# On Cyber Security for Networked Control Systems

by

Saurabh Amin

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Civil and Environmental Engineering

and the Designated Emphasis

in

Communication, Computation, and Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Alexandre M. Bayen, Co-Chair
Professor S. Shankar Sastry, Co-Chair
Professor Mark Stacey
Professor Pravin P. Varaiya

Spring 2011

On Cyber Security for Networked Control Systems

# Abstract

On Cyber Security for Networked Control Systems

by

Saurabh Amin

Doctor of Philosophy in Engineering – Civil and Environmental Engineering

and the Designated Emphasis in Communication, Computation, and Statistics

University of California, Berkeley

Professor Alexandre M. Bayen, Co-Chair
Professor S. Shankar Sastry, Co-Chair

The instrumentation of infrastructure systems by embedded sensors, computation, and communication networks has enabled significant advances in their management. Examples include monitoring of structural health, traffic congestion, environmental hazards, and energy usage. The use of homogeneous (especially, commercially available off-the-shelf) information technology (IT) solutions makes infrastructure systems subject to correlated hardware malfunctions and software bugs. Over the past decade, many concerns have been raised about the vulnerabilities of infrastructure systems to both random failures and security attacks. Cyber-security of Supervisory Control and Data Acquisition (SCADA) systems is especially important, because these systems are employed for sensing and control of large physical infrastructures. So far, the existing research in robust and fault-tolerant control does not account for cyber attacks on networked control system (NCS) components. Also, the existing research in computer security neither considers the attacks targeting NCS components nor accounts for their interactions with the physical system. The goal of this thesis is to bridge this gap by focusing on (1) security threat assessment, (2) model-based attack diagnosis, and (3) resilient control design.

First, cyber-security assessment for SCADA systems is performed based on well-defined attacker and defender objectives. The mathematical model of SCADA systems considered in this work has two control levels: regulatory control using distributed proportional-integral (PI) controllers, and supervisory fault diagnosis based on approximate dynamical system models. The performance of a PI control based regulatory scheme and a model-based supervisory diagnostic scheme is studied under a class of deception attacks. In order to test the system resilience, a class of stealthy attacks which can evade detection by SCADA systems is presented.

Second, design of attack diagnosis schemes that incorporate the knowledge of physical dynamics of the system is presented. For SCADA systems used to manage water canal networks, an observer-based attack diagnostic scheme, in which each observer estimates the state of a reduced-order flow model, is presented. The observer parameters are computed

using a convex optimization method, and the performance of this scheme is tested on a number of attack scenarios. An application of the theoretical results is illustrated by a field operational test performed on the SCADA system of the Gignac water canal system, located in Montpellier, France. A successful experimental cyber-attack on the sensors and actuators of this canal network revealed new vulnerabilities of the current SCADA system implementation.

Another illustration includes security analysis of two benchmark scenarios: the Tennessee Eastman process control system (TE-PCS) and a power system state estimator (PSSE). In both these cases, model-based statistical detection schemes are used to study stealthy deception attacks. For the case of TE-PCS, design of practically implementable attack-detection and response mechanisms to maintain operational safety is presented. For the case of PSSE, it is assumed that the attacker only has a partial knowledge of the actual system model. For a set of attacker objectives, the trade-off between the attacker knowledge and possible impact of a successful attack on the performance of false data detection schemes is studied.

Third, the stability of linear hyperbolic systems of PDEs when the boundary control actions and the system parameters switch discontinuously between a finite set of modes is studied. Switched PDE models can describe a class of fault and attack scenarios resulting from intermittent withdrawals through offtake nodes and compromise of sensor-control data. Motivated by such scenarios, a new condition for stability of linear hyperbolic systems of PDEs under arbitrary switching of boundary control actions and system parameters is derived. A class of switching attack strategies is presented, which violate the stability condition and result in unstable flow dynamics.

Fourth, the problem of controlling stochastic linear systems for networked control settings is considered when the sensor-control data is prone to packet loss and jamming. For a class of packet drop models, feedback control policies which minimize a given objective function subject to safety constraints are synthesized. For marginally stable systems, under mild hypotheses on the noise introduced by the control channel and large enough control authority, the synthesis of a control policy that render the state of the closed-loop system mean-square bounded is presented.

Finally, a class of games involving discrete interdependent risks is considered when each player is a NCS, and their security is interdependent due to the exposure to network induced risks. The problem of security decisions of individual players is formulated as a two-stage non-cooperative game defined as follows: in the first stage, the players decide whether to invest in security or not; and in the second stage, they apply control inputs to minimize the average operational costs. The characterization of the equilibria of the game is presented, which includes the determination of the individually optimal security levels. The presence of interdependent security causes a negative externality, and the individual players tend to under invest in security relative to the social optimum. From these results, for a wide parameter range, public policy incentivising higher security investments is desirable.

न चौरहार्यम् न च राज हार्यम् ।
न भ्रातृभाज्यम् न च भारकारी ।।
व्यये कृते वर्धत् एव नित्यं ।
विद्या धनम् सर्व धन प्रधानम् ।।
विद्या ददाति विनयम् विनयाद्याति पात्रताम् ।
पात्रत्वा धनमाप्नोति धनात्धर्मम् ततः सुखम् ।।

For Richa

# Contents

# List of Figures

# Acknowledgments

It is a pleasure to acknowledge the contributions of many people who have made this thesis possible. I am very grateful to my Ph.D. supervisors Professor S. Shankar Sastry and Professor Alexandre M. Bayen for their continuous support of my study and research, and for sharing their enthusiasm and knowledge with me. I thank Prof. Sastry for teaching me control theory, for having faith in my ability, and for providing me the intellectual freedom to explore new ideas. I have benefited from his exceptional vision for identifying new research directions. I am equally grateful to Prof. Bayen for teaching me infrastructure systems engineering, for his dedicated mentorship, and for his technical guidance on numerous occasions. I am inspired by his exceptional leadership in the Mobile Millennium and Floating Sensor Network projects. It has been an honor to be co-advised by Prof. Sastry and Prof. Bayen.

Besides my supervisors, I have immensely enjoyed several discussions with Professor Pravin Varaiya through many insightful comments he made during our meetings. I thank him for being on my thesis committee and for providing useful feedback. I also thank Prof. Mark Stacey for providing useful pointers on estimation and control of water networks, and for being on my thesis committee. I thank Prof. Raja Sengupta, Prof. Steven Glaser, Prof. Laurent El Ghaoui, Prof. Mark Stacey, and Prof. Samer Madanat for being in my qualifying and prelim exam committees. I am grateful to Prof. El Ghaoui for sharing his knowledge of convex optimization with me on several occasions. I was also fortunate to be a student in the classes offered by Prof. Venkat Anantharam and Prof. Ole Hald. I would like to thank the faculty of the Department of Civil and Environmental Engineering. In particular, I thank Prof. Raja Sengupta and Prof. Steven Glaser for admitting me to the CEE systems Ph.D. program.

I have had the terrific mentors and friends throughout my graduate school. Galina A. Schwartz has taught me game theory and has provided me much needed support. I am very grateful to Alexandr A. Kurzhanskiy for sharing his knowledge of transportation systems, and for being a true friend and mentor. Alvaro A. Cárdenas was instrumental in starting the secure control research at the TRUST center, and is responsible for the genesis of many ideas presented in this thesis. Thanks are also due to Falk M. Hante for his collaboration on the topic of stability of switching partial differential equations. I am grateful to Debasish Chatterjee and Peter Hokayem for sharing their research on stochastic model predictive control with me. I look forward to many more exciting interactions with all these researchers. I am indebted to Dr. Xavier Litrico for hosting me at the Cemagref Research Institute, France in October 2009. His support and guidance was instrumental for the success of the experimental cyber-attack on the Gignac water SCADA system.

I am grateful Prof. Manfred Morari and Prof. John Lygeros for inviting me to the Automatic Control Laboratory at ETH Zurich in January 2010. I also enjoyed my visit to the Automatic Control Laboratory in April 2011, and I thank Prof. Karl Henrik Johansson and Prof. Henrik Sandberg for their hospitality. I have enjoyed collaborating with Andre Teixeira on security of power system state estimators, and with Prof. Y. L. Huang and her students on security of process control systems.

I am grateful to the following mentors, colleagues, and friends who have either shared interesting research ideas with me, provided me with academic information, or who have clarified important concepts for me: Alessandro Abate, Parvez Ahammad, Anil Aswani, Sam Burden, Sebastian Blandin, Phoebus Chen, Christian Claudel, Andrew Godbehere, Humberto Gonzalez, Juan-Carlos Herrera, Ryan Herring, Aude Hofleitner, Nikhil Naikal, Tarek Rabbani, Dheeraj Singaraju, Issam Strub, Dengfeng Sun, Andrew Tinka, Qingfang Wu, Bonnie Zhu.

I would like to acknowledge the financial support of the TRUST Science and Technology Center, which receives support from the National Science Foundation (NSF award number CCF-0424422) and the following organizations: AFOSR (#FA9550-06-1-0244), BT, Cisco, DoCoMo USA Labs, EADS, ESCHER, HP, IBM, iCAST, Intel, Microsoft, ORNL, Pirelli, Qualcomm, Sun, Symantec, TCS, Telecom Italia, and United Technologies. In addition, financial support of the Center for Hybrid and Embedded Software Systems (CHESS) at UC Berkeley, and the France-Berkeley fund is gratefully acknowledged.

I am also grateful to the staff of the TRUST Science and Technology Center, in particular, Larry Rohrbough, Jessica Gamble, Gladys Khoury, Mary Stewart, and Gary Given for providing much needed help and support to carry out this research. In addition, I am indebted to our Graduate Student Adviser Ms. Shelley Okimoto for her providing unwavering support to all CEE graduate students.

My friends Marghoob Mohiyuddin, Deepak Kumar, and Sanyam Dangayach deserve special thanks for being there for me when I needed them the most.

I am grateful to my parents, Aruna and Ashwin Patel, and sister Kriti Patel for their support, love and understanding. Lastly, and most importantly, I wish to thank my wife Richa. Nothing I can say will do justice to how I feel for her love and support. To her I dedicate this thesis.

# Chapter 1

# Introduction

Control systems are computer-based systems that monitor and control physical processes. These systems represent a wide variety of networked information technology (IT) systems connected to the physical world. Depending on the application, these control systems are also called Process Control Systems (PCS), Supervisory Control and Data Acquisition (SCADA) systems (in industrial control or in the control of the critical infrastructures), Distributed Control Systems (DCS) or Cyber-Physical Systems (CPS) (to refer to embedded sensor and actuator networks).

Control systems are usually composed of a set of networked agents, consisting of sensors, actuators, control processing units such as programmable logic controllers (PLCs), and communication devices. For example, the oil and gas industry use integrated control systems to manage refining operations at plant sites, remotely monitor the pressure and flow of gas pipelines, and control the flow and pathways of gas transmission. Water utilities can remotely monitor well levels and control the wells pumps; monitor flows, tank levels, or pressure in storage tanks; monitor pH, turbidity, and chlorine residual; and control the addition of chemicals to the water.

Modern day industrial control systems have a multi-layer structure Quin and Badgwell [2003]. The overall objectives of such a control structure are: (1) to maintain safe operational goals by limiting the probability of undesirable behavior, (2) to meet the production demands by keeping certain process values within prescribed limits, (3) to maximize production profit.

Several control applications can be labeled as *safety-critical*: their failure can cause irreparable harm to the physical system being controlled and to the people who depend on it. SCADA systems, in particular, perform vital functions in national critical infrastructure systems, such as electric power distribution, oil and natural gas distribution, water and waste-water treatment, and transportation systems. They are also at the core of health-care devices, weapons systems, and transportation management. The disruption of these control systems could have a significant impact on public health, safety and lead to large economic losses.

Control systems have been at the core of critical infrastructures, manufacturing and industrial plants for many decades, and yet, there have been few confirmed cases of cyber attacks. Control systems, however, are now at a higher risk to computer attacks because their vulnerabilities are increasingly becoming exposed and available to an ever-growing set of motivated and highly-skilled attackers.

No other attack demonstrates the threat to control systems as the Stuxnet worm. The ultimate goal of Stuxnet is to sabotage that facility by reprogramming controllers to operate, most likely, out of their specified boundaries Falliere et al. [2010]. Stuxnet demonstrates that the motivation and capability exists for creating computer attacks capable to achieve military goals Bellovin [2010]. Not only can Stuxnet cause devastating consequences, but it is also very difficult to detect. Because Stuxnet used zero-day vulnerabilities, antivirus software would not have prevented the attack. In fact, the level of sophistication of the attack prevented some well known security companies such as Kaspersky to detect it initially Peterson [2010]. In addition, victims attempting to detect modifications to their embedded controllers would not see any rogue code as Stuxnet hides its modifications with sophisticated PLC rootkits, and validated its drivers with trusted certificates.

In this thesis it is argued that attackers may be able to hide the specific information technology methods used to exploit the system and reprogram their computers; however, they cannot hide their final goal: the need to cause an adverse effect on the physical system by sending malicious sensor or controller data that will not match the control behavior expected by a diagnostic system or an an anomaly detection system at the supervisory control layer. In order to address this problem, this thesis explores attack detection mechanisms that detect attacks by monitoring the physical system under control. Our goal is to detect modifications to the sensed or controlled data as soon as possible, before the attack causes irreversible damages to the system (such as violating safety margins and causing instability).

In the rest of the chapter, we first summarize the vulnerability of control systems by discussing known attacks. We then discuss the efforts for securing control systems solely from an information technology perspective and identify the new and unique research problems that can be formulated by including a model of the physical system under control.

## 1.1 The Vulnerability of Control Systems and Stuxnet

There have been many computer-based incidents in control systems. Computer-based accidents can be caused by any unanticipated software error, like the power plant shutdown caused by a computer rebooting after a patch Krebs [2008]. Non-targeted attacks are incidents caused by the same attacks that any computer connected to the Internet may suffer, such as the Slammer worm infecting the Davis-Besse nuclear power plant Turk [2005], or the case of a controller being used to send spam in a water filtering plant.

However, the biggest threat to control systems are targeted attacks. These attacks are the ones where the miscreants know that they are targeting control systems, and therefore, *they tailor their attack strategy with the aim of damaging the physical system under control.* Targeted attacks against control systems are not new. Physical attacks–for extortion and terrorism–are a reality in some countries CCTV [2002]. Cyber-attacks are a natural progression to physical attacks: they are cheaper, less risky for the attacker, are not constrained by distance, and are easier to replicate and coordinate.

A classic computer-based targeted attack to SCADA systems is the attack on Maroochy Shire Council's sewage control system in Queensland, Australia Slay and Miller [2007]. There are many other reported targeted attacks Attorney [2007]; Greenberg [2008]; Kravets [2009]; Leyden [2008]; Quinn-Judge [2002]; Reed [2004]; however, no other attack has demonstrated the threats that control systems are subject to as well as the Stuxnet worm Falliere et al. [2010]; Langner [2010]. Stuxnet has made it clear that there are groups with the motivation and skills to mount sophisticated computer-based attacks to critical infrastructures, and that these attacks are not just speculations or belong only in Hollywood movies.

Stuxnet intercepts routines to read, write and locate blocks on a Programmable Logic Controller (PLC). By intercepting these requests, Stuxnet is able to modify the data sent to or returned from the PLC without the operator of the PLC ever realizing it Falliere et al. [2010]. Stuxnet was discovered on systems in June 2010 by researchers from Belarus–from the company VirusBlokAda; however, it is believed to have been released more than a year before. Stuxnet is a worm that spreads by infecting Windows computers. It uses multiple methods and zero-day exploits to spread itself via LANs or USB sticks. It is likely that propagation by LAN served as the first step, and propagation through removable drives was used to reach PCs not connected to other networks–therefore being isolated from the Internet or other networks is not a complete defense.

Once Stuxnet infects a computer, It installs its own driver into Windows computers. Because these drivers have to be signed, Stuxnet used two stolen certificates. Stuxnet also installs a rootkit to hide itself. The goal of the worm in a Windows computer is to search for WinCC/Step 7, a type of software used to program and monitor PLCs. (PLCs are the embedded systems attached to sensors and actuators that run control algorithms to keep the physical system operating correctly. They are typically programmed with a ladder logic program: a logic traditionally used to design control algorithms for panels of electromechanical relays.)

If Stuxnet does not find the WinCC/Step 7 software in the infected Windows machine, it does nothing; however, if it finds the software, it infects the PLC with another zero-day exploit, and then reprograms it. Stuxnet also attempts to hide the PLC changes with a PLC rootkit. The reprogramming is done by changing only particular parts of the code–overwriting certain process variables every five seconds and inserting rouge ladder logic–therefore it is impossible to predict the effects of this change without knowing exactly how the PLC is originally programmed and what it is connected to, since the PLC program depends on the physical system under control, and typically, physical system parameters are unique to each individual facility. This means that the attackers were targeting a very specific PLC program and configuration (i.e., a very specific control system deployment).

Many security companies, including Symantec and Kaspersky have said that Stuxnet is the most sophisticated attack they have ever analyzed, and it is not difficult to see the reasons. Stuxnet uses four zero-day exploits, a Windows rootkit, the first known PLC rootkit, antivirus evasion techniques, peer-to-peer updates, and stolen certificates from trusted certification authorities (CAs). There is evidence that Stuxnet kept evolving since its initial deployment as attackers upgraded the infections with encryption and exploits, apparently adapting to conditions they found on the way to their target. The command and control architecture used two servers if the infected machines were able to access the Internet, or a peer to peer messaging system that could be used for machines that are offline. In addition, the attackers had a good level of intelligence about their target; they knew all the details of the control system configuration and its programs. The sophistication of this attack has led many to believe Stuxnet is the creation of a state-level sponsored attack.

This thesis puts forth the viewpoint that a threat like the Stuxnet worm must be dealt with defense-in-depth mechanisms like anomaly detection schemes. While traditional anomaly detection mechanisms may have some drawbacks like false alarms, it is shown in this thesis that for certain control systems, anomaly detection schemes focusing on the physical system–instead of using software or network models–can provide good detection capabilities with negligible false alarm rates.

## 1.2 New Security Problems for Control Systems

### 1.2.1 Efforts for Securing Control Systems

Most of the efforts for protecting control systems (and in particular SCADA) have focused on safety and reliability (the protection of the system against random and/or independent faults). Traditionally, control systems have not dealt with intentional actions or systematic failures. There is, however, an urgent growing concern for protecting control systems against malicious cyberattacks Byres and Lowe [2004]; Eisenhauer et al. [2006]; Geer [2006]; Igure et al. [2006]; Oman et al. [2000]; Turk [2005]; US-CERT [2008].

There are several industrial and government-led efforts to improve the security of control systems. Several sectors–including chemical, oil and gas, and water–are currently developing programs for securing their infrastructure. The electric sector is leading the way with the North American Electric Reliability Corporation (NERC) cybersecurity standards for control systems NERC-CIP [2008]. NERC is authorized to enforce compliance to these standards, and it is expected that all electric utilities are fully compliant with these standards by the end of 2010.

NIST has also published a guideline for security best practices for general IT in Special Publication 800-53. Federal agencies must meet NIST SP800-53. To address the security of control systems, NIST has also published a Guide to Industrial Control System (ICS) Security Stouffer et al. [2006], and a guideline to smart grid security in NIST-IR 7628. Although these recommendations are not enforceable, they can provide guidance for analyzing the security of most utility companies. ISA (a society of industrial automation and control systems) is developing ISA-SP 99: a security standard to be used in manufacturing and general industrial controls.

The Department of Energy has also led security efforts by establishing the national SCADA test bed program INL [2010] and by developing a 10-year outline for securing control systems in the energy sector Eisenhauer et al. [2006]. The report–released in January 2006–identifies four main goals (in order from short-term goals to long-term goals): (1) measure the current security posture of the power grid, (2) develop and integrate protective measures, (3) implement attack detection and response strategies; and (4) sustain security improvements.

The use of wireless sensor networks in SCADA systems is becoming pervasive, and thus we also need to study their security. A number of companies have teamed up to bring wireless sensor network technology in the field of process control systems, and currently, there are two working groups to standardize their communications Hart [2007]; ISA [2007]. Their wireless communication proposal has options to configure hop-by-hop and end-to-end confidentiality and integrity mechanisms. Similarly, they provide the necessary protocols for access control and key management.

All these efforts have essentially three goals: (1) create awareness of security issues with control systems, (2) help control systems operators and IT security officers design a security policy, and (3) recommend basic security mechanisms for prevention (authentication, access controls, etc), detection, and response to security breaches.

While these recommendations and standards have placed significant importance on survivability of control systems (their ability to operate while they are under attack), this thesis explores some new research problems that arise when control systems are under attack.

### 1.2.2 Control System Security vs. IT Security

It is clear that the security of control systems has become an active area in recent years. However, there is a pressing need to articulate what is new and fundamentally different in this field from a research point of view when compared to traditional IT security. In this section, we would like to start this discussion by summarizing some previously identified differences and by proposing some new problems.

The property of control systems that is most commonly brought up as a distinction with IT security is that software patching and frequent updates, are not well suited for control systems. For example, upgrading a system may require months of advance in planning how to take the system offline; it is, therefore, economically difficult to justify suspending the operation of an industrial computer on a regular basis to install new security patches. Some security patches may even violate the certification of control systems, or–as previously mentioned–cause accidents to control systems Krebs [2008].

Patching, however, is not a fundamental limitation to control systems. A number of companies have demonstrated that a careful antivirus and patching policy (e.g., the use of tiered approaches) can be used successfully Cosman [2006]. Also, most of the major control equipment vendors now offer guidance on both patch management and antivirus deployment for their control products. Thus there is little reason for SCADA system operators not to have good patch and antivirus programs in place today Byres et al. [2007].

Large industrial control systems also have a large amount of legacy systems. Several research efforts have tried to provide lightweight cryptographic mechanisms to ensure data integrity and confidentiality Tsang and Smith [2008]; Wright et al. [2004]. The recent IEEE P1711 standard is designed for providing security in legacy serial links Hurd et al. [2008]. Having some small level of security is better than having no security at all; however, it is widely believed that most of the efforts done for legacy systems can only be considered as short-term solutions. For properly securing critical control systems the underlying technology must satisfy some minimum performance requirements to allow the implementation of well tested security mechanisms and standards.

Another property of control systems that is commonly mentioned is the real-time requirements of control systems. Control systems are autonomous decision making agents which need to make decisions in real time. While availability is a well studied problem in information security, real-time availability provides a stricter operational environment than most traditional IT systems. In this thesis it is shown that real-time availability requirements depend on the dynamics of the physical system.

Not all operational differences are more severe in control systems than in traditional IT systems. By comparison to enterprise systems, control systems exhibit comparatively simpler network dynamics: Servers change rarely, there is a fixed topology, a stable user population, regular communication patterns, and a limited number of protocols. Therefore, implementing network intrusion detection systems, anomaly detection, and white listing may be easier than in traditional enterprise systems Cheung et al. [2007].

### 1.2.3 What is new and fundamentally different?

While all these differences are important, the major distinction of control systems with respect to other IT systems is the interaction of the control system with the physical dynamics. While current tools from information security can give *necessary* mechanisms for securing control systems, these mechanisms alone are not *sufficient* for defense-in-depth of control systems. When attackers bypass basic security defenses they may be able to affect the physical world. In particular, research in computer security has focused traditionally on the protection of information; but it has not considered how attacks affect estimation and control algorithms–and ultimately, how attacks affect the physical world. This thesis proposes that a systematic framework for securing control systems should focus on three fundamentally new areas:

1. Better understand the consequences of an attack for *risk assessment*: While there has been previous risk assessment studies on cyber security for SCADA systems Craig et al. [2008]; Hamoud et al. [2003]; Oman et al. [2000]; Ralston et al. [2007], currently, there are few studies on identifying the attack strategy of an adversary, once it has obtained unauthorized access to some control network devices. One notable exception is the study of false data-injection attacks to power grids Liu et al. [2009]. Further research is needed to understand the threat model in order to design appropriate defenses and to invest in securing the most critical sensors or actuators.

2. Design new *attack-detection algorithms*: By monitoring the behavior of the physical system under control, one should be able to detect a wide range of attacks by compromised measurements. The work presented in Rrushi [2009] is worth mentioning here althouth it does not consider dynamical models of the process control system. This thesis introduces dynamical system models used in control theory as a tool for specification-based intrusion detection systems (regardless of how an attacker obtained its unauthorized privileges).

3. Design new *attack-resilient algorithms*: A resilient control system is one that maintains an accepted level of operational normalcy in response to disturbances, including random disturbances and malicious attacks. The design goal is then to develop control algorithms where even if attackers manage to bypass some basic security mechanisms, they will still face several control-specific security devices that will minimize the damage done to the system. Thus, there is a particular need to investigate how control systems can be reconfigured and adapted when they are under an attack. Prior work has not fully addressed the design of new control algorithms or reconfiguration algorithms which are able to withstand attacks, or that reconfigure their operations based on detected attacks. There is previous work on fault detection and isolation; however, as we explain in this thesis, these systems are not enough for a complete diagnosis of deception attacks launched by an intelligent attacker with knowledge on how to evade fault detection methods used by the system.

In the next chapters, the ideas, experiments, and results for each of the three areas are presented, i.e., (1) risk-assessment, (2) attack diagnosis, and (3) resilient control mechanisms. We first present a general theory for approaching the topic, and then implement our ideas to various experimental scenarios.

# Chapter 2

# Attacks on Hierarchically Structured Water SCADA Systems

The goal of this chapter is to perform security risk assessment for hierarchically structured supervisory control and data acquisition (SCADA) systems used to monitor and control water distribution networks. The analysis presented in this chapter includes the performance assessment of a proportional-integral (PI) control based regulation method and a model-based supervisory scheme for fault detection and isolation (FDI), under deception attacks on water canal distribution systems. These systems typically use IT-enabled communications and therefore, are representative of SCADA systems used to operate physical infrastructures. In order to test the resiliency of control methods, this work adopts a conservative approach by assuming that the attacker has knowledge of: 1) the approximate system dynamics, 2) the parameters of FDI scheme, and 3) the sensor-control signals. A deception attack to enable water pilfering from the canal system is proposed, and it is demonstrated that the attack is realizable in practice by implementing it on the Gignac canal system in Southern France.

## 2.1 Introduction

Security of water SCADA systems has become an area of considerable focus Weiss [2010]. The question then arises as to what security mechanisms for water SCADA can make them resilient against cyber-attacks (and enable them to degrade gracefully under very powerful attacks). One of the goals of this chapter is to highlight that only a sustained progress in risk assessment and mitigation for NCS security can achieve this goal. In a noteworthy government-industry initiative, a ten-year roadmap to secure control systems in the water sector was released in March, 2008 WSCC-CSWG [2008]. This roadmap advocates development of risk assessment and mitigation measures for water NCS/SCADA systems so that they continue to operate with no loss of critical function during and after a cyber event. In the context of NCS, security risk assessment will involve: 1) determination of the likelihood that an attacker will obtain unauthorized access to one or more NCS components and will successfully compromise their function, and 2) the computation of (physical and operational) losses associated with that particular compromise. Risk mitigation for NCS will involve: 1) the development of real-time state-monitoring systems for intrusions to NCS, and 2) the design of attack-resilient control methods which can reconfigure and adapt to maintain critical NCS functionality under attack.

The aim of this chapter is to perform threat assessment for the Gignac water SCADA system located in Southern France. The performance of regulatory and supervisory control methods under deception attacks on the sensor-control data is analyzed. Although the topic of water contamination through cyber means is an important one, it is not the focus of this chapter and has been studied elsewhere (e.g., Krause and Guestrin [2009]). The main contributions of this chapter are as follows:

- The effect of cyber attacks to sensor measurements on the performance of a commonly used regulation method, which uses distributed proportional-integral (PI) controllers, is investigated. The performance of a model-based supervisory scheme for fault detection and isolation (FDI) under a class of deception attacks is also analyzed. The scheme chosen here is one of the several available FDI methods, all of which use model generated residuals, e.g., Bedjaoui and Weyer [2011]. The performance assessment of other detection methods under attacks can be done in a similar manner, e.g., Cárdenas et al. [2011].

- Next, the results from a field operational test in which deception attacks were implemented on the Gignac water SCADA system are presented. These attacks model the attacker as an intelligent insider who is resourceful enough to obtain access to sensor-control data and has knowledge to evade the FDI scheme. The field operational test shows that such an attack enables water pilfering from the canal system thereby increasing water loss, decreasing operational efficiency.

This chapter is organized as follows: In Section 2.2, we present a taxonomy of cyber-attacks on hierarchically structured SCADA systems which typically manage the operations of automated canal networks. A model of cyber-attacks on level sensors is also specified. The regulatory control method and the supervisory FDI scheme, which we use to analyze the effect of cyber-attacks, are presented in Section 2.3. The performance of the supervisory FDI scheme under a stealthy deception attack is also investigated by way of simulation. In Section 2.4, salient features of the Gignac SCADA system are presented. The results from our field experiment, in which a deception attack was implemented to enable water pilfering from the canal, are also discussed. Finally, salient points of our analysis are summarized in Section 2.5. In Chapter 3, we use the insights gained in this chapter to develop a diagnostic scheme to better detect and isolate deception attacks and suggest some ways to defend against them. The diagnostic method presented in Chapter 3 uses an enhanced hydrodynamic model, and performs well in a range of security scenarios.

## 2.2 Cyber-Attacks Against Water SCADA Systems

Modern water SCADA systems have a hierarchical structure with at least two levels of control: regulatory control and supervisory control. The regulatory control layer directly interacts with the hydrodynamics of the physical canal network through sensors and actuators. These field devices are connected via a *field area network* to PLCs or remote terminal units (RTUs), which in turn implement local control actions (regulatory control). Under a decentralized regulatory control policy, a PLC may also interact with the neighboring PLCs via the field area network. A *control network* carries (real-time) data between regulatory controllers (or PLCs) and supervisory workstations. These workstations are used for data logging, diagnostic functions such as fault diagnosis or FDI, and supervisory control computations such as set-point control and controller reconfigurations (e.g., see Section 2.4.1 about Gignac SCADA implementation). Finally, authorized remote users (e.g., canal managers) can access information about the canal network and provide specifications to the supervisory layer via a *corporate network*.

Attacks on cyber-infrastructure of water SCADA systems can result in partial or complete loss of operational performance such as closed-loop stability, safety with respect to over-topping, or performance loss. Cyber-attacks to water SCADA systems, and in general to networked control systems (NCS), can be broadly classified as either deception attacks or denial-of-service (DoS) attacks. Integrity of sensor and control data packets refers to their trustworthiness, and lack of integrity results in deception. Availability refers to the ability of all the system components of being accessible and usable when needed, and lack of availability results in DoS. While confidentiality, which refers to the system's ability to keep information secret from unauthorized users, is an important security attribute for IT systems, integrity and availability take a natural precedence for security of SCADA systems. We now explain the characteristics of deception and DoS attacks in the context of water SCADA systems.

