\infty ,$$

$$\|d_i\|_\infty \leq \alpha.$$

The above formulation guarantees that the upper bound on the attack impact (i.e. μ_i) is non-increasing with each iteration.

5.6 Example - Controller Design for Resiliency

In this section we build on the AVR example in Section 5.4.1 sampled at $T = 0.1$ sec and seek to design a controller that minimizes the performance variable z while possibly minimizing the impact of the worst attack d. Similar to Section 5.4.1 we start with a simple PI controller represented by the transfer function

$$K_1 = \frac{0.1z - 0.09}{z - 1}.$$

We use controller parametrization for stable transfer functions to set up the controller optimization problem. As a result, the maps Φ_{zd} and $\Phi_{\psi d}$ are given by

$$\begin{bmatrix} \Phi_{zd} \\ \Phi_{\psi d} \end{bmatrix} =: d \mapsto \begin{bmatrix} z \\ y \\ u \end{bmatrix} = \begin{bmatrix} \dfrac{P_F}{1+PK} \\ \dfrac{P}{1+PK} \\ \dfrac{PK}{1+PK}. \end{bmatrix} = \begin{bmatrix} P_F(1-PQ) \\ P(1-PQ) \\ PQ \end{bmatrix},$$

where

$$Q = \frac{K}{1+PK}$$

and P is the open loop transfer function of the AVR system given in (5.8) along with P_F, both sampled at $T = 0.1$ sec. The controller synthesis problem is carried on the affine parameter Q in the time domain [68, 69] using the following formulation:

Table 5.1: Outcomes of γ_i and μ_i For Each Iteration

Iteration i	$\gamma_i = \|\Phi_{zd}(K_i)\|_1$	$\mu_i = \|\Phi_{zd}(K_i)d_i\|_\infty$
1	64.6449	3.8008
2	37.0244	2.0107
3	36.9109	0.5704
4	36.9109	0.5704

- Given K_i with $\|\Phi_{zd}(K_i)\|_1 = \gamma_i$ find d_i from:

$$\begin{aligned}\mu_i = & \max_{d,n\in\{0,1,\ldots,t_a+t_{zd}\}} \sum_{k=0}^{n} \Phi_{zd}(n-k)d(k)\\ \text{s.t.} & \left|\sum_{k=0}^{\tau} \Phi_{\psi d}(\tau-k)d(k)\right| \leq \theta,\ \tau = 0,1,\ldots,t_a+t_{\psi d},\\ & |d(k)| \leq \alpha,\ k=0,1,\ldots,t_a,\\ & d(k) = 0, \quad k = t_a+1,\ldots\ .\end{aligned}$$

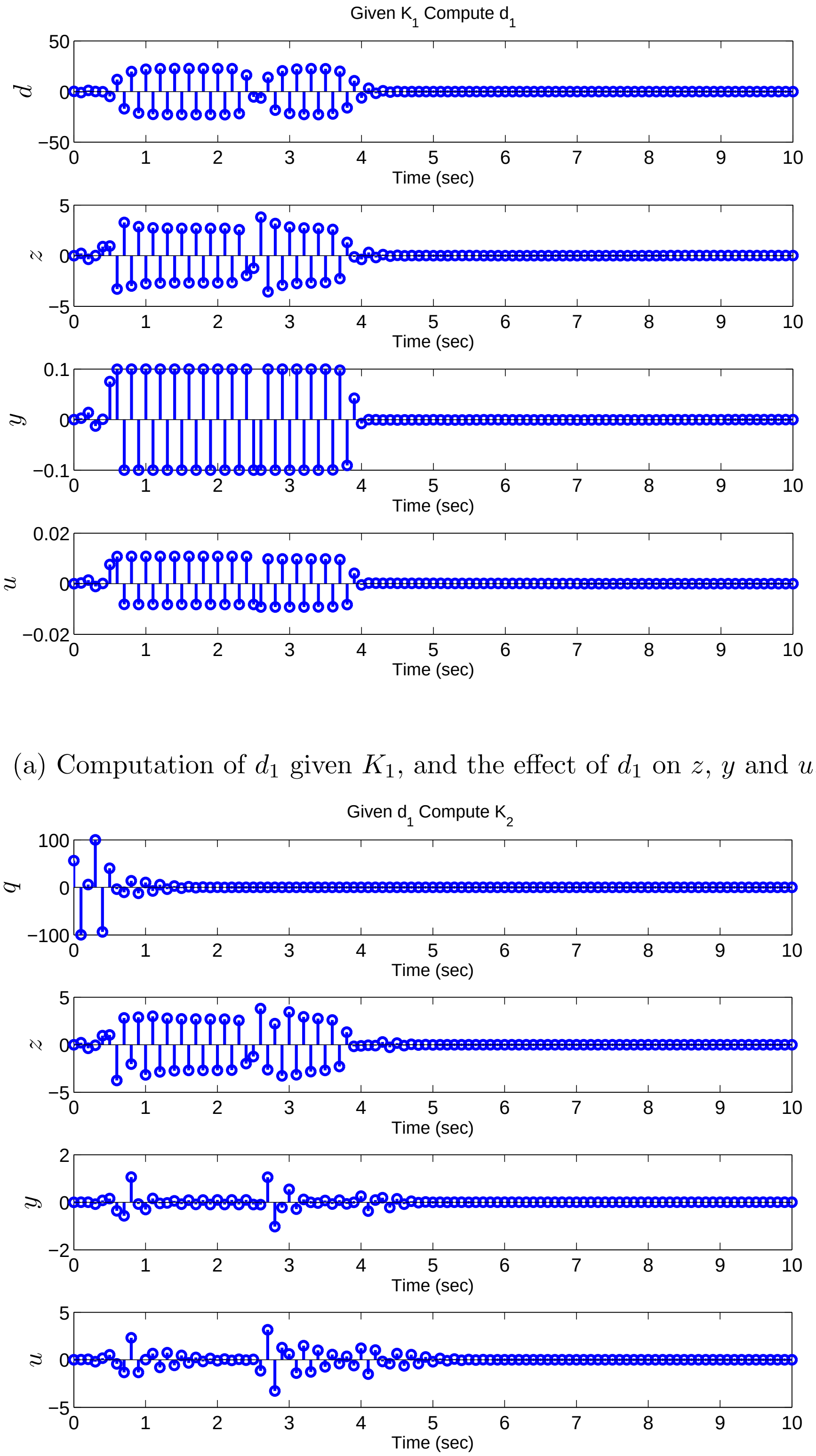
- Given d_i find K_{i+1} from:

$$\begin{aligned}\gamma_{i+1} := \quad & \min_q \|p_F * (1-p*q)\|_1\\ & \text{s.t.} \quad \|p_F*(1-p*q)*d_i\|_\infty \leq \mu_i\\ & \qquad \|q\|_\infty \leq \beta\\ & \qquad q(t) = 0, \quad t \geq t_q,\end{aligned}$$

where $p = \{p(k)\}$, $p_F = \{p_F(k)\}$ and $q = \{q(k)\}$ are the pulse responses of P, P_F and Q respectively, and t_q and β are design constraints for shaping the controller. The problem is solved for the following parameters: $t_a = 500$, $t_q = 500$, $\theta = 0.1$, $\alpha = 100$, and $\beta = 100$. The optimal Q is finite impulse response (FIR). Table 5.1 shows the outcome of the controller synthesis iterative procedure. From the table we see that at each iteration we improved the performance and reduced the impact of the worst d until no further improvement is feasible. Figures 5.7-5.10 show the results of the first and last iterations of the controller design process. Figures 5.7a, 5.8a, 5.9a, 5.10b plot the computed worst attack d and its impact on the variables z, y and u. Figures 5.7b, 5.8b, 5.10a plot the optimized controller parameter impulse response q, and the effect of the previous d on the variables z, y and u governed by the new controller (i.e. $\Phi_{zd}(K_{i+1})d_i$ and $\Phi_{\psi d}(K_{i+1})d_i$). We note that for iteration 1, although $\|\Phi_{zd}(K_1)d_1\|_\infty = \|\Phi_{zd}(K_2)d_1\|_\infty$, d_1 is no longer optimal for the next iteration because $\|\Phi_{\psi d}(K_2)d_1\|_\infty \geq \theta$ as seen in Figure 5.7b.