*Integrity* for automated water SCADA systems can be defined as the ability to maintain operational goals by preventing, detecting, or surviving deception attacks in the information sent and received by the sensors (e.g., water level measurements), the controllers (e.g., desired discharges, set-points), and the actuators (e.g., commanded gate openings). Deception or false information can include an incorrect sensor measurement or command input, a time stamp which is different from the actual time, or a wrong identity of the sending device. The adversary can launch these attacks by obtaining the secret keys used by the sending devices, or by compromising some of the sensors and actuators. During compromise of the field area network, the adversary may send false measurement data to the regulatory controllers, thereby affecting the performance of the closed-loop system. Similarly, manipulation of actual gate openings can result in unintended gate movements. However, during a compromise of the control network, multiple sensor and controller signals can be compromised. In addition to reduction in the control performance, such deception attacks can also cause the FDI schemes to report false alarms (when there is no malfunction) and missed/delayed detection (when there is an actual attack). False alarms can result in waste of maintenance resources, and may ultimately result in loss of canal manager's confidence in the SCADA system. However, the case of missed or delayed detection is more problematic because the attacker can almost arbitrarily affect the control functions, which may result in considerable losses.

*Availability* in the context of water SCADA systems can be defined as the ability to main-

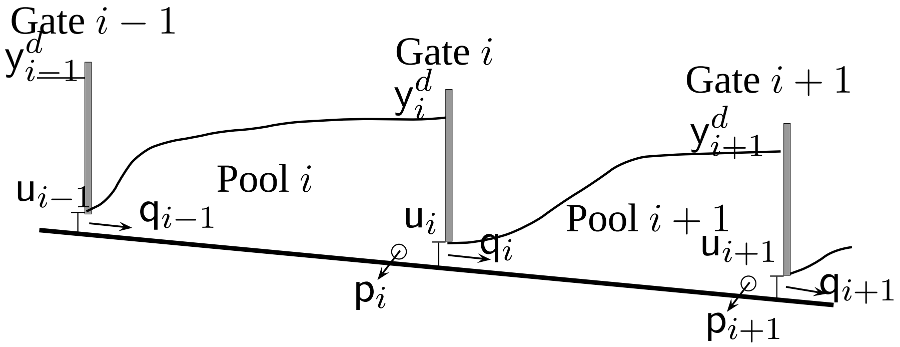


Figure 2.1: Schematic view of a multiple canal system with free-flow gates.

tain operational goals by preventing or surviving DoS attacks to information collected by the sensors, commands given by controllers, and the actions implemented by the actuators. To launch a DoS attack, the adversary can jam the communication channels, compromise field devices and prevent them from sending data, attack the communication protocols used by field or control networks, flood the network with random data etc. While in numerous computer systems a temporary DoS attack may not result in long-term compromise of their service (these system will operate normally again after DoS), water SCADA operation is often subject to real-time constraints. For example, a DoS on sensor measurements may prevent the regulatory controllers from maintaining water level fluctuations within safety bounds. This can compromise the control performance, and may even cause instability. In a worst case, a prolonged DoS may even render the SCADA system unusable for a prolonged time period.

### 2.2.1 Attack Models

We now discuss the possible cyber attacks against a generic hierarchically controlled water SCADA system. Fig. 2.2 shows the hierarchical control structure and possible attacks for a cascade canal system of canal pools. Here we assume that the regulation gates are in free-flow condition at their downstream end (see Fig. 2.1).

Under these conditions, the flow downstream of the gate is super-critical, and the transition to sub-critical flow happens due to a hydraulic jump. In Section 2.3, we elaborate on the regulatory control and supervisory FDI methods. For the $i-$th pool, we denote the discharge (m$^3$/s) at the upstream end (resp. downstream end) by $q_{i-1}$ (resp. $q_i$), the water-level (m) at the downstream end by $y_i^d$, and the offtake water withdrawal (m$^3$/s) at the occurring at the downstream end by $p_i$. We will assume that $q_{i-1}$ and $q_i$ are the control input variables, $y_i^d$ is the measurement variable, and $p_i$ is the disturbance variable. These variables are the respective deviations around a steady state flow.

Six possible types of cyber-attacks **A0**–**A6** are illustrated in Fig. 2.2. The attack **A0** denotes physical attack against the physical infrastructure (gates, offtakes) or the field devices (sensors and actuators). Deterrence and prevention of this attack can be achieved by implementing physical security mechanisms such as fences, surveillance cameras, etc. Since

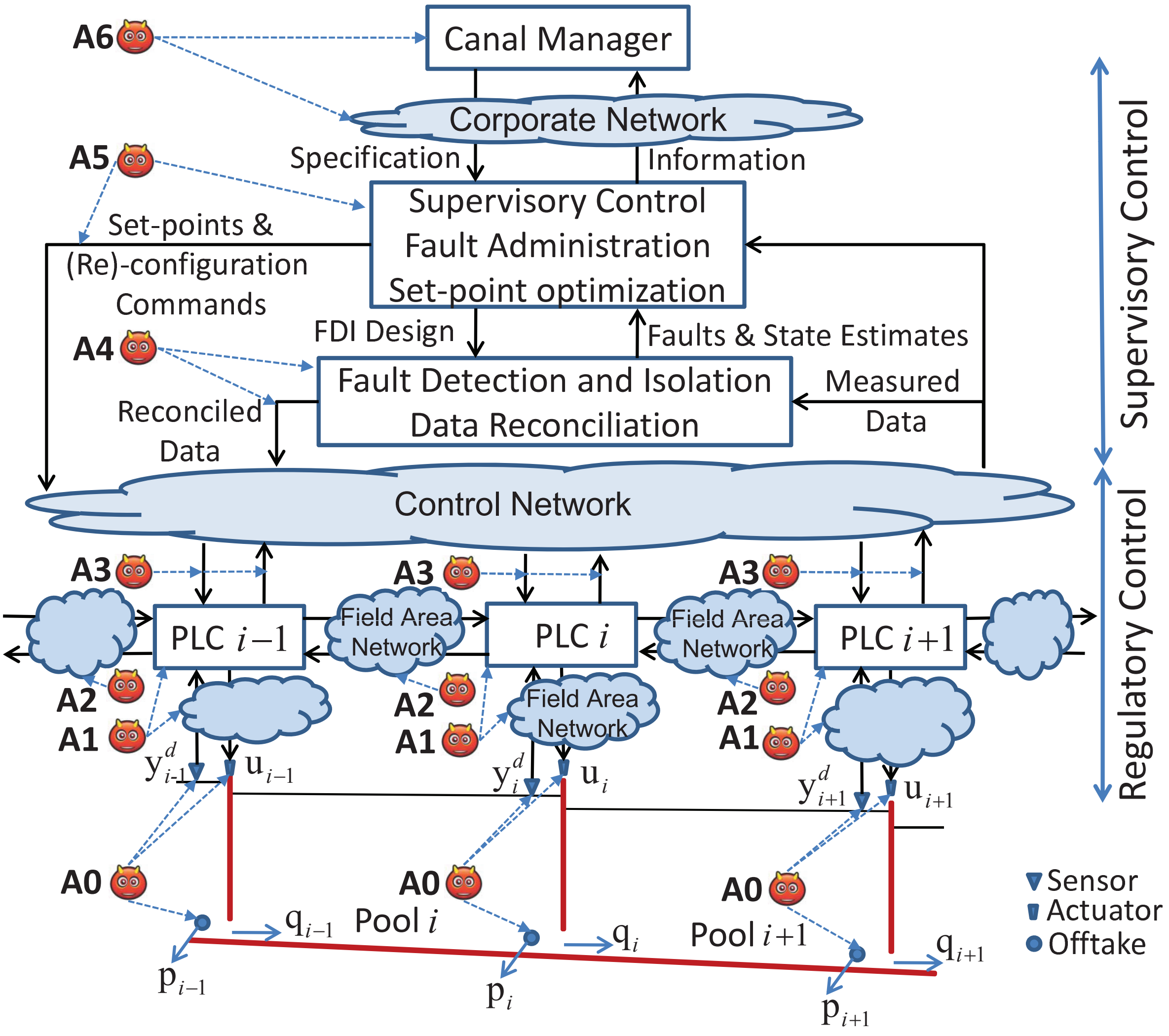


Figure 2.2: Cyber-Attacks on hierarchically structured water SCADA systems.

Table 2.1: Taxonomy of Cyber-attacks.

| | Control Layer | |
|---|---|---|
| | Regulatory Control | Supervisory Control |
| Deception Attacks | spoofing, replay | set-point change |
| | measurement substitution | tuning parameter substitution |
| DoS Attacks | physical jamming | network flooding |
| | increase comm. latency | disrupt process operation |

a physical attack requires access to the canal infrastructure, a risk averse attacker is more likely to launch cyber-attacks **A1**–**A6** on SCADA system components and communications. We thus do not consider physical attacks in the rest of this chapter.

Attack **A1** denotes the DoS attacks (via jamming or increasing communication latency) on the communication between the PLCs and the field devices, or the deception attack (via spoofing or replaying) of the sensor measurements $\mathsf{y}_i^d$ and control actuations $\mathsf{u}_i$. Attack **A2** denotes similar DoS or deception attacks on inter-PLC communication. This may adversely affect the interaction between the canal pools and result in amplification of disturbances across the canal cascade. By **A3** we mean cyber-attacks on the control network which enables communication between the regulatory and the supervisory control layers. This network transmits 1) water level measurements $\mathsf{y}_i^d$, gate openings $\mathsf{u}_i$, and discharge readings $\mathsf{q}_i$ from the PLCs to the supervisory control layer, and 2) set-points for levels, target dispatches through offtakes, the reconciled data and commands needed for control loop reconfiguration, and tuning of controller parameters from the supervisory control layer to the PLCs. Thus, compromise of control network (via man-in-the-middle attacks, flooding attacks) may result in wrong inputs to the fault diagnosis scheme, unintended set-point changes, incorrect tuning parameters, etc.

Attacks **A4** and **A5** denote the attacks on the supervisory control layer, which has state estimators for data reconciliation and observers for attack or fault diagnosis. The state estimates and diagnostic information is used to generate set-points and control configurations by using an optimization based procedure, or more commonly by employing human expertise. Possible attacks here could be manipulation of state estimators and observers for A/FDI so that incorrect estimates and alerts are generated. As a consequence, one or more supervisory control functions may behave in a bad manner. Of course, attacks **A1**–**A3** on the regulatory layer will also affect the performance of supervisory layer, since the latter could be fed with bad data (deception attacks) or no data at all (DoS attacks) when the former is under attack. Finally, **A6** denotes the attack by a malicious insider who can assume the role of the canal manager. In Table 2.1, a summary of DoS and deception cyber-attacks on SCADA control layers is presented.

We now present the model of cyber-attacks specific to level sensor $\mathsf{y}_i^d$ measurements; attacks on control signals can be similarly modeled. Each sensor measurement can be assumed to have a nominal range $\mathcal{Y}_i$ which captures all operating conditions, i.e., $\mathsf{y}_i^d(t) \in \mathcal{Y}_i$, for all $t$. We also assume that each sensor is uniquely authenticated via a cryptographic key. The notation $\tilde{\mathsf{y}}_i^d(t)$ denotes the measurements received by the regulatory and supervisory control system at time $t$. If the $i-$th sensor is under attack, $\tilde{\mathsf{y}}_i^d(t)$ may be different from the real measurement $\mathsf{y}_i^d(t)$; however, it can be assumed that the attacked signals $\tilde{\mathsf{y}}_i^d(t)$ also lie within $\mathcal{Y}_i$ (signals outside this range can be easily detected by standard data reconciliation methods). Furthermore, once the attack is successful, the attacker is likely to continue the attack until he/she exhausts available resources (e.g., battery power used for jamming) or achieves the final goal (e.g., over topping or water pilfering). Thus, it is reasonable to assume *block attacks* of duration $\mathcal{T} := [\tau^s, \tau^e]$; between the start time $\tau^s$ and stop time $\tau^e > \tau^s$. Under these assumptions, a general model for attacks on the sensor signals is the following:

$$\tilde{\mathsf{y}}_i^d(t) = \begin{cases} \mathsf{y}_i^d(t), & \text{for } t \notin \mathcal{T} \\ \mathsf{g}_i(t), & \text{for } t \in \mathcal{T}, \mathsf{g}_i(t) \in \mathcal{Y}_i \end{cases} \tag{2.1}$$

where $\mathsf{g}_i(t)$ is the attack signal.

The above sensor attack model can be used to represent both deception and DoS attacks. In an integrity attack, we can assume that the sensor is compromised and an arbitrary false value $\mathsf{g}_i(t)$ is injected. The goal of SCADA system's A/FDI scheme is to detect the attack as fast as possible, and identify the compromised sensor $i$. In a DoS attack, it can be assumed that lack of available measurements will be detected by the SCADA system, and it uses $\tilde{\mathsf{y}}_i^d(t) = 0$ (no signal) or $\tilde{\mathsf{y}}_i^d(t) = \tilde{\mathsf{y}}_i^d(\tau^s)$ (last available measurement) to generate control inputs.

## 2.3 Flow Models & Hierarchical System Architecture

In this section, first, a dynamical model commonly used in control design for cascaded canal systems is presented; and second, a frequency-domain controller for regulatory control layer and a model-based FDI scheme for the supervisory control layer are developed.

### 2.3.1 Model of Canal Cascade

The following frequency domain input-output relationship has been obtained by Litrico and Fromion by taking the Laplace transform of the linearized shallow water equations (see Chapter 3 in Litrico and Fromion [2009*b*]):

$$\hat{\mathsf{y}}_i^d(s) = p_{i,21}(s)\hat{\mathsf{q}}_{i-1}(s) + p_{i,22}(s)\left(\hat{\mathsf{q}}_i(s) + \hat{\mathsf{p}}_i(s)\right), \tag{2.2}$$

where $s$ is the Laplace variable, and $p_{i,21}(s)$ (resp. $p_{i,22}(s)$) denotes the infinite-dimensional transfer function from $\mathsf{q}_{i-1}$ (resp. $\mathsf{q}_i$ and $\mathsf{p}_i$) to $\mathsf{y}_i^d$. For uniform flow regime, the transfer functions $p_{i,21}(s)$ and $p_{i,22}(s)$ belong to an algebra of irrational transfer function called the *Callier-Desoer class*; powerful methods for direct controller design exist for such systems Litrico and Fromion [2009*a*]. Similar results can also be proven for the non-uniform flow regime; however, this is beyond the scope of our chapter. For low frequencies, these transfer functions can be approximated by the following integrator-delay (ID) model:

$$p_{i,21}(s) \approx \frac{a_i^d}{s} e^{-\tau_i s}, \quad p_{i,22}(s) \approx -\frac{a_i^d}{s} \tag{2.3}$$

where $a_i^d$ is the inverse of equivalent backwater area ($\mathrm{m}^{-2}$) and $\tau_i$ is the propagation delay (s). Using equation (2.2), the multi-pool representation of the canal cascade is obtained as

$$\hat{\mathsf{y}}^d(s) = \mathcal{G}(s)\hat{\mathsf{q}}(s) + \tilde{\mathcal{G}}(s)\hat{\mathsf{p}}(s), \tag{2.4}$$

where

$$\mathsf{y}^d = \begin{pmatrix} \mathsf{y}_1^d & \dots & \mathsf{y}_m^d \end{pmatrix}, \quad \mathsf{q} = \begin{pmatrix} \mathsf{q}_0 & \dots & \mathsf{q}_m \end{pmatrix}, \quad \mathsf{p} = \begin{pmatrix} \mathsf{p}_1 & \dots & \mathsf{p}_m \end{pmatrix},$$

and $\mathcal{G}(s) = (g_{jk}(s))$ is a $m \times (m+1)$ dimensional bidiagonal matrix, and $\tilde{\mathcal{G}}(s) = (\tilde{g}_{jk}(s))$ is a $m \times m$ dimensional diagonal matrix. For example, for single canal pool ($i = 1$),

$$\mathcal{G}(s) = \begin{pmatrix} \frac{a_1^d}{s} e^{-\tau_1 s} & -\frac{a_1^d}{s} \end{pmatrix}, \quad \tilde{\mathcal{G}}(s) = -\frac{a_1^d}{s}, \tag{2.5}$$

and for 2−pool system,

$$\mathcal{G}(s) = \begin{pmatrix} \frac{a_1^d}{s} e^{-\tau_1 s} & -\frac{a_1^d}{s} & 0 \\ 0 & \frac{a_2^d}{s} e^{-\tau_2 s} & -\frac{a_2^d}{s} \end{pmatrix}, \quad \tilde{\mathcal{G}}(s) = \begin{pmatrix} -\frac{a_1^d}{s} & 0 \\ 0 & -\frac{a_2^d}{s} \end{pmatrix}. \tag{2.6}$$

We will henceforth consider a 2−pool system, noting that our analysis can be easily extended to multi-pool system.

Taking the inverse Laplace transform of (2.4) for $m = 2$, we obtain the following time-domain model with delayed inputs:

$$\begin{aligned} \dot{\mathsf{y}}_1^d(t) &= a_1^d \mathsf{q}_0(t - \tau_1) - a_1^d \left[\mathsf{q}_1(t) + \mathsf{p}_1(t)\right], \\ \dot{\mathsf{y}}_2^d(t) &= a_2^d \mathsf{q}_1(t - \tau_2) - a_2^d \left[\mathsf{q}_2(t) + \mathsf{p}_2(t)\right]. \end{aligned} \tag{2.7}$$

Each regulation gate is represented by the following linearized model around the steady state:

$$\mathsf{q}_i(t) = b_i^d \mathsf{y}_i^d(t) + k_i \mathsf{u}_i(t), \quad i = 1, 2 \tag{2.8}$$

where $\mathsf{u}_i(t)$ is the controlled gate opening, and the constant $b_i^d$ (resp. $k_i$) denotes the gain of the upstream level $\mathsf{y}_i^d$ (resp. gate opening $\mathsf{u}_i(t)$). Combining (2.8) and (2.7), we obtain the state-space representation of the 2−pool system with delayed state and inputs:

$$\begin{aligned} \dot{\mathsf{x}}(t) &= \sum_{i=0}^{2} A_i \mathsf{x}(t - \tau_i) + \sum_{i=0}^{2} B_i \mathsf{u}(t - \tau_i) \\ \mathsf{y}(t) &= C\mathsf{x}(t), \end{aligned} \tag{2.9}$$

where $\mathsf{x} := \begin{pmatrix} \mathsf{y}_1^d, & \mathsf{y}_2^d \end{pmatrix}^\mathsf{T} \in \mathbb{R}^2$ is the state, $\mathsf{u} := \begin{pmatrix} \mathsf{u}_0, & \mathsf{u}_1, & \mathsf{p}_1, & \mathsf{p}_2 \end{pmatrix}^\mathsf{T} \in \mathbb{R}^4$ denotes the known input, $\mathsf{y} := \begin{pmatrix} \mathsf{y}_1^d, & \mathsf{y}_2^d \end{pmatrix}^\mathsf{T} \in \mathbb{R}^2$ is the measured output (perfect state measurements); $\tau_0 = 0$,

$\tau_1 = \underline{\tau}_1$, $\tau_2 = \underline{\tau}_2$. The system matrices are respectively given by $C = \operatorname{diag}\begin{pmatrix}1, & 1\end{pmatrix}$, and

$$
\begin{aligned}
A_0 &= \begin{pmatrix} -a_1^d b_1^d & 0 \\ 0 & -a_2^d b_2^d \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 \\ a_2^d b_1^d & 0 \end{pmatrix}, \\
B_0 &= \begin{pmatrix} 0 & -a_1^d k_1 & -a_1^d & 0 \\ 0 & 0 & 0 & -a_2^d \end{pmatrix}, \quad B_1 = \begin{pmatrix} a_1^d k_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & a_2^d k_1 & 0 & 0 \end{pmatrix}.
\end{aligned}
\tag{2.10}
$$

Let $\tau_{\max}$ denote the upper bound of the time delays $\tau_i$, $i = 0, 1, 2$. In practice, controller design for canal regulation is based on frequency-domain models while supervisory level fault diagnosis is based on time-domain models, as explained next.

### 2.3.2 Regulatory Control

The regulatory control layer is responsible for maintaining the operational performance of the canal cascade by implementing dynamic feedback control actions Litrico et al. [2007]. Two performance objectives are commonly specified at this layer: 1) water is efficiently delivered to the end users, and 2) unknown perturbations (or disturbances) are rejected. Here we briefly discuss the structure of frequency-domain based controllers. Following the approach in Litrico and Fromion [2005], we choose frequency-domain controllers since they have been classically used for managing canal systems; however, recently, model-predictive control (MPC) designs have been also been proposed, e.g., Negenborn et al. [2009]. Let $\mathsf{y}_i^r$ denote the set-point or the reference level for pool $i$, which is typically obtained as a result of an optimization problem by the supervisory layer. The aim of the regulatory control is to regulate $\mathsf{y}_i^d$ to set-point $\mathsf{y}_i^r$. Let the output error be defined as $\epsilon_i := (\mathsf{y}_i^r - \mathsf{y}_i^d)$, and $\mathsf{y}^r := \begin{pmatrix}\mathsf{y}_1^r & \dots & \mathsf{y}_m^r\end{pmatrix}$, $\epsilon := \begin{pmatrix}\epsilon_1 & \dots & \epsilon_m\end{pmatrix}$. Let $\mathcal{K}(s)$ denote the Laplace transform of the multi-variable controller $\mathcal{K}$, i.e.,

$$\hat{\mathsf{q}}(s) = \mathcal{K}(s)\hat{\epsilon}(s). \tag{2.11}$$

From (2.4) and (2.11), we see that the control input vector $\hat{\mathsf{q}}(s)$ is given by:

$$\hat{\mathsf{q}}(s) = \mathcal{S}_{\mathsf{q}}(s)\mathcal{K}(s)\hat{\mathsf{y}}^r(s) - \mathcal{S}_q(s)\mathcal{K}(s)\tilde{\mathcal{G}}(s)\hat{\mathsf{p}}(s) \tag{2.12}$$

with $\mathcal{S}_q(s) := (I + \mathcal{K}(s)\mathcal{G}(s))^{-1}$ the input sensitivity function and the output error $\epsilon$ by:

$$\hat{\epsilon}(s) = \mathcal{S}_\epsilon(s)\hat{\mathsf{y}}^r(s) - \mathcal{S}_\epsilon(s)\tilde{\mathcal{G}}(s)\hat{\mathsf{p}}(s) \tag{2.13}$$

with $\mathcal{S}_\epsilon(s) := (1 + \mathcal{G}(s)\mathcal{K}(s))^{-1}$ the output sensitivity function. The closed-loop transfer matrix $\mathcal{M}(s) := -\mathcal{S}_\epsilon(s)\tilde{\mathcal{G}}(s)$ governs the disturbance rejection performance. A *semi-decentralized* controller design, which can be implemented in a PLC for regulatory control of a multi-pool canal system, is presented in the Section 2.A of the Appendix.

### 2.3.3 Supervisory Control (FDI)

Detection and isolation of faults is an important function for canal operations and is usually carried out by the supervisory control layer. Without proper diagnosis of sensor and actuator faults due to random events (and component malfunctions carried by a malicious attacker), the supervisory control functions such as set-point control and control reconfiguration might loose their effectiveness. Thus, correct FDI is also a pre-requisite for achieving an efficient operation of the closed-loop system and for performing reconfiguration and maintenance tasks. In the following, we describe a model-based scheme for detection and isolation of unknown withdrawals from the canal offtakes. We note that the choice of the FDI scheme presented here is based on its conceptual elegance; other FDI schemes for canal SCADA systems essentially share the same features Bedjaoui and Weyer [2011]. In contrast to the decentralized regulatory control scheme described in the previous section, the FDI scheme is *centralized*, i.e., it requires all the measured sensor signals and control commands to be assembled at the base station.

Let us consider faults $\mathsf{f}_i(t) := \delta\mathsf{p}_i(t), \quad i = 1, 2$, which represent the unmeasured or unscheduled water withdrawals occurring *non simultaneously* through the offtakes (located at the downstream end of the respective canal pools). Extending model (2.9) to include such faults, we obtain the fault model:

$$\begin{aligned} \dot{\mathsf{x}}(t) &= \sum_{i=0}^{2} A_i \mathsf{x}(t - \tau_i) + \sum_{i=0}^{2} B_i \mathsf{u}(t - \tau_i) + \sum_{j=1}^{2} E_j \mathsf{f}_i(t) \\ \mathsf{y}(t) &= C\mathsf{x}(t), \end{aligned} \tag{2.14}$$

with $A_i, B_i, C$ given by (2.10), and

$$E_1 = \begin{pmatrix} -a_1^d \\ 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 \\ -a_2^d \end{pmatrix}. \tag{2.15}$$

*Example* 2.3.1. Consider a system (2.14) of two identical pools with parameters: downstream propagation delays $\tau_i = \underline{\tau}_i = 647$ s, inverse equivalent backwater areas $a_i^d = (3.21)^{-1} \times 10^5$ m$^{-2}$, and coefficients of linearized gate equations $b_i^d = 29.05$, $k_i = 18.11$, $i = 1, 2$ (methods for estimating these parameters are discussed in Litrico and Fromion [2009*b*]). Assume that $\mathsf{u}(t) = 0$ for $t \in [-\underline{\tau}_1, \infty)$ and $\mathsf{x}(t) = 0$ for $t \in [-\underline{\tau}_1, 0]$. Water at rate 0.1 m$^3$/s is withdrawn from offtake of pool 1 (resp. pool 2) during the interval $2.5 - 5.0$ hr (resp. $15 - 17.5$ hr). Fig. 2.3 shows the upstream and downstream water level deviations (cm) under the effect of unmeasured withdrawals during a 24 hr simulation.

In order to detect and isolate faults $\mathsf{f}_j$, $(i = 1, 2)$, we now describe a model-based diagnostic scheme. The scheme consists of generating a bank of 2 observers, which are designed as follows: The observer 1 (resp. observer 2) is designed to be insensitive to $\mathsf{f}_1(t)$ (resp. $\mathsf{f}_2(t)$). The residual $\mathsf{r}_j$ of the $j-$th observer is defined as follows

$$\mathsf{r}_j(t) := \mathsf{y}_j(t) - C\hat{\mathsf{x}}_j(t), \tag{2.16}$$

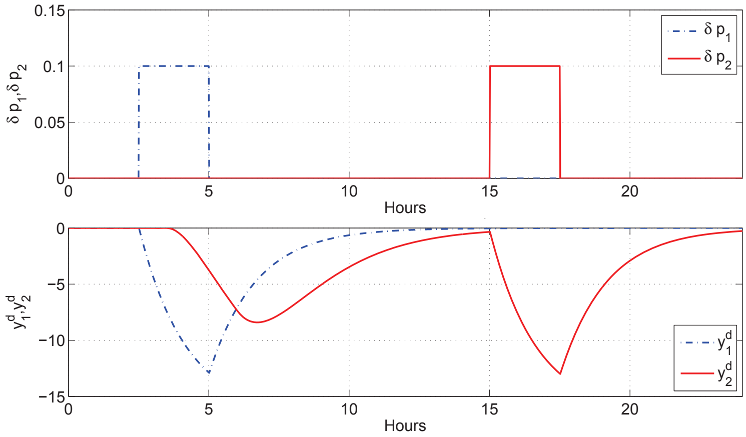


Figure 2.3: Example 2-pool system: Withdrawals (top), Downstream levels (bottom).

where $\hat{\mathsf{x}}_j(t)$ is the $j-$th observer's output denoting the state of the following fault model:

$$\begin{aligned}\dot{\mathsf{x}}_j(t) &= \sum_{i=0}^{2} A_i \mathsf{x}_j(t-\tau_i) + \sum_{i=0}^{2} B_i \mathsf{u}_j(t-\tau_i) + E_j \mathsf{f}_j(t) + E_{-j}\mathsf{f}_{-j}(t) \\ \mathsf{y}_j(t) &= C\mathsf{x}_j(t).\end{aligned} \tag{2.17}$$

Here the matrices $A_i, B_i$ $i = 0, 1, 2$ and $C$ are given by (2.10), and vectors $E_j, E_{-j}$ are given by (2.15) with $-j := (3-j)$, $j = 1, 2$.

The following (full-order) model:

$$\begin{aligned}\dot{\mathsf{z}}_j(t) &= \sum_{i=0}^{2} F_{ij} \mathsf{z}_j(t-\tau_i) + \sum_{i=0}^{2} T_j B_i u_j(t-\tau_i) + \sum_{i=0}^{2} G_{ij} \mathsf{y}_j(t-\tau_i) \\ \hat{\mathsf{x}}_j(t) &= \mathsf{z}_j(t) + N_j \mathsf{y}_j(t),\end{aligned} \tag{2.18}$$

with initial state $\mathsf{z}_j(\theta) = \rho(\theta)$, $\forall\theta \in [-\tau_{\max}, 0]$, describes the dynamics of the $j-$th observer for the fault model (2.17), and $F_{ij}$, $G_{ij}$, $i = 0, 1, 2$, $T_j$, and $N_j$ are unknown parameter matrices with real-valued elements. The design of observers is based on the following proposition:

**Proposition 2.3.2.** *If the parameter matrices $F_{ij}$, $G_{ij}$, $i = 0, 1, 2$, $T_j$, and $N_j$ in the $j-$th observer (2.18), are such that the residuals $\mathsf{r}_j(t) = (\mathsf{y}_j(t) - C\hat{\mathsf{x}}_j(t))$, $j = 1, 2$ satisfy the following properties:*

*1) $\mathsf{r}_j(t)$ is insensitive to $\mathsf{f}_j(t)$,*

*2) $\mathsf{r}_j(t)$ asymptotically converges to zero if $\mathsf{f}_{-j}(t) = 0$ for every $t$,*

*3) $\|\mathsf{r}_j(t)\| \neq 0$ when $\mathsf{f}_{-j}(t) \neq 0$,*

*then the diagnosis of faults can be achieved using the decision rule presented in Table 2.2.*

Table 2.2: Decision table for 2−pool system under offtake withdrawals.

| If | $\|\mathsf{r}_1\|$ | $\|\mathsf{r}_2\|$ |
|---|---|---|
| $\mathsf{f}_1 \neq 0$ | $\approx 0$ | $\neq 0$ |
| $\mathsf{f}_2 \neq 0$ | $\neq 0$ | $\approx 0$ |

In order to achieve this observer design objective, we define the state estimation error $\mathsf{e}_j(t)$ as:

$$\mathsf{e}_j(t) := \mathsf{x}_j(t) - \hat{\mathsf{x}}_j(t),$$

and observe from (2.16)–(2.18) that the residual $\mathsf{r}_j(t)$ can be written as output of the error dynamic:

$$\begin{cases} \dot{\mathsf{e}}_j(t) & = \sum_{i=0}^{2} F_{ij}\mathsf{e}_j(t-\tau_i(t)) + T_jE_j\mathsf{f}_j(t) + T_{-j}E_{-j}\mathsf{f}_{-j}(t) \\ & \quad - \sum_{i=0}^{2}\left(F_{ij} + \bar{G}_{ij}C - T_jA_i\right)\mathsf{x}_j(t-\tau_i(t)) \\ \mathsf{r}_j(t) & = C\mathsf{e}_j(t). \end{cases} \tag{2.19}$$

where we define

$$\bar{G}_{ij} := (G_{ij} - F_{ij}N_j), \quad i = 0,1,2 \tag{2.20}$$

$$T_j := (\mathsf{I}_n - N_jC). \tag{2.21}$$

where $\mathsf{I}_n$ denotes the $n$ (here 2) dimensional identity matrix. Consider the following conditions:

$$F_{ij} = T_jA_i - \bar{G}_{ij}C, \quad i = 0,1,2 \tag{2.22}$$

$$T_jE_j = 0, \tag{2.23}$$

$$\dot{\mathsf{e}}_j(t) = \sum_{i=0}^{2} F_{ij}\mathsf{e}_j(t-\tau_i(t)) \text{ is asymptotically stable.} \tag{2.24}$$

Let (2.20)–(2.24) hold, and note from (2.15) that $E_1$ and $E_2$ are linearly independent. Then it can be concluded that $T_jE_{-j} \neq 0,\ j = 1,2$. Thus, the residuals $\mathsf{r}_1$ and $\mathsf{r}_2$ satisfy the conditions of Proposition 2.3.2. The computation of observer parameter matrices is presented in the Section 2.B of the Appendix.

*Example* 2.3.3. Consider the fault model (2.17) for the 2−pool system with parameters as in Example 2.3.1, zero known input signal $\mathsf{u}(t) = 0$, and unknown withdrawals (faults) from pool 1 (resp. pool 2) during the interval $2.5 - 5.0$ hr (resp. $15 - 17.5$ hr) be the fault signal $\mathsf{f}_1(t)$ (resp. $\mathsf{f}_2(t)$). The LMI conditions presented in Proposition 2.B.1 (Section 2.B of the Appendix) are feasible for $\epsilon_1 = 10, \epsilon_2 = \epsilon_3 = \epsilon_5 = \epsilon_6 = -1, \epsilon_4 = -10,\ \bar{\epsilon}_1 = -1, \bar{\epsilon}_2 = -1$, and the parameter matrices of observers (2.18) are obtained according to the procedure outlined

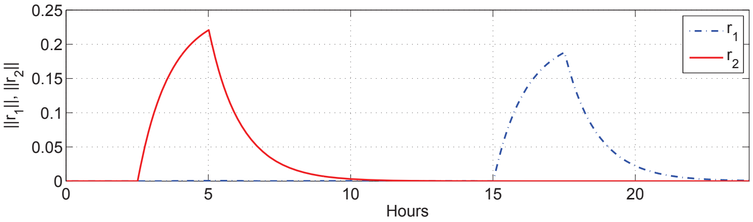


Figure 2.4: Norms of residuals $\mathsf{r}_1$ and $\mathsf{r}_2$ [simulated results].

above. From the computed observer matrices $T_1$ and $T_2$ we obtain:

$$T_1E_1 = 10^{-15}\begin{pmatrix}-0.341\\-0.0622\end{pmatrix} \approx 0, \quad T_1E_2 = \begin{pmatrix}-0.002\\-0.554\end{pmatrix} \neq 0$$

$$T_2E_1 = \begin{pmatrix}-0.651\\0.0\end{pmatrix} \neq 0, \quad T_1E_2 = 10^{-16}\begin{pmatrix}-0.002\\-0.548\end{pmatrix} \approx 0.$$

From Fig. 2.4 we observe that the generated residuals $\mathsf{r}_j(t)$ $j = 1, 2$ in Example 2.3.3 satisfy the condition of Proposition 2.3.2:

- $\mathsf{r}_1(t)$ (resp. $\mathsf{r}_2(t)$) is insensitive to $\mathsf{f}_1(t)$ ($\mathsf{f}_2(t)$),
- The residual $\mathsf{r}_j(t)$ defined by (2.19), is asymptotically zero when $\mathsf{f}_{-j}(t) = 0$ for every $t$ (note that $T_1E_1 = T_2E_2 \approx 0$),
- $\|\mathsf{r}_j(t)\| \neq 0$ when $\mathsf{f}_{-j}(t) \neq 0$ since $T_jE_{-j} \neq 0,\ j = 1, 2$.

Hence, the diagnosis of faults can be achieved using the decision rule presented in Table 2.2.

*Remark* 2.3.4. Notice that the error dynamics (2.19) and hence the observer residuals do not depend on the known control input $\mathsf{u}$ and so, the behavior of FDI scheme does not change when $\mathsf{u}$ is manipulated by the regulatory control layer.

### 2.3.4 Simulation of a Stealthy Attack

Let us consider the attack model (2.1), where the attack duration $\mathcal{T}$ and attack signal $\mathsf{g}_i(t)$ are chosen by an attacker. We assume that a PI-based regulatory controller and an observer based FDI scheme has been deployed after proper tuning, and that the attacker has full knowledge of the regulatory control as well as the FDI schemes. Equivalently, it amounts to assuming that the attacker has the knowledge of 1) the approximate system dynamics, 2) the parameters of FDI scheme, and 3) the sensor-control signals. Indeed, such a powerful attacker may be unrealistic for many SCADA systems with some IT security in place; however by adopting this conservative approach, we can better test the resiliency of the regulatory control and FDI schemes. Moreover, this assumption also covers the case of an adaptive attacker who will attempt to evade detection by the SCADA system.

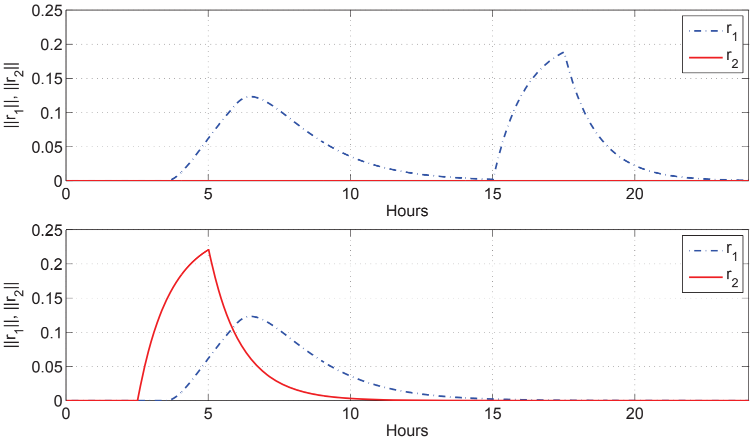


Figure 2.5: Residuals under attack on $\mathsf{y}_1^d$ (top) and $\mathsf{y}_2^d$ (bottom) [simulated results].

The attacker's intent is to steal water from the canal system by attacking the downstream level sensor measurements $\mathsf{y}^d$. Our goal is then to synthesize an attack strategy such that the compromise is not detected by the regulatory and supervisory control layers. We call such an attack, a *stealthy attack*. By detection at regulatory control level, we mean that the regulatory controllers should react as expected to reduce deviations with respect to set-point targets. Let us recall that performance of regulatory control (resp. supervisory FDI scheme) essentially depends on the output error $\epsilon$ (resp. observer residual $\mathsf{r}$) as defined in (2.11) defined in Sec. 2.3.2 (resp. (2.16) defined in Sec. 2.3.3); the stealthy attack $\mathsf{g}$ should then be aimed at manipulating these quantities to avoid detection.

In the following example, we propose a stealthy attack strategy which evades detection by the supervisory FDI scheme. A stealthy attack which evades detection by the regulatory controller is proposed in Section 2.4.3.

*Example* 2.3.5. Consider the FDI scheme in Example 2.3.3 which achieve FDI for non-simultaneous withdrawals for 2−pool system. We consider two cases when true sensor measurements are spoofed with a deception attack: $\mathsf{y}_1^d$ (resp. $\mathsf{y}_2^d$) is compromised, and spoofed with the signal $\mathsf{g}_1(t) = 0$ (resp. $\mathsf{g}_2(t) = 0$) for $t \in [0, 24]$; see top (resp. bottom) of Fig. 2.5.

As shown in Fig. 2.5 (top), when $\mathsf{y}_1^d$ is compromised, the fault $\mathsf{f}_2$ is correctly diagnosed according to Table 2.2; however, $\mathsf{f}_1$ can no longer be diagnosed correctly since $\mathsf{r}_1$ (resp. $\mathsf{r}_2$) which was only sensitive to $\mathsf{f}_2$ (resp. $\mathsf{f}_1$) in the case of no attack, is now sensitive (resp. insensitive) to $\mathsf{f}_1$. More interestingly, when $\mathsf{y}_2^d$ is compromised, both $\mathsf{r}_1$ and $\mathsf{r}_2$ are sensitive to $\mathsf{f}_1$ and could lead to mis-detection, but $\mathsf{f}_2$ goes completely undetected since neither residual is sensitive to $\mathsf{f}_2$. This is due to the fact that, in the framework presented in this chapter, the effect of water withdrawal in pool 2 does not propagate upstream due to free-flow condition of the gates. Thus, under the proposed FDI scheme, compromising $\mathsf{y}_2^d$ fully achieves the objective of a stealthy attacker, i.e., the water can be stolen via the second canal pool's offtake without being detected by the FDI scheme. The losses due to water withdrawals can be assessed by estimating the amount of water withdrawn from the canal system before the attack is detected, which can be considerable in the case of stealthy attacks.

*Remark* 2.3.6. (Stealthy attack for multiple pools) The above example also hints toward a stealthy attack strategy for the case of multiple canal pools when all the level sensor measurements $\mathsf{y}_i^d$, $i = 1, \dots, m$ are vulnerable to compromise by the attacker. In this case, the stealthy attack strategy is to first compromise the most downstream sensor measurement $\mathsf{y}_m^d$ and systematically proceed to compromise measurements of upstream canal pools $\mathsf{y}_{m-1}^d, \mathsf{y}_{m-2}^d, \dots, \mathsf{y}_1^d$.

In the next section, we present results from a field experiment in which deception attacks were implemented on the Gignac SCADA system.

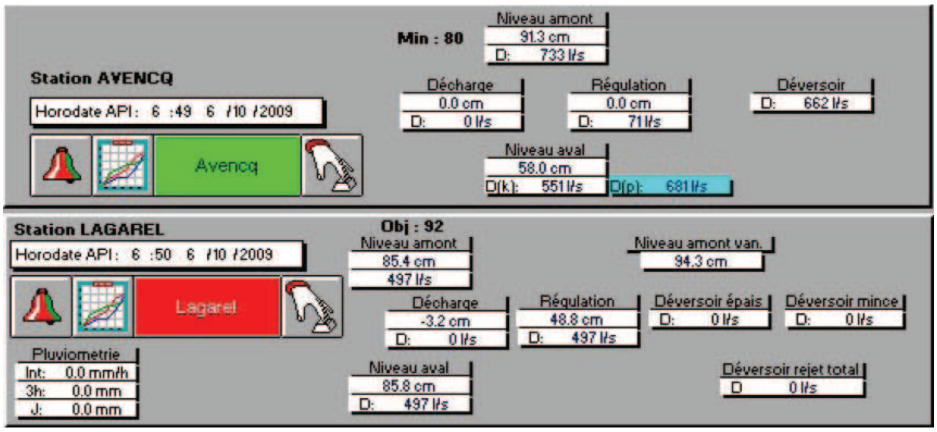
## 2.4 Field Operational Test Attacks

### 2.4.1 Gignac canal network and SCADA system

We now discuss the main components of the Gignac canal network and the SCADA system which manages this network. The Gignac canal network is located about 40 km north-west of Montpellier in South France, and irrigates about 3000 ha of agricultural land. The main canal network is comprised of a 8 km feeder canal which emerges from the Hérault river, and bifurcates at the diversion structure (or Partiteur in French) into two branches, to the left and right banks of the river. These branches are of lengths 27 km and 15 km respectively. The design flow of the canal is about 3.5 $\mathrm{m}^3$/s. The canal is equipped with sensors at different sites to collect water levels, gate openings, and discharge data, as well as automated structures to control water flow.

The components of the Gignac SCADA system include a centralized control station with several host computers which communicate with remote local processors and field devices (or PLCs) operating the sensors and actuators. These components communicate over standard channels (including the Internet and public-switched telephone networks). The time-step of data acquisition from field devices can be chosen as low as a few seconds. The SCADA system is capable of implementing a variety of control-loops for regulatory control, and also enable supervisory capabilities such as human-machine interfaces (HMIs), remote diagnostics and maintenance. For a snapshot of supervisory interface at the Avencq and Lagarel cross-regulators, see Fig. 2.6. The functionalities of the SCADA system include: 1) Real-time hydraulic state monitoring, data-logging, alarming and diagnostic functions for handling faults, 2) Activating local slave controllers and sending gate position targets in real-time, 3) Changing parameters of local slave controllers, and 4) Modifying operational objectives by specifying desired discharges or water levels. Over the past decade, researchers have developed a suite of automatic control methods for canal management ranging from simple PI based regulatory controllers to more advanced methods based on $\mathcal{H}_\infty$, $\ell_1$, and model predictive control Litrico and Fromion [2009*b*]. Implementation of FDI schemes has also been investigated Bedjaoui and Weyer [2011].

The Gignac SCADA system periodically suffers attacks on SCADA system components, which we now summarize. Information regarding these incidents was obtained from the news reported in the French media and through our personal interaction with the canal manager. First, the solar panels that power radio communication systems used for data



Station AVENCQ

Horodate API : 6 :49 6 /10 /2009

Min : 80

Niveau amont: 91.3 cm, D: 733 l/s

Décharge: 0.0 cm, D: 0 l/s

Régulation: 0.0 cm, D: 71 l/s

Déversoir: D: 662 l/s

Niveau aval: 58.0 cm, D(k): 551 l/s, D(p): 681 l/s

Station LAGAREL

Horodate API : 6 :50 6 /10 /2009

Obj : 92

Niveau amont: 85.4 cm, 497 l/s

Niveau amont van.: 94.3 cm

Décharge: -3.2 cm, D: 0 l/s

Régulation: 48.8 cm, D: 497 l/s

Déversoir épais: D: 0 l/s

Déversoir mince: D: 0 l/s

Pluviometrie: Int: 0.0 mm/h, 3h: 0.0 mm, J: 0.0 mm

Niveau aval: 85.8 cm, D: 497 l/s

Déversoir rejet total: D 0 l/s

Figure 2.6: Gignac SCADA supervisory interface.

transmission from sensors to the base station were stolen. This resulted in loss of sensor data and hindered canal operations for days. In a second attack, social miscreants damaged the monitoring bridge on which a local gate controller was supported soon after it was repaired. Finally, farmers who use the canal water for irrigation have made repeated attempts to steal water from the canal by tampering water offtakes and installing additional pumps to withdraw water. Such threats remain a challenge for the management agency. Although these incidents were mainly physical, they directly affected the functioning of cyber components of the SCADA system.

### 2.4.2 Field Operational Test Setup

We now demonstrate the feasibility of a cyber-attack on water SCADA systems by conducting a field operational test on the Gignac canal and assessing the losses due to undetected water withdrawals. In our experiment, we consider a two pool system situated on the canal branch which diverts from the Partiteur device to the right bank of the Hérault river; see Fig. 2.7.

The first pool is the 4.8 km canal reach between the Partiteur device and the Avencq cross-regulator, and the second pool is the 5.2 km reach between the Avencq and Lagarel cross-regulators. The ID model parameters for the respective pools are given by: $a_1^d = 1.105 \times 10^{-4}$ m$^{-2}$, $a_2^d = 2.597 \times 10^{-5}$ m$^{-2}$, $\tau_1 = 45$ min, $\tau_2 = 40$ min. During our experiment, the Lagarel gate was submerged and therefore, the linearized gate equation (2.8) for free-flow gate cannot be used. The linearized gate equation for submerged gate is given by:

$$\mathsf{q}_i(t) = b_i^d \mathsf{y}_i^d(t) + b_{i+1}^u \mathsf{y}_{i+1}^u(t) + k_i \mathsf{u}_i(t), \quad i = 1, 2 \tag{2.25}$$

$b_{i+1}^u$ is the gain from the water level $\mathsf{y}_{i+1}^u$ downstream of the gate. Both Avencq and Lagarel regulators are equipped with motorized gates and level sensors, and communicate with the base station via radio communication. The discharge required to regulate the upstream water level in response to perturbations caused due to offtake withdrawals is achieved by

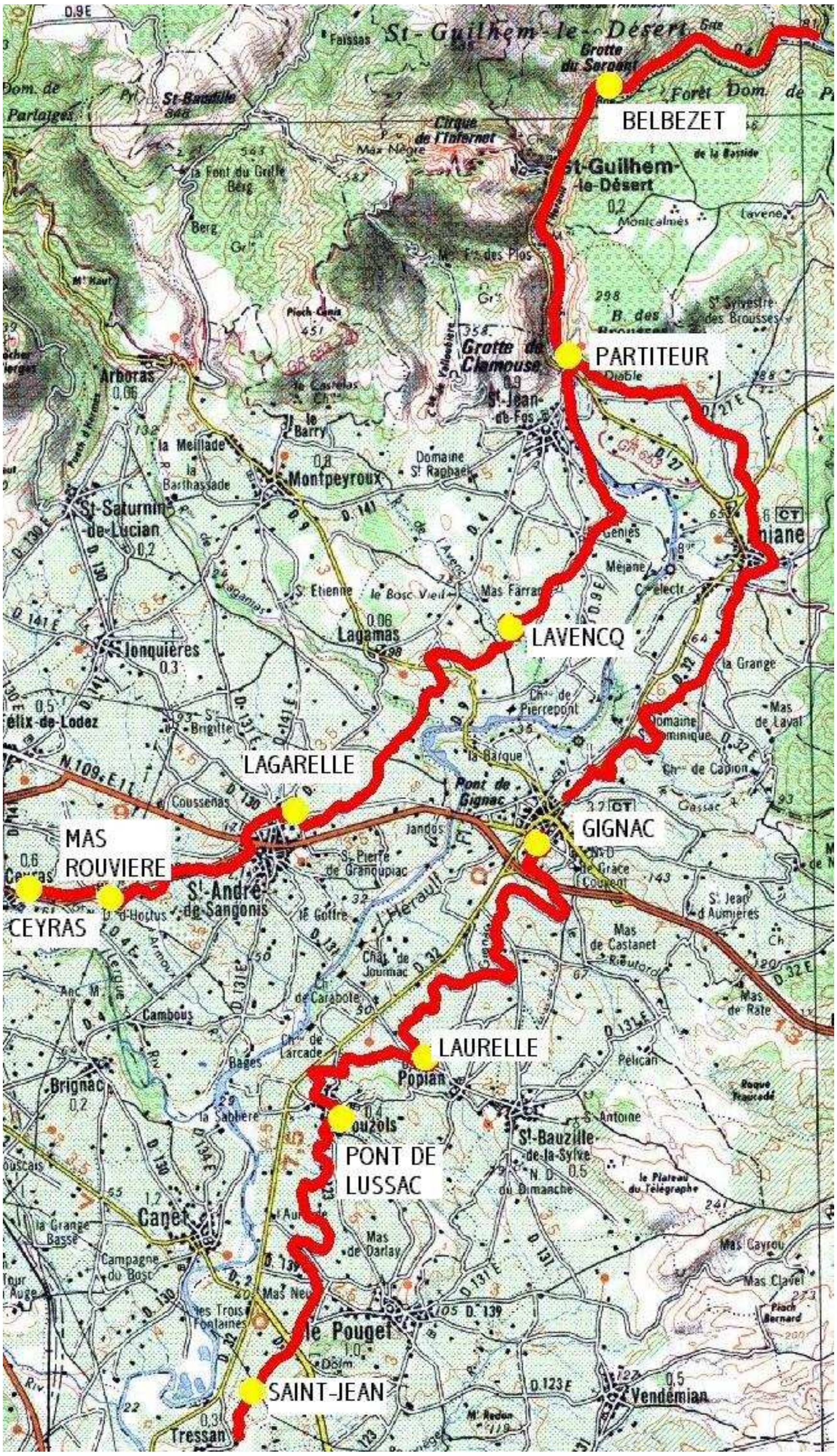


Figure 2.7: Map of the Gignac canal system.

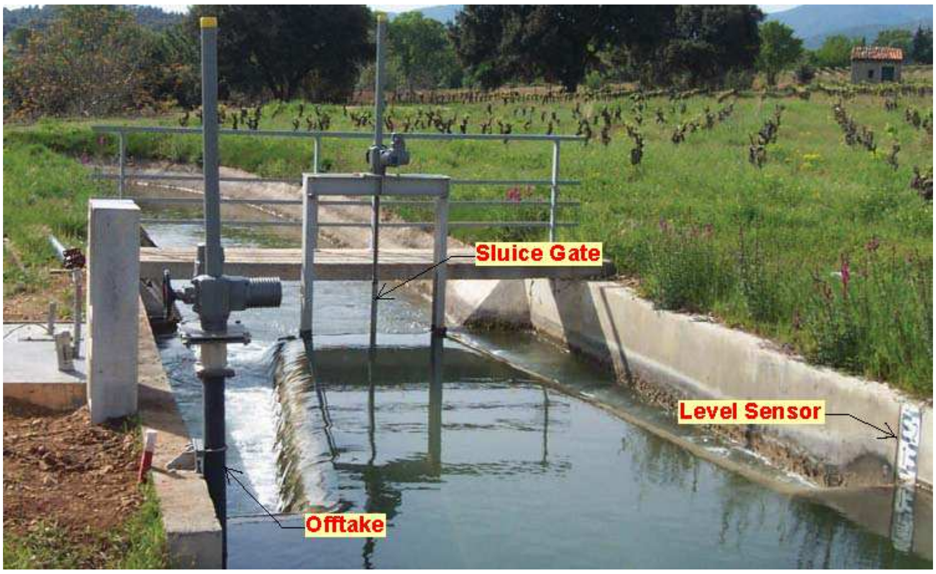


Figure 2.8: Upstream of the Avencq station.

the slave controller (PLC) via movement of a 1 m wide sluice gate; see Fig. 2.8. We now implement a deception attack on the Avencq regulator; this attack corresponds to attack **A**1 in Fig. 2.2 and spoofing attack in Table 2.1. Under our attack model, the attacker's intention is to steal water from an offtake located upstream of Avencq gate, and he/she tries to achieve this goal by compromising the level sensor measurements at Avencq.

### 2.4.3 Effect of cyber-attack on regulatory control

We assess the performance loss of regulatory control under compromise of $\mathsf{y}^d$ at the Avencq gate first in simulation, and then in a field operational test on the Gignac canal. In our setting, the steady state water level is $\bar{\mathsf{Y}} = 79$ cm. The regulatory control aim is to stabilize $\mathsf{y}^d$ (which is the deviation from $\bar{\mathsf{Y}}$) to 0, i.e., a set-point $\mathsf{y}^r = 0$ cm. The upstream water level is measured every 2 min, and a PI controller

$$\kappa(s) = k\left(1 + \frac{1}{Ts}\right),$$

with the proportional gain $k = -2.9$ and the integral time $T = 360$ s is used to regulate $\mathsf{y}^d$ at Avencq gate. Now, if the attack signal $\mathsf{g}(t)$ in the attack model (2.1) is chosen such that error under attack $\tilde{\epsilon} = (\tilde{\mathsf{y}}^d - \mathsf{y}^r)$ is close to zero, then from (2.11) it follows that the regulatory controller will not react correctly to reject water level deviations from the set-point $\mathsf{y}^r$. Thus, $\mathsf{g}(t) \approx \mathsf{y}^r(t)$ achieves a stealthy attack for regulatory control layer.

We now describe a stealthy deception attack scenario on Avencq gate using the SIC software as a simulator. The SIC software developed by CEMAGREF provides us with following capabilities: 1) performance testing of any regulatory control method on a fully nonlinear shallow water equation simulator, and 2) a direct implementation of the tested controller on the physical canal via a software interface to the SCADA system Malaterre and Chateau

[2007]. In the attack scenario shown in Fig. 2.9, the offtake is opened to about 3 cm at time $t = 15$ min after the beginning of the test. The PI controller reacts rapidly by closing the sluice-gate and rejects the perturbation in about 40 min. At $t = 75$ min, the offtake is closed. The controller achieves good closed-loop performance and rejects the perturbation in about 45 min by opening the sluice gate as shown in Fig. 2.9 (bottom right). The offtake is again opened and closed at $t = 255$ min and $t = 315$ min respectively, this time under the influence of attacker's action; see Fig. 2.9 (top left). The attacker compromises $\mathsf{y}^d$ and injects a deception attack $\mathsf{g} \approx \mathsf{y}^r$; see Fig. 2.9 (bottom left). Therefore, the PI control does not react to the opening of the offtake. The effect of this attack on the performance of local upstream controller is shown in Fig. 2.9 (top right).

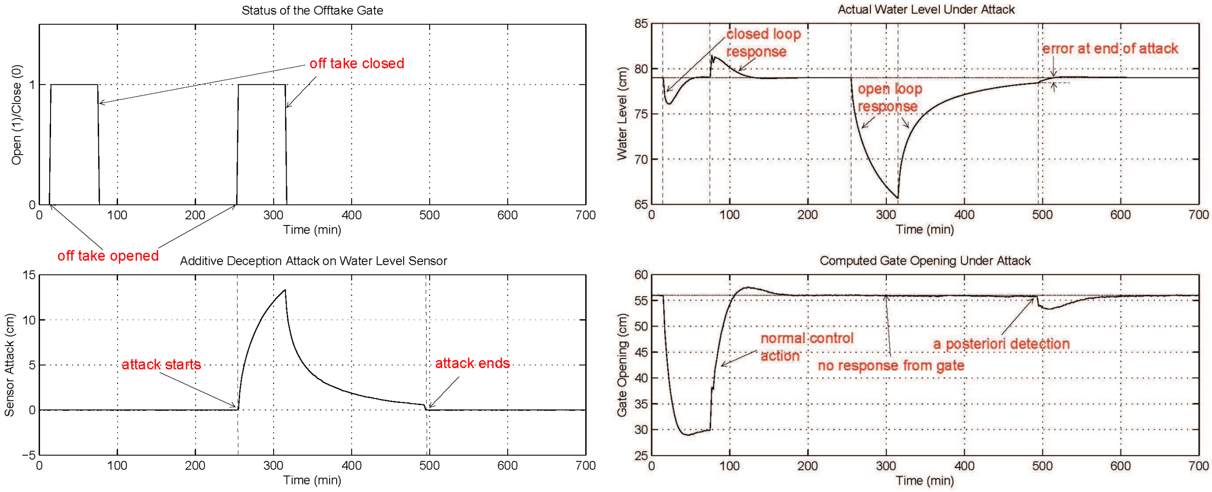


Figure 2.9: Performance of PI controller under attack [simulation results].

The duration of attack is determined as follows: even after the closing of the offtake at $t = 315$ min, the attacker continues the deception attack until $t = 495$ min when the water level – evolving in open-loop – comes close to the set point $\mathsf{y}^r = 0$ cm. At $t = 495$ min the attacker stops the deception attack and PI controller reacts to the residual error. This may signal an a posteriori detection; however, it may be still difficult to distinguish between a residual error resulting from an attack from an error resulting from small (random) perturbations in $\mathsf{y}^d$. The amount of water the attacker manages to withdraw from the offtake between $t = 255$ min and $t = 315$ min can be computed by integrating the gate discharge equation

$$\mathsf{Q}(t) = C_g L_g \mathsf{U}\sqrt{2g\left(\mathsf{y}^d(t) + \bar{\mathsf{Y}}\right)}$$

where $C_g \approx 0.6$ denotes the discharge coefficient, $L_g = 1$ m the gate width, $\mathsf{U} = 0.03$ m the offtake opening, $(\mathsf{y}^d(t) + \bar{\mathsf{Y}})$ the actual water level.

We now demonstrate the feasibility of deception attacks with a field operational test on Avencq. The experiment was performed on October $12^{th}$, 2009 during which we carried out the attack directly by modifying the sensor measurements sent from the real-time SCADA interface of the SIC software to the Matlab code which implemented the PI controller.

Although we played the attacker's role in this experiment, the resulting effect is same as that of a deception attack on the Avencq water level sensor (attack **A**1 in Fig. 2.2). At the start of experiment, the PI controller reacts by changing set-points every few minutes and then letting the water level stabilize close to set-point in closed-loop. As shown in Fig. 2.10, at $t = 90$ min, the offtake is opened and the attacker injects false data to water level measurement such that the PI controller fails to react to perturbation.

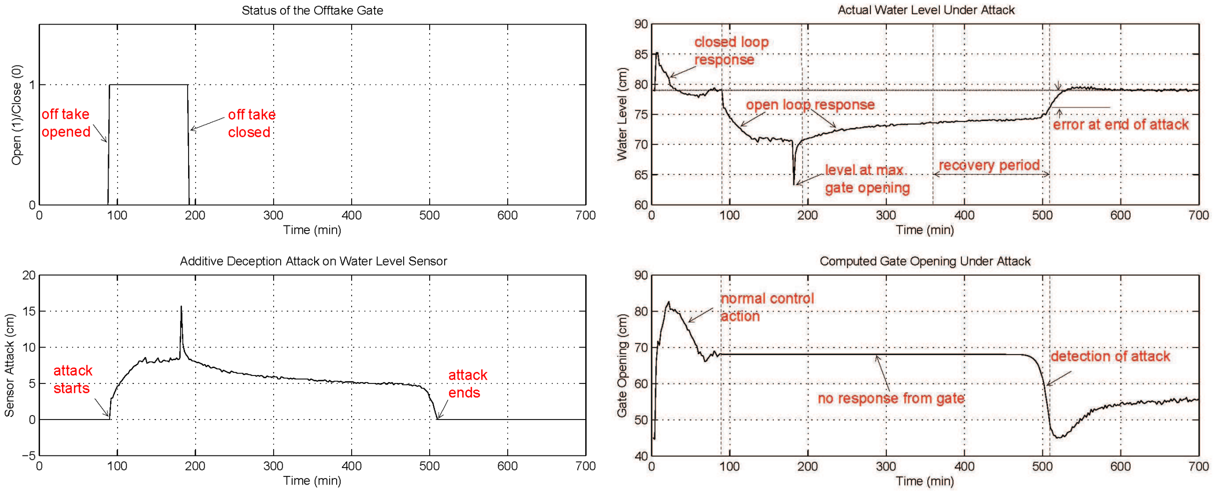


Figure 2.10: Performance of PI controller under attack [field operational test].

At around $t = 184$ min the offtake was fully opened and then fully closed at around $t = 190$ min by a physical intervention at the Avencq cross-regulator; see Fig. 2.10 (top left). This effect is captured in the sudden drop in the actual water level as shown in Fig. 2.10 (top right). From $t = 190$ min until $t = 510$ min, the attacker continues the attack; see Fig. 2.10 (bottom left). This results in open-loop response of actual water level. However, a residual error still remains after the end of the attack, and the PI controller reacts to this error as seen in Fig. 2.10 (bottom right). It can be concluded that the response of the PI control after the attack ends and the response to random perturbations can be difficult to distinguish at the regulatory control level.