5.7 Conclusions

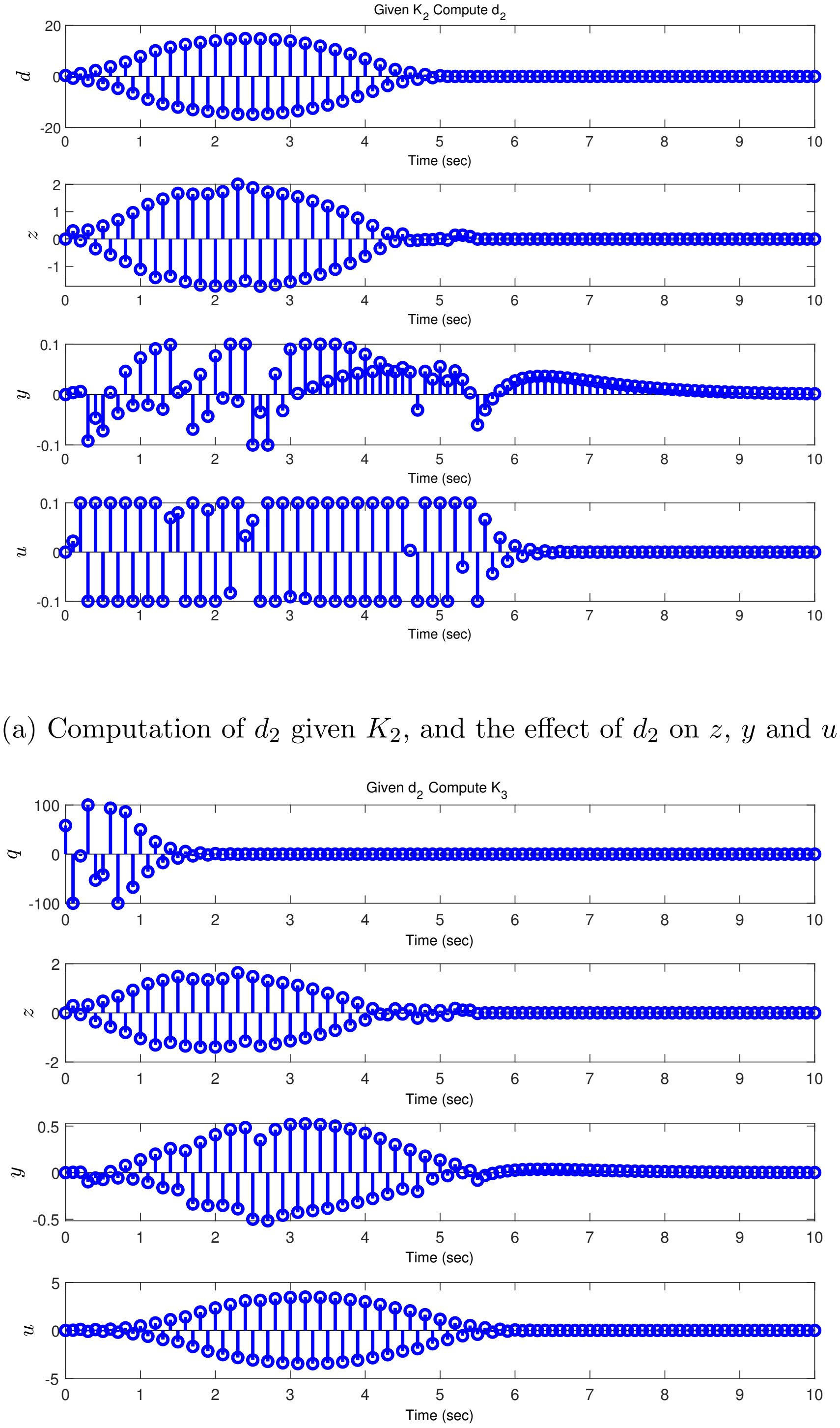
We considered the problem of computing worst case bounded stealthy false data injection attacks for LTI systems. We considered different attack resource constraints and stealthiness intervals. This problem involves a maximization of a convex function subject to convex constraints, and it was shown that it can be cast as a series of LP problems under the ℓ_∞ framework. A search algorithm is constructed to solve the set of LPs and was used to compute the worst stealthy attacks on AVR systems. Furthermore, we provided an iterative controller synthesis procedure that alternates between computing worst attacks and designing optimal controllers that enhance performance and minimize the impact of worst attacks. We used this method to design a controller for the AVR system that resulted in a substantial decrease in the worst impact inflicted by the worst attack.



(a) Computation of d_1 given K_1, and the effect of d_1 on z, y and u.

(b) Computation of q_2, and the effect of d_1 on z, y and u controlled by K_2.

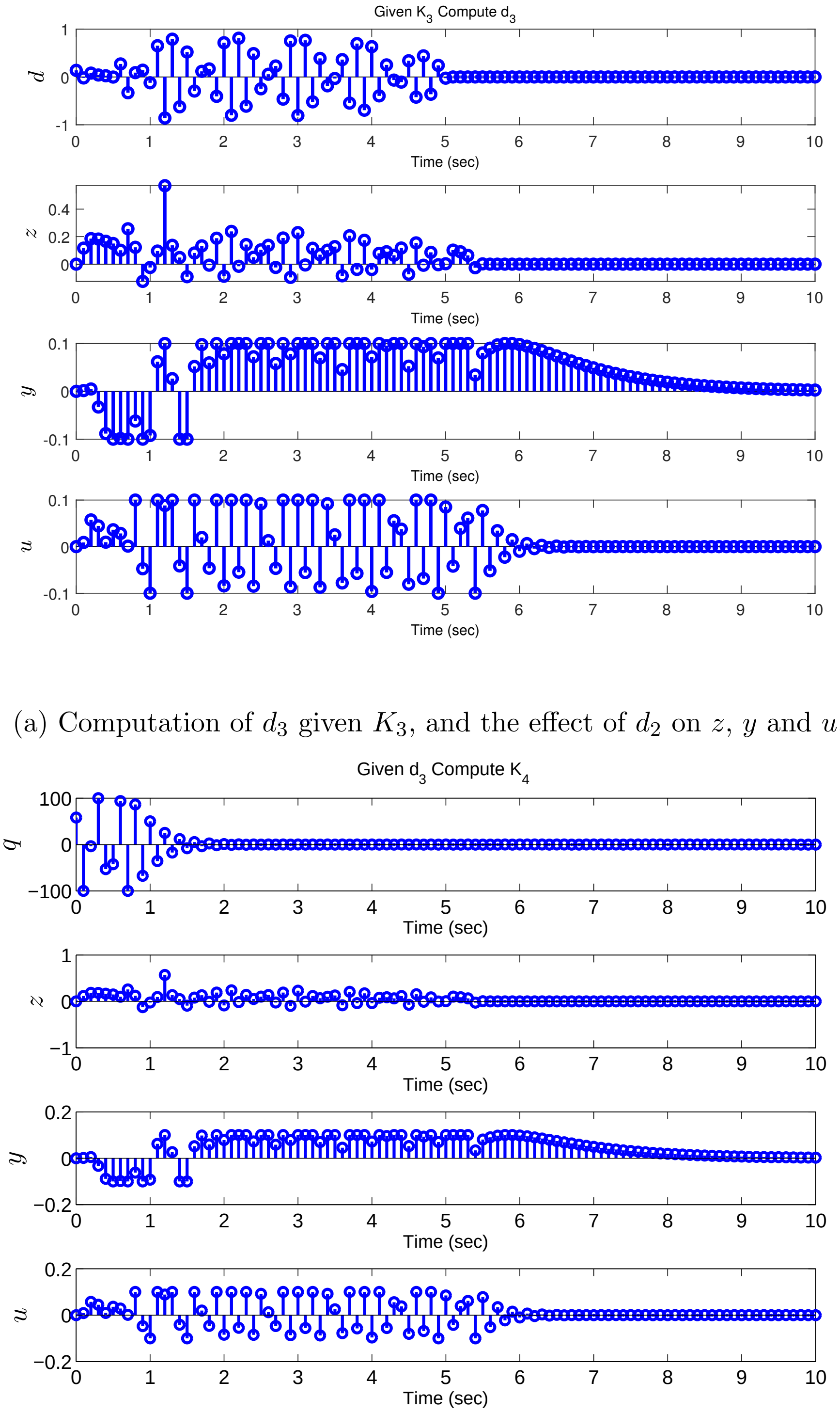
Figure 5.7: Controller synthesis using K-d iteration. Iteration 1.



(a) Computation of d_2 given K_2, and the effect of d_2 on z, y and u.

(b) Computation of q_3, and the effect of d_2 on z, y and u controlled by K_2.

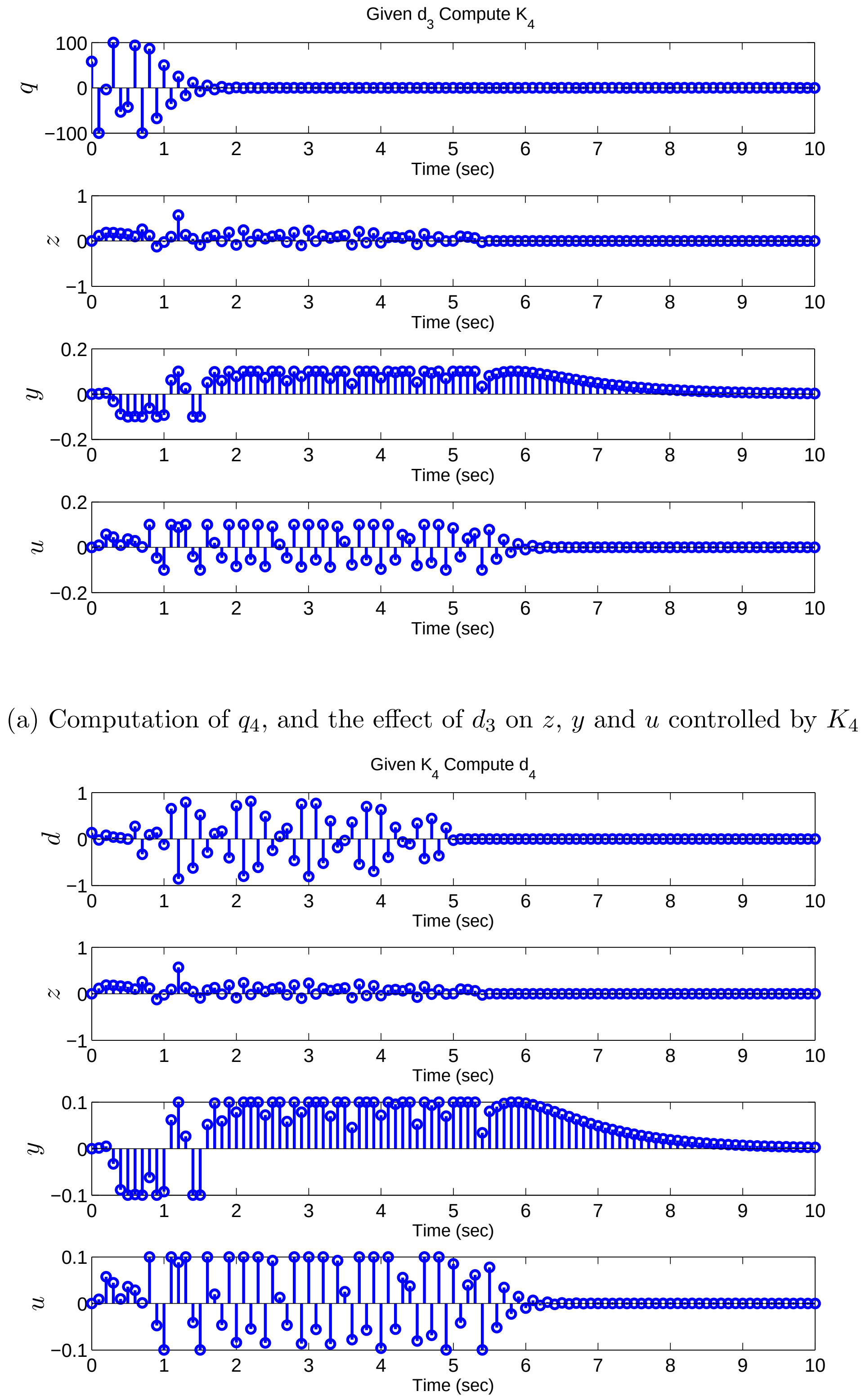
Figure 5.8: Controller synthesis using K-d iteration. Iteration 1.