### 2.4.4 Effect of cyber-attack on supervisory FDI scheme

In order to assess the effect of attacks on supervisory FDI scheme proposed in Section 2.3.3, we collected the 15 min archived data from the Gignac SCADA system. The data includes the upstream and downstream water levels, gate openings, and discharges. We used the FDI scheme based on the observer design assuming that the withdrawal through offtake from pool 1 during $90 - 190$ min is an unknown withdrawal. The LMI conditions in Proposition 2.B.1 are found to be feasible, and the parameter matrices of observers (2.18) are obtained according to the procedure in Section 2.3.3. From the computed observer matrices $T_1$ and $T_2$ we obtain:

$$T_1E_1 = 10^{-15}\begin{pmatrix}0.247\\-112\end{pmatrix} \approx 0, \quad T_1E_2 = \begin{pmatrix}-0.001\\-0.7595\end{pmatrix} \neq 0$$
$$T_2E_1 = \begin{pmatrix}-0.729\\0.0\end{pmatrix} \neq 0, \quad T_1E_2 = 10^{-15}\begin{pmatrix}0.00\\-0.125\end{pmatrix} \approx 0.$$

Similar to example 2.3.3, we can check that the residuals satisfy the conditions of Proposition 2.3.2, and thus, the fault diagnosis can be achieved by Table 2.2. Indeed, Fig. 2.11 (top left) shows that under no attack on sensor measurements, the residual of observer 2 is sensitive to fault occurring in the form of lateral withdrawal in pool 1. However, when Avencq's sensor measurements are compromised and false data $\mathbf{g} = 0$ is injected, the observer residuals no longer indicate a correct diagnosis as shown in Fig 2.11 (bottom left). The actual upstream water level at the Avencq regulator and the computed gate opening are shown in Fig 2.11 (top right) and (bottom right), respectively. Recall that FDI scheme uses 15 min archived data from the SCADA system. Hence, the spike in actual water level, that is clearly visible with a 2 min sampling period (see Fig. 2.10 (top right)), is not visible here and is not reflected in the residual computed from the observers.

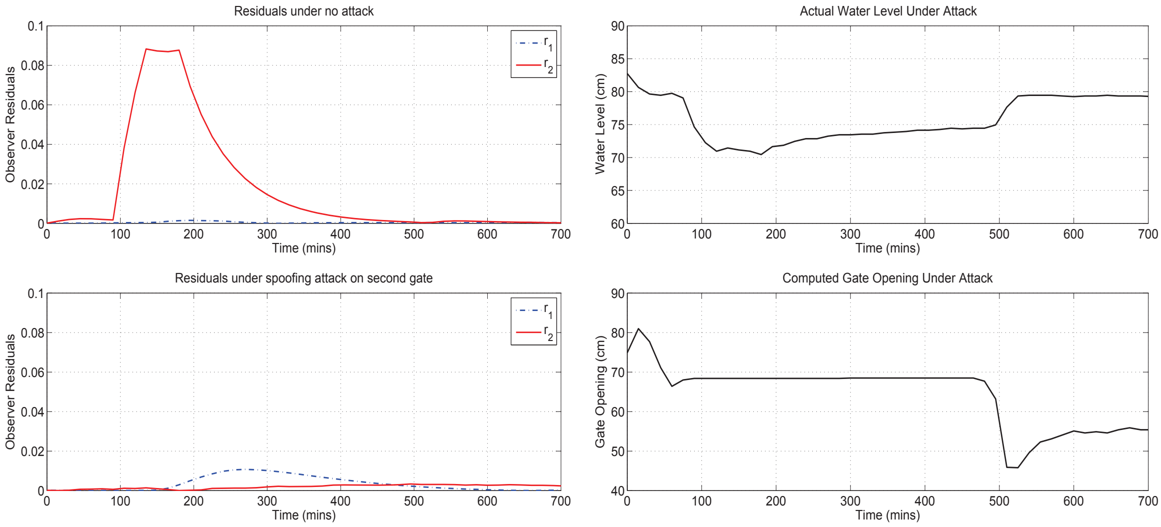


Figure 2.11: Performance of FDI scheme under attack [field operational test].

## 2.5 Discussion

In this chapter, we present a taxonomy of deception and denial-of-service (DoS) attacks on hierarchical water SCADA systems. To demonstrate the effect of cyber attacks on an actual SCADA system, we discuss results from a field operational test conducted on the Gignac water SCADA system. This test illustrates the effect of deception attacks on the performance of a PI control based regulation method and a model-based supervisory FDI scheme based on a low-frequency approximation. Our results indicate that it is possible for the attacker to stealthily withdraw water from the canal pool without getting detected.

Our synthesis of deception attacks can be extended to the case of multiple canal pools. This could be done by approximating the effect of water withdrawal on the downstream level sensor readings and subsequently manipulating the sensors such that regulatory and supervisory controllers react to wrong level deviations. An interesting research question is then to characterize the relation between the resources required by the attacker to manipulate multiple sensors versus the impact of the resulting attack in terms of water loss and operational inefficiencies.

Such analyses have many practical and theoretical implications. In particular, they provide a framework to assess the robustness of detection and regulation methods under cyber attacks. From a computer security viewpoint, they also provide novel insights on securing SCADA systems. Finally, we note that the cyber security of SCADA systems managing other infrastructures (e.g., oil and natural gas distribution networks) can be studied in a similar manner.

# Appendix 2.A Regulatory Control Design

Three classical controller designs are most common for canal regulation: local upstream (denoted lu) control, distant downstream (denoted dd) control, and mixed control. We first discuss controller designs for a single canal pool and then illustrate the extension to multiple canal pools. For a detailed analysis of stability and performance guarantees of these controllers, the reader is referred to Chapters 7 and 8 in Litrico and Fromion [2009*b*].

## 2.A.1 Regulatory control of single canal pool

Local upstream control (denoted lu) of a canal pool consists of controlling the downstream water level $\mathsf{y}_i^d$ using the downstream discharge $\mathsf{q}_i$ as control action variable. Distant downstream control (denoted dd) consists of controlling $\mathsf{y}_i^d$ using the upstream discharge $\mathsf{q}_{i-1}$ as control action variable. Let the transfer functions of the dd controller and the lu controller be defined as $\kappa_{i-1i}(s)$ and $\kappa_{ii}(s)$ respectively. Thus, we have $\hat{\mathsf{q}}_{i-1}(s) = \kappa_{i-1i}(s)\hat{\epsilon}_i(s)$, $\hat{\mathsf{q}}_i(s) = 0$ (resp. $\hat{\mathsf{q}}_{i-1}(s) = 0$, $\hat{\mathsf{q}}_i(s) = \kappa_{ii}(s)\hat{\epsilon}_i(s)$) for the dd (resp. lu) control. Using (2.5) and (2.13) the tracking error $\epsilon_i$ can be expressed as

$$\hat{\epsilon}_i(s) = \begin{cases} \left(1 + \frac{a_i^d}{s} e^{-\tau_i s} \kappa_{i-1i}(s)\right)^{-1} [\hat{\mathsf{y}}_i^r + \frac{a_i^d}{s}\hat{\mathsf{p}}_i(s)] & \text{dd} \\ \left(1 - \frac{a_i^d}{s}\kappa_{ii}(s)\right)^{-1} [\hat{\mathsf{y}}_i^r + \frac{a_i^d}{s}\hat{\mathsf{p}}_i(s)] & \text{lu} \end{cases} \tag{2.26}$$

The disturbance rejection is then characterized by the modulus $|\frac{a_i^d}{s}(1+\frac{a_i^d}{s}e^{-\tau_i s}\kappa_{i-1i}(s))^{-1}|$ for dd control and by the modulus $|\frac{a_i^d}{s}(1-\frac{a_i^d}{s}\kappa_{ii}(s))^{-1}|$ for lu control. The control objective is to choose the linear controllers $\kappa_{i-1i}(s)$ and $\kappa_{ii}(s)$ such that the respective moduli are close to 0 over largest frequency bandwidth. Note that while dd control has low performance due to presence of time-delay in $\frac{a_i^d}{s}e^{-\tau_i s}$ limiting the achievable frequency bandwidth; the lu control has a higher performance since there is no time-delay in $-\frac{a_i^d}{s}$ and the achievable bandwidth is only limited by actuator's limitation. On the other hand, dd control has high water efficiency because the controller regulates upstream water supply leading to parsimonious water management; however, lu control has low water efficiency because it propagates all perturbations downstream of the canal pool without managing the upstream discharge.

To address limitations the mixed control policy uses both $\mathsf{q}_{i-1}$ and $\mathsf{q}_i$ to control $\mathsf{y}_i^d$, where a (fast) lu control $\mathsf{q}_i$ is used to regulate water level $\mathsf{y}_i^d$ to set-point $\mathsf{y}_i^r$ and a (slow) dd control $\mathsf{q}_{i-1}$ is used to regulate $\mathsf{q}_i$ to set-point $\mathsf{q}_i^r$. The mixed controller structure can be specified as

$$\begin{aligned}\hat{\mathsf{q}}_{i-1}(s) &= \tilde{\kappa}_{i-1i}(s)[\hat{\mathsf{q}}_i^r(s) - \hat{\mathsf{q}}_i(s)]\\ \hat{\mathsf{q}}_i(s) &= \kappa_{ii}(s)[\hat{\mathsf{y}}_i^r(s) - \hat{\mathsf{y}}_i^d(s)],\end{aligned}$$

or equivalently,

$$\begin{pmatrix}\hat{\mathsf{q}}_{i-1}(s)\\ \hat{\mathsf{q}}_i(s)\end{pmatrix} = \begin{pmatrix}\kappa_{i-1i}(s)\\ \kappa_{ii}(s)\end{pmatrix}\hat{\epsilon}_i(s) + \begin{pmatrix}\tilde{\kappa}_{i-1i}(s)\\ 0\end{pmatrix}\hat{\mathsf{q}}_i^r(s) \tag{2.27}$$

where $\kappa_{i-1i}(s) = -\tilde{\kappa}_{i-1i}(s)\kappa_{ii}(s)$. For steady state conditions, i.e., when $\mathsf{q}_i^r = 0$, from (2.5) and (2.13) we obtain

$$\hat{\epsilon}_i(s) = \left(1 + \frac{a_i^d}{s}e^{-\tau_i s}\kappa_{i-1i}(s) - \frac{a_i^d}{s}\kappa_{ii}(s)\right)^{-1}[\hat{\mathsf{y}}_i^r + \frac{a_i^d}{s}\hat{\mathsf{p}}_i(s)].$$

Comparing this with (2.26), it is obvious that the structured mixed controller corresponds to an addition of lu and dd controllers.

### 2.A.2 Regulatory control of multiple canal pools

For simplicity but with not loss of generality, we focus on the case of two canal pools. The controller matrix $\mathcal{K}$ in (2.11) has the following structure for dd and lu controllers:

$$\mathcal{K}^{dd} = \begin{pmatrix}\kappa_{i-1i} & \kappa_{i-1i+1}\\ 0 & \kappa_{ii+1}\\ 0 & 0\end{pmatrix}, \quad \mathcal{K}^{lu} = \begin{pmatrix}0 & 0\\ \kappa_{ii} & 0\\ \kappa_{i+1i} & \kappa_{i+1i+1}\end{pmatrix} \tag{2.28}$$

with $\hat{\mathsf{q}}(s) = \begin{pmatrix}\hat{\mathsf{q}}_{i-1}(s), & \hat{\mathsf{q}}_i(s), & \hat{\mathsf{q}}_{i+1}(s)\end{pmatrix}^\intercal$, and $\hat{\epsilon}(s) = \begin{pmatrix}\hat{\epsilon}_i(s), & \hat{\epsilon}_{i+1}(s)\end{pmatrix}^\intercal$, and where $\kappa_{i-1i}$ and $\kappa_{ii+1}$ (resp. $\kappa_{ii}$ and $\kappa_{i+1i+1}$) are SISO dd (resp. lu) controllers, and $\kappa_{i-1i+1}$ (resp. $\kappa_{i+1i}$) is the decoupling term. Using (2.5) and (2.28) and for simplicity considering $\hat{\mathsf{y}}^r(s) = 0$, the tracking error (2.13) can be expressed as:

$$\hat{\epsilon}(s) = \mathcal{M}\hat{\mathsf{p}}(s),$$

where the closed-loop transfer matrix $\mathcal{M}$ has the following structure

$$\mathcal{M} = \begin{cases} \begin{pmatrix} \mu_{i-1i} & \mu_{i-1i+1} \\ 0 & \mu_{ii+1} \end{pmatrix}, & \text{dd} \\ \begin{pmatrix} \mu_{ii} & 0 \\ \mu_{i+1i} & \mu_{i+1i+1} \end{pmatrix}, & \text{lu} \end{cases} \tag{2.29}$$

with

$$\begin{aligned} \mu_{i-1i} &= \frac{a_i^d}{s}\left(1 + \frac{a_i^d}{s} e^{-\tau_i s} \kappa_{i-1i}\right)^{-1} \\ \mu_{ii+1} &= \frac{a_{i+1}^d}{s}\left(1 + \frac{a_{i+1}^d}{s} e^{-\tau_{i+1} s} \kappa_{ii+1}\right)^{-1} \\ \mu_{i-1i+1} &= \mu_{i-1i}\kappa_{ii+1}\mu_{ii+1}\left(1 - \kappa_F^{dd} e^{-\tau_i s}\right) \\ \mu_{ii} &= \frac{a_i^d}{s}\left(1 - \frac{a_i^d}{s}\kappa_{ii}\right)^{-1} \\ \mu_{i+1i+1} &= \frac{a_{i+1}^d}{s}\left(1 - \frac{a_{i+1}^d}{s}\kappa_{i+1i+1}\right)^{-1} \\ \mu_{i+1i} &= \mu_{ii}\kappa_{ii}\left(\kappa_F^{lu} - e^{-\tau_{i+1} s}\right)\mu_{i+1i+1}, \end{aligned}$$

and

$$\kappa_F^{dd} = \kappa_{i-1i+1}\kappa_{ii+1}^{-1}, \quad \kappa_F^{lu} = \kappa_{i+1i}\kappa_{ii}^{-1},$$

the feed-forward terms which govern the interactions between the pools.

From (2.29) we see that the closed-loop transfer matrix $\mathcal{M}$ is structurally upper triangular (resp. structurally lower triangular) for the dd (resp. lu) control because $\mu_{ii}$ (resp. $\mu_{ii+1}$) is null. Therefore, both dd and lu multivariable controllers (2.28) inherit the stability and performance properties of monovariable systems. In both cases, the interactions between the two pools decrease the performance of the overall closed-loop systems; these interactions can be reduced by appropriate choices of the decoupling terms $\kappa_{i-1i+1}$ and $\kappa_{i+1i}$. In contrast to dd, the lu control design allows for perfect decoupling of the canal pools; however, in practice, the presence of model uncertainties governs the achievable decoupling.

The mixed control policy for multi-pool system is multi-variable controller and can be designed to satisfy the global objective of low frequency (resp. high frequency) perturbation rejection using upstream (resp. downstream) discharge. The stability and performance of the mixed controller are related to those of individual pools. However, the closed-loop transfer matrix $\mathcal{M}$ is not structural upper-triangular in this case, and the robustness with respect to dynamic uncertainties can only be evaluated a posteriori.

In order to achieve a fault-tolerant design, the mixed control policy can be implemented in a structured semi-decentralized fashion, i.e., each local controller (PLC) communicates only with neighboring upstream and downstream controllers. Here we illustrate the design

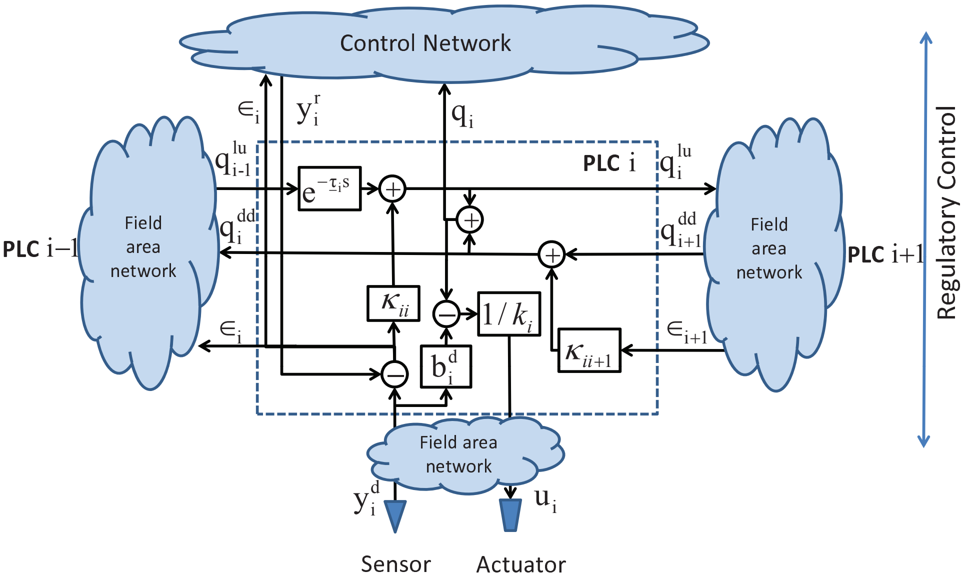


Figure 2.12: Control design for a programmable logic controller (PLC).

reported in Litrico and Fromion [2005]; see Cantoni et al. [2007] for a similar analysis where the disturbance propagation behavior is also investigated. Extending the single pool mixed controller design (2.27) to the case of multiple-pools, we decompose each control variable $\hat{\mathsf{q}}_i(s)$ into a dd control $\hat{\mathsf{q}}_i^{dd}$ and a lu control $\hat{\mathsf{q}}_i^{lu}(s)$.

$$\hat{\mathsf{q}}_i(s) = \hat{\mathsf{q}}_i^{dd}(s) + \hat{\mathsf{q}}_i^{lu}(s), \tag{2.30}$$

where

$$\hat{\mathsf{q}}_i^{dd}(s) := \kappa_{ii+1}(s)\hat{\epsilon}_{i+1}(s), \quad \hat{\mathsf{q}}_i^{lu}(s) = \kappa_{ii}(s)\hat{\epsilon}_i(s)$$

The off-diagonal elements of controller matrix $\mathcal{K}$ can be chosen according to the following rules:

$$\kappa_{ij}(s) = \begin{cases} \kappa_{j-1j}, & \forall i < j-1 \\ e^{-\sum_{k=j+1}^{i} \tau_k s}\kappa_{jj}(s), & \forall i > j. \end{cases} \tag{2.31}$$

This entails choosing $\kappa_F^{dd} = 1$ and $\kappa_F^{lu} = e^{-\sum_{k=j+1}^{i} \tau_k s}$ (the aggregate propagation delay). In the hierarchical control structure of Fig 2.2, the design of $i-$th PLC is specified by (2.30), (2.31), along with the equation (2.8) for determining gate opening. This is further illustrated by Fig 2.12.

In the following example, we state (robust) PI controller tuning rules for lu control of a canal pool.

We will use this setting in experimental results of Section 2.4.

*Example* 2.A.1. Consider a canal pool $i = 1$ with ID model:

$$\mathsf{y}_1^d(s) = \frac{a_1^d}{s}e^{-\tau_1 s}\mathsf{q}_0(s) - \frac{a_1^d}{s}[\mathsf{q}_1(s) + \mathsf{p}_1(s)].$$

For lu control we have $\hat{\mathsf{q}}_0(s) = 0$, $\hat{\mathsf{q}}_1(s) = \kappa_{11}(s)\hat{\epsilon}_1(s)$, where $\epsilon_1 = (\mathsf{y}_i^r - \mathsf{y}_1^d)$ is the output error, $\mathsf{y}_1^r$ is the reference level. The PI controller is given by:

$$\kappa_{11}(s) = k_1 \left(1 + \frac{1}{T_1 s}\right),$$

with $k_1$ the proportional gain and $T_1$ the integral time. These parameters can be determined for a gain margin $\Delta G$ dB and a phase margin $\Delta\Theta°$ using following tuning rules Litrico et al. [2007]. We recall that the gain margin (resp. phase margin) is the maximum multiplicative (resp. additive) increase in the gain (resp. phase) of the system such that the system remains closed-loop stable. These robustness margins in frequency domain are directly related to the time domain performance of the system.

$$\begin{aligned} k_1 &= \frac{\pi}{2a_1^d \tau_1} 10^{-\Delta G/20} \sin\left(\frac{\pi}{180}\Delta\Theta + \frac{\pi}{2}10^{-\Delta G/20}\right) \\ T_1 &= \frac{2\tau_1}{\pi} 10^{\Delta G/20} \tan\left(\frac{\pi}{180}\Delta\Theta + \frac{\pi}{2}10^{-\Delta G/20}\right) \end{aligned} \tag{2.32}$$

where the phase margin satisfies $\Delta\Theta < 90(1 - 10^{-\Delta G/20})$, and the parameters $\tau_1$ and $a_1^d$ are obtained by the relay-feedback auto-tuning method proposed by Åström and Hägglund [1995]. The method uses a single relay experiment to determine the frequency response of the canal pool at phase lag of 180°.

## Appendix 2.B Computation of Observer Parameters

The computation of observer parameter matrices $F_{ij}$, $G_{ij}$ (or equivalently $\bar{G}_{ij}$), $i = 0, 1, 2$, $T_j$, and $N_j$ for $j = 1, 2$ proceeds in the following two steps: In the first step, we check that the fault model (2.17) satisfies conditions for the existence of observer parameter matrices that are compatible with (2.21)–(2.23). In the second step, the observer parameter matrices are chosen by determining a free parameter matrix, such that the asymptotic stability condition (2.24) holds. For notational convenience, we will henceforth drop the observer index $j$ from the subscripts.

Step 1 Conditions (2.21)–(2.23) can be written as

$$S\Theta = \Psi, \tag{2.33}$$

where

$$\begin{aligned} S &= \begin{pmatrix} T, & N, & F_0, & \bar{G}_0, & F_1, & \bar{G}_1, & F_2 & \bar{G}_2 \end{pmatrix}, \\ \Theta &= \begin{pmatrix} \mathsf{I}_n & E & A_0 & A_1 & A_2 \\ C & 0 & 0 & 0 & 0 \\ 0 & 0 & -\mathsf{I}_n & 0 & 0 \\ 0 & 0 & -C & 0 & 0 \\ 0 & 0 & 0 & -\mathsf{I}_n & 0 \\ 0 & 0 & 0 & -C & 0 \\ 0 & 0 & 0 & 0 & -\mathsf{I}_n \\ 0 & 0 & 0 & 0 & -C \end{pmatrix}, \\ \Psi &= \begin{pmatrix} \mathsf{I}_n & 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Under the condition that $\operatorname{rank}(CE) = \operatorname{rank}(E)$, the general solution of (2.33) is

$$S = \Psi\Theta^{+} - K(I - \Theta\Theta^{+}), \tag{2.34}$$

where $K$ is an arbitrary matrix of appropriate dimension, and $\Theta^{+}$ is the generalized inverse matrix of $\Theta$. By inserting the solution (2.34) in (2.22), the matrices $F_i$ can now be expressed as

$$F_i = \chi_i - K\beta_i, \quad i = 0, 1, \ldots, 2, \tag{2.35}$$

where

$$\begin{aligned}
\chi_0 &= \Psi\Theta^{+}\begin{pmatrix} A_0 & 0 & 0 & -C & 0 & 0 & 0 & 0 \end{pmatrix}^{\mathsf{T}} \\
\chi_1 &= \Psi\Theta^{+}\begin{pmatrix} A_0 & 0 & 0 & 0 & 0 & -C & 0 & 0 \end{pmatrix}^{\mathsf{T}} \\
\chi_2 &= \Psi\Theta^{+}\begin{pmatrix} A_0 & 0 & 0 & 0 & 0 & 0 & 0 & -C \end{pmatrix}^{\mathsf{T}} \\
\beta_0 &= (I - \Theta\Theta^{+})\begin{pmatrix} A_0 & 0 & 0 & -C & 0 & 0 & 0 & 0 \end{pmatrix}^{\mathsf{T}} \\
\beta_1 &= (I - \Theta\Theta^{+})\begin{pmatrix} A_0 & 0 & 0 & 0 & 0 & -C & 0 & 0 \end{pmatrix}^{\mathsf{T}} \\
\beta_2 &= (I - \Theta\Theta^{+})\begin{pmatrix} A_0 & 0 & 0 & 0 & 0 & 0 & 0 & -C \end{pmatrix}^{\mathsf{T}}.
\end{aligned}$$

From above results, the condition (2.24) can be written as

$$\dot{\mathsf{e}}(t) = \sum_{i=0}^{2}(\chi_i - K\beta_i)\mathsf{e}(t - \tau_i(t)) \text{ is asymptotically stable.} \tag{2.36}$$

Step 2 Note that once the free-parameter matrix $K$ is computed such that conditions (2.36) holds, then all the observer parameter matrices can be computed. The following Proposition gives conditions under which $K$ can be computed:

**Proposition 2.B.1.** *System in (2.36) is asymptotically stable if for some scalars, $\epsilon_0, \ldots, \epsilon_6$ and $\bar{\epsilon}_1, \bar{\epsilon}_2$, there exist matrices $S_i > 0$, $Z_i > 0$, $Q_i > 0$, $R_i > 0$, $U_i$, $W_i$, i=1,...,2, and matrices $H_i$, $i = 1, \ldots, 6$, $U$ and $P > 0$ such that the following linear matrix inequalities (LMIs) are satisfied:*

$$\begin{pmatrix} Q_i & U_i \\ U_i^{\mathsf{T}} & R_i \end{pmatrix} \geqslant 0, \quad i = 1, \ldots, 4, \qquad \begin{pmatrix} \Phi & h_1\bar{H}_1 & h_2\bar{H}_2 \\ * & -h_1\bar{Z}_1 & 0 \\ * & * & -h_2\bar{Z}_2 \end{pmatrix} < 0,$$

*where*

$$\bar{Z}_i := \begin{pmatrix} S_i & W_i \\ * & Z_i \end{pmatrix}, \quad \bar{H}_i := \begin{pmatrix} -\bar{\epsilon}_i(P\chi_1 - U\beta_1)^{\mathsf{T}} & H_1 \\ -\bar{\epsilon}_i(P\chi_2 - U\beta_2)^{\mathsf{T}} & H_2 \\ -\bar{\epsilon}_i(P\chi_3 - U\beta_3)^{\mathsf{T}} & H_3 \\ \bar{\epsilon}_i P & H_4 \\ 0 & H_5 \\ 0 & H_6 \end{pmatrix},$$

*for $i = 1, 2$, and $\Phi = (\phi_{ij})$ is a symmetric matrix with block elements $\phi_{ij}$ given by the following:*

$$
\begin{aligned}
\phi_{11} &= \sum_{i=1}^{2}(Q_i + h_i S_i) + \epsilon_1 \operatorname{sym}(P\chi_0 - U\beta_0) + 2\operatorname{sym}(H_1),\\
\phi_{12} &= \epsilon_1(P\chi_1 - U\beta_1) + \epsilon_2(P\chi_0 - U\beta_0)^{\top} + 2H_2^{\top} - H_1,\\
\phi_{13} &= \epsilon_3(P\chi_0 - U\beta_0)^{\top} + \epsilon_1(P\chi_2 - U\beta_2) + 2H_3^{\top} - H_1,\\
\phi_{14} &= P + \sum_{i=1}^{2}(U_i + h_i W_i) + \epsilon_4(P\chi_0 - U\beta_0)^{\top} + 2H_4^{\top} - \epsilon_1 P,\\
\phi_{15} &= 2H_5^{\top} + \epsilon_5(P\chi_0 - U\beta_0)^{\top},\\
\phi_{16} &= 2H_6^{\top} + \epsilon_6(P\chi_0 - U\beta_0)^{\top},\\
\phi_{22} &= -Q_1 - \operatorname{sym}(H_2) + \epsilon_2 \operatorname{sym}(P\chi_1 - U\beta_1),\\
\phi_{23} &= -H_3^{\top} - H_2 + \epsilon_2 \operatorname{sym}(P\chi_2 - U\beta_2) + \epsilon_3 \operatorname{sym}(P\chi_1 - U\beta_1)^{\top},\\
\phi_{24} &= -H_4^{\top} + \epsilon_4 \operatorname{sym}(P\chi_1 - U\beta_1)^{\top} - \epsilon_2 P,\\
\phi_{25} &= -U_1 - H_5^{\top} + \epsilon_5 \operatorname{sym}(P\chi_1 - U\beta_1)^{\top},\\
\phi_{26} &= -H_6^{\top} + \epsilon_6(P\chi_1 - U\beta_1)^{\top},\\
\phi_{33} &= -Q_2 + \epsilon_3 \operatorname{sym}(P\chi_2 - U\beta_2) - \operatorname{sym}(H_3),\\
\phi_{34} &= -\epsilon_3 P + \epsilon_4(P\chi_2 - U\beta_2)^{\top} - H_4^{\top},\\
\phi_{35} &= +\epsilon_5(P\chi_2 - U\beta_2)^{\top} - H_5^{\top},\\
\phi_{36} &= -U_2 + \epsilon_6(P\chi_2 - U\beta_2)^{\top} - H_6^{\top},\\
\phi_{44} &= \sum_{i=1}^{2}(R_i + h_i Z_i) - \epsilon_4 \operatorname{sym}(P),\\
\phi_{45} &= -\epsilon_5 P^{\top},\\
\phi_{46} &= -\epsilon_6 P^{\top},\\
\phi_{55} &= -R_1,\\
\phi_{56} &= 0,\\
\phi_{66} &= -R_2
\end{aligned}
$$

*where $h_i = \tau_i$, and $\operatorname{sym}(M) := M + M^{\top}$. The parameter matrix $K$ is given by $K = P^{-1}U$.*

The proof is a simplification of a more general case considered in Chapter 3.

# Chapter 3

# Detection of Deception Attacks on Water SCADA Systems

This chapter investigates the problem of detection and isolation of attacks in a water distribution network of cascaded canal pools when the measured sensor signals are compromised by an adversary. An approach based on a bank of delay-differential observer systems is proposed, in which each observer has the same general form as an analytically approximate model of canal hydrodynamics. Each observer is insensitive to one fault/attack mode and sensitive to other modes. Design of observers is achieved by using a delay-dependent linear matrix inequality (LMI) method. The performance of the proposed model-based detection scheme is tested on a class of adversarial scenarios resulting from a generalized fault/attack model. This model can represent classical sensor-actuator faults as well as communication network-induced deception attacks. A particular focus of this chapter is on stealthy deception attacks in which the attacker's goal is to pilfer water through canal offtakes. From the viewpoint of canal system operations, this analysis reveals the advantages of using more advanced physics based models in detecting physical faults and cyber-induced attacks in automated canal systems. The criticality of measured sensor signals for the purpose of detection is also investigated. Finally, from an attacker's viewpoint, the knowledge and effort required to carry out a successful deception attack is discussed.

## 3.1 Introduction

Modernization of irrigation systems is often viewed as a solution for improving their performance. In numerous countries networked fully gated irrigation systems have been upgraded with supervisory control and data acquisition (SCADA) systems to enable communications, sensing, and control. Real-time knowledge of the system state and the ability to remotely control flows at critical points can vastly improve performance of irrigation systems (see for e.g., Litrico et al. [2007]; Rijo and Arranja [2010]). In numerous developed countries (e.g., Australia, France, United States) well-defined rules for demand regulation, proper maintenance plans, and a required legislative framework are already in place to sustain modernization plans of irrigation systems. Several emerging and developing countries (e.g., Brazil, India, Morocco) are also in the process of modernizing their irrigation systems. A significant body of research work now exists focusing on automatic control methods for regulation of discharges and levels in networks of irrigation canals. We refer the reader to Cantoni et al. [2007]; Litrico and Fromion [2009*b*], and the references therein for a survey of these methods.

However, modernization does *not* always imply reliable service Plusquellec [2009]. Even in developed countries, automated irrigation systems are experiencing large amounts of water losses due to management and distribution related inefficiencies. Such issues are more challenging to address in developing countries. Clemmens Clemmens [2005] has argued that instability of canal flows and large deviations from target levels at downstream ends can lead to inefficient water distribution. This can lead to interference from the end users, e.g., water *pilfering* by farmers. An increased reliance of communication systems to transmit and receive control data, has added new concerns of cyber attacks Attorney [2007]; Slay and Miller [2007] (in addition to existing issues of physical faults).

Chapter 2 highlighted the ways in which simultaneous and uncoupled cyber-physical faults (or cyber *attacks*) in automated irrigation canal systems can be achieved by an intelligent adversary. By presenting the results from a field operational test, we showed in Chapter 2 that it is possible for an attacker to stealthily withdraw water from an automated canal without getting detected. This motivates an improved understanding of hydrodynamic principles which can assist the design of better fault/attack detection mechanisms. The focus of this chapter is on the design a fault/attack detection and isolation (F/ADI) scheme which is based on an accurate hydrodynamic model, and uses the theory of robust observer design in the presence of unknown inputs. The generalized fault/attack model which we consider here can model both random sensor-actuator faults and a class of cyber attacks.

### 3.1.1 Related Work

A wide body of work has been reported during the past few years to address the problem of fault detection and isolation (FDI) of unknown water withdrawals (or leaks) Bedjaoui et al. [2009]; Weyer and Bastin [2008], and random sensor-actuator faults in canal systems Bedjaoui et al. [2008]. The authors in Bedjaoui et al. [2008] use data reconciliation based on static and dynamic models to isolate unknown water withdrawals and sensor-actuator faults. A simple finite-dimensional model of canal flow is used in Weyer and Bastin [2008] to generate differences (i.e., residuals) between the model and observed data. The residuals are aggregated over time by a cumulative sum (CUSUM) algorithm (based on the theory of change-point detection Basseville and Nikiforov [1993]). An alert for a leak is generated when the CUSUM statistic reaches a given threshold. Under the assumption that the size of the leak and the time of start are known, Bedjaoui et al. [2009] uses a bank of Luenberger observers based on the shallow water equations to localize the leaks. The authors of Bedjaoui et al. [2009] also discuss the use of observed time-difference between the effect of leaks seen at the upstream and downstream of canal pools to localize the leaks. Results from stability of hyperbolic conservation laws Aamo et al. [2006]; deHalleux et al. [2003] is used to prove observer stability in Bedjaoui et al. [2009]. Some response mechanisms to address sensor-actuator faults are presented in Choy and Weyer [2008].

The most closely related works to the approach presented in this chapter are Koenig et al. [2005] and Bedjaoui and Weyer [2011]. The article by Bedjaoui and Weyer [2011] provides a comparison of different methods of residual generation based on finite and infinite dimensional models. The authors propose that a properly tuned CUSUM algorithm can achieve leak detection. An estimation of water leakage is generated from residuals based on a simple conversion formula. A technique to isolate a single sensor fault from a single leak is also presented based on monitoring of canal pools located upstream and downstream of the suspect pool. The article Koenig et al. [2005] uses unknown input observers (UIO) for time-delay systems, as developed in systems theory (see for e.g., Conte and Perdon [2006]; Darouach et al. [1994]) to design a FDI scheme for a single canal reach. This approach was extended to multiple pools when only downstream levels are measured in Bedjaoui et al. [2006].

The aforementioned work indicates that the problem of isolating sensor-actuator faults from unknown water withdrawals can be difficult because both these faults have similar effect on the observer residuals. Moreover, to the best our knowledge, the performance of available detection schemes where sensor-actuator faults and unknown water withdrawals occur simultaneously has been not been investigated in the literature. From the viewpoint of security of automated canal systems, such *simultaneous faults form an interesting class of attacks*. These attacks have been recently shown to be feasible for an intelligent attacker who is interested in water pilfering, or has malicious intentions to harm canal operations (see Chapter 2).

### 3.1.2 Contributions for Fault and Attack Diagnosis

Based on the discussion above, the contributions of this chapter are as follows:

- This chapter presents conditions for detectability and isolability of faults due to non-simultaneous (and uncoupled) withdrawals/leaks and sensor disturbances in cascade of canal pools using a bank of UIOs. The UIO design uses an analytic approximation of the canal hydrodynamics (Theorem 3.4.1). This model captures the effect of both upstream and downstream flow variations. The detection scheme can be designed provided that a feasible solution to delay-dependent observer stability criterion exists (Proposition 3.4.2), and observer decoupling conditions are satisfied (Definition 3.4.3).

- A F/ADI scheme is proposed based on residuals generated from the bank of UIO, and its performance is analyzed under simultaneous and uncoupled faults (called attacks), for e.g., simultaneous compromise of one or more sensor measurements and water pilfering using offtake structures. This analysis points toward fundamental limitations of model-based detection schemes in isolating attacks caused by a malicious attacker on distributed physical infrastructure systems. Implications of these findings on the security of canal SCADA systems are also discussed.

This chapter is organized as follows. In Section 3.2 we first introduce infinite-dimensional models for a cascade of canal pools, and propose the use of an analytically approximate finite-dimensional model. This model is used to design a UIO based scheme for detecting faults entering in state and measurement equations in Section 3.4. In Section 3.5, we first present a generalized fault/attack model which captures attack scenarios such as simultaneous water pilfering through offtakes and sensor compromise. Next, we analyze the advantages and limitations of the proposed detection scheme under the generalized fault/attack model. We also discuss security implications of attack scenarios which can result in such failures. Some concluding remarks are drawn in Section 3.6.

## 3.2 Models of Canal Pool Cascade

### 3.2.1 Model of Flow Dynamics

Consider an irrigation system consisting of a cascade of $m$ canal pools. Each canal pool is represented by a portion of canal in between two automated hydraulic structures; thus, the cascade has $(m+1)$ hydraulic structures. We assume that pool $i$, where $i = 1, \ldots, m$ has prismatic cross-section and is of length $l_i$ (m). Let the space variable be denoted by $x \in [0, l_i]$ and time variable by $t \in \mathbb{R}_+$. The unsteady flow dynamics of each canal pool are classically modeled by the one-dimensional shallow water equations (SWE) Litrico and Fromion [2009*b*]. The SWE is a model of coupled hyperbolic PDEs with $\mathsf{A}_i(t,x)$ the wetted cross-sectional area (m$^2$) and $\mathsf{Q}_i(t,x)$ the discharge (m$^3$/s) across cross-section $\mathsf{A}_i$ as the dependent variables, and $t$ and $x$ as independent variables. The SWE for pool $i$ is given by

$$\partial_t \begin{pmatrix} \mathsf{A}_i \\ \mathsf{Q}_i \end{pmatrix} + \mathbf{F}(\mathsf{A}_i, \mathsf{Q}_i)\partial_x \begin{pmatrix} \mathsf{A}_i \\ \mathsf{Q}_i \end{pmatrix} = \mathbf{H}(\mathsf{A}_i, \mathsf{Q}_i), \tag{3.1}$$

on the domain $x \in (0, l_i)$, $t > 0$ with

$$\mathbf{F}(\mathsf{A}_i, \mathsf{Q}_i) = \begin{pmatrix} 0 & 1 \\ g\mathsf{A}_i\partial_{\mathsf{A}_i}\mathsf{Y}_i(\mathsf{A}_i) - \frac{\mathsf{Q}_i^2}{\mathsf{A}_i^2} & 2\frac{\mathsf{Q}_i}{\mathsf{A}_i} \end{pmatrix}, \quad \mathbf{H}(\mathsf{A}_i, \mathsf{Q}_i) = \begin{pmatrix} 0 \\ g\mathsf{A}_i(\mathsf{S}_{bi} - \mathsf{S}_{fi}(\mathsf{A}_i, \mathsf{Q}_i)) \end{pmatrix}.$$

Here the notation $\partial_t$, $\partial_x$, and $\partial_{\mathsf{A}_i}$ denote the partial derivatives with respect to $t$, $x$, and $\mathsf{A}_i$ respectively. The function $\mathsf{S}_{fi}(\mathsf{A}_i, \mathsf{Q}_i)$ denotes the friction slope (m/m), $\mathsf{S}_{bi}$ the bed slope (m/m), $\mathsf{Y}_i(\mathsf{A}_i)$ the water depth (m) in section $\mathsf{A}_i$, and $g$ the acceleration due to gravity (m$^2$/s). We model the friction slope as $\mathsf{S}_{fi} := \frac{\mathsf{Q}_i^2 n_i^2}{\mathsf{A}_i^2 R_i(\mathsf{A}_i)^{4/3}}$, where $n_i$ is the Manning roughness coefficient (sm$^{-1/3}$), $R_i(\mathsf{A}_i) := \frac{P_i}{\mathsf{A}_i}$ is the hydraulic radius (m), $P_i$ is the wetted perimeter (m), $\mathsf{V}_i(t,x) := \frac{\mathsf{Q}_i(t,x)}{\mathsf{A}_i(t,x)}$ is the average velocity (m/s) in section $\mathsf{A}_i$, $\mathsf{C}_i(t,x) := \sqrt{\frac{g\mathsf{A}_i(t,x)}{\mathsf{T}_i(t,x)}}$ is the celerity (m/s), and $\mathsf{T}_i$ is the top width (m).

We assume that $\mathsf{V}_i < \mathsf{C}_i$ (sub-critical flow), and therefore, one boundary condition must be specified at each boundary. The initial and boundary conditions are given by:

$$\mathsf{Q}_i(t,0) = \mathsf{Q}_i^u(t), \quad \mathsf{Q}_i(t,l_i) = \mathsf{Q}_i^d(t) + \mathsf{P}_i(t), \quad t \geqslant 0, \tag{3.2}$$

$$\mathsf{A}_i(0,x) = \mathsf{A}_{0,i}(x), \quad \mathsf{Q}_i(0,x) = \mathsf{Q}_{0,i}(x), \quad x \in (0,l_i). \tag{3.3}$$

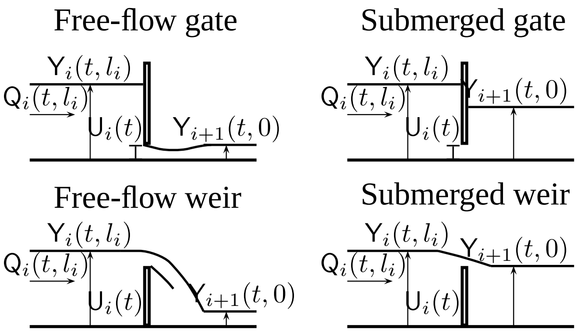


Figure 3.1: Free-Flow and submerged hydraulic structures.

Here $\mathsf{Q}_i^u(t)$ and $\mathsf{Q}_i^d(t)$ denote the controllable upstream and downstream boundary discharges (m$^3$/s) for pool $i$ respectively, and $\mathsf{P}_i(t)$ denote the withdrawals through lateral offtakes (m$^3$/s). The boundary discharges are constrained as

$$\mathsf{Q}_i^d(t) = \mathsf{Q}_{i+1}^u(t), \quad t \geqslant 0 \quad i = 0, \dots m. \tag{3.4}$$

We will assume the following: a) the effect of offtakes along the canal pool can be lumped into a single perturbation $\mathsf{P}_i(t)$ acting near the downstream end of the pool[1]; b) the conversion of the boundary discharges into automated movement of hydraulic structures is handled by the respective controllers located at these structures; c) the water levels $\mathsf{Y}_i(t,0)$ and $\mathsf{Y}_i(t,l_i)$ (or equivalently, the areas $\mathsf{A}_i(t,0)$ and $\mathsf{A}_i(t,l_i)$) are measured variables, the boundary discharges $\mathsf{Q}_i^u(t)$ and $\mathsf{Q}_i^d(t)$ are control variables, and the offtake withdrawals $\mathsf{P}_i(t)$ are disturbance variables.

### 3.2.2 Model of Hydraulic Structures

Overflow weirs and underflow gates are the most commonly used hydraulic structures for regulating flows in canal networks. These structures can be in free-flow or submerged condition (see Fig. 3.1).

In submerged condition (resp. free-flow condition), the downstream level influences (resp. does not influence) the flow through the structure. We define $\mathsf{Y}_0(t,l_0) := \mathsf{Y}_{\text{up}}(t)$ and $\mathsf{Y}_{m+1}(t,0) := \mathsf{Y}_{\text{do}}(t)$, where $\mathsf{Y}_{\text{up}}(t)$ (resp. $\mathsf{Y}_{\text{do}}(t)$) are the upstream (resp. downstream) water levels of the first (resp. last) canal pool in the cascade. The flow through structure $i$ is modeled by a static nonlinear relation $\mathbf{G}_i$ with following general form :

$$\mathsf{Q}_i(t,l_i) = \mathbf{G}_i(\mathsf{Y}_i(t,l_i), \mathsf{Y}_{i+1}(t,0), \mathsf{U}_i(t)) \tag{3.5}$$

for $i = 0, \dots, m$, where $\mathsf{U}_i(t)$ denotes opening of the structure (m) at time $t$. The level-discharge relations for overflow weirs and underflow gates under both free-flow and submerged conditions are presented below:

[1] Distributed offtake withdrawal models have been considered elsewhere (see for e.g., Bedjaoui et al. [2009] and Amin et al. [2010]). The FDI scheme presented in Section 3.4 can be readily extended to the case of distributed withdrawals by suitable expansion of the observer bank.

*Remark* 3.2.1. (Level-discharge relations) For underflow gates, we have

$$\mathsf{Q}_i(t,l_i)=\begin{cases}C_{g,i}L_{g,i}\mathsf{U}_i(t)\sqrt{2g(\mathsf{Y}_i(t,l_i)-\mathsf{Y}_{i+1}(t,0))} & \text{(sf)}\\ C_{g,i}L_{g,i}\mathsf{U}_i(t)\sqrt{2g(\mathsf{Y}_i(t,l_i))} & \text{(ff)},\end{cases}$$

where (ff) and (sf) respectively denote the free-flow and submerged flow conditions, $C_{g,i}$ denotes the discharge coefficient (generally close to 0.6), $L_{g,i}$ the lateral width (m) of the gate. For overflow weirs, we have

$$\mathsf{Q}_i(t,l_i)=\begin{cases}C_{w,i}\sqrt{2g}(\mathsf{Y}_i(t,l_i)-\mathsf{Y}_{i+1}(t,0))^{3/2} & \text{(sf)}\\ C_{w,i}\sqrt{2g}(\mathsf{Y}_i(t,l_i)-\mathsf{U}_i(t))^{3/2} & \text{(ff)},\end{cases}$$

where $C_{w,i}$ denotes the discharge coefficient for weirs (generally close to 0.4). ◁

### 3.2.3 Linearized Models

Under compatible and constant openings $\mathsf{U}_i(t)=\bar{\mathsf{U}}_i$, withdrawals $\mathsf{P}_i(t)=\bar{\mathsf{P}}_i$, and levels $\mathsf{Y}_{\text{up}}(t)=\bar{\mathsf{Y}}_{\text{up}}$, $\mathsf{Y}_{\text{do}}(t)=\bar{\mathsf{Y}}_{\text{do}}$, the system (3.1)–(3.4) achieves a steady state. Let the corresponding wetted area and discharge in steady state be denoted by $\bar{\mathsf{A}}_i(x)$ and $\bar{\mathsf{Q}}_i(x)$ respectively; similarly for other variables. We henceforth omit the dependence on $x$. Following Litrico and Fromion [2009*b*], SWE (3.1) can be linearized around a steady state $(\bar{\mathsf{A}}_i,\bar{\mathsf{Q}}_i)$ using the approximation:

$$f(\mathsf{A}_i,\mathsf{Q}_i)\approx f(\bar{\mathsf{A}}_i,\bar{\mathsf{Q}}_i)+\left(\overline{\partial_{\mathsf{A}_i}f}\right)\mathsf{a}_i+\left(\overline{\partial_{\mathsf{Q}_i}f}\right)\mathsf{q}_i,$$

where $\mathsf{a}_i(t,x):=(\mathsf{A}_i(t,x)-\bar{\mathsf{A}}_i(x))$, $\mathsf{q}_i(t,x):=(\mathsf{Q}_i(t,x)-\bar{\mathsf{Q}}_i(x))$ are the deviations from steady state. The notation $\overline{(\cdot)}$ indicates that all quantities are evaluated at steady state. The linearized shallow water equations (LWSE) are given by

$$\frac{\partial}{\partial t}\begin{pmatrix}\mathsf{a}_i\\ \mathsf{q}_i\end{pmatrix}+\bar{\mathbf{F}}_i(x)\frac{\partial}{\partial x}\begin{pmatrix}\mathsf{a}_i\\ \mathsf{q}_i\end{pmatrix}+\bar{\mathbf{G}}_i(x)\begin{pmatrix}\mathsf{a}_i\\ \mathsf{q}_i\end{pmatrix}=0, \tag{3.6}$$

on the domain $x\in(0,l_i), t\geqslant 0$, where $\begin{pmatrix}\mathsf{a}_i(t,x), & \mathsf{q}_i(t,x)\end{pmatrix}^{\mathsf{T}}$ is the state of canal pool $i$, and

$$\bar{\mathbf{F}}_i(x):=\begin{pmatrix}0 & 1\\ \alpha_i(x)\beta_i(x) & \alpha_i(x)-\beta_i(x)\end{pmatrix},\quad \bar{\mathbf{G}}_i(x):=\begin{pmatrix}0 & 0\\ -\gamma_i(x) & \delta_i(x)\end{pmatrix}.$$

Omitting the dependence on $x$, and defining $\kappa_i:=\frac{7}{3}-\frac{4\bar{\mathsf{A}}_i}{3\bar{\mathsf{T}}_i\bar{P}_i}\frac{\partial\bar{P}_i}{\partial\bar{\mathsf{Y}}_i}$, we have

$$\alpha_i=\bar{\mathsf{C}}_i+\bar{\mathsf{V}}_i,\quad \beta_i=\bar{\mathsf{C}}_i-\bar{\mathsf{V}}_i,\quad \delta_i=\frac{2g}{\bar{V}_i}\left(\bar{\mathsf{S}}_{fi}-\frac{\bar{\mathsf{V}}_i^2\bar{\mathsf{T}}_i}{g\bar{\mathsf{A}}_i}\frac{d\bar{\mathsf{Y}}_i}{dx}\right),$$
$$\gamma_i=\frac{\bar{\mathsf{C}}_i^2}{\bar{\mathsf{T}}_i}\frac{d\bar{\mathsf{T}}_i}{dx}+g\left[(1+\kappa_i)\mathsf{S}_{bi}-(1+\kappa_i-(\kappa_i-2)\frac{\bar{\mathsf{V}}_i^2\bar{\mathsf{T}}_i}{g\bar{\mathsf{A}}_i})\frac{d\bar{\mathsf{Y}}_i}{dx}\right].$$

System (3.6), along with the initial and boundary conditions

$$\mathsf{a}_i(0,x)=\mathsf{a}_{0,i}(x)\text{ and }\mathsf{q}_i(0,x)=\mathsf{q}_{0,i}(x),\quad x\in(0,l_i), \tag{3.7}$$
$$\mathsf{q}_i(t,0)=\mathsf{q}_i^u(t)\text{ and }\mathsf{q}_i(t,l_i)=\mathsf{q}_i^d(t)+\mathsf{p}_i(t),\quad t\geqslant 0, \tag{3.8}$$

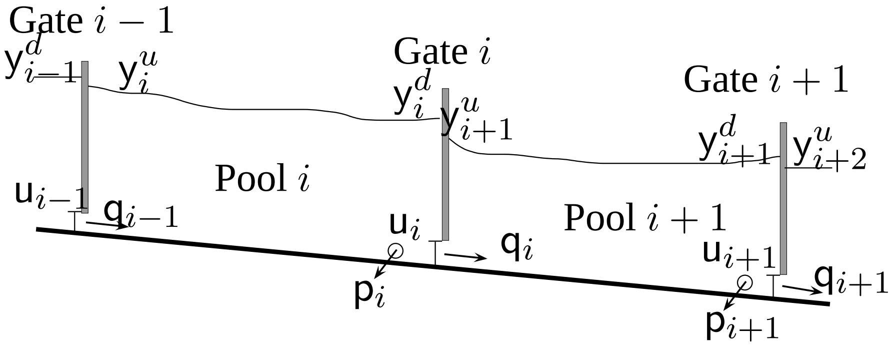


Figure 3.2: Schematic view of a multiple pool canal system with submerged gates.

and the constraint

$$\mathsf{q}_i^d(t) = \mathsf{q}_{i+1}^u(t), \quad t \geqslant 0, \tag{3.9}$$

form the linearized model for canal pool $i$, where $\mathsf{q}_i^u(t) = \mathsf{Q}_i(t,0) - \bar{\mathsf{Q}}_i(0)$ and $\mathsf{q}_i^d(t) = \mathsf{Q}_i(t,l_i) - \bar{\mathsf{Q}}_i(l_i)$ denote the boundary discharge deviations, and $\mathsf{p}_i(t) = \mathsf{P}_i(t) - \bar{\mathsf{P}}_i$ the withdrawal deviations from the respective steady states. We note that for rectangular cross-sections, the linearized model with $\mathsf{y}_i(t,x)$ and $\mathsf{a}_i(t,x)$ as state can be deduced by using

$$\mathsf{a}_i(t,x) = \bar{\mathsf{T}}(x)\mathsf{y}_i(t,x).$$

With a slight abuse of notation, we define (see Fig. 3.2):

$$\mathsf{q}_{i-1}(t) := \mathsf{q}_i^u(t), \quad \mathsf{q}_i(t) := \mathsf{q}_i^d(t), \quad \mathsf{y}_i^u(t) := \mathsf{y}_i(t,0), \quad \mathsf{y}_i^d(t) := \mathsf{y}_i(t,l_i). \tag{3.10}$$

Finally, linearizing (3.5) about the steady state we obtain

$$\mathsf{q}_i(t) = b_i^d \mathsf{y}_i^d(t) + b_{i+1}^u \mathsf{y}_{i+1}^u(t) + k_i \mathsf{u}_i(t), \tag{3.11}$$

where $\mathsf{u}_i(t) = (\mathsf{U}_i(t) - \bar{\mathsf{U}}_i)$ denotes the deviation in the structure opening, the coefficients $b_i^d = \left(\overline{\partial_{\mathsf{Y}_i}\mathbf{G}_i}\right)$ and $b_{i+1}^u = \left(\overline{\partial_{\mathsf{Y}_{i+1}}\mathbf{G}_i}\right)$ are the feedback gains of the upstream and downstream levels, and $k_i = \left(\overline{\partial_{\mathsf{U}_i}\mathbf{G}_i}\right)$ is the gain of structure opening. These coefficients can be computed by linearizing the respective level-discharge relations for the case of gate and weir. Note that $b_{i+1}^u$ is strictly negative (resp. zero) for submerged (resp. free-flow) condition, and $b_i^d$ and $k_i$ are positive.

*Remark* 3.2.2. (Uniform Flow) The following condition corresponds to the uniform flow regime:

$$\bar{\mathsf{S}}_{fi}(\bar{\mathsf{Q}}_i(x), \bar{\mathsf{A}}_i(x)) = \mathsf{S}_{bi}, \quad x \in (0, l_i),$$

Let the steady state in uniform flow be denoted by $\bar{\mathsf{A}}_i$ and $\bar{\mathsf{Q}}_i$. The matrix functions in (3.6) now become independent of $x$, and the constants $\gamma_i$ and $\delta_i$ simplify to $\gamma_i = g\mathsf{S}_{bi}\left(\frac{10}{3} - \frac{4\bar{\mathsf{A}}_i}{3\bar{\mathsf{T}}_i\bar{P}_i}\frac{d\bar{P}_i}{d\bar{\mathsf{Y}}_i}\right)$, and $\delta_i = \frac{2g\mathsf{S}_{bi}}{\bar{\mathsf{V}}_i}$. ◁

## 3.3 Characteristic form and IDZ model

### 3.3.1 Characteristic Form

The matrix function $\bar{\mathbf{F}}_i(x)$ in (3.6) can be diagonalized as

$$\bar{\mathbf{F}}_i(x) = \mathbf{X}_i(x)^{-1}\mathbf{A}_i(x)\mathbf{X}_i(x),$$

where

$$\mathbf{X}_i(x) = \begin{pmatrix} \beta_i(x) & 1 \\ -\alpha_i(x) & 1 \end{pmatrix}, \quad \mathbf{A}_i(x) = \begin{pmatrix} \alpha_i(x) & 0 \\ 0 & -\beta_i(x) \end{pmatrix},$$

Using the following change of coordinates

$$\xi_i(t,x) := \begin{pmatrix} \xi_{1,i}(t,x) \\ \xi_{2,i}(t,x) \end{pmatrix} = \mathbf{X}_i(x) \begin{pmatrix} \mathsf{a}_i(t,x) \\ \mathsf{q}_i(t,x) \end{pmatrix}, \tag{3.12}$$

where $\xi_{1,i}(t,x)$ and $\xi_{2,i}(t,x)$ denote the characteristic variables, and applying the transformation (3.12), the system (3.6) can be expressed as

$$\frac{\partial \xi_i}{\partial t} + \mathbf{A}_i(x)\frac{\partial \xi_i}{\partial x} + \mathbf{B}_i(x)\xi_i = 0, t \geqslant 0, \tag{3.13}$$

with $\mathbf{B}_i(x) = \left[\mathbf{X}_i(x)\bar{\mathbf{G}}_i(x) - \bar{\mathbf{A}}_i(x)\partial_x \mathbf{X}_i(x)\right]\mathbf{X}_i(x)^{-1}$. The initial conditions are

$$\xi_{1,i}(0,x) = \xi^0_{1,i}(x) \text{ and } \xi_{2,i}(0,x) = \xi^0_{2,i}(x), \quad x \in (0, l_i), \tag{3.14}$$

where $\xi^0_{1,i}(x) := +\beta_i(x)\mathsf{a}_{0,i}(x) + \mathsf{q}_{0,i}(x)$ and $\xi^0_{2,i}(x) := -\alpha_i(x)\bar{\mathsf{a}}_{0,i}(x) + \bar{\mathsf{q}}_{0,i}(x)$. Under the assumption that the boundary control actions are linear functions of the local state variables, the boundary conditions can be expressed as

$$\begin{pmatrix} \hat{\xi}_{1,i}(s,0) \\ \hat{\xi}_{2,i}(s,l_i) \end{pmatrix} = \mathcal{K}_i \begin{pmatrix} \hat{\xi}_{1,i}(s,l_i) \\ \hat{\xi}_{2,i}(s,0) \end{pmatrix}, \tag{3.15}$$

where $\mathcal{K}_i$ is the controller matrix. One can observe that along the characteristic curves defined by the ODEs

$$\frac{dx_1(t)}{dt} = \alpha_i(x_1(t)), \text{ and } \frac{dx_2(t)}{dt} = -\beta_i(x_2(t)),$$

the characteristic variables verify

$$\begin{aligned} \frac{d\xi_{1,i}}{dt}(t,x_1) &= -\, b_{11,i}(x_1)\xi_{1,i}(t,x_1) - b_{12,i}(x_1)\xi_{2,i}(t,x_1), \\ \frac{d\xi_{2,i}}{dt}(t,x_2) &= -\, b_{21,i}(x_2)\xi_{1,i}(t,x_2) - b_{22,i}(x_2)\xi_{2,i}(t,x_2), \end{aligned}$$

where the dependence of $x_1$ and $x_2$ on $t$ is omitted, and $(b_{jk,i}(x))$ denote the elements of $\mathbf{B}_i(x)$. From the above characteristic form, we observe that the state at any point $(t,x)$

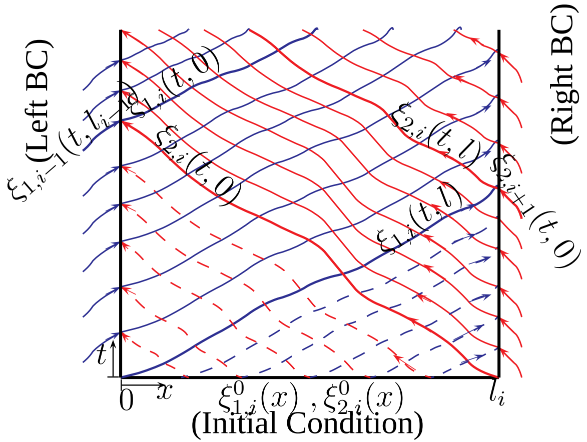


Figure 3.3: Characteristic curves for $i-$th pool.

depends on information from both upstream and downstream propagating characteristic curves (see Fig 3.3).

The elements of the matrix function $\mathbf{B}_i(x) = (b_{jk,i}(x))$ can be expressed as

$$\mathbf{B}_i(x) = \begin{pmatrix} \frac{-\gamma_i+\alpha_i\delta_i-\alpha_i\partial_x\beta_i}{\alpha_i+\beta_i} & \frac{\gamma_i+\beta_i\delta_i+\alpha_i\partial_x\beta_i}{\alpha_i+\beta_i} \\ \frac{-\gamma_i+\alpha_i\delta_i-\beta_i\partial_x\alpha_i}{\alpha_i+\beta_i} & \frac{\gamma_i+\beta_i\delta_i+\beta_i\partial_x\alpha_i}{\alpha_i+\beta_i}, \end{pmatrix}$$

where the dependence on $x$ in the right-hand-side is omitted for notational convenience.