(a) Computation of d_3 given K_3, and the effect of d_2 on z, y and u.

(b) Computation of q_4, and the effect of d_3 on z, y and u controlled by K_2.

Figure 5.9: Controller synthesis using K-d iteration. Iteration 1.



(a) Computation of q_4, and the effect of d_3 on z, y and u controlled by K_4.

(b) Computation of d_4 given K_4, and the effect of d_4 on z, y and u.

Figure 5.10: Controller synthesis using K-d iteration. Iteration 4.

CHAPTER 6

ON THE ESTIMATION OF SIGNAL ATTACKS

6.1 Introduction

We consider the problem of estimating signal attacks on actuators and/or sensors of control systems using the available measurements. The estimated attack signal will help the operator decide whether it is a persistent intelligent attack or just a nominal disturbance. First, we show that the design of controller for estimation and controller for rejection are coupled, and that a trade-off exists between their individual performances. The quality of the estimate depends on the performance of the attack rejection controller. In particular, the faster and better we reject the attack, the worse is the attack estimate. This is of course assuming the attack can be detected or seen from the outputs used for estimation.

Next we consider multirate (MR) sampling to estimate the injected attack d. In particular, we consider the case where we have two sets of sensors measuring the output. The first set is sampled at the same rate of the hold device, and is used to provide input for the feedback controller creating a single rate control system. The second set is secure and is sampled at a higher frequency, and is used for attack detection and estimation. This architecture is practical for different applications such as wireless networked control systems, where the sensor measurements are sent over wireless (unsecured) networks to the control center, and the control signals are sent back to the physical plant again over wireless networks. A local estimator that has access to some of the measurements over hard-wired secure lines can be built to generate the attack estimates in this kind of scenario. The faster sampling loop is needed so that all unbounded attacks are detectable (i.e. removes the unstable zeros) [70] (Chapter 4), and to allow for the design of a certain class of observers as will be discussed later. Furthermore, we want to estimate the attack at a faster rate than control so that we can isolate the attack and limit the damage as fast as possible. In addition to detecting unbounded attacks, removing the unstable zeros is essential because they limit the achievable estimation performance. The attack estimation problem is similar to the unknown input observer (UIO) problem discussed in [71, 39, 72, 73, 74, 75, 41, 40] in which such an observer exists if and only if the system is strongly detectable, i.e., all zeros are strictly stable. Multirate sampling guarantees that the system has at most one non-minimum phase zero and is located at $\lambda = 1$, and under specific conditions, multrirate sampling can remove all zeros in the lifted domain. Conditions when a zero at $\lambda = 1$ exists in the MR scheme can be found in [70]. After introducing dual rate sampling for attack estimation, we introduce a few estimator design methods utilizing the dual rate property. In particular, we show that UIOs always exist if the dual rate system does not have a zero at $\lambda = 1$. In addition, the observer provides an estimate of the attack with a delay of a single time-step only. This result is significant because single rate observers do not exist most of the time due to the hard conditions for their existence [72], or they may exist but estimation is delayed (the system must be strongly detectable) [71].

6.2 Problem Formulation

In the absence of zero dynamics attack possibilities, we investigate the tradeoff between the ability to control the damage that an attack d inflicts versus the ability to estimate d. In other words, we would like to investigate whether one can trade control performance for extra ability to estimate the attack signal d. A relevant problem to study how to design a controller K jointly

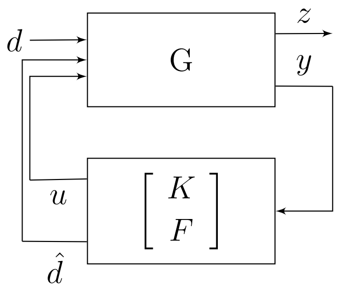


Figure 6.1: General block diagram to reject and estimate d.

with a filter F to reject as well as estimate d can be cast as

$$\min_{K,F} \|d \mapsto z\|, \quad z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix},$$

where z_1 relates to performance in terms of disturbance rejection, e.g., $z_1 = \begin{bmatrix} W_1 y \\ u \end{bmatrix}$ and z_2 relates to attack estimation, i.e., $z_2 = W_2(d - \hat{d})$, where $\hat{d}$ is the estimated attack and $W_{1,2}$ are weights, as seen in the general block diagram in Figure 6.1, where G is a general discrete LTI system. This type of problem is convex in any norm when the Youla parametrization of all stabilizing controllers [68, 69] is employed. The estimated signal will help the system operator decide whether the rejected signal is a carefully designed attack or just a random disturbance.

The input-output map of the system in Figure 6.1 can be described as:

$$\begin{bmatrix} z_1 \\ z_2 \\ y \end{bmatrix} = \left[\begin{array}{c|cc} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ \hline G_{31} & G_{32} & G_{33} \end{array}\right] \begin{bmatrix} d \\ u \\ \hat{d} \end{bmatrix},$$

where G_{32} is the open loop discrete time LTI plant. For z_2 defined as $z_2 = W(d - \hat{d})$ and z_1 does not depend on $\hat{d}$, we have $G_{13} = G_{22} = G_{33} = 0, \; G_{21} =$

$W,\ G_{23} = -W$. For actuator-only attacks, we have

$$G_{31} = G_{32},$$

while for sensor-only attacks

$$G_{31} = I.$$

The remaining maps G_{11} and G_{12} depend on how z_1 is defined. The input-output map is now more sparse and can be described by:

$$\begin{bmatrix} z_1 \\ z_2 \\ y \end{bmatrix} = \left[\begin{array}{c|cc} G_{11} & G_{12} & 0 \\ W & 0 & -W \\ \hline G_{31} & G_{32} & 0 \end{array}\right] \begin{bmatrix} d \\ u \\ \hat{d} \end{bmatrix}.$$

The closed loop map T_{zd} can then be found and is

$$\begin{bmatrix} T_{z_1 d} \\ T_{z_2 d} \end{bmatrix} = \begin{bmatrix} G_{11} \\ W \end{bmatrix} + \begin{bmatrix} G_{12} & 0 \\ 0 & -W \end{bmatrix} \begin{bmatrix} K \\ F \end{bmatrix} \left(I - \begin{bmatrix} G_{32} & 0 \end{bmatrix} \begin{bmatrix} K \\ F \end{bmatrix} \right)^{-1} G_{31}$$

$$= \begin{bmatrix} G_{11} + G_{12}K(I - G_{32}K)^{-1}G_{31} \\ W - WF(I - G_{32}K)^{-1}G_{31} \end{bmatrix}.$$

It is easy to see that minimizing $\|T_{z_1 d}\|$ depends only on finding the optimal K and can be solved as a model matching problem. On the other hand, minimizing $\|T_{z_2 d}\|$ depends on finding the optimal K and F simultaneously. By inspecting T_{zd}, keeping $\|T_{z_2 d}\|$ small is achieved by making $|F(I - G_{32}K)^{-1}G_{31}| \approx I$ for all frequencies. On the other hand, it is well known that $|K|$ has to be large for good disturbance rejection [38]. As a result, a trade-off between good estimation and good disturbance rejection exists. The following example demonstrates this trade-off.

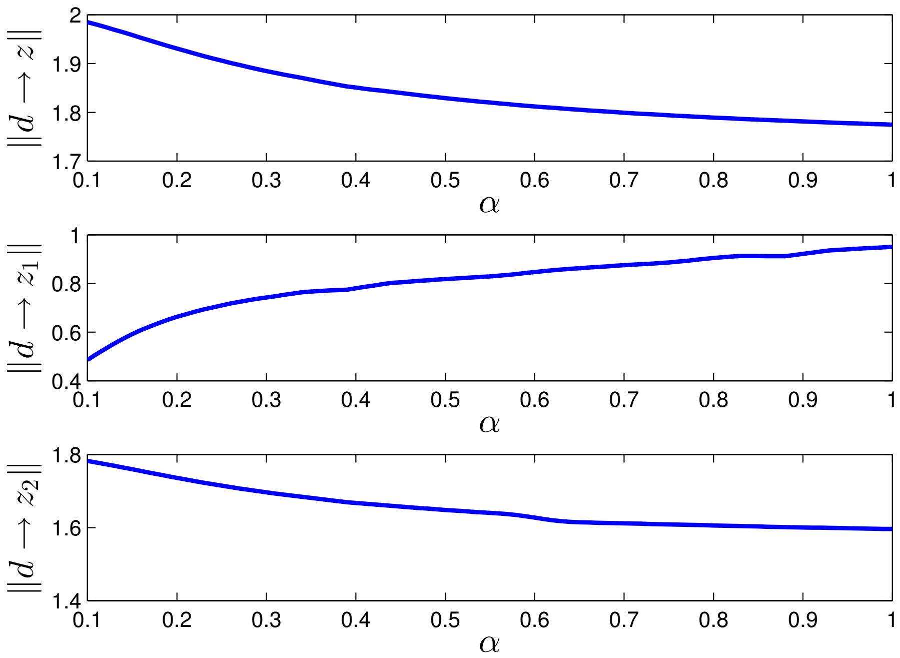


Figure 6.2: Tradeoff in performance between rejection and estimation.