For uniform flow, the matrix functions $\mathbf{X}_i(x)$ in coordinate transformation given by equation (3.12), and $\mathbf{A}_i(x)$, $\mathbf{B}_i(x)$ and $\mathbf{C}_i(x)$ in linear SWE expressed in the $\chi_i$ variables (3.13) also become independent of the space variable $x$, and thus, $\mathbf{X}_i(x) = \mathbf{X}_i$, $\mathbf{A}_i(x) = \mathbf{A}_i$, $\mathbf{B}_i(x) = \mathbf{B}_i$, and $\mathbf{C}_i(x) = \mathbf{C}_i$ . In particular, the elements of the matrix $\mathbf{B}_i = (b_{jk,i})$ are given by

$$\mathbf{B}_i = \begin{pmatrix} \frac{-\gamma_i+\alpha_i\delta_i}{\alpha_i+\beta_i} & \frac{\gamma_i+\beta_i\delta_i}{\alpha_i+\beta_i} \\ \frac{-\gamma_i+\alpha_i\delta_i}{\alpha_i+\beta_i} & \frac{\gamma_i+\beta_i\delta_i}{\alpha_i+\beta_i} \end{pmatrix}$$

### 3.3.2 Integrator-Delay Model

Using analytic approximation in the frequency domain, Litrico and Fromion have derived a finite-dimensional input-output model which accounts for the effect of both upstream and downstream variations (see also Section 5.3 in Litrico and Fromion [2009*b*]). In low-frequencies, this approximate model is given by the integrator-delay (ID) model[2]:

$$\begin{pmatrix} \hat{\mathsf{y}}^u_i(s) \\ \hat{\mathsf{y}}^d_i(s) \end{pmatrix} = \begin{pmatrix} \frac{a^u_i}{s} & -\frac{a^u_i}{s}e^{-\bar{\tau}_i s} \\ \frac{a^d_i}{s}e^{-\tau_i s} & -\frac{a^d_i}{s} \end{pmatrix} \begin{pmatrix} \hat{\mathsf{q}}_{i-1}(s) \\ \hat{\mathsf{q}}_i(s) + \mathsf{p}_i(s) \end{pmatrix}. \tag{3.16}$$

[2] The integrator-delay-zero (IDZ) model, as presented in Litrico and Fromion [2004*a*], also accounts for high frequencies by using a constant gain (in addition to an integrator and a delay).

The parameter $a_i^u$ (resp. $a_i^d$) corresponds to the inverse of the equivalent backwater area for the upstream (resp. downstream) water level, and the parameter $\bar{\tau}_i$ (resp. $\underline{\tau}_i$) is the upstream (resp. downstream) propagation time delays, i.e., the minimum time for a change in the downstream (resp. upstream) discharge to have an effect on the upstream (resp. downstream) water level. For uniform flow, these parameters can be obtained analytically Litrico and Fromion [2009*b*]:

$$a_i^u = \frac{\gamma_i}{\alpha_i\beta_i\bar{\mathsf{T}}_i\left(e^{\frac{\gamma_i l_i}{\alpha_i\beta_i}}-1\right)}, \quad a_i^d = \frac{\gamma_i}{\alpha_i\beta_i\bar{\mathsf{T}}_i\left(1-e^{-\frac{\gamma_i l_i}{\alpha_i\beta_i}}\right)}, \quad \underline{\tau}_i = \frac{l_i}{\alpha_i}, \quad \bar{\tau}_i = \frac{l_i}{\beta_i}.$$

For non-uniform regime, these parameters can be computed via a numerical scheme which connects several (virtual) uniform flow pools Litrico and Fromion [2009*b*]. Notice from (3.16) that the ID model accounts for the influence of both upstream and downstream discharge and thus, captures the input-output behavior in backwater flow configurations. In the time-domain, we have the following ODE with delayed inputs:

$$\begin{aligned} \dot{\mathsf{y}}_i^u(t) &= a_i^u\mathsf{q}_{i-1}(t) - a_i^u\left[\mathsf{q}_i(t-\bar{\tau}_i) + \mathsf{p}_i(t-\bar{\tau}_i)\right], \\ \dot{\mathsf{y}}_i^d(t) &= a_i^d\mathsf{q}_{i-1}(t-\underline{\tau}_i) - a_i^d\left[\mathsf{q}_i(t) + \mathsf{p}_i(t)\right]. \end{aligned} \tag{3.17}$$

Applying Laplace transform to (3.6), we obtain the following ODE system in $x$ variable, parameterized by the Laplace variable $s$:

$$\frac{\partial}{\partial x}\begin{pmatrix}\hat{\mathsf{a}}_i(s,x)\\ \hat{\mathsf{q}}_i(s,x)\end{pmatrix} = \mathcal{A}_i(s,x)\begin{pmatrix}\hat{\mathsf{a}}_i(s,x)\\ \hat{\mathsf{q}}_i(s,x)\end{pmatrix} + \mathcal{B}_i(x)\begin{pmatrix}\mathsf{a}_i(0,x)\\ \mathsf{q}_i(0,x)\end{pmatrix}, \tag{3.18}$$

where

$$\mathcal{A}_i(s,x) = \begin{pmatrix}\frac{(\alpha_i(x)-\beta_i(x))s+\gamma_i(x)}{\alpha_i(x)\beta_i(x)} & -\frac{s+\delta_i(x)}{\alpha_i(x)\beta_i(x)}\\ -s & 0\end{pmatrix},$$

$$\mathcal{B}_i(x) = \begin{pmatrix}\frac{\beta_i(x)-\alpha_i(x)}{\alpha_i(x)\beta_i(x)} & \frac{1}{\alpha_i(x)\beta_i(x)}\\ 1 & 0\end{pmatrix}.$$

In general, solution of (3.18) cannot be obtained analytically; however, its general solution exists, is unique (see Chapter 3, Section 4 in Litrico and Fromion [2009*b*]), and can be expressed as

$$\begin{pmatrix}\hat{\mathsf{a}}_i(s,x)\\ \hat{\mathsf{q}}_i(s,x)\end{pmatrix} = \Psi_i(s,x)\left[\begin{pmatrix}\hat{\mathsf{a}}_i(s,0)\\ \hat{\mathsf{q}}_i(s,0)\end{pmatrix} + \begin{pmatrix}\bar{\mathsf{a}}_{0,i}(s,x)\\ \bar{\mathsf{q}}_{0,i}(s,x)\end{pmatrix}\right], \tag{3.19}$$

where $\Psi_i(s,x) = (\psi_{i,jk}(s,x))$ is the state transition matrix for (3.18), and

$$\begin{pmatrix}\bar{\mathsf{a}}_{0,i}(s,x)\\ \bar{\mathsf{q}}_{0,i}(s,x)\end{pmatrix} := \int_0^x \Psi_i(s,v)^{-1}\mathcal{B}_i(v)\begin{pmatrix}\mathsf{a}_i(0,v)\\ \mathsf{q}_i(0,v)\end{pmatrix}dv.$$

For the uniform flow regime $\Psi_i(s,x)$ can be obtained analytically.

$$\Psi_i(s,x) = \begin{pmatrix} \frac{\lambda_{1,i}e^{\lambda_{1,i}x} - \lambda_{2,i}e^{\lambda_{2,i}x}}{\lambda_{1,i}-\lambda_{2,i}} & \frac{\lambda_{1,i}\lambda_{2,i}(e^{\lambda_{1,i}x} - e^{\lambda_{2,i}x})}{s(\lambda_{1,i}-\lambda_{2,i})} \\ \frac{s(e^{\lambda_{2,i}x} - e^{\lambda_{1,i}x})}{\lambda_{1,i}-\lambda_{2,i}} & \frac{\lambda_{1,i}e^{\lambda_{2,i}x} - \lambda_{2,i}e^{\lambda_{1,i}x}}{\lambda_{1,i}-\lambda_{2,i}} \end{pmatrix}.$$

For nonuniform flow regime, $\Psi_i(s,x)$ can be computed numerically using an efficient numerical scheme proposed in Litrico and Fromion [2004*b*]. Since we are interested in obtaining an analytically approximate model, we only consider uniform flow regime here, and note that our framework also extends to non-uniform regime via the numerical scheme which interconnects several (virtual) uniform flow pools Litrico and Fromion [2009*b*].

Using (3.19) at $x = l_i$ and assuming that $\psi_{21,i}(s,l_i) \neq 0$, we obtain[3]

$$\bar{\mathsf{T}}_i\hat{\mathsf{y}}_i(s,0) = -\frac{\psi_{22,i}(s,l_i)}{\psi_{21,i}(s,l_i)}\hat{\mathsf{q}}_i(s,0) + \frac{1}{\psi_{21,i}(s,l_i)}\hat{\mathsf{q}}_i(s,l_i). \tag{3.20}$$

Let $\mathcal{P}_i(s) = (p_{i,jk}(s))$ denote the input-output transfer matrix in the Laplace domain relating the inputs $\hat{\mathsf{q}}_{i-1}(s)$ and $(\hat{\mathsf{q}}_i(s) + \hat{\mathsf{p}}_i(s))$ to the outputs $\hat{\mathsf{y}}_i^u(s)$ and $\hat{\mathsf{y}}_i^d(s)$. Assuming zero initial conditions for simplicity, substituting (3.20) in (3.19), and using the notation (3.10), we obtain the following relation:

$$\begin{pmatrix} \hat{\mathsf{y}}_i(s,x) \\ \hat{\mathsf{q}}_i(s,x) \end{pmatrix} = \mathcal{G}_i(s) \begin{pmatrix} \hat{\mathsf{q}}_{i-1}(s) \\ \hat{\mathsf{q}}_i(s) + \hat{\mathsf{p}}_i(s) \end{pmatrix}, \tag{3.21}$$

$$\begin{pmatrix} \hat{\mathsf{y}}_i^u(s) \\ \hat{\mathsf{y}}_i^d(s) \end{pmatrix} = \mathcal{P}_i(s) \begin{pmatrix} \hat{\mathsf{q}}_{i-1}(s) \\ \hat{\mathsf{q}}_i(s) + \hat{\mathsf{p}}_i(s) \end{pmatrix}, \tag{3.22}$$

The transfer matrices $\mathcal{G}_i(s)$ and $\mathcal{P}_i(s)$ are given by

$$\mathcal{G}_i(s,x) = \begin{pmatrix} \frac{1}{\bar{\mathsf{T}}_i(x)}\left(\psi_{12,i}(s,x) - \psi_{11,i}(s,x)\frac{\psi_{22,i}(s,l_i)}{\psi_{21,i}(s,l_i)}\right) & \frac{1}{\bar{\mathsf{T}}_i(x)}\left(\frac{\psi_{11,i}(s,x)}{\psi_{21,i}(s,l_i)}\right) \\ \psi_{22,i}(s,x) - \psi_{21,i}(s,x)\frac{\psi_{22,i}(s,l_i)}{\psi_{21,i}(s,l_i)} & \frac{\psi_{21,i}(s,x)}{\psi_{21,i}(s,l_i)} \end{pmatrix},$$

and

$$\mathcal{P}_i(s) = \begin{pmatrix} \frac{1}{\bar{\mathsf{T}}_i(0)}\left(\psi_{12,i}(s,0) - \psi_{11,i}(s,0)\frac{\psi_{22,i}(s,l_i)}{\psi_{21,i}(s,l_i)}\right) & \frac{1}{\bar{\mathsf{T}}_i(0)}\frac{\psi_{11,i}(s,0)}{\psi_{21,i}(s,l_i)} \\ \frac{1}{\bar{\mathsf{T}}_i(l_i)}\left(\psi_{12,i}(s,l_i) - \psi_{11,i}(s,l_i)\frac{\psi_{22,i}(s,l_i)}{\psi_{21,i}(s,l_i)}\right) & \frac{1}{\bar{\mathsf{T}}_i(l_i)}\frac{\psi_{11,i}(s,l_i)}{\psi_{21,i}(s,l_i)} \end{pmatrix}.$$

In uniform flow regime, the matrix functions $\mathcal{A}_i(s,x)$ and $\mathcal{B}_i(x)$ in (3.18) become independent of $x$, and the state transition matrix $\Psi_i(s,x)$ in (3.19) can be expressed analytically

$$\Psi_i(s,x) = \begin{pmatrix} \frac{\lambda_{1,i}e^{\lambda_{1,i}x} - \lambda_{2,i}e^{\lambda_{2,i}x}}{\lambda_{1,i}-\lambda_{2,i}} & \frac{\lambda_{1,i}\lambda_{2,i}(e^{\lambda_{1,i}x} - e^{\lambda_{2,i}x})}{s(\lambda_{1,i}-\lambda_{2,i})} \\ \frac{s(e^{\lambda_{2,i}x} - e^{\lambda_{1,i}x})}{\lambda_{1,i}-\lambda_{2,i}} & \frac{\lambda_{1,i}e^{\lambda_{2,i}x} - \lambda_{2,i}e^{\lambda_{1,i}x}}{\lambda_{1,i}-\lambda_{2,i}} \end{pmatrix}, \tag{3.23}$$

[3] The poles of the system correspond to the values of $s$ such that $\psi_{21,i}(s,l_i) = 0$.

The distributed transfer matrix $\mathcal{G}_i(s)$ in (3.21), and the input-output transfer matrix $\mathcal{P}_i(s)$ in (3.22) are given by

$$\mathcal{G}_i(s) = \begin{pmatrix} \frac{\lambda_{2,i}e^{(\lambda_{1,i}-\lambda_{2,i})(l_i-x)}-\lambda_{1,i}}{\bar{\mathsf{T}}_i s(1-e^{(\lambda_{1,i}-\lambda_{2,i})l_i})} e^{\lambda_{1,i}x} & \frac{\lambda_{1,i}e^{(\lambda_{1,i}-\lambda_{2,i})(x)}-\lambda_{2,i}}{\bar{\mathsf{T}}_i s(1-e^{(\lambda_{1,i}-\lambda_{2,i})l_i})} e^{-\lambda_{2,i}(l_i-x)} \\ \frac{1-e^{(\lambda_{1,i}-\lambda_{2,i})(l_i-x)}}{1-e^{(\lambda_{1,i}-\lambda_{2,i})l_i}} e^{\lambda_{1,i}x} & \frac{1-e^{(\lambda_{1,i}-\lambda_{2,i})x}}{1-e^{(\lambda_{1,i}-\lambda_{2,i})l_i}} e^{-\lambda_{2,i}(l_i-x)} \end{pmatrix},$$

$$\mathcal{P}_i(s) = \begin{pmatrix} \frac{\lambda_{2,i}e^{(\lambda_{1,i}-\lambda_{2,i})l_i}-\lambda_{1,i}}{\bar{\mathsf{T}}_i s(1-e^{(\lambda_{1,i}-\lambda_{2,i})l_i})} & \frac{(\lambda_{1,i}-\lambda_{2,i})e^{-\lambda_{2,i}l_i}}{\bar{\mathsf{T}}_i s(1-e^{(\lambda_{1,i}-\lambda_{2,i})l_i})} \\ \frac{(\lambda_{2,i}-\lambda_{1,i})e^{\lambda_{1,i}l_i}}{\bar{\mathsf{T}}_i s(1-e^{(\lambda_{1,i}-\lambda_{2,i})l_i})} & \frac{\lambda_{1,i}e^{(\lambda_{1,i}-\lambda_{2,i})l_i}-\lambda_{2,i}}{\bar{\mathsf{T}}_i s(1-e^{(\lambda_{1,i}-\lambda_{2,i})l_i})} \end{pmatrix}.$$

Here the $\lambda_{1,i}(s)$ and $\lambda_{2,i}(s)$ (dependence on $s$ is omitted in the above for notational simplicity) are the eigenvalues of $\mathcal{A}_i(s)$.

*Remark* 3.3.1. The eigenvalues $\lambda_{1,i}(s)$ and $\lambda_{2,i}(s)$ are

$$\begin{aligned} \lambda_{1,i}(s) &= -r_{1,i} - \frac{s}{\alpha_i} + \theta_i \hat{F}_i(s) \\ \lambda_{2,i}(s) &= r_{2,i} + \frac{s}{\beta_i} - \theta_i \hat{F}_i(s), \end{aligned}$$

where

$$\hat{F}_i(s) = s + b_i - \sqrt{(s+b_i)^2 - a_i^2}, \quad a_i^2 = \frac{4\alpha_i\beta_i(\alpha_i\delta_i - \gamma_i)(\gamma_i + \beta_i\delta_i)}{(\alpha_i+\beta_i)^4}, \quad \theta_i = \frac{(\alpha_i+\beta_i)}{2\alpha_i\beta_i},$$

and

$$r_{1,i} = \frac{\alpha_i\delta_i - \gamma_i}{\alpha_i(\alpha_i+\beta_i)}, \quad r_{2,i} = \frac{\beta_i\delta_i + \gamma_i}{\beta_i(\alpha_i+\beta_i)}, \quad b_i = \frac{(\alpha_i-\beta_i)\gamma_i + 2\alpha_i\beta_i\delta_i}{(\alpha_i+\beta_i)^2}.$$

◁

The input-output transfer matrix $\mathcal{P}_i(s)$ can approximated to obtain the integrator-delay-zero (IDZ) model:

$$\begin{pmatrix} \hat{\mathsf{y}}_i^u(s) \\ \hat{\mathsf{y}}_i^d(s) \end{pmatrix} = \mathcal{Q}_i(s) \begin{pmatrix} \hat{\mathsf{q}}_{i-1}(s) \\ \hat{\mathsf{q}}_i(s) + \hat{\mathsf{p}}_i(s) \end{pmatrix}, \tag{3.24}$$

where

$$\mathcal{Q}_i(s) = \left( \underbrace{\begin{pmatrix} \frac{a_i^u}{s} & -\frac{a_i^u}{s} e^{-\bar{\tau}_i s} \\ \frac{a_i^d}{s} e^{-\underline{\tau}_i s} & -\frac{a_i^d}{s} \end{pmatrix}}_{\text{low freq. approx.}} + \underbrace{\begin{pmatrix} \bar{c}_i^u & -\tilde{c}_i^u e^{-\bar{\tau}_i s} \\ \tilde{c}_i^d e^{-\underline{\tau}_i s} & -\bar{c}_i^d \end{pmatrix}}_{\text{high freq. approx.}} \right).$$

The IDZ model is an analytically approximate model of the input-output relation (3.22). It accounts for low frequencies by an integrator and a delay, and high frequencies by a constant gain and a delay. The parameters $\bar{c}_i^u$, $\tilde{c}_i^u$, $\tilde{c}_i^d$, and $\bar{c}_i^d$ are obtained as the mean values of $|p_{i,jk}(s)|$ as $s \to \infty$:

$$
\begin{aligned}
\bar{c}_i^u &= \frac{1}{\bar{\mathsf{T}}_i \alpha_i} \sqrt{\frac{1 + \frac{\alpha_i^2}{\beta_i^2} e^{-2(r_{1,i}+r_{2,i})l_i}}{1 + e^{-2(r_{1,i}+r_{2,i})l_i}}}, \\
\bar{c}_i^d &= \frac{1}{\bar{\mathsf{T}}_i \beta_i} \sqrt{\frac{1 + \frac{\beta_i^2}{\alpha_i^2} e^{-2(r_{1,i}+r_{2,i})l_i}}{1 + e^{-2(r_{1,i}+r_{2,i})l_i}}}, \\
\tilde{c}_i^u &= \frac{\alpha_i + \beta_i}{\bar{\mathsf{T}}_i \alpha_i \beta_i} \frac{e^{-r_{2,i}l_i}}{\sqrt{1 + e^{-2(r_{1,i}+r_{2,i})l_i}}}, \\
\tilde{c}_i^d &= \frac{\alpha_i + \beta_i}{\bar{\mathsf{T}}_i \alpha_i \beta_i} \frac{e^{-r_{1,i}l_i}}{\sqrt{1 + e^{-2(r_{1,i}+r_{2,i})l_i}}}.
\end{aligned}
$$

Taking the inverse laplace transform of (3.24), we obtain the following time-domain model for pool $i$:

$$
\begin{aligned}
\dot{\mathsf{h}}_i^u(t) &= a_i^u \mathsf{q}_{i-1}(t) - a_i^u \left[\mathsf{q}_i(t - \bar{\tau}_i) + \mathsf{p}_i(t - \bar{\tau}_i)\right] \\
\dot{\mathsf{h}}_i^d(t) &= a_i^d \mathsf{q}_{i-1}(t - \underline{\tau}_i) - a_i^d \left[\mathsf{q}_i(t) + \mathsf{p}_i(t)\right] \\
\mathsf{y}_i^u(t) &= \mathsf{h}_i^u(t) + \bar{c}_i^u \mathsf{q}_{i-1}(t) - \tilde{c}_i^u \left[\mathsf{q}_i(t - \bar{\tau}_i) + \mathsf{p}_i(t - \bar{\tau}_i)\right] \\
\mathsf{y}_i^d(t) &= \mathsf{h}_i^d(t) + \tilde{c}_i^d \mathsf{q}_{i-1}(t - \underline{\tau}_i) - \bar{c}_i^d \left[\mathsf{q}_i(t) + \mathsf{p}_i(t)\right],
\end{aligned}
\tag{3.25}
$$

where $\mathsf{h}_i^d(t)$ and $\mathsf{h}_i^u(t)$ are intermediate variables initialized by $\mathsf{h}_i^d(t) = 0$ and $\mathsf{h}_i^u(t) = 0$.

Let us recall that, under our assumptions, the local slave controllers are responsible to deliver the required input discharge. In practice, only low frequency compensation can be achieved due to the bandwidth limitation imposed by the digital implementation of the slave controllers; the high-frequency control is achieved in a passive manner. This feature leads to recovery of only the low frequency part of (3.24). With this justification, we now consider the following low-frequency approximation (the integrator-delay (ID) model).

Combining (3.11) and (3.17), we obtain the delay-differential equation:

$$
\begin{aligned}
\dot{\mathsf{y}}_i^u(t) &= a_i^u \left[b_{i-1}^d \mathsf{y}_{i-1}^d(t) + b_i^u \mathsf{y}_i^u(t) + k_{i-1}\mathsf{u}_{i-1}(t)\right] \\
&\quad - a_i^u \left[b_i^d \mathsf{y}_i^d(t - \bar{\tau}_i) + b_{i+1}^u \mathsf{y}_{i+1}^u(t - \bar{\tau}_i) + k_i \mathsf{u}_i(t - \bar{\tau}_i) - \mathsf{p}_i(t - \bar{\tau}_i)\right] \\
\dot{\mathsf{y}}_i^d(t) &= a_i^d \left[b_{i-1}^d \mathsf{y}_{i-1}^d(t - \underline{\tau}_i) + b_i^u \mathsf{y}_i^u(t - \underline{\tau}_i) + k_{i-1}\mathsf{u}_{i-1}(t - \underline{\tau}_i)\right] \\
&\quad - a_i^d \left[b_i^d \mathsf{y}_i^d(t) + b_{i+1}^u \mathsf{y}_{i+1}^u(t) + k_i \mathsf{u}_i(t) + \mathsf{p}_i(t)\right].
\end{aligned}
\tag{3.26}
$$

We now consider the specific case of a two pools ($m = 2$) canal with three submerged hydraulic gates (see Fig. 3.2 and consider $i = 1$). For sake of simplicity, we will assume that the upstream level at gate 0 and downstream level at gate 2 are constant, i.e., $\mathsf{y}_0^d = 0$ and $\mathsf{y}_3^u = 0$, and moreover, the opening of gate 2 is fixed, i.e., $\mathsf{u}_2 = 0$. The full model for the

2-pool system can be written in state-space form as follows

$$\begin{aligned}\dot{\mathsf{x}}(t) &= \sum_{i=0}^{4} A_i \mathsf{x}(t-\tau_i) + \sum_{i=0}^{4} B_i \mathsf{u}(t-\tau_i) \\ \mathsf{y}(t) &= C\mathsf{x}(t),\end{aligned} \tag{3.27}$$

where $\mathsf{x} := \begin{pmatrix}\mathsf{y}_1^u, & \mathsf{y}_2^u, & \mathsf{y}_1^d, & \mathsf{y}_2^d\end{pmatrix}^{\mathsf{T}} \in \mathbb{R}^4$ is the state, $\mathsf{u} := \begin{pmatrix}\mathsf{u}_0, & \mathsf{u}_1, & \mathsf{p}_1, & \mathsf{p}_2\end{pmatrix}^{\mathsf{T}} \in \mathbb{R}^4$ denotes the known input, $\mathsf{y} := \begin{pmatrix}\mathsf{y}_1^u, & \mathsf{y}_2^u, & \mathsf{y}_1^d, & \mathsf{y}_2^d\end{pmatrix}^{\mathsf{T}} \in \mathbb{R}^4$ is the measured output; $\tau_0 = 0$, $\tau_1 = \bar{\tau}_1$, $\tau_2 = \underline{\tau}_1$, $\tau_3 = \bar{\tau}_2$, $\tau_4 = \underline{\tau}_2$. The matrices $C$, $A_i$, $B_i$ are known matrices in $\mathbb{R}^{4\times 4}$ which are respectively given by $C = \operatorname{diag}\big(1, 1, 1, 1\big)$, and

$$A_0 = \begin{pmatrix} a_1^u b_1^u & 0 & 0 & 0 \\ 0 & a_2^u b_2^u & a_2^u b_1^d & 0 \\ 0 & -a_1^d b_2^u & -a_1^d b_1^d & 0 \\ 0 & 0 & 0 & -a_2^d b_2^d \end{pmatrix}, \quad B_0 = \begin{pmatrix} a_1^u k_0 & 0 & 0 & 0 \\ 0 & a_2^u k_1 & 0 & 0 \\ 0 & -a_1^d k_1 & -a_1^d & 0 \\ 0 & 0 & 0 & -a_2^d \end{pmatrix},$$

$$A_1 = \begin{pmatrix} 0 & -a_1^u b_2^u & -a_1^u b_1^d & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 0 & -a_1^u k_1 & -a_1^u & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ a_1^d b_1^u & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ a_1^d k_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$A_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -a_2^u b_2^d \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -a_2^u \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$A_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & a_2^d b_2^u & a_2^d b_1^d & 0 \end{pmatrix}, \quad B_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & a_2^d k_1 & 0 & 0 \end{pmatrix}.$$

Consider the case of unmeasured water withdrawals (denoted $\delta\mathsf{p}_i(t)$) occurring through the offtakes, located at the downstream ends (see Fig. 3.2). Model (3.27) now becomes

$$\begin{aligned}\dot{\mathsf{x}}(t) &= \sum_{i=0}^{4} A_i \mathsf{x}(t-\tau_i) + \sum_{i=0}^{4} B_i \mathsf{u}(t-\tau_i) + \sum_{i=1}^{2} E_i \mathsf{f}_i(t) \\ \mathsf{y}(t) &= C\mathsf{x}(t),\end{aligned} \tag{3.28}$$

where

$$\begin{aligned} &\mathsf{f}_i(t) = \begin{pmatrix}\delta\mathsf{p}_i(t) & \delta\tilde{\mathsf{p}}_i(t)\end{pmatrix}^{\mathsf{T}}, \quad i = 1, 2 \\ &E_1 = \begin{pmatrix} 0 & -a_1^u & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -a_1^d & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -a_2^u & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -a_2^d & 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned} \tag{3.29}$$

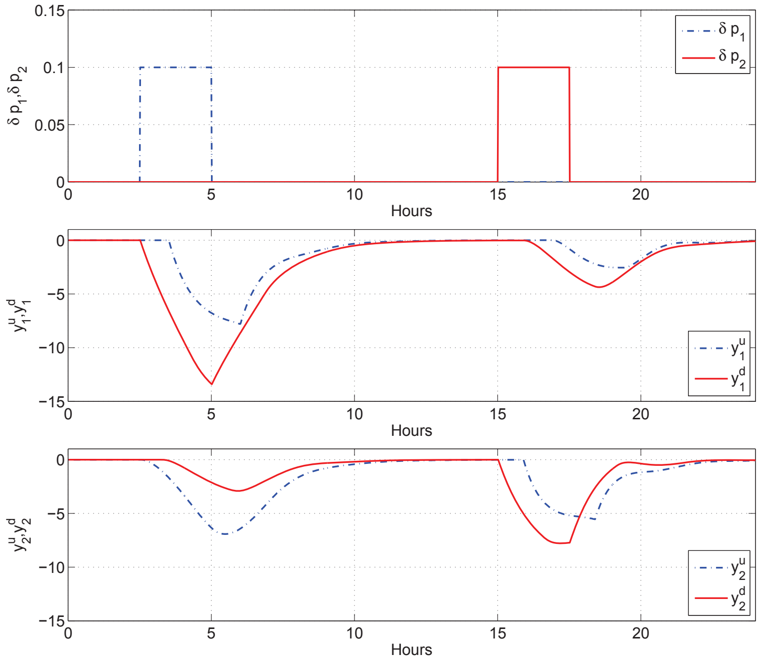


Figure 3.4: Withdrawals (top), Pool 1 (middle) and Pool 2 (bottom) levels.

with $\delta\tilde{\mathsf{p}}_i(t) := \big(\delta\mathsf{p}_i(t-\tau_1) \quad \ldots \quad \delta\mathsf{p}_i(t-\tau_4)\big)$.

We will consider the following numerical example of a 2−pool system throughout the chapter:

*Example* 3.3.2. (Two Pool System) Consider (3.28) with following parameters: upstream (resp. downstream) propagation delays $\bar{\tau}_1 = 846.5$ s, $\bar{\tau}_2 = 750.5$ s (resp. $\underline{\tau}_1 = 707.5$ s, $\underline{\tau}_2 = 647.5$ s), equivalent inverse backwater areas for upstream (resp. downstream) water levels $a_1^u = 3.975 \times 10^{-5}$ m$^{-2}$, $a_2^u = 3.675 \times 10^{-5}$ m$^{-2}$ (resp. $a_1^d = 3.21 \times 10^{-5}$ m$^{-2}$, $a_2^d = 3.115 \times 10^{-5}$ m$^{-2}$) . Let the coefficients of linearized gate equations $b_1^d = 20.0, b_2^d = 29.0,\ b_1^u = -21.36, b_2^u = -25.36,\ k_0 = 18.1, k_2 = 12.1$. Assume that $\mathsf{u}(t) = 0$ for $t \in [-\underline{\tau}_1, \infty)$ and $\mathsf{x}(t) = 0$ for $t \in [-\underline{\tau}_1, 0]$. Water at the rate 0.1 m$^3$/s is withdrawn from offtake of pool 1 (resp. pool 2) during the interval 2.5 − 5.0 hr (resp. 15 − 17.5 hr). Fig. 3.4 shows the upstream and downstream water level deviations (cm) under the the effect of unmeasured withdrawals during a 24 hr simulation. △

## 3.4 UIO Based Fault Detection and Isolation

In this section we present the design of unknown input observers (UIO) for linear time delay systems when unknown inputs are present in *both* state and measurement equations. A bank of UIO observers so designed are then used for detection and isolation under coupled disturbance/fault signals.

### 3.4.1 Unknown Input Observer Design

Consider the following linear, time-invariant, delay differential system (DDS) with unknown inputs

$$\begin{aligned}\dot{\mathsf{x}}(t) &= \sum_{i=0}^{r} A_i \mathsf{x}(t-\tau_i(t)) + \sum_{i=1}^{r} B_i \mathsf{u}(t-\tau_i(t)) + E\mathsf{f}(t)\\ \mathsf{x}(\theta) &= \rho_1(\theta), \mathsf{u}(\theta) = \rho_2(\theta), \quad \theta \in [-\tau_{\max}, 0]\\ \mathsf{y}(t) &= C\mathsf{x}(t) + H\mathsf{f}(t),\end{aligned} \tag{3.30}$$

where $\mathsf{x}(t) \in \mathbb{R}^n$ is the state vector, $\mathsf{u}(t) \in \mathbb{R}^m$ is the known input vector, $\mathsf{f} \in \mathbb{R}^q$ the unknown input vector, $\mathsf{y} \in \mathbb{R}^p$ the measurement output vector, and $\rho_1 \in \mathbb{R}^n$ and $\rho_2 \in \mathbb{R}^m$ are continuous initial vector functions for the state and input. The matrices $A_i$, $B_i$, $C$, and $E$ are known real matrices of appropriate dimensions. The matrix $E$ (resp. $H$) is called the disturbance distribution matrix for state (resp. observation) equation, and $H\mathsf{f}(t)$ (resp. $E\mathsf{f}(t)$) determines the unknown sensor disturbance (resp. unknown input uncertainty). The time delays $\tau_i(t)$ are bounded but possibly time varying, and satisfy[4]

$$\begin{aligned}&\tau_i(t) \leqslant h_i, \quad \dot{\tau}_i(t) \leqslant d_i < 1, \quad i = 1, \ldots, r,\\ &\tau_{\max} := \max\{h_1, \ldots, h_r\}\end{aligned} \tag{3.31}$$

where $h_i$ and $d_i$ are known constants.

Consider the following full-order observer for system (3.30):

$$\begin{aligned}\dot{\mathsf{z}}(t) &= \sum_{i=0}^{r} F_i \mathsf{z}(t-\tau_i) + \sum_{i=0}^{r} TB_i u(t-\tau_i) + \sum_{i=0}^{r} G_i \mathsf{y}(t-\tau_i)\\ \mathsf{z}(\theta) &= \rho_3(\theta), \quad \theta \in [-\tau_{\max}, 0]\\ \hat{\mathsf{x}}(t) &= \mathsf{z}(t) + N\mathsf{y}(t),\end{aligned} \tag{3.32}$$

where $\mathsf{z}(t) \in \mathbb{R}^n$ is the observer state vector, $\rho_3 \in \mathbb{R}^n$ the initial vector function, and $\hat{\mathsf{x}}(t)$ the estimate of $\mathsf{x}(t)$. The matrices $F_i$, $G_i$, $T$ and $N$ are constant matrices of appropriate dimensions which must be determined such that $\hat{\mathsf{x}}(t)$ asymptotically converges to $\mathsf{x}(t)$, regardless of the presence of unknown inputs $\mathsf{f}(t)$. Such an observer, if it exists, achieves perfect decoupling from unknown inputs. We define the error between $\mathsf{x}(t)$ and its estimate $\hat{\mathsf{x}}(t)$ as

$$\mathsf{e}(t) = \hat{\mathsf{x}}(t) - \mathsf{x}(t) = \mathsf{z}(t) - T\mathsf{x}(t) + NH\mathsf{f}(t),$$

where $T = \mathsf{I}_n - NC$. The error dynamics is given by

$$\begin{aligned}\dot{\mathsf{e}}(t) = &\sum_{i=0}^{r} F_i \mathsf{e}(t-\tau_i) + (F_i - TA_i + (G_i - F_i N)C)\,\mathsf{x}(t-\tau_i)\\ &- (TE + F_0 NH - G_0 H)\,\mathsf{f}(t) - \sum_{i=1}^{r} (F_i N - G_i)\, H\mathsf{f}(t-\tau_i) + NH\dot{\mathsf{f}}(t)\end{aligned} \tag{3.33}$$

Then it is straightforward to obtain the following result

[4] For e.g., time varying delays in automated canal systems can result via a communication network which transmits the sensor-control data packets.

**Theorem 3.4.1.** *The full order observer* (3.32) *will asymptotically estimate* $\mathsf{x}(t)$ *if the following conditions hold*

1. $\dot{\mathsf{e}}(t) = \sum_{i=0}^{r} F_i \mathsf{e}(t - \tau_i)$ *is asymptotically stable,*
2. $\mathsf{I}_n = T + NC$,
3. $\bar{G}_i = G_i - F_i N, \quad i = 0, \dots, r$,
4. $F_i = TA_i - \bar{G}_i C, \quad i = 0, \dots, r$,
5. $\bar{G}_0 H = TE$,
6. $\bar{G}_i H = 0, \quad i = 1, \dots, r$,
7. $NH = 0$.

Thus, the observer design problem is reduced to finding the matrices $T, N$, and $F_i, \bar{G}_i$, $i = 0, \dots, r$ such that the conditions in Theorem 3.4.1 are satisfied. For $r = 4$ (for e.g., this is the case for 2−pool system), the conditions (2)–(7) in Theorem 3.4.1 can be written as follows:

$$S\Theta = \Psi, \tag{3.34}$$

where

$$\begin{aligned} S &= \begin{pmatrix} T & N & F_0 & \bar{G}_0 & \dots & F_4 & \bar{G}_4 \end{pmatrix} \in \mathbb{R}^{n \times (6n+6p)}, \\ \Theta &= \begin{pmatrix} \Theta_1 & \Theta_2 & \Theta_3 \end{pmatrix} \in \mathbb{R}^{(6n+6p) \times (6n+6q)}, \\ \Psi &= \begin{pmatrix} \mathsf{I}_n & 0 \end{pmatrix} \in \mathbb{R}^{n \times (6n+6q)}, \end{aligned}$$

and

$$\Theta_1 = \begin{pmatrix} \mathsf{I}_n & E \\ C & 0 \\ 0 & 0 \\ 0 & H \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \Theta_2 = \begin{pmatrix} A_0 & A_1 & A_2 & A_3 & A_4 \\ 0 & 0 & 0 & 0 & 0 \\ -\mathsf{I}_n & 0 & 0 & 0 & 0 \\ -C & 0 & 0 & 0 & 0 \\ 0 & -\mathsf{I}_n & 0 & 0 & 0 \\ 0 & -C & 0 & 0 & 0 \\ 0 & 0 & -\mathsf{I}_n & 0 & 0 \\ 0 & 0 & -C & 0 & 0 \\ 0 & 0 & 0 & -\mathsf{I}_n & 0 \\ 0 & 0 & 0 & -C & 0 \\ 0 & 0 & 0 & 0 & -\mathsf{I}_n \\ 0 & 0 & 0 & 0 & -C \end{pmatrix}, \quad \Theta_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ H & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & H & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & H & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & H & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & H \end{pmatrix}.$$

Following the general solution of a set of linear matrix equations (see for e.g., Darouach et al. [1994]), there exists a solution to (3.34) if and only if:

$$\operatorname{rank} \begin{pmatrix} \Theta \\ \Psi \end{pmatrix} = \operatorname{rank} \left( \Theta \right),$$

or equivalently,

$$\operatorname{rank}\begin{pmatrix} CE \\ H \end{pmatrix} = \operatorname{rank}\begin{pmatrix} E \\ H \end{pmatrix}. \tag{3.35}$$

Under the above rank condition, the general solution of (3.34) is

$$S = \Psi\Theta^{+} - K(I - \Theta\Theta^{+}), \tag{3.36}$$

where $K$ is an arbitrary matrix of appropriate dimension, and $\Theta^{+}$ is the generalized inverse matrix of $\Theta$. The choice of matrix $K$ is important in determining the asymptotic stability of the observer. This can be seen by inserting the solution (3.36) in condition (4) of theorem 3.4.1. The matrices $F_i$ can now be expressed as

$$F_i = \chi_i - K\beta_i, \quad i = 0, 1, \ldots, 4, \tag{3.37}$$

where

$$\begin{aligned}
\chi_0 &= \Psi\Theta^{+}\begin{pmatrix} A_0 & 0 & 0 & -C & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^{\mathsf{T}} \\
\chi_1 &= \Psi\Theta^{+}\begin{pmatrix} A_0 & 0 & 0 & 0 & 0 & -C & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^{\mathsf{T}} \\
\chi_2 &= \Psi\Theta^{+}\begin{pmatrix} A_0 & 0 & 0 & 0 & 0 & 0 & 0 & -C & 0 & 0 & 0 & 0 \end{pmatrix}^{\mathsf{T}} \\
\chi_3 &= \Psi\Theta^{+}\begin{pmatrix} A_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -C & 0 & 0 \end{pmatrix}^{\mathsf{T}} \\
\chi_4 &= \Psi\Theta^{+}\begin{pmatrix} A_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -C \end{pmatrix}^{\mathsf{T}} \\
\beta_0 &= \tilde{\Theta}\begin{pmatrix} A_0 & 0 & 0 & -C & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^{\mathsf{T}} \\
\beta_1 &= \tilde{\Theta}\begin{pmatrix} A_0 & 0 & 0 & 0 & 0 & -C & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^{\mathsf{T}} \\
\beta_2 &= \tilde{\Theta}\begin{pmatrix} A_0 & 0 & 0 & 0 & 0 & 0 & 0 & -C & 0 & 0 & 0 & 0 \end{pmatrix}^{\mathsf{T}} \\
\beta_3 &= \tilde{\Theta}\begin{pmatrix} A_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -C & 0 & 0 \end{pmatrix}^{\mathsf{T}} \\
\beta_4 &= \tilde{\Theta}\begin{pmatrix} A_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -C \end{pmatrix}^{\mathsf{T}}
\end{aligned}$$

with $\tilde{\Theta} := (I - \Theta\Theta^{+})$. Under condition (3.35), and from above results, the error dynamics (3.33) for $r = 4$ can be written as

$$\dot{\mathsf{e}}(t) = \sum_{i=0}^{4} (\chi_i - K\beta_i)\mathsf{e}(t - \tau_i(t)). \tag{3.38}$$

Thus the problem of observer (3.32) design reduces to the determination of the matrix parameter $K$ such that the stability condition (1) of theorem 3.4.1 holds. We now give *delay-dependent* conditions for the stability of the observer under the delay bounds (3.31). By extension, similar conditions can be determined for any $r$.

**Proposition 3.4.2.** *Suppose that condition* (3.35) *is satisfied, and let* $r = 4$. *Then there exists an asymptotically stable unknown input observer* (3.32), *if for some scalars* $\epsilon_0, \ldots, \epsilon_9$

*and $\bar{\epsilon}_1, \dots, \bar{\epsilon}_4$, there exist matrices $S_i > 0$, $Z_i > 0$, $Q_i > 0$, $R_i > 0$, $U_i$, $W_i$, i=1,…,4, and matrices $H_i$, $i = 0, \dots, 9$, $U$ and $P > 0$ such that the following linear matrix inequalities (LMIs) are satisfied:*

$$\begin{pmatrix} Q_i & U_i \\ U_i^{\mathsf{T}} & R_i \end{pmatrix} \geqslant 0, \quad i = 1, \dots, 4, \tag{3.39}$$

$$\begin{pmatrix} \Phi & h_1\bar{H}_1 & h_2\bar{H}_2 & h_3\bar{H}_3 & h_4\bar{H}_4 \\ * & -h_1\bar{Z}_1 & 0 & 0 & 0 \\ * & * & -h_2\bar{Z}_2 & 0 & 0 \\ * & * & * & -h_3\bar{Z}_3 & 0 \\ * & * & * & * & -h_4\bar{Z}_4 \end{pmatrix} < 0, \tag{3.40}$$

*where*

$$\bar{Z}_i := \begin{pmatrix} S_i & W_i \\ W_i^{\mathsf{T}} & Z_i \end{pmatrix}, \quad \bar{H}_i := \begin{pmatrix} -\bar{\epsilon}_i(P\chi_0 - U\beta_0)^{\mathsf{T}} & H_0 \\ -\bar{\epsilon}_i(P\chi_1 - U\beta_1)^{\mathsf{T}} & H_1 \\ -\bar{\epsilon}_i(P\chi_2 - U\beta_2)^{\mathsf{T}} & H_2 \\ -\bar{\epsilon}_i(P\chi_3 - U\beta_3)^{\mathsf{T}} & H_3 \\ -\bar{\epsilon}_i(P\chi_4 - U\beta_4)^{\mathsf{T}} & H_4 \\ \bar{\epsilon}_i P & H_5 \\ 0 & H_6 \\ 0 & H_7 \\ 0 & H_8 \\ 0 & H_9 \end{pmatrix}, \tag{3.41}$$

*for $i = 1, \dots, 4$, and $\Phi = (\phi_{jk})$ is a symmetric matrix of the form (3.56) with block elements $\phi_{jk}$ presented in (3.57) (see Appendix 3.6). The parameter matrix $K$ is given by $K = P^{-1}U$.*

The proof is presented in the Appendix 3.6. We now present our FDI scheme for delay-differential system of the form (3.30) which uses the LMI method of Proposition 3.4.2.

### 3.4.2 Residual Generation

Consider $j-$th DDS, $j = 1, \dots, s$, with $s$ candidate fault signals:

$$\begin{aligned} \dot{\mathsf{x}}_j(t) &= \sum_{i=0}^{r} A_i \mathsf{x}_j(t - \tau_i) + \sum_{i=1}^{r} B_i \mathsf{u}_j(t - \tau_i) + \sum_{i=1}^{s} E_i \mathsf{f}_i(t) \\ \mathsf{y}_j(t) &= C\mathsf{x}_j(t) + \sum_{i=1}^{s} H_i \mathsf{f}_i(t). \end{aligned} \tag{3.42}$$

The FDI scheme we consider here is required to detect the occurrence as well as isolate an unknown signal $\mathsf{f}_j(t)$ from other unknown signals $\mathsf{f}_k(t)$ $k \neq j$. Each unknown signal models a coupled disturbance/fault in the state and measurement equations. Following Conte and Perdon [2006], we consider the problem of residual generation according to following definition:

**Definition 3.4.3.** (Residual Generation Problem) The problem consists of finding residuals $\mathsf{r}_j(t)$ defined as follows:

$$\mathsf{r}_j(t) := \mathsf{y}_j(t) - C\hat{\mathsf{x}}_j(t), \quad j = 1, \ldots, s, \tag{3.43}$$

where $\hat{\mathsf{x}}_j(t)$ is the output of the $j-$th UIO of the form (3.32), and $\mathsf{y}_j(t)$ is the output of system (3.42), with the following properties:

1. $\mathsf{r}_j(t)$ is insensitive (i.e., robust) to $\mathsf{f}_j(t)$,
2. $\mathsf{r}_j(t)$ converges to zero asymptotically if $\mathsf{f}_k(t) = 0, k \neq j$ for every $t$,
3. $\exists p \geqslant 0$ such that $\frac{d}{d\mathsf{f}_k}\left(\frac{d^p \mathsf{r}_j(t)}{dt^p}\right) \neq 0$ for $k \neq j$.

If the residuals $\mathsf{r}_i(t)$ $i = 1, \ldots, s$ satisfy the properties of Definition 3.4.3, fault diagnosis can be successfully achieved based on the following decision rule:

$$\mathsf{f}_j(t) \neq 0 \text{ if } \|\mathsf{r}_j(t)\| \approx 0, \text{ and } \|\mathsf{r}_k(t)\| \neq 0, k \neq j. \tag{3.44}$$

We now discuss the FDI scheme for non-simultaneous withdrawals for the 2-pool system.

*Example* 3.4.4. (FDI Scheme for Unknown Withdrawals) System (3.42) models a $2-$pool system with $r = 4$, $s = 2$. Assume $E_1$ and $E_2$ are of the form (3.29), $H_1 = H_2 = 0$, all other parameters as in Example 3.3.2, and zero known input signal $\mathsf{u}(t) = 0$ (the system evolves in open-loop). Let the unknown withdrawal from pool 1 (resp. pool 2) during the interval $2.5 - 5.0$ hr (resp. $15 - 17.5$ hr) be the fault signal $\mathsf{f}_1(t)$ (resp. $\mathsf{f}_2(t)$). Assume the bounds of the time delays $\tau_i(t)$ to be 1.1 times their nominal values, for e.g., $h_1 = 1.1 \times \bar{\tau}_1$, and so on; and the time derivatives of the delays all less than 0.1, i.e., $d_i < 0.1$. Two observers are designed as follows:

Observer 1 (resp. observer 2) is designed to be insensitive to $\mathsf{f}_1(t)$ (resp. $\mathsf{f}_2(t)$). Residual $\mathsf{r}_j(t)$ $j = 1, 2$ of the $j-$th observer is defined by (3.43), and $\hat{\mathsf{x}}_j(t)$ is the output of $j-$th UIO designed for the following model:

$$\begin{aligned} \dot{\mathsf{x}}_j(t) &= \sum_{i=0}^{4} A_i \mathsf{x}_j(t-\tau_i) + \sum_{i=0}^{4} B_i \mathsf{u}_j(t-\tau_i) + E_j \mathsf{f}_j(t) + E_{-j}\mathsf{f}_{-j}(t) \\ \mathsf{y}_j(t) &= C\mathsf{x}_j(t). \end{aligned} \tag{3.45}$$

where $-j := (3-j)$. In (3.45) $\mathsf{f}_2(t) = 0$ (resp. $\mathsf{f}_1(t) = 0$) for observer 1 (resp. observer 2). The LMI conditions in Proposition 3.4.2 are feasible for $\epsilon_0 = 10$, $\epsilon_1 = \cdots = \epsilon_9 = -1$, and $\bar{\epsilon}_1 = \cdots = \bar{\epsilon}_4 = -1$, and the parameter matrices $F_{ij}$, $G_{ij}$, $T_j$ and $N_j$ $(i = 0, \ldots, 4)$ are obtained for the observers:

$$\begin{aligned} \dot{\mathsf{z}}_j(t) &= \sum_{i=0}^{4} F_{ij}\mathsf{z}_j(t-\tau_i) + \sum_{i=0}^{4} T_j B_i u_j(t-\tau_i) + \sum_{i=0}^{4} G_{ij}\mathsf{y}_j(t-\tau_i) \\ \hat{\mathsf{x}}_j(t) &= \mathsf{z}_j(t) + N_j \mathsf{y}_j(t). \end{aligned}$$

From the computed observer matrices $T_1$ and $T_2$ we obtain:

$$T_1E_1 = 10^{-15} \times \begin{pmatrix} 0.040 & 0.041 & 0 & 0 & 0 \\ -0.286 & -0.054 & 0 & 0 & 0 \\ 0.241 & 0.010 & 0 & 0 & 0 \\ -0.388 & -0.330 & 0 & 0 & 0 \end{pmatrix} \approx 0,$$

$$T_1E_2 = \begin{pmatrix} -0.000 & 0 & 0 & -0.000 & 0 \\ 0.288 & 0 & 0 & 0.149 & 0 \\ -0.383 & 0 & 0 & -0.021 & 0 \\ 0.044 & 0 & 0 & 0.289 & 0 \end{pmatrix} \neq 0,$$

$$T_2E_1 = \begin{pmatrix} 0.523 & -0.106 & 0 & 0 & 0 \\ -0.077 & 0.074 & 0 & 0 & 0 \\ -0.026 & 0.479 & 0 & 0 & 0 \\ 0.000 & 0.000 & 0 & 0 & 0 \end{pmatrix} \neq 0,$$

$$T_2E_2 = 10^{-14} \times \begin{pmatrix} -0.014 & 0 & 0 & -0.007 & 0 \\ 0.008 & 0 & 0 & -0.006 & 0 \\ 0.002 & 0 & 0 & -0.001 & 0 \\ 0.150 & 0 & 0 & -0.227 & 0 \end{pmatrix} \approx 0.$$

$\Delta$

We can check that the residuals $\mathsf{r}_j(t)$ $j = 1, 2$ in Example 3.4.4 satisfy the properties of Definition 3.4.3:

- $\mathsf{r}_1(t)$ (resp. $\mathsf{r}_2(t)$) is insensitive to $\mathsf{f}_1(t)$ ($\mathsf{f}_2(t)$) (follows from UIO property of observers 1 and 2),
- The residual dynamics defined by

$$\dot{\mathsf{r}}_j(t) = C\left(\sum_{i=0}^{4} F_{ij}\mathsf{e}_j(t-\tau_i)\right),$$

  converges to zero asymptotically when $\mathsf{f}_{-j}(t) = 0$ for every $t$ because the conditions of Theorem 3.4.1 are satisfied (e.g., $T_1E_1 = T_2E_2 = 0$),
- $\frac{d}{df_{-j}}\frac{d\mathsf{r}_j(t)}{dt} = T_jE_{-j} \neq 0.$

Hence, the FDI scheme for the above example can be achieved using the decision rule presented in Table 3.1. From Fig. 3.5 we can observe that the generated residuals successfully achieve FDI for 2−pool system.

## 3.5 Attack Detection and Isolation

In this section, we study the performance of the FDI scheme designed in Section 3.4 on a generalized fault/attack model. This model allows the modeling of many adversarial

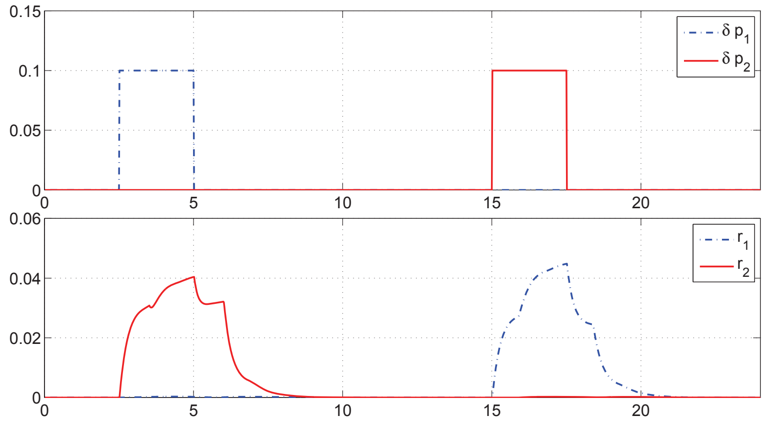


Figure 3.5: Fault signals $\delta \mathsf{p}_1$ and $\delta \mathsf{p}_2$ (top), and norms of residuals $\mathsf{r}_1$ and $\mathsf{r}_2$ (bottom).

Table 3.1: Decision table for $2-$pool system.

| If | $\Vert\mathsf{r}_1\Vert$ | $\Vert\mathsf{r}_2\Vert$ |
|---|---|---|
| $\mathsf{f}_1 \neq 0$ | $\approx 0$ | $\neq 0$ |
| $\mathsf{f}_2 \neq 0$ | $\neq 0$ | $\approx 0$ |

scenarios in which, differently from faults, the failure signals in the state and measurement equations are uncoupled. For the sake of simplicity, we will only consider the $2-$pool system, noting that similar analysis can be performed for multi-pool systems.

### 3.5.1 Generalized Fault/Attack Model for Two Pool System

Consider the DDS when fault/disturbances signals in the input and sensor measurements appear in uncoupled forms:

$$\Sigma_a = \begin{cases} \dot{\mathsf{x}}(t) & = \sum_{i=0}^{4} A_i \mathsf{x}(t-\tau_i) + \sum_{i=0}^{4} B_i \mathsf{u}(t-\tau_i) + \sum_{i=0}^{s} E_i \mathsf{f}_i(t) \\ \mathsf{y}(t) & = C\mathsf{x}(t) + \sum_{i=0}^{s} H_i \mathsf{g}_i(t), \end{cases} \tag{3.46}$$

where, $\mathsf{f}_i(t)$ and $\mathsf{g}_i(t)$ with $i = 1, \ldots, s$ are fault/disturbance signals affecting the state and measurement equations. Notice that this is in contrast to (3.42) where these signals are linearly coupled. We now show that the model (3.46) can represent traditional faults such as non-simultaneous discharge withdrawals (leaks) or sensor-actuator faults, and many adversarial scenarios when these disturbances can be manifested simultaneously.

#### Leaks and Sensor-Actuator Faults

Unmeasured withdrawals or leaks (denoted $\delta \mathsf{p}_i(t)$) may be caused by random faults or deliberate tampering of offtakes Bedjaoui and Weyer [2011]. For system (3.46), such discharge withdrawals can be modeled by considering $s = 2$, $H_1 = 0$, $H_2 = 0$, and $E_1$ and $E_2$ given by (3.29) (see Example 3.3.2). Similarly, we can model the actuator fault (denoted

$\delta \mathsf{u}_i(t)$) caused due to blockage of hydraulic structures or intentional manipulation of control actions. Consider, for example, $H_1 = 0$, and $H_2 = 0$, and

$$\mathsf{f}_i(t) = \begin{pmatrix} \delta \mathsf{u}_i(t) & \delta \tilde{\mathsf{u}}_i(t) \end{pmatrix}^\intercal ,$$
$$E_1 = \begin{pmatrix} a_1^u k_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_1^d k_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 & -a_1^u k_1 & 0 & 0 & 0 \\ a_2^u k_1 & 0 & 0 & 0 & 0 \\ -a_1^d k_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_2^d k_1 \end{pmatrix},$$

with $\delta \tilde{\mathsf{u}}_i(t) \coloneqq \begin{pmatrix} \delta \mathsf{u}_i(t-\tau_1) & \dots & \delta \mathsf{u}_i(t-\tau_4) \end{pmatrix}$. The sensor signals $\mathsf{y}_i^u(t)$ and $\mathsf{y}_i^d(t)$ may be subjected to random faults Choy and Weyer [2008] (e.g., effect of temperature variations in pressure sensors, malfunction of electronic circuitry in ultrasonic sensors), or b) adversarial biases which distort the true sensor signals (e.g., false-data injection attack Amin et al. [2010]). Sensor failures (denoted $\delta \mathsf{y}_i(t)$) in system (3.46) can be modeled by considering $s = 2$, $E_1 = 0$, $E_2 = 0$, and

$$\mathsf{g}_i(t) = \begin{pmatrix} \delta \mathsf{y}_i^u(t) & \delta \mathsf{y}_i^d(t) \end{pmatrix}^\intercal , i = 1, 2$$
$$H_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.47}$$

In many situations, faults/disturbance signals can appear in both measurement and state evolution equations in a linearly coupled manner, i.e., $\mathsf{f}_i(t) = \mathsf{g}_i(t)$ and the system (3.46) takes the same form as (3.30). For example, when a level sensor measurement is subjected to an additive bias and is injected in the system via output feedback control, the same bias will enter in the state equation as well.

Finally, note that the scheme proposed in Section 3.4 can be extended to achieve detection and isolation of faults in all the above mentioned scenarios under the assumption of non-simultaneous faults (i.e., if $\mathsf{f}_i(t) \neq 0$, then $\mathsf{f}_j(t) = 0$ where $j \neq i$).

### Simultaneous and Uncoupled Attacks

In many adversarial scenarios, the faults or disturbances on inputs and measurements can enter in an uncoupled manner (i.e., $\mathsf{f}_i(t) \neq \mathsf{g}_i(t)$ in (3.46)). Moreover, they can manifest simultaneously. Consider an adversarial scenario for system (3.46) when a deception attack simultaneously causes distortion of true sensor signals and unknown water withdrawal from the offtake. This scenario can be modeled with $\mathsf{f}_i(t)$, $E_1$ and $E_2$ (resp. $\mathsf{g}_i(t)$, $H_1$ and $H_2$) given by (3.29) (resp. (3.47)). This attack was the main focus of Chapter 2, where it was shown that a deception attack on sensor signals prevented correct isolation of unknown withdrawals through offtakes.

In general, without assuming any prior knowledge of attack signals, the FDI scheme of Section 3.4 cannot be extended to such adversarial scenarios. In the following example, we evaluate the performance of this scheme on different adversarial scenarios.

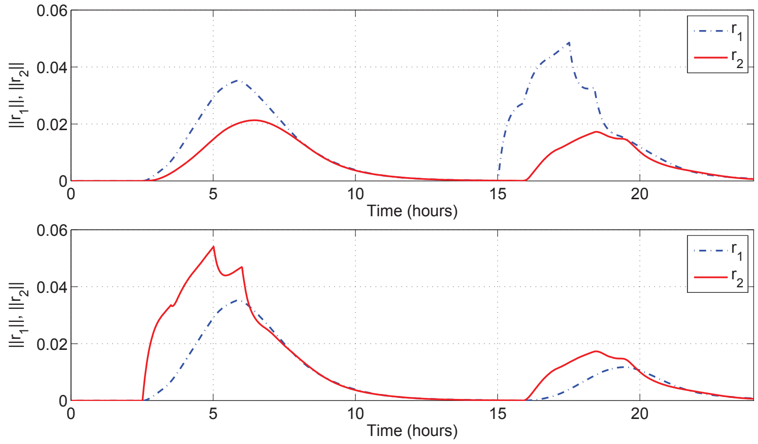


Figure 3.6: Residuals under attack on $y_1^u$, $y_1^d$ (top), and $y_2^u$, $y_2^d$ (bottom).

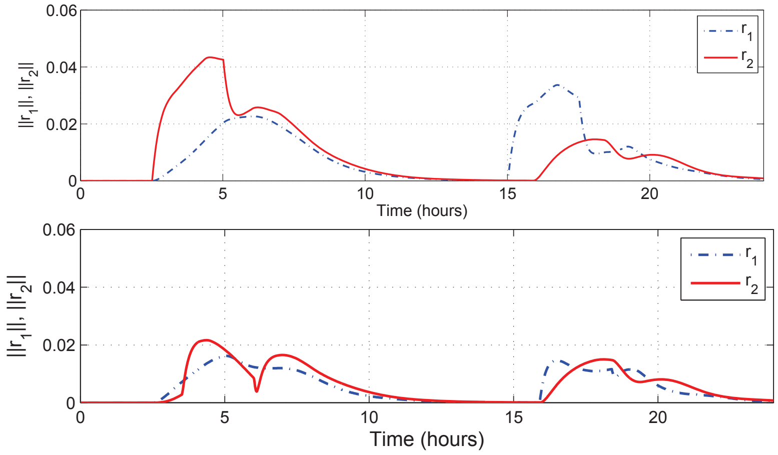


Figure 3.7: Residuals under attack on $y_1^u$, $y_2^u$ (top), and $y_1^d$, $y_2^d$ (bottom).

*Example* 3.5.1. Consider the FDI scheme designed in Example 3.4.4 which generated correct residuals to detect and isolate non-simultaneous withdrawals for 2−pool system. To evaluate the performance of this scheme when the true sensor measurements are *spoofed* with an additive deception attack, we consider four cases: 1) For each pool $i$, $y_i^u$ and $y_i^d$ are spoofed simultaneously (Fig. 3.6), 2) Both $y_1^u$ and $y_2^u$ are spoofed simultaneously; similarly for $y_1^d$ and $y_2^d$ (Fig. 3.7), 3) Middle gate measurements $y_1^d$, $y_2^u$ are spoofed (Fig. 3.8), 4) All $y_1^u$, $y_1^d$ and $y_2^u$ are spoofed simultaneously; similarly for $y_1^d$, $y_2^u$ and $y_2^d$ (Fig. 3.9). In all the four cases, it is assumed that the attacker injects an additive attack such that the targeted level sensor measurement signal does not deviate from zero. For e.g., for case 1), $g_i(t) := \begin{pmatrix} -y_i^u(t) & -y_i^d(t) \end{pmatrix}^\mathsf{T}$, where $y_i^u(t)$ and $y_i^d(t)$ are true measurement signals, and $H_i$ is given by (3.47); similarly for other cases. $\Delta$

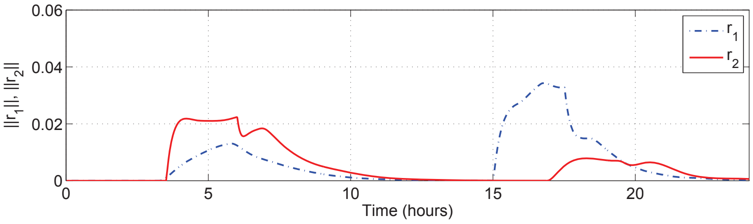


Figure 3.8: Residuals under attack on $y_1^d$, $y_2^u$.