6.3 Example

Control design for rejection and estimation of the attack was carried out for the AVR system in Figure 4.3. The attack is assumed to be a step input. The controller is synthesized using the method of $\mathcal{H}_\infty$ control [43, 42], using the tools in MATLAB. The chosen weights are

$$W_1 = \frac{0.009995}{\lambda - 1.001}, \quad W_2 = \alpha\frac{0.009995}{\lambda - 1.001},$$

where α represents the emphasis on the estimation performance. We can deduce from Figures 6.2 and 6.3 that for this example, as α increases, the rejection performance deteriorates while the estimation performance is enhanced ($\hat{d}$ converges faster to the true attack value). Hence, a trade-off exists between control and estimation.

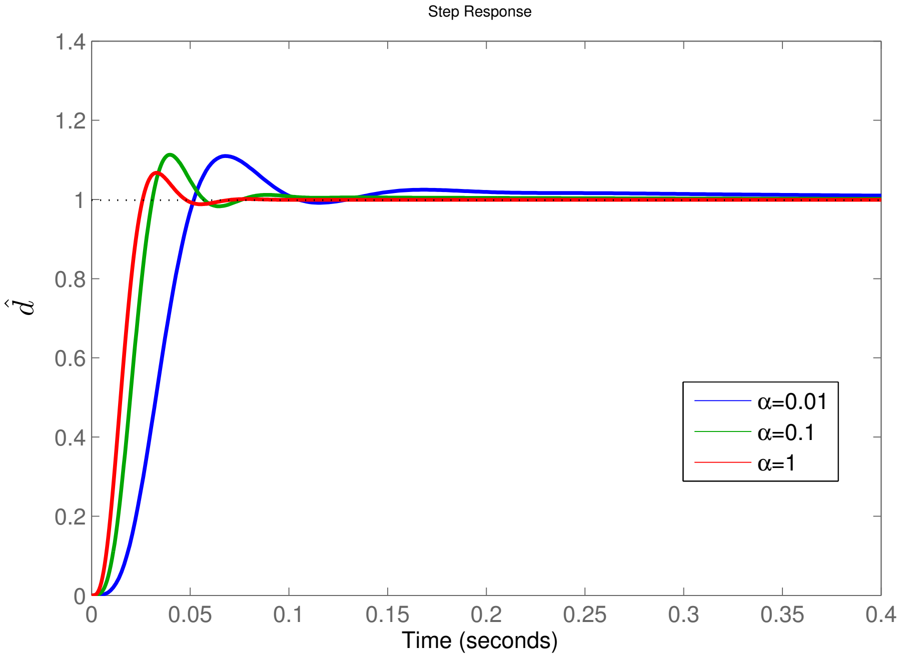


Figure 6.3: Attack estimation for different values of α.

6.4 Estimation Via Multirate Sampling

6.4.1 Motivation and Control Loop Architecture

Next we consider multirate (MR) sampling to estimate the injected attack d. In particular, we consider the case where we have two sets of sensors measuring the output. The first set is sampled at the same rate of the hold device, and is used to provide input for the feedback controller creating a single rate control system. The second set is secure and is sampled at a higher frequency, and is used for attack detection and estimation as seen in Figure 6.4, where G is the continuous-time LTI general input-output map. This architecture is practical for different applications such as wireless networked control systems, where the sensor measurements are sent over wireless (unsecured and delayed) networks to the control center, and the control signals are sent back to the physical plant again over wireless networks. A local estimator that has access to some of the measurements over hard-wired secure lines can be built

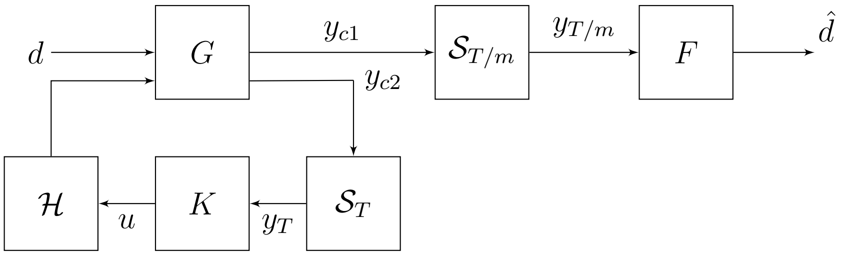


Figure 6.4: General block diagram to reject and estimate d with secured sensors.

to generate the attack estimates in this kind of scenario. Higher sampling rate for the detection loop is recommended to detect stealthy unbounded attacks, and to make it harder for attackers to design stealthy bounded attacks. In addition, MR sampling removes unstable zeros (except for possibly one zero at $\lambda = 1$) in the map from the attack signal d to the monitored signals (y and possibly u).

Remark 33. *The control loop in the architecture in Figure 6.4 can also be dual rate. What is important is to have the output feeding the estimation loop sampled at a sufficiently higher rate than that at which the attack is injected into the system. This helps in detecting the attack faster, and ensures that the system detects unbounded stealthy actuator or sensor attacks.*

6.4.2 Estimator Design

In this section we present a few control methods for the design of the estimator F for a fixed controller K, for the architecture in Figure 6.4. First we find the mapping from d to the measurements

$$\begin{bmatrix} \tilde{y} \\ y_T \end{bmatrix} = G_d \begin{bmatrix} d \\ u \end{bmatrix} = \begin{bmatrix} L\mathcal{S}_{T/m}G_{11}\mathcal{H} & L\mathcal{S}_{T/m}G_{12}\mathcal{H} \\ L\mathcal{S}_T G_{21}\mathcal{H} & L\mathcal{S}_T G_{22}\mathcal{H} \end{bmatrix} \begin{bmatrix} d \\ u \end{bmatrix},$$

where $\tilde{y}(k) = [y'_{c1}(kT/m)\ y'_{c1}((k+1)T/m) \dots y'_{c1}((k+m-1)T/m)]'$, $y_T(k) = y_{c2}(kT)$, G_{11} is the mapping from d to y_{c1}, G_{12} is the mapping from u to

y_{c1} and G_{21} is the mapping from d to y_{c2}, G_{22} is the mapping from u to y_{c2}, y_{c1} and y_{c2} are the continuous-time measurements feeding $\mathcal{S}_T/m$ and $\mathcal{S}_T$ respectively, and L is the lifting operator. G_{11} may represent actuator attacks or sensor attacks as explained in section 6.2. In view of the above, let G be controllable, observable and have the following representation:

$$G = \left[\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & 0 \end{array}\right], \tag{6.1}$$

then

$$G_d = \left[\begin{array}{c|cc} A_d & B_{1d} & B_{2d} \\ \hline \tilde{C}_1 & \tilde{D}_{11} & \tilde{D}_{12} \\ C_2 & D_{21} & 0 \end{array}\right],$$

where

$$\tilde{C}_1 = \begin{bmatrix} C_1 \\ C_1 A_f \\ \vdots \\ C_1 A_f^{m-1} \end{bmatrix}, \quad \tilde{D}_{11} = \begin{bmatrix} D_{11} \\ D_{11} + C_1 B_{1f} \\ \vdots \\ D_{11} + C_1 \sum_{k=0}^{m-2} A_f^k B_{1f} \end{bmatrix},$$

$$\tilde{D}_{12} = \begin{bmatrix} D_{12} \\ \\ D_{12} + C_1 B_{2f} \\ \\ \vdots \\ \\ D_{12} + C_1 \sum_{k=0}^{m-2} A_f^k B_{2f} \end{bmatrix},$$

and

$$\begin{aligned} A_d &:= e^{AT}, & A_f &:= e^{AT/m}, \\ B_{1d} &:= \int_0^T e^{A\tau} B_1 \mathrm{d}\tau, & B_{2d} &:= \int_0^T e^{A\tau} B_2 \mathrm{d}\tau, \\ B_{1f} &:= \int_0^{T/m} e^{A\tau} B_1 \mathrm{d}\tau, & B_{2f} &:= \int_0^{T/m} e^{A\tau} B_2 \mathrm{d}\tau \end{aligned}$$

Now for a given controller K with state space

$$K = \left[\begin{array}{c|c} A_K & B_K \\ \hline C_K & D_K \end{array}\right],$$

the input-output map from d to $y_{T/m}$ is described as

$$\tilde{P} = \left[\begin{array}{cc|c} A_d + B_{2d} D_K C_2 & B_{2d} C_K & B_{1d} + B_{2d} D_K D_{21} \\ B_K C_2 & A_K & B_K D_{21} \\ \hline \tilde{C}_1 + \tilde{D}_{12} D_K C_2 & \tilde{D}_{12} C_K & \tilde{D}_{11} + \tilde{D}_{12} D_K D_{21} \end{array}\right], \qquad (6.2)$$

as seen in Figure 6.5, where n is sensor noise.

In Chapter 4 we showed that if $\tilde{P}$ has a non-minimum phase zero, then

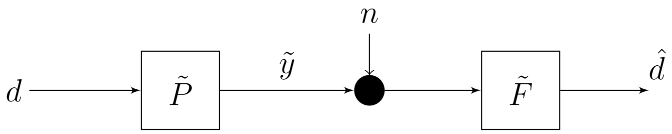


Figure 6.5: Block diagram for estimator design in the lifted domain. $\tilde{P}$ is dual rate, lifted and augmented with a controller for stabilization.

this zero is located at $\lambda = 1$, and its multiplicity is 1. As a result, dual rate control renders the system secure against unbounded stealthy actuator attacks. This applies to any G_{11} of any structure (as long as K does not introduce a zero at $\lambda = 1$). For the case when G_{11} is tall and has no zeros at the origin, [76] (Theorem 1) states that $\tilde{P}$ has no zeros at all for almost all $m \in \mathcal{R}$ such that $m > 1$. In our MR scheme in Chapter 4 and in this chapter, we only consider m to be an integer.

6.4.2.1 Model Matching

The problem of finding the best $\hat{d}$ (in some sense) can now be cast as

$$\min_{F} \left\| W - W\tilde{F}\tilde{P} \right\|,$$

or in the case of noisy measurements

$$\min_{F} \left\| W(I \quad 0) - W\tilde{F}(\tilde{P} \quad I) \right\|$$

such that F is stable (to minimize noise amplification) and causal. Since we are solving the problem in the lifted domain, the causality of F is guaranteed by enforcing the constraint that $F(0)$ is block lower triangular. Several methods to solve this synthesis problem can found in [49, 50, 51, 52, 60, 77, 61, 78].

6.4.2.2 Unknown Input Reconstruction

In this section we seek to exploit dual rate sampling to accurately reconstruct the unknown input (attack) d injected in the system in Figures 6.4 and 6.5, as well as the initial condition $x(0)$. In particular, we consider the relationship between the states and input from one end and the output of the system from another end. This relationship has been studied for single rate systems in the context of strong observability in the literature [39, 40, 41].

We consider the state space description $\tilde{P}$ in (6.2). We assume for now (without loss of generality) that $K = 0$; we also assume that the measurements are noise free, as a result $\tilde{P}$ reduces to

$$\tilde{P} = \left[\begin{array}{c|c} A_d & B_{1d} \\ \hline \tilde{C}_1 & \tilde{D}_{11} \end{array}\right]. \tag{6.3}$$

The lifted output of $\tilde{P}$ can be described as

$$\begin{bmatrix} y(0) \\ y(T/m) \\ y(2T/m) \\ \vdots \\ y((m-1)T/m) \end{bmatrix} = \begin{bmatrix} C_1 \\ C_1 A_f \\ C_1 A_f^2 \\ \vdots \\ C_1 A_f^{m-1} \end{bmatrix} x(0) + \begin{bmatrix} D_{11} \\ D_{11} + C_1 B_{1f} \\ D_{11} + C_1 B_{1f} + C_1 A_f B_{1f} \\ \vdots \\ D_{11} + C_1 \sum_{k=0}^{m-2} A_f^k B_{1f} \end{bmatrix} d(0). \tag{6.4}$$

From the above equation, we can deduce that $x(0)$ and $d(0)$ can be recovered without delay with respect to the original single rate system if and only if

$$\begin{bmatrix} \tilde{C}_1 & \tilde{D}_{11} \end{bmatrix} \tag{6.5}$$

is full column rank. A necessary condition for (6.5) to have full column rank is that $\tilde{P}$ be strongly observable, i.e., have no invariant zeros [39, 40].

Strong observability of $\tilde{P}$ is guaranteed if G_{11} is tall and does not have a zero at the origin for a sufficiently large m (Theorem 1 in [76]), given that (A_f, C_1) is observable. Choosing m to satisfy Assumptions 17 and 18 of section 2.1 is one choice. Strong observability does not imply that (6.5) is full column rank; however, (6.5) can be made to have full column rank by choosing m sufficiently large [48]. The idea is to add more linearly independent rows to (6.5) by sampling faster until (6.5) is tall. The added rows are linearly independent because m satisfies Assumptions 17 and 18, assuming $|C_1 B_{1d}| \neq 0$ as mentioned in section 2.1.

Remark 34. *The advantages of MR sampling here are (1) It makes $\tilde{P}$ strongly observable, and (2) It makes* (6.5) *full column rank. Hence, MR sampling guarantees that such an observer always exists (assuming G_{11} is tall and does not have a zero at the origin), which may not be the case for single rate systems.*

Remark 35. *The attack and the states are reconstructed with no delay with respect to original single rate period T. Still, m samples are needed within T, so in actual continuous-time the delay is T sec (or one sample period of the original single rate). In contrast, for single rate systems the delay can be up to nT where n is the dimension A_d (assuming the observer exists for single rate, i.e., assuming the single rate system is strongly observable).*

Remark 36. *As long as we choose m to make $\tilde{P}$ strongly observable, then we can still reconstruct d using delayed measurements (i.e., $\tilde{y}(0), \tilde{y}(1), \ldots, \tilde{y}(n)$, where n is the dimension of A_d) even if* (6.5) *is not full column rank. The amount of delay would be smaller than for single rate systems (provided that the single rate system is strongly observable). Details about this scheme for single rate systems can be found in [39].*

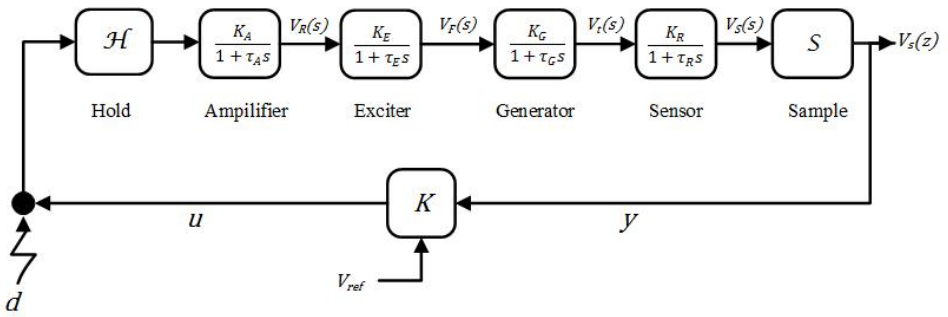


Figure 6.6: A simplified automatic voltage regulator block diagram.

6.4.2.2.1 Example - Automatic Voltage Regulator

We revisit the AVR example of section 4.3.2. A simplified block diagram of a linearized AVR is shown in Figure 6.6 [58]. The open loop state space representation of the single rate system after discretization at a sample rate $T = 0.5$ sec is

$$A_d = \begin{bmatrix} 0.0105 & 0.3949 & 3.86 & 2.869 \\ -0.0057 & -0.1817 & -1.369 & -0.587 \\ 0.00117 & 0.03359 & 0.1793 & -0.4597 \\ 0.00092 & 0.03197 & 0.3163 & 0.8918 \end{bmatrix},$$

$$B_d = \begin{bmatrix} -0.005738 \\ 0.001174 \\ 0.0009193 \\ 0.0002165 \end{bmatrix}, C_d = \begin{bmatrix} 0 & 0 & 0 & 5000 \end{bmatrix}, D_d = \begin{bmatrix} 0 \end{bmatrix},$$

which has an unstable zero at $\lambda = -0.7045$. Since the system has an unstable zero, we know that we cannot reconstruct attacks even if we use an arbitrary large number of measurements. Next we know from section 4.3.2 that dual rate sampling at a rate of $T/m = 0.25$ sec removes the unstable zero when viewed from the lifted domain. The resulting open loop state space representation after lifting is

$$\tilde{A} = A_d,\ \tilde{B} = B_d, \tilde{C} = \begin{bmatrix} 0 & 0 & 0 & 5000 \\ 2.185 & 86.13 & 1092 & 4902 \end{bmatrix}, \tilde{D} = \begin{bmatrix} 0 \\ 0.196 \end{bmatrix}.$$

Although the open loop system is strongly observable using $m = 2$ for dual rate sampling, $[\tilde{C} \quad \tilde{D}]$ is not full column rank, and we cannot reconstruct actuator without delay. Next if we select $m = 5$, the resulting $\tilde{C}$ and $\tilde{D}$ matrices become

$$\tilde{C} = \begin{bmatrix} 0 & 0 & 0 & 5000 \\ 0.38 & 21.38 & 491.24 & 4994.4 \\ 1.53 & 64.35 & 917.65 & 4948.5 \\ 2.80 & 106.05 & 1238.6 & 4839.5 \\ 3.86 & 138.22 & 1454.3e & 4671.5 \end{bmatrix}, \tilde{D} = \begin{bmatrix} 0 \\ 0.011 \\ 0.10 \\ 0.32 \\ 0.66 \end{bmatrix}.$$

Now $[\tilde{C} \quad \tilde{D}]$ is full column rank and the attack along with $x(0)$ can be reconstructed without delay.

This concludes how to reconstruct d using dual rate sampling. The case where K is augmented in $\tilde{P}$ as in (6.2) can be handled similarly as long as K does not introduce a zero at $\lambda = 1$ in the closed loop map from d to $y_{T/m}$.

6.4.2.3 Unknown Input Observer

In the previous section we saw how to reconstruct d and $x(0)$ given that $\tilde{P}$ is sampled faster than the rate at which the input is feeding the system. However, the method involved inverting a matrix with high dimensions, which might be computationally expensive. A cheaper alternative is to design a dual rate unknown input observer that estimates the states of the system asymptotically, and then estimates the attack d using the state estimates.

The theory for single rate unknown input observers is well studied and can be found in [39, 71, 73, 72, 74, 75] and the references therein. In this section, we will extend the theory to design a dual rate observer to estimate the attack in Figure 6.4. Dual rate unknown input observers were briefly mentioned in [76], however, the authors assumed D_{11} and D_{12} to be equal to zero in (6.1), which changes the analysis and the conditions for existence of the observer and how the attack is estimated. In addition, they invert the matrices $|C_1 B_{1f}|$ which we are trying to avoid.

We consider the state space description $\tilde{P}$ in (6.2), assuming without loss of generality that $K = 0$ and that the measurements are noise free. $\tilde{P}$ is then represented by

$$\begin{aligned} x(k+1) &= A_d x(k) + B_{1d} d(k) \\ \tilde{y}(k) &= \tilde{C}_1 x(k) + \tilde{D}_{11} d(k). \end{aligned} \tag{6.6}$$

We assume that

$$\begin{bmatrix} B_{1d} \\ \tilde{D}_{11} \end{bmatrix}$$

is full column rank. We consider an observer of the form

$$\hat{x}(k+1) = E\hat{x}(k) + L\tilde{y}(k), \tag{6.7}$$

where E and L are matrices to be designed.

Definition 37. *The system* (6.7) *is said to be an unknown input dual rate observer with rate T/m if $\hat{x}(k) - x(k) \to 0$ as $k \to \infty$, regardless of the input.*

We note that the observer in (6.7) does not depend on the input to the system (6.6). To choose the observer matrices E and L, we examine the estimation error

$$\begin{aligned}
e(k+1) &= \hat{x}(k+1) - x(k+1) \\
&= E\hat{x}(k) + L\tilde{y}(k) - A_d x(k) - B_{1d}d(k) \\
&= Ee(k) + L\tilde{y}(k) + (E - A_d)x(k) - B_{1d}d(k) \\
&= Ee(k) + (E - A_d + L\tilde{C}_1)x(k) + (L\tilde{D}_{11} - B_{1d})d(k).
\end{aligned}$$

In order to force the error to go to zero, regardless of the values of $x(k)$ and $d(k)$, E and L must simultaneously satisfy

$$L\tilde{D}_{11} = B_{1d} \tag{6.8}$$

$$E = A_d - L\tilde{C}_1 \tag{6.9}$$

such that E is stable. There exists a matrix L that satisfies (6.8) if and only if B_{1d} is in the space spanned by the rows of $\tilde{D}_{11}$, which is equivalent to

$$rank\left(\begin{bmatrix} B_{1d} \\ \tilde{D}_{11} \end{bmatrix}\right) = rank(\tilde{D}_{11}). \tag{6.10}$$

Necessary and sufficient conditions for the existence of E and L that satisfy (6.8) and (6.9) are that $\tilde{P}$ is strongly detectable (i.e., all zeros of $\tilde{P}$ are strictly stable), and that (6.10) holds.

The strong detectability condition is inherited from the conditions of existence of UIO for single rate systems. We know that using dual rate sampling guarantees that at most one zero exists and is at $\lambda = 1$; therefore, checking $\tilde{P}$ for this zero is sufficient to check for the strong detectability of $\tilde{P}$, as long as Assumptions 17 and 18 are met. Furthermore, strong observability of $\tilde{P}$, which is a more strict property, is guaranteed if G_{11} is tall and does not have a zero at the origin, as long as Assumptions 17 and 18 are met. Now to ensure the solvability of (6.8), m is chosen long enough until (6.10) holds.

Once a state observer is constructed, we can obtain an estimate of the attack by first rearranging (6.6) as

$$\begin{bmatrix} x(k+1) - A_d x(k) \\ \tilde{y}(k) - \tilde{C}_1 x(k) \end{bmatrix} = \begin{bmatrix} B_{1d} \\ \tilde{D}_{11} \end{bmatrix} d(k). \tag{6.11}$$

Since $\begin{bmatrix} B_{1d} \\ \tilde{D}_{11} \end{bmatrix}$ is full column rank, there exists a matrix R such that

$$R \begin{bmatrix} B_{1d} \\ \tilde{D}_{11} \end{bmatrix} = I,$$

where I has the appropriate dimension. Left multiplying (6.11) by R and using $\hat{x}(k)$ instead of $x(k)$, we find the estimate of the attack to be

$$\hat{d}(k) = R \begin{bmatrix} \hat{x}(k+1) - A_d \hat{x}(k) \\ \tilde{y}(k) - \tilde{C}_1 \hat{x}(k) \end{bmatrix}.$$

Since

$$\hat{x}(k) - x(k) \to 0 \text{ as } k \to \infty,$$

$\hat{d}(k)$ will asymptotically approach $d(k)$.

Note that there is a single step delay in computing the attack estimate. In case of single rate sampling, there will be at most $n+1$ steps delay where n is the dimension of the vector x in (6.6), if the observer exists (i.e. if the single rate system is strongly detectable) [39]. Single rate observers have to accumulate several measurements to be able to estimate the attack, which during several instants of the attack signal are injected into the system. Therefore, dual rate sampling provides a much more secure framework of control, and guarantees that UIOs always exist if G_{11} is tall and has no zeros at the origin.

6.4.2.3.1 Examples

We provide two examples to illustrate the design of unknown input observers for dual rate systems.

Example 1- Simple first-order system

We consider the following stable, non-minimum phase first-order system

$$G = \frac{s-1}{s+2}.$$

We use dual rate control to remove the unstable zero at the sampling rates $T = 1$ sec and $T = 0.5$ sec. The state space representation of the lifted system is

$$A_d = 0.1353,\ B_d = 0.4323,\ \tilde{C} = \begin{bmatrix} -3 \\ -1.104 \end{bmatrix},\ \tilde{D} = \begin{bmatrix} 1 \\ -0.9482 \end{bmatrix},$$

which removes the unstable zero and makes $rank\left(\begin{bmatrix} B_{1d} \\ \tilde{D}_{11} \end{bmatrix}\right) = rank(\tilde{D}_{11})$. Since $B \in \mathbb{R}^{1\times1}$ and $D \in \mathbb{R}^{2\times1}$, $L \in \mathbb{R}^{1\times2}$. Let L be defined as $L = \begin{bmatrix} L_1 & L_2 \end{bmatrix}$. Solving for $L\tilde{D} = B$ we get

$$\begin{bmatrix} L_1 & L_2 \end{bmatrix} \begin{bmatrix} 1 \\ -0.9482 \end{bmatrix} = 0.4323,$$

which gives $L_1 = 0.4323 + 0.9482L_2$. Plugging this back and solving for $A_d - L\tilde{C}$ to be stable, we get

$$A_d - L\tilde{C} = 0.1353 - \begin{bmatrix} 0.4323 + 0.9482L_2 & L_2 \end{bmatrix} \begin{bmatrix} -3 \\ -1.104 \end{bmatrix}$$

$$= -1.2969 - 2.8446L_2 - 1.104L_2 = -1.2969 - 3.95L_2$$

$$|A_d - L\tilde{C}| < 1 \implies -0.58 < L_2 < -0.075.$$

Example 2 - Automatic Voltage Regulator

We revisit the AVR example of section 4.3.2. A simplified block diagram of a linearized AVR is shown in Figure 6.7 [58]. The open loop state space representation of the single rate system after discretization at a sample rate $T = 0.5$ sec is

$$A_d = \begin{bmatrix} 0.0105 & 0.3949 & 3.86 & 2.869 \\ -0.0057 & -0.1817 & -1.369 & -0.587 \\ 0.00117 & 0.03359 & 0.1793 & -0.4597 \\ 0.00092 & 0.03197 & 0.3163 & 0.8918 \end{bmatrix},$$

$$B_d = \begin{bmatrix} -0.005738 \\ 0.001174 \\ 0.0009193 \\ 0.0002165 \end{bmatrix}, C_d = \begin{bmatrix} 0 & 0 & 0 & 5000 \end{bmatrix}, D_d = \begin{bmatrix} 0 \end{bmatrix},$$

which has an unstable zero at $\lambda = -0.7045$. We note that although the continuous system has no unstable zeros, sampling at the relatively slow rate of 0.5 sec per sample created an unstable zero. Since the system has an unstable zero, we know that we cannot construct a single rate unknown input observer of the form

$$\hat{x}(k+1) = E\hat{x}(k) + Ly(k : k+n)$$

to estimate actuator attacks even if we use an arbitrary large number of measurements. Next we know from section 4.3.2 that dual rate sampling at a rate of $T/m = 0.25$ sec removes the unstable zero when viewed from the lifted domain. The resulting open loop state space representation after lifting

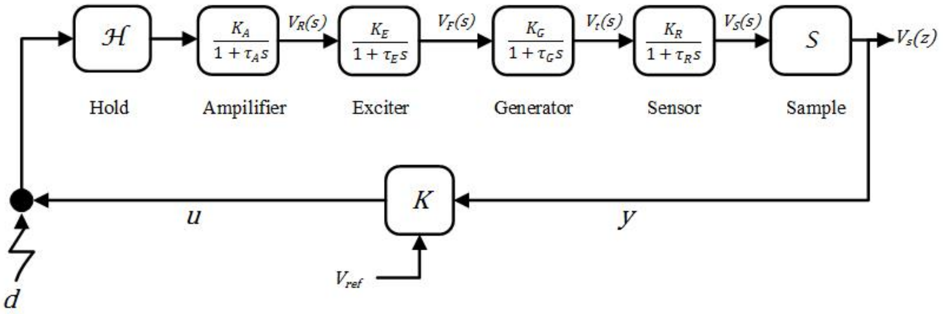


Figure 6.7: A simplified automatic voltage regulator block diagram.

is

$$\tilde{A} = A_d,\ \tilde{B} = B_d, \tilde{C} = \begin{bmatrix} 0 & 0 & 0 & 5000 \\ 2.185 & 86.13 & 1092 & 4902 \end{bmatrix}, \tilde{D} = \begin{bmatrix} 0 \\ 0.196 \end{bmatrix}. \tag{6.12}$$

And the condition $rank\left(\begin{bmatrix} B_d \\ \tilde{D} \end{bmatrix}\right) = rank(\tilde{D})$ is satisfied. We construct a dual rate unknown input observer of the form (6.7), i.e.,

$$\hat{x}(k+1) = E\hat{x}(k) + L\tilde{y}(k),$$

where E and L satisfy (6.8), (6.9) and (6.12). Since (6.12) has no unstable zeros and (6.10) is satisfied, then we know such an E and L exist. Using MATLAB solver, we find E and L to be

$$E = \begin{bmatrix} 0.074 & 2.91 & 35.70 & -3413.41 \\ -0.013 & -0.47 & -5.00 & 9003.32 \\ 0.013 & 0.51 & 6.21 & -734.21 \\ 0.0063 & 0.25 & 3.02 & -5.42 \end{bmatrix}, \quad L = \begin{bmatrix} 0.7118 & -0.0292 \\ -1.8040 & 0.0033 \\ 0.1522 & -0.0055 \\ 0.0037 & -0.0025 \end{bmatrix}.$$

We note that for the above AVR example, sampling faster using $m = 2$ was sufficient to estimate the states and the attack asymptotically, while in section 6.4.2.2.1, we saw that $m = 5$ was needed to accurately reconstruct the states and the attack for each period T. This observation makes sense as it means more measurements are needed for accurate estimation for each period T vs. asymptotic estimation.

6.5 Conclusion

In this chapter, we posed the problem of estimating signal attacks injected into the actuators or sensors of control systems. We showed that there exists a trade-off between attack rejection and estimation, and that the estimator design depends on the controller used. We used dual rate sampling to enhance the detectability of the attack and we provided three methods to design the estimator. Method 1 solves a model matching problem subject to causality constraints. Method 2 exploits dual rate sampling to accurately reconstruct the unknown input. Method 3 uses a dual rate unknown input observer. Using dual rate sampling, necessary and sufficient rank and zero location conditions to check the existence of the observers in methods 2 and 3 are provided. Once these conditions are satisfied, then the attack can be estimated with at most a single step delay. This work shows again the importance of studying the security problem in the SD framework, and the power of using dual rate sampling to design observers for the detection and estimation of signal attacks. Dual rate sampling ensures that UIOs always exist if the continuous-time map from the attack to the output is tall and has no zero at the origin, which may not be the case for single rate observers. A future research direction is to study dual rate unknown input observers discussed in the last section when noise is present in the measurements. Optimal estimators for the single rate delayed UIOs were discussed in [73] (minimizing mean square error), but to our knowledge no results exist for the dual rate UIOs.

CHAPTER 7

SUMMARY AND FUTURE RESEARCH

In this thesis, we introduced a sampled-data framework to study the effect of attacks on cyber-physical systems. We defined the attack detection mechanism and derived the input-output maps for actuator and sensor attacks on the monitoring signals. We showed that unbounded stealthy actuator attacks are related to the open-loop discrete plant unstable zeros, while unbounded stealthy sensor attacks are related to the open-loop discrete plant unstable poles. We also showed that coordinated actuator and sensor attacks can always be designed to be stealthy and unbounded regardless of the locations of poles and zeros of the plant. Next we presented a dual rate sampled data scheme which guarantees detectability of unbounded actuator and/or sensor attacks, if a secure output that maintains observability of the open loop modes is available. The main observation is that the sampled data nature in the implementation of the cyber-physical system cannot be ignored as sampling can generate additional vulnerabilities due to the extra unstable zeros it may introduce, particularly if high rates are necessary to achieve certain performance level. The proposed method solves this issue by the use of multirate sampling that ensures that zeros exist only in harmless locations in the lifted domain. A few examples were presented that show how the multirate scheme detects the unbounded actuator attacks. In addition, the examples incorporated a dual rate controller cost comparison based on LQG control. After that we studied the problem of computing worst case bounded stealthy false data injection attacks for LTI systems. We considered different attack resource constraints and stealthiness intervals. This problem involves a maximization of a convex function subject to convex constraints, and it was shown that it can be cast as a series of LP problems under the ℓ_∞ framework. A search algorithm is constructed to solve the set of LPs and was used to compute the worst stealthy attacks on AVR systems. Furthermore, we provided an iterative controller synthesis procedure that alternates between computing worst attacks and designing optimal controllers that enhance performance and minimize the impact of worst attacks. We used this method to design a controller for the AVR system that resulted in a substantial decrease in the worst impact inflicted by the worst attack. Lastly, we posed the problem of estimating signal attacks injected into the actuators or sensors of control systems, assuming the attack is detectable (can be seen at the output). We showed that their exists a trade-off between attack rejection and control, and that the estimator design depends on the controller used. We used dual rate sampling on the secured sensors to enhance detectability of the attack, and we provided different methods to design the estimator. The first method is by solving a model matching problem subject to causality constraints. The second method exploited dual rate sampling to accurately reconstruct the unknown input. The third method is using an unknown input observer. Using dual rate sampling, necessary and sufficient conditions to check the existence of the observers are provided. The work shows the importance of studying the security problem in the SD framework, and shows the power of using dual rate sampling to design observers to detect and estimate attacks. Dual rate sampling ensures that UIOs always exist if the continuous-time map from the attack to the output is tall and has no zero at the origin, which may not be the case for single rate observers.

Several future research directions can be investigated. One direction is to study methods to detect unbounded stealthy sensor attacks described in Chapter 3 in the absence of secure sensors. We showed in Chapter 4 that multirate sampling removes the unstable zeros in the lifted domain and can detect unbounded actuator attacks. However, this method cannot detect sensor attacks related to the unstable poles if the attacker has knowledge of the multirate scheme. One idea is to use stochastic sampling. It is well known that the poles of the discrete plant are related to the sampling and hold rate. Therefore, by changing the sample-hold rate in a random fashion (e.g. Markov process with a known transition matrix), the attacker will not be able to exactly determine the location of the unstable poles for each instant of time, and hence will not be able to inject a stealthy attack. The problem can be studied in the context of stochastic linear switched systems such as the work in [79], or in the context of optimal control [80], and shows that no stealthy attack is possible under certain stochastic/switching sampling conditions.

Another future research direction is to quantify the minimum number of sensors that need to be sampled faster to detect unbounded stealthy actuator attacks. In Chapter 3 we showed that dual rate sampling can detect unbounded stealthy actuator attacks under certain conditions. However, the method and related proofs assumed that all outputs will be sampled faster. The interesting question is whether we can remove the unstable zeros and detect the attacks if only a subset of the outputs (in the case of MIMO systems) is sampled faster, and how we decide on which outputs to be selected for faster sampling. Another way to state the problem would be to find the minimum number of outputs that must be sampled faster to detect stealthy actuator attacks.

Another future research direction is to study dual rate unknown input observers discussed in Chapter 6 when noise is present in the measurements. Optimal estimators for the single rate delayed UIOs were discussed in [73] (minimizing mean square error), but to our knowledge no results exist for the dual rate UIOs.

REFERENCES

[1] S. Karnouskos, “Stuxnet worm impact on industrial cyber-physical system security,” in *IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society*, Nov 2011, pp. 4490–4494.

[2] J. Slay and M. Miller, “Lessons learned from the Maroochy water breach,” in *International Conference on Critical Infrastructure Protection*. Springer, 2007, pp. 73–82.

[3] H. Sandberg, S. Amin, and K. H. Johansson, “Cyberphysical security in networked control systems: An introduction to the issue,” *IEEE Control Systems*, vol. 35, no. 1, pp. 20–23, 2015.

[4] Y. Shoukry, P. Martin, P. Tabuada, and M. Srivastava, “Non-invasive spoofing attacks for anti-lock braking systems,” in *Cryptographic Hardware and Embedded Systems - CHES 2013*, G. Bertoni and J.-S. Coron, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 55–72.

[5] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, “A survey of fault detection, isolation, and reconfiguration methods,” *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 636–653, May 2010.

[6] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, “Cyber security analysis of state estimators in electric power systems,” in *49th IEEE Conference on Decision and Control*, December 2010, pp. 5991–5998.

[7] K. C. Sou, H. Sandberg, and K. H. Johansson, “On the exact solution to a smart grid cyber-security analysis problem,” *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 856–865, June 2013.

[8] J. M. Hendrickx, K. H. Johansson, R. M. Jungers, H. Sandberg, and K. C. Sou, “Efficient computations of a security index for false data attacks in power networks,” *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3194–3208, Dec 2014.

[9] K. C. Sou, H. Sandberg, and K. H. Johansson, “Data attack isolation in power networks using secure voltage magnitude measurements,” *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 14–28, Jan 2014.

[10] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on dc state estimation," in *Workshop on Secure Control Systems*, 2010.

[11] Z. Mao, T. Xu, and T. J. Overbye, "Real-time detection of malicious pmu data," in *2017 19th International Conference on Intelligent System Application to Power Systems (ISAP)*, Sept 2017, pp. 1–6.

[12] Y. Zhao, A. Goldsmith, and H. V. Poor, "A polynomial-time method to find the sparsest unobservable attacks in power networks," in *American Control Conference*, July 2016, pp. 276–282.

[13] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, June 2014.

[14] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *2015 American Control Conference (ACC)*, July 2015, pp. 2439–2444.

[15] Y. Nakahira and Y. Mo, "Attack-resilient h2, h-infinity, and l1 state estimator," *IEEE Transactions on Automatic Control*, pp. 1–1, 2018.

[16] S. Mishra, Y. Shoukry, N. Karamchandani, S. N. Diggavi, and P. Tabuada, "Secure state estimation against sensor attacks in the presence of noise," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 49–59, March 2017.

[17] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2079–2091, Aug 2016.

[18] S. Z. Yong, M. Q. Foo, and E. Frazzoli, "Robust and resilient estimation for cyber-physical systems under adversarial attacks," in *American Control Conference*, July 2016, pp. 308–315.

[19] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems*, vol. 35, no. 1, pp. 93–109, Feb 2015.

[20] H. Liu, J. Yan, Y. Mo, and K. H. Johansson, "An On-line Design of Physical Watermarks," *ArXiv e-prints*, Sep. 2018.

[21] M. Hosseini, T. Tanaka, and V. Gupta, "Designing optimal watermark signal for a stealthy attacker," in *2016 European Control Conference (ECC)*, June 2016, pp. 2258–2262.

[22] F. Pasqualetti, F. Dorfler, and F. Bullo, “Attack detection and identification in cyber-physical systems,” *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, November 2013.

[23] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, “A secure control framework for resource-limited adversaries,” *Automatica*, vol. 51, pp. 135–148, 2015.

[24] R. Tan, H. H. Nguyen, E. Y. S. Foo, X. Dong, D. K. Y. Yau, Z. Kalbarczyk, R. K. Iyer, and H. B. Gooi, “Optimal false data injection attack against automatic generation control in power grids,” in *2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS)*, April 2016, pp. 1–10.

[25] C. Kwon, W. Liu, and I. Hwang, “Analysis and design of stealthy cyber attacks on unmanned aerial systems,” *Journal of Aerospace Information Systems*, vol. 11, no. 8, pp. 525–539, August 2014.

[26] R. Zhang and P. Venkitasubramaniam, “Stealthy control signal attacks in vector lqg systems,” in *2016 American Control Conference*, July 2016, pp. 1179–1184.

[27] Y. Chen, S. Kar, and J. M. F. Moura, “Cyber physical attacks constrained by control objectives,” in *2016 American Control Conference (ACC)*, July 2016, pp. 1185–1190.

[28] G. Wu and J. Sun, “Optimal data integrity attack on actuators in cyber-physical systems,” in *2016 American Control Conference (ACC)*, July 2016, pp. 1160–1164.

[29] S. Liu, X. Feng, D. Kundur, T. Zourntos, and K. L. Butler-Purry, “Switched system models for coordinated cyber-physical attack construction and simulation,” in *2011 IEEE First International Workshop on Smart Grid Modeling and Simulation (SGMS)*, Oct 2011, pp. 49–54.

[30] E. Hammad, A. M. Khalil, A. Farraj, D. Kundur, and R. Iravani, “A class of switching exploits based on inter-area oscillations,” *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4659–4668, Sept 2018.

[31] N. Bezzo, J. Weimer, Y. Du, O. Sokolsky, S. H. Son, and I. Lee, “A stochastic approach for attack resilient uav motion planning,” in *2016 American Control Conference (ACC)*, July 2016, pp. 1366–1372.

[32] M. Wakaiki, A. Cetinkaya, and H. Ishii, “Quantized output feedback stabilization under dos attacks,” in *2018 Annual American Control Conference (ACC)*, June 2018, pp. 6487–6492.

[33] B. Ramasubramanian, M. A. Rajan, and M. G. Chandra, "Structural resilience of cyberphysical systems under attack," in *2016 American Control Conference (ACC)*, July 2016, pp. 283–289.

[34] C. Barreto, A. A. Cárdenas, and N. Quijano, "Controllability of dynamical systems: Threat models and reactive security," in *Decision and Game Theory for Security*, S. K. Das, C. Nita-Rotaru, and M. Kantarcioglu, Eds. Cham: Springer International Publishing, 2013, pp. 45–64.

[35] J. Bhatti and T. E. Humphreys, "Hostile control of ships via false gps signals: Demonstration and detection," *Navigation*, vol. 64, no. 1, pp. 51–66.

[36] S. H. Kafash, J. Giraldo, C. Murguia, A. A. Cárdenas, and J. Ruths, "Constraining attacker capabilities through actuator saturation," *CoRR*, vol. abs/1710.02576, 2017. [Online]. Available: http://arxiv.org/abs/1710.02576

[37] T. Chen and B. Francis, *Optimal Sampled-Data Control Systems.* Springer, 1995.

[38] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design.* USA: John Wiley & Sons, Inc., 2005.

[39] S. Sundaram, *Fault-Tolerant and Secure Control Systems.* Online, Univesity of Waterloo., 2012.

[40] M. Hautus, "Strong detectability and observers," *Linear Algebra and Its Applications*, vol. 50, pp. 353 – 368, 1983.

[41] W. Kratz, "Characterization of strong observability and construction of an observer," *Linear Algebra and Its Applications*, vol. 221, pp. 31 – 40, 1995.

[42] M. A. Dahleh and I. Diaz-Bobillo, *Control of Uncertain Systems: A Linear Programming Approach.* Prentice-Hall, 1995.

[43] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control.* Prentice-Hall, 1995.

[44] J. Tokarzewski, *Finite Zeros in Discrete Time Control Systems.* Springer-Verlag Berlin Heidelberg, 2006.

[45] J. A. Torres and S. Roy, "Connecting network graph structure to linear-system zero structure," in *2013 American Control Conference*, June 2013, pp. 6114–6119.

[46] K. Astrom, P. Hagander, and J. Sternby, "Zeros of sampled systems," *Automatica*, vol. 20, no. 1, pp. 31 – 38, 1984.

[47] P. Antsaklis and A. N. Michel, *Linear Systems*. McGraw-Hill, 2006.

[48] T. Hagiwara and M. Araki, "Design of a stable state feedback controller based on the multirate sampling of the plant output," *IEEE Transactions on Automatic Control*, vol. 33, no. 9, pp. 812–819, September 1988.

[49] P. G. Voulgaris, M. A. Dahleh, and L. S. Valavani, "H_∞ and H_2 optimal controllers for periodic and multi-rate systems," in *Proceedings of the 30th IEEE Conference on Decision and Control*, December 1991, pp. 214–216.

[50] P. G. Voulgaris and B. Bamieh, "Optimal H_∞ control of hybrid multirate systems," in *Proceedings of the 31st IEEE Conference on Decision and Control*, December 1992, pp. 457–462.

[51] T. Chen and L. Qiu, "H_∞ design of general multirate sampled-data control systems," in *Proceedings of the 32nd IEEE Conference on Decision and Control*, December 1993, pp. 315–320.

[52] L. Qiu and T. Chen, "H_2 optimal design of multirate sampled-data systems," *IEEE Transactions on Automatic Control*, vol. 39, no. 12, pp. 2506–2511, December 1994.

[53] P. Colaneri, R. Scattolini, and N. Schiavoni, "Stabilization of multirate sampled-data linear systems," *Automatica*, vol. 26, no. 2, pp. 377–380, 1990.

[54] D. G. Meyer, "A parametrization of stabilizing controllers for multirate sampled-data systems," *IEEE Transactions on Automatic Control*, vol. 35, no. 2, pp. 233–236, February 1990.

[55] R. Ravi, P. P. Khargonekar, K. D. Minto, and C. N. Nett, "Controller parametrization for time-varying multirate plants," *IEEE Transactions on Automatic Control*, vol. 35, no. 11, pp. 1259–1262, 1990.

[56] K. H. Johansson, "The quadruple-tank process: a multivariable laboratory process with an adjustable zero," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 3, pp. 456–465, May 2000.

[57] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Revealing stealthy attacks in control systems," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, October 2012, pp. 1806–1813.

[58] H. Saadat, *Power System Analysis*. McGraw-Hill, 2009.

[59] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*. John Wiley & Sons, 2005.

[60] P. Voulgaris, "Control of asynchronous sampled data systems," *IEEE Transactions on Automatic Control*, vol. 39, no. 7, pp. 1451–1455, July 1994.

[61] M. F. Sgfors and H. T. Toivonen, "H_∞ and lqg control of asynchronous sampled-data systems," *Automatica*, vol. 33, no. 9, pp. 1663 – 1668, 1997.

[62] P. Kabamba, "Control of linear systems using generalized sampled-data hold functions," *IEEE Transactions on Automatic Control*, vol. 32, no. 9, pp. 772–783, September 1987.

[63] J. S. Freudenberg, R. H. Middleton, and J. H. Braslavsky, "Robustness of zero shifting via generalized sampled-data hold functions," *IEEE Transactions on Automatic Control*, vol. 42, no. 12, pp. 1681–1692, Dec 1997.

[64] G. Dullerud, *Control of Uncertain Sampled-Data Systems.* Prentice Hall Professional Technical Reference, 1996.

[65] A. Teixeira, H. Sandberg, and K. H. Johansson, "Strategic stealthy attacks: The output-to-output l2-gain," in *2015 54th IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 2582–2587.

[66] D. Umsonst, H. Sandberg, and A. A. Cardenas, "Security analysis of control system anomaly detectors," in *2017 American Control Conference (ACC)*, May 2017, pp. 5500–5506.

[67] S. D. Bopardikar, A. Speranzon, and J. P. Hespanha, "An H_∞ approach to stealth-resilient control design," in *2016 Resilience Week (RWS)*, Aug 2016, pp. 56–61.

[68] X. Qi, M. V. Salapaka, P. G. Voulgaris, and M. Khammash, "Structured optimal and robust control with multiple criteria: a convex solution," *IEEE Transactions on Automatic Control*, vol. 49, no. 10, pp. 1623–1640, Oct 2004.

[69] M. Khammash, "A new approach to the solution of the ℓ_1 control problem: the scaled-q method," *IEEE Transactions on Automatic Control*, vol. 45, no. 2, pp. 180–187, Feb 2000.

[70] M. Naghnaeian, N. Hirzallah, and P. G. Voulgaris, "Dual rate control for security in cyber-physical systems," in *2015 54th IEEE Conference on Decision and Control*, December 2015, pp. 1415–1420.

[71] S. Sundaram and C. N. Hadjicostis, "Delayed observers for linear systems with unknown inputs," *IEEE Transactions on Automatic Control*, vol. 52, no. 2, pp. 334–339, Feb 2007.

[72] J. Chen and R. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Springer Publishing Company, Incorporated, 2012.

[73] S. Sundaram and C. N. Hadjicostis, "Optimal state estimators for linear systems with unknown inputs," in *Proceedings of the 45th IEEE Conference on Decision and Control*, Dec 2006, pp. 4763–4768.

[74] B. Shafai and M. Saif, "Proportional-integral observer in robust control, fault detection, and decentralized control of dynamic systems," in *Control and Systems Engineering: A Report on Four Decades of Contributions*, A. El-Osery and J. Prevost, Eds. Cham: Springer International Publishing, 2015, pp. 13–43.

[75] H. Lee, S. Snyder, and N. Hovakimyan, "An adaptive unknown input observer for fault detection and isolation of aircraft actuator faults," *AIAA Guidance, Navigation, and Control Conference*, 2014.

[76] T. Mita, Y. Chida, Y. Kaku, and H. Numasato, "Two-delay robust digital control and its applications-avoiding the problem on unstable limiting zeros," *IEEE Transactions on Automatic Control*, vol. 35, no. 8, pp. 962–970, Aug 1990.

[77] M. F. Sagfors, H. T. Toivonen, and B. Lennartson, "H_∞ control of multirate sampled-data systems: A state-space approach," *Automatica*, vol. 34, no. 4, pp. 415 – 428, 1998.

[78] P. G. Voulgaris, "Optimal ℓ_∞ to ℓ_∞ estimation for periodic systems," *IEEE Transactions on Automatic Control*, vol. 41, no. 9, pp. 1392–1396, Sept 1996.

[79] M. Naghnaeian and P. G. Voulgaris, "Stochastic l_∞ performance optimization for markovian linear switched systems," in *2015 54th IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 2686–2690.

[80] W. L. D. Koning, "Equivalent discrete optimal control problem for randomly sampled digital control systems," *International Journal of Systems Science*, vol. 11, no. 7, pp. 841–850, 1980.