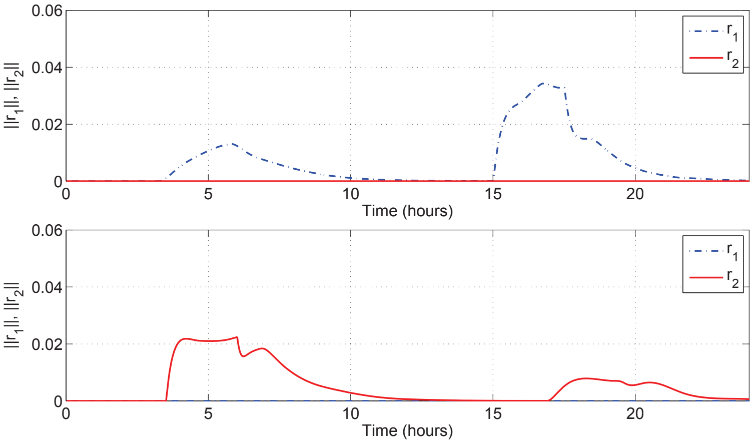


Figure 3.9: Residuals under attack on $y_1^u$, $y_1^d$ and $y_2^u$ (top); $y_1^d$, $y_2^u$ and $y_2^d$ (bottom).

### 3.5.2 Implications for Water Security

Based on the performance of our FDI scheme on adversarial scenarios from the generalized attack model (3.46), and in particular from the deception attack scenarios of Example 3.5.1, we can make several interesting observations. Firstly, the diagnosis rule presented in Table 3.1 can no longer be used in the presence of deception attacks. In general, the residuals will not satisfy the conditions of Definition 3.4.3 and hence, (3.44) is not guaranteed to achieve a correct diagnosis. However, in certain adversarial scenarios (for e.g., the case when $y_1^u$ and $y_2^u$ are spoofed in Fig. 3.7 (top), a correct diagnosis can still be achieved using the following fault/attack detection and isolation (F/ADI) rule:

$$f_j(t) \neq 0 \text{ if } \|r_j(t)\| < \vartheta_{f_j}, \text{ and } \|r_k(t)\| \geqslant \vartheta_{f_k}, k \neq j, \tag{3.48}$$

where the parameters $\vartheta_{f_i}$ $i = 1, \ldots, s$ are the *isolation threshold parameters* of the F/ADI scheme. These parameters can be constant or time-varying depending on the nature adversarial scenarios which are likely to be encountered in practice[5].

The F/ADI rule (3.48) may not successfully isolate unknown withdrawals in a pool (say $i$) when both $y_i^u$ and $y_i^d$ are compromised. For example, in Fig. 3.6 (top), observer 1 which was designed to be insensitive to $f_1$ is no longer able to maintain $r_1$ to zero (whereas,

[5] An elaborate discussion regarding the tuning of these parameters is outside the scope of our work. However, we recommend that, similar to the case of CUSUM implementation, parameter tuning can be achieved by testing the performance based on desired false-alarm and missed-detection rates.

$\mathsf{r}_2$ generated by observer 2 is still sensitive to $\mathsf{f}_1$). However, note that in this case $\mathsf{f}_2$ can be still be correctly isolated using (3.48). From this observation, it can be concluded that when both upstream and downstream measurements of a canal pool are compromised, it is difficult to isolate the *local* faults in the pool; however, faults in other pools can still be isolated.

Another observation is that the location of compromised sensor measurements relative to the location of the fault is an important factor for achieving successful diagnosis. We recall that, under our setting, the offtakes are located near the downstream ends (see Fig. 3.2). From Fig. 3.6 (bottom) it can be seen that, in contrast to Fig. 3.6 (top), the attack on downstream measurements is more detrimental to the performance of residuals in detecting unknown withdrawals from offtakes. Since our diagnosis scheme is based on the physics-based ID model (see model (3.26) in Section 3.2), the effect of water withdrawals is captured by both upstream and downstream level sensors; however, the effect is more pronounced at the downstream level sensors. This insight can also be applied when both measurements of a single gate are compromised. See Fig. 3.8 when attack on $\mathsf{y}_1^d$ and $\mathsf{y}_2^u$ of the middle gate makes the diagnosis of fault $\mathsf{f}_1$ located near the gate difficult, while $\mathsf{f}_2$ can still be diagnosed successfully based on (3.48).

Last but perhaps the most interesting observation is that when sensor measurements of multiple pools are accessible to a strategic attacker, the deception attack can be perfectly stealthy, i.e., the attack can result in wrong diagnosis or may not be even detected! Consider, for e.g., Fig. 3.9 (top) (resp. Fig. 3.9 (bottom)) when $\mathsf{y}_1^u$, $\mathsf{y}_1^d$ and $\mathsf{y}_2^u$ (resp. $\mathsf{y}_1^d$, $\mathsf{y}_2^u$ and $\mathsf{y}_2^d$) are compromised. Residual $\mathsf{r}_1$ (resp. $\mathsf{r}_2$) which was only sensitive to fault $\mathsf{f}_2$ (resp. $\mathsf{f}_1$) in the case of no attack, now reacts to both faults, whereas $\mathsf{r}_2$ (resp. $\mathsf{r}_1$) is not sensitive to either faults. Following (3.48), this leads to incorrect diagnosis, i.e., $\mathsf{f}_1$ is detected when $\mathsf{f}_2$ is presented and vice versa. Moreover, from a practical viewpoint, the norms of residuals in the case of such attacks may not be high enough to enable the F/ADI rule (3.48) to distinguish these faults from random disturbances.

By comparing this stealthy attack with the stealthy attack reported in Chapter 2, the following remarks can be made: 1) From an attacker's point-of-view more sensor measurements (three sensors as opposed to a single sensor in Chapter 2) need to be compromised to achieve perfect stealthiness when the F/ADI scheme proposed herewith is used, 2) The attacker requires strategic knowledge (and perhaps more resources) to carry out such an attack; for e.g., only a particular choice of compromised measurements result in a stealthy attack, 3) In contrast to Chapter 2 where the $\mathsf{f}_2$ under the compromise of $\mathsf{y}_2^d$ went completely undetected since neither residuals reacted to the fault, here the residual $\mathsf{r}_2$ shows a delayed response (see Fig. 3.9 (bottom)). Thus detection is not completely evaded in this case, although the diagnosis is incorrect. The observed delay is the delay in propagation of disturbance due to offtake withdrawal in the second pool to reach the upstream of first pool.

## 3.6 Discussion

In this chapter, we investigated the applicability of a model-based scheme for detection and isolation of a wide class of faults and attacks in automated canal systems. The scheme is based on a bank of UIO designed for a linear delay-differential system obtained as an analytically approximate model of the linearized SWE. Our approach is based on a simplified model of canal hydrodynamics which captures the influence of both upstream and downstream variations. We present conditions for the existence of a UIO when failure signals in the state and measurement equations are coupled. These conditions are delay-dependent, and can also incorporate communication network induced time-delays in the sensor-control data. A residual generation procedure is used to detect and isolate such failure signals.

The performance of the UIO-based FDI scheme is investigated on scenarios when the fault signals in the state and measurement equations are uncoupled. Such scenarios can result from the actions of an attacker which simultaneously compromises sensor-control data and offtakes for the purpose of water pilfering (or even for causing damage to the canal system). For a class of attack scenarios, we also propose a simple modification of the UIO based FDI scheme to a threshold-based A/FDI scheme. While practical tuning rules of the proposed A/FDI scheme is a topic of further investigation, an interesting theoretical open question is to adapt these threshold parameters to be sensitive to attacks.

From the viewpoint of cyber-security of canal automation systems, we find that sensor redundancy (i.e., installation of multiple sensors for each candidate fault/attack), and making critical sensors more resilient to manipulation and tampering is a reasonable cyber-defense strategy. For e.g., for the cases when offtake withdrawals are located near the downstream end, the downstream level sensors are more critical for successful isolation of failures and hence, more investment should be made to make them tamper resistant.

When the compromise of sensor measurements is restricted to a given pool, the diagnosis of faults that are local to the pool is the most severely affected. The effect is also propagated to neighboring pools, although to a lesser extent. However, when sensor measurements from multiple pools are compromised by a strategic and resourceful attacker, the F/ADI scheme can result in an incorrect diagnosis (or even perfect stealthiness). Thus reducing the chance that multiple and coordinated compromises occur should be prioritized for cyber-security of water SCADA systems.

Finally, we believe that the insights presented in this chapter motivates further investigation of novel model-based attack detection schemes which are not based on the assumptions made by classical fault detection and isolation schemes (for e.g., the assumption of non-simultaneous failure signals). From our analysis we conclude that a proper selection of internal model, and increased emphasis on securing critical sensor measurements could lead to better performance of F/ADI schemes under deception attacks. Such attack-sensitive detection schemes will also assist in the development of automatic controller response schemes which are resilient to a broad class of physical faults and cyber-attack signals.

## Appendix 3.A Conditions for observer design

*Proof.* Under (3.40), we note that $\bar{Z}_i$ defined in (3.41) satisfies $\bar{Z}_i > 0$, $i = 1, \ldots, 4$. Inspired by the work of Lin et.al. Lin et al. [2006], under (3.39) and $P > 0$, we consider the following Lyapunov-Krasovskii functional:

$$\begin{aligned} V(\mathsf{e}(t)) &= \mathsf{e}(t)^{\mathsf{T}} P \mathsf{e}(t) \\ &\quad + \sum_{i=1}^{4} \int_{t-\tau_i(t)}^{t} \begin{pmatrix} \mathsf{e}(s) \\ \dot{\mathsf{e}}(s) \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} Q_i & U_i \\ U_i^{\mathsf{T}} & R_i \end{pmatrix} \begin{pmatrix} \mathsf{e}(s) \\ \dot{\mathsf{e}}(s) \end{pmatrix} ds \\ &\quad + \sum_{i=1}^{4} \int_{0}^{h_i} \int_{t-\theta}^{t} \begin{pmatrix} \mathsf{e}(s) \\ \dot{\mathsf{e}}(s) \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} S_i & W_i \\ W_i^{\mathsf{T}} & Z_i \end{pmatrix} \begin{pmatrix} \mathsf{e}(s) \\ \dot{\mathsf{e}}(s) \end{pmatrix} ds \, d\theta. \end{aligned} \tag{3.49}$$

Let us define the following vectors:

$$\eta(t)^{\mathsf{T}} := \left( \tilde{\mathsf{e}}(t)^{\mathsf{T}}, \tilde{\dot{\mathsf{e}}}(t)^{\mathsf{T}} \right), \quad \zeta(s)^{\mathsf{T}} := \left( \mathsf{e}(s)^{\mathsf{T}}, \dot{\mathsf{e}}(s)^{\mathsf{T}} \right).$$

where $\tilde{\mathsf{e}}(t)^{\mathsf{T}} := \left( \mathsf{e}(t)^{\mathsf{T}}, \mathsf{e}(t-\tau_1(t))^{\mathsf{T}}, \ldots, \mathsf{e}(t-\tau_4(t))^{\mathsf{T}} \right)$, and

$$\tilde{\dot{\mathsf{e}}}(t)^{\mathsf{T}} := \left( \dot{\mathsf{e}}(t)^{\mathsf{T}}, \dot{\mathsf{e}}(t-\tau_1(t))^{\mathsf{T}}, \ldots, \dot{\mathsf{e}}(t-\tau_4(t))^{\mathsf{T}} \right).$$

We make the following two observations: First, using the Leibnitz rule,

$$\sum_{i=1}^{4} \mathsf{e}(t-\tau_i(t)) = 4\mathsf{e}(t) - \sum_{i=1}^{4} \int_{t-\tau_i(t)}^{t} \dot{\mathsf{e}}(s) ds,$$

we obtain for any matrices $H_i$, with appropriate dimensions, and $i = 0, \ldots, 9$,

$$\begin{aligned} 0 = 2 &\left( \sum_{i=0}^{4} \mathsf{e}(t-\tau_i(t))^{\mathsf{T}} H_i + \sum_{i=5}^{9} \dot{\mathsf{e}}(t-\tau_i(t))^{\mathsf{T}} H_i \right) \\ \times &\left( 4\mathsf{e}(t) - \sum_{i=1}^{4} \mathsf{e}(t-\tau_i(t)) - \sum_{i=1}^{4} \int_{t-\tau_i(t)}^{t} \dot{\mathsf{e}}(s) ds \right), \end{aligned} \tag{3.50}$$

or equivalently,

$$0 = 2\eta(t)^{\mathsf{T}} H \Delta_1 \eta(t) - 2 \sum_{i=1}^{4} \int_{t-\tau_i(t)}^{t} \eta(t)^{\mathsf{T}} \begin{pmatrix} 0 \\ H^{\mathsf{T}} \end{pmatrix}^{\mathsf{T}} \zeta(s) ds, \tag{3.51}$$

where

$$H^{\mathsf{T}} := \begin{pmatrix} H_0^{\mathsf{T}} & H_1^{\mathsf{T}} & \ldots & H_9^{\mathsf{T}} \end{pmatrix}, \quad \Delta_1 := \begin{pmatrix} 4 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Second, using $\sum_{i=0}^{4} F_i \mathsf{e}(t-\tau_i) - \dot{\mathsf{e}}(t) = 0$, we obtain for a matrix P with appropriate dimensions and scalars $\epsilon_0, \dots, \epsilon_9$, $\bar{\epsilon}_1, \dots, \bar{\epsilon}_4$

$$\begin{aligned} 0 =& 2\left(\sum_{i=0}^{4} \mathsf{e}(t-\tau_i(t))^{\mathsf{T}}\epsilon_i + \sum_{i=5}^{9} \dot{\mathsf{e}}(t-\tau_i(t))^{\mathsf{T}}\epsilon_i + \sum_{i=1}^{4}\int_{t-\tau_i(t)}^{t} \mathsf{e}^{\mathsf{T}}(s)ds\bar{\epsilon}_i\right) P \\ &\times \left(\sum_{i=0}^{4} F_i \mathsf{e}(t-\tau_i) - \dot{\mathsf{e}}(t)\right), \end{aligned} \tag{3.52}$$

or equivalently,

$$0 = 2\eta(t)^{\mathsf{T}}\Upsilon\Delta_2\eta(t) - 2\sum_{i=1}^{4}\int_{t-\tau_i(t)}^{t} \eta(t)\begin{pmatrix} -\bar{\epsilon}_i\Delta_2^{\mathsf{T}}P^{\mathsf{T}} & 0 \end{pmatrix}\zeta(s)ds, \tag{3.53}$$

where

$$\Upsilon^{\mathsf{T}} := P^{\mathsf{T}}\begin{pmatrix} \epsilon_0 & \epsilon_1 & \dots & \epsilon_9 \end{pmatrix}, \quad \Delta_2 := \begin{pmatrix} F_0 & \dots & F_4 & -I & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Adding (3.51) and (3.53) to the time derivative of $V(\mathsf{e}(t))$ along the solution of (3.33), we can write:

$$\begin{aligned} \dot{V}(\mathsf{e}(t)) = 2\mathsf{e}(t)^{\mathsf{T}}P\dot{\mathsf{e}}(t) &+ \sum_{i=1}^{4}\begin{pmatrix} \mathsf{e}(t) \\ \dot{\mathsf{e}}(t) \end{pmatrix}^{\mathsf{T}}\begin{pmatrix} Q_i & U_i \\ U_i^{\mathsf{T}} & R_i \end{pmatrix}\begin{pmatrix} \mathsf{e}(t) \\ \dot{\mathsf{e}}(t) \end{pmatrix} \\ &- \sum_{i=1}^{4}(1-\dot{\tau}_i(t))\begin{pmatrix} \mathsf{e}(t-\tau_i(t)) \\ \dot{\mathsf{e}}(t-\tau_i(t)) \end{pmatrix}^{\mathsf{T}}\begin{pmatrix} Q_i & U_i \\ U_i^{\mathsf{T}} & R_i \end{pmatrix}\begin{pmatrix} \mathsf{e}(t-\tau_i(t)) \\ \dot{\mathsf{e}}(t-\tau_i(t)) \end{pmatrix} \\ &+ \sum_{i=1}^{4} h_i\begin{pmatrix} \mathsf{e}(t) \\ \dot{\mathsf{e}}(t) \end{pmatrix}^{\mathsf{T}}\begin{pmatrix} S_i & W_i \\ W_i^{\mathsf{T}} & Z_i \end{pmatrix}\begin{pmatrix} \mathsf{e}(t) \\ \dot{\mathsf{e}}(t) \end{pmatrix} \\ &- \sum_{i=1}^{4}\int_{t-h_i(t)}^{t}\begin{pmatrix} \mathsf{e}(s) \\ \dot{\mathsf{e}}(s) \end{pmatrix}^{\mathsf{T}}\begin{pmatrix} S_i & W_i \\ W_i^{\mathsf{T}} & Z_i \end{pmatrix}\begin{pmatrix} \mathsf{e}(s) \\ \dot{\mathsf{e}}(s) \end{pmatrix} ds \\ &+ 2\eta(t)^{\mathsf{T}}[H\Delta_1 + \Upsilon\Delta_2]\eta(t) - 2\sum_{i=1}^{4}\int_{t-h_i(t)}^{t}\eta(t)^{\mathsf{T}}\bar{H}_i\zeta(s)ds \\ &+ \sum_{i=1}^{4}\left(\tau_i(t)\eta(t)^{\mathsf{T}}\bar{H}_i\bar{Z}_i\bar{H}_i^{\mathsf{T}}\eta(t) - \int_{t-\tau_i(t)}^{t}\eta(t)^{\mathsf{T}}\bar{H}_i\bar{Z}_i\bar{H}_i^{\mathsf{T}}\eta(t)ds\right) \end{aligned} \tag{3.54}$$

where $\bar{Z}_i$ and $\bar{H}_i$ are given by:

$$\bar{Z}_i := \begin{pmatrix} S_i & W_i \\ W_i^{\mathsf{T}} & Z_i \end{pmatrix}, \quad \bar{H}_i := \begin{pmatrix} -\bar{\epsilon}_i(PF_0)^{\mathsf{T}} & H_0 \\ -\bar{\epsilon}_i(PF_1)^{\mathsf{T}} & H_1 \\ -\bar{\epsilon}_i(PF_2)^{\mathsf{T}} & H_2 \\ -\bar{\epsilon}_i(PF_3)^{\mathsf{T}} & H_3 \\ -\bar{\epsilon}_i(PF_4)^{\mathsf{T}} & H_4 \\ \bar{\epsilon}_i P^{\mathsf{T}} & H_5 \\ 0 & H_6 \\ 0 & H_7 \\ 0 & H_8 \\ 0 & H_9 \end{pmatrix},$$

for $i = 1, 2, 3, 4$. Using the fact that $\tau_i(t) \leqslant h_i$, and $\dot{\tau}_i(t) \leqslant d_i < 1$, for $i = 1, 2, 3, 4$,

$$\dot{V}(\mathsf{e}(t)) \leqslant \eta(t)^{\mathsf{T}} \left( \Phi + \sum_{i=1}^{4} h_i \bar{H}_i \bar{Z}_i^{-1} \bar{H}_i^{\mathsf{T}} \right) \eta(t) - \sum_{i=1}^{4} \int_{t-h_i(t)}^{t} \Gamma_i(t,s)^{\mathsf{T}} \bar{Z}_i^{-1} \Gamma_i(t,s) ds, \tag{3.55}$$

where $\Gamma_i(t,s) := \left( \bar{H}_i^{\mathsf{T}} \eta(t) + \bar{Z}_i \zeta(s) \right)$, and

$$\Phi := \begin{pmatrix} \phi_{00} & \phi_{01} & \phi_{02} & \phi_{03} & \phi_{04} & \phi_{05} & \phi_{06} & \phi_{07} & \phi_{08} & \phi_{09} \\ * & \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} & \phi_{15} & \phi_{16} & \phi_{17} & \phi_{18} & \phi_{19} \\ * & * & \phi_{22} & \phi_{23} & \phi_{24} & \phi_{25} & \phi_{26} & \phi_{27} & \phi_{28} & \phi_{29} \\ * & * & * & \phi_{33} & \phi_{34} & \phi_{35} & \phi_{36} & \phi_{37} & \phi_{38} & \phi_{39} \\ * & * & * & * & \phi_{44} & \phi_{45} & \phi_{46} & \phi_{47} & \phi_{48} & \phi_{49} \\ * & * & * & * & * & \phi_{55} & \phi_{56} & \phi_{57} & \phi_{58} & \phi_{59} \\ * & * & * & * & * & * & \phi_{66} & \phi_{67} & \phi_{68} & \phi_{69} \\ * & * & * & * & * & * & * & \phi_{77} & \phi_{78} & \phi_{79} \\ * & * & * & * & * & * & * & * & \phi_{88} & \phi_{89} \\ * & * & * & * & * & * & * & * & * & \phi_{99} \end{pmatrix}, \tag{3.56}$$

with block elements $\phi_{jk}$ given by

$$\begin{aligned}
&\phi_{00} = \sum_{i=1}^{4}(Q_i + h_i S_i) + \epsilon_0 \operatorname{sym}(PF_0) + 4\operatorname{sym}(H_0)\\
&\phi_{01} = \epsilon_0 PF_1 + \epsilon_1 (PF_0)^{\top} + 4H_1^{\top} - H_0, \quad \phi_{02} = \epsilon_0 PF_2 + \epsilon_2 (PF_0)^{\top} + 4H_2^{\top} - H_0\\
&\phi_{03} = \epsilon_0 PF_3 + \epsilon_3 (PF_0)^{\top} + 4H_3^{\top} - H_0, \quad \phi_{04} = \epsilon_0 PF_4 + \epsilon_4 (PF_0)^{\top} + 4H_4^{\top} - H_0\\
&\phi_{05} = P + \sum_{i=1}^{4}(U_i + h_i W_i) - \epsilon_0 P + \epsilon_5 (PF_0)^{\top} + 4H_5^{\top}, \quad \phi_{06} = \epsilon_6 (PF_0)^{\top} + 4H_6^{\top},\\
&\phi_{07} = \epsilon_7 (PF_0)^{\top} + 4H_7^{\top}, \quad \phi_{08} = \epsilon_8 (PF_0)^{\top} + 4H_8^{\top}, \quad \phi_{09} = \epsilon_9 (PF_0)^{\top} + 4H_9^{\top},\\
&\phi_{11} = \epsilon_1 \operatorname{sym}(PF_1) - (1-d_1)Q_1 - \operatorname{sym}(H_1)\\
&\phi_{12} = \epsilon_1 PF_2 + \epsilon_2 (PF_1)^{\top} - H_1 - H_2^{\top}, \quad \phi_{13} = \epsilon_1 PF_3 + \epsilon_3 (PF_1)^{\top} - H_1 - H_3^{\top},\\
&\phi_{14} = \epsilon_1 PF_4 + \epsilon_4 (PF_1)^{\top} - H_1 - H_4^{\top}, \quad \phi_{15} = -\epsilon_1 P + \epsilon_5 (PF_1)^{\top} - H_5^{\top},\\
&\phi_{16} = +\epsilon_6 (PF_1)^{\top} - (1-d_1)U_1 - H_6^{\top}, \quad \phi_{17} = +\epsilon_7 (PF_1)^{\top} - H_7^{\top},\\
&\phi_{18} = +\epsilon_8 (PF_1)^{\top} - H_8^{\top}, \quad \phi_{19} = +\epsilon_9 (PF_1)^{\top} - H_9^{\top}\\
&\phi_{22} = +\epsilon_2 \operatorname{sym}(PF_2) - (1-d_2)Q_2 - \operatorname{sym}(H_2)\\
&\phi_{23} = +\epsilon_2 PF_3 + \epsilon_3 (PF_2)^{\top} - H_2 - H_3^{\top}, \quad \phi_{24} = +\epsilon_2 PF_4 + \epsilon_4 (PF_2)^{\top} - H_2 - H_4^{\top}\\
&\phi_{25} = -\epsilon_2 P + \epsilon_5 (PF_2)^{\top} - H_5^{\top}, \quad \phi_{26} = +\epsilon_6 (PF_2)^{\top} - H_6^{\top}\\
&\phi_{27} = -(1-d_2)U_2 + \epsilon_7 (PF_2)^{\top} - H_7^{\top}, \quad \phi_{28} = +\epsilon_8 (PF_2)^{\top} - H_8^{\top}, \quad \phi_{29} = +\epsilon_9 (PF_2)^{\top} - H_9^{\top}\\
&\phi_{33} = -(1-d_3)Q_3 + \epsilon_3 \operatorname{sym}(PF_3) - \operatorname{sym}(H_3)\\
&\phi_{34} = +\epsilon_3 PF_4 + \epsilon_4 (PF_3)^{\top} - H_3 - H_4^{\top}, \quad \phi_{35} = -\epsilon_3 P + \epsilon_5 (PF_3)^{\top} - H_5^{\top},\\
&\phi_{36} = +\epsilon_6 (PF_3)^{\top} - H_6^{\top}, \quad \phi_{37} = +\epsilon_7 (PF_3)^{\top} - H_7^{\top}\\
&\phi_{38} = +\epsilon_8 (PF_3)^{\top} - (1-d_3)U_3 - H_8^{\top}, \quad \phi_{39} = +\epsilon_9 (PF_3)^{\top} - H_9^{\top}\\
&\phi_{44} = -(1-d_4)Q_4 + \epsilon_4 \operatorname{sym}(PF_4)^{\top} - \operatorname{sym}(H_4)\\
&\phi_{45} = -\epsilon_4 P + \epsilon_5 (PF_4)^{\top} - H_5^{\top}, \quad \phi_{46} = +\epsilon_6 (PF_4)^{\top} - H_6^{\top},\\
&\phi_{47} = +\epsilon_7 (PF_4)^{\top} - H_7^{\top}, \quad \phi_{48} = +\epsilon_8 (PF_4)^{\top} - H_8^{\top}, \quad \phi_{49} = -(1-d_4)U_4 + \epsilon_9 (PF_4)^{\top} - H_9^{\top}\\
&\phi_{55} = \sum_{i=1}^{4}(R_i + h_i Z_i) - \epsilon_5 \operatorname{sym}(P),\\
&\phi_{56} = -\epsilon_6 P^{\top}, \quad \phi_{57} = -\epsilon_7 P^{\top}, \quad \phi_{58} = -\epsilon_8 P^{\top}, \quad \phi_{59} = -\epsilon_9 P^{\top},\\
&\phi_{66} = -(1-d_1)R_1, \quad \phi_{67} = 0, \quad \phi_{68} = 0, \quad \phi_{69} = 0\\
&\phi_{77} = -(1-d_2)R_2, \quad \phi_{78} = 0, \quad \phi_{79} = 0,\\
&\phi_{88} = -(1-d_3)R_3, \quad \phi_{89} = 0, \quad \phi_{99} = -(1-d_4)R_4
\end{aligned}$$

where $\operatorname{sym}(M) := M + M^{\top}$. From (3.55), we see that if $\left(\Phi + \sum_{i=1}^{4} h_i \bar{H}_i \bar{Z}_i^{-1} \bar{H}_i^{\top}\right) < 0$ (equivalently, using Schur complements if LMI (3.40) holds), then $\dot{V}(\mathsf{e}(t)) < 0$. Following stability theory of delay differential equations Hale and Lunel [1993], the error dynamic (3.38) is asymptotically stable. Using (3.37) and defining $U \coloneqq PK$, we obtain $\bar{H}_i$. □

Finally, from (3.37) and using $U = PK$, we obtain the terms $\phi_{jk}$:

$$
\begin{aligned}
\phi_{00} &= \sum_{i=1}^{4}(Q_i + h_i S_i) + \epsilon_0 \operatorname{sym}(P\chi_0 - U\beta_0) + 4\operatorname{sym}(H_0)\\
\phi_{01} &= \epsilon_0(P\chi_1 - U\beta_1) + \epsilon_1(P\chi_0 - U\beta_0)^\top + 4H_1^\top - H_0,\\
\phi_{02} &= \epsilon_0(P\chi_2 - U\beta_2) + \epsilon_2(P\chi_0 - U\beta_0)^\top + 4H_2^\top - H_0\\
\phi_{03} &= \epsilon_0(P\chi_3 - U\beta_3) + \epsilon_3(P\chi_0 - U\beta_0)^\top + 4H_3^\top - H_0,\\
\phi_{04} &= \epsilon_0(P\chi_4 - U\beta_4) + \epsilon_4(P\chi_0 - U\beta_0)^\top + 4H_4^\top - H_0\\
\phi_{05} &= P + \sum_{i=1}^{4}(U_i + h_i W_i) - \epsilon_0 P + \epsilon_5(P\chi_0 - U\beta_0)^\top + 4H_5^\top,\\
\phi_{06} &= \epsilon_6(P\chi_0 - U\beta_0)^\top + 4H_6^\top,\\
\phi_{07} &= \epsilon_7(P\chi_0 - U\beta_0)^\top + 4H_7^\top,\\
\phi_{08} &= \epsilon_8(P\chi_0 - U\beta_0)^\top + 4H_8^\top,\\
\phi_{09} &= \epsilon_9(P\chi_0 - U\beta_0)^\top + 4H_9^\top\\
\phi_{11} &= \epsilon_1 \operatorname{sym}(P\chi_1 - U\beta_1) - (1 - d_1)Q_1 - \operatorname{sym}(H_1)\\
\phi_{12} &= \epsilon_1(P\chi_2 - U\beta_2) + \epsilon_2(P\chi_1 - U\beta_1)^\top - H_1 - H_2^\top,\\
\phi_{13} &= \epsilon_1(P\chi_3 - U\beta_3) + \epsilon_3(P\chi_1 - U\beta_1)^\top - H_1 - H_3^\top\\
\phi_{14} &= \epsilon_1(P\chi_4 - U\beta_4) + \epsilon_4(P\chi_1 - U\beta_1)^\top - H_1 - H_4^\top,\\
\phi_{15} &= -\epsilon_1 P + \epsilon_5(P\chi_1 - U\beta_1)^\top - H_5^\top\\
\phi_{16} &= +\epsilon_6(P\chi_1 - U\beta_1)^\top - (1 - d_1)U_1 - H_6^\top,\\
\phi_{17} &= +\epsilon_7(P\chi_1 - U\beta_1)^\top - H_7^\top,\\
\phi_{18} &= +\epsilon_8(P\chi_1 - U\beta_1)^\top - H_8^\top,\\
\phi_{19} &= +\epsilon_9(P\chi_1 - U\beta_1)^\top - H_9^\top\\
\phi_{22} &= +\epsilon_2 \operatorname{sym}(P\chi_2 - U\beta_2) - (1 - d_2)Q_2 - \operatorname{sym}(H_2)\\
\phi_{23} &= +\epsilon_2(P\chi_3 - U\beta_3) + \epsilon_3(P\chi_2 - U\beta_2)^\top - H_2 - H_3^\top,\\
\phi_{24} &= +\epsilon_2(P\chi_4 - U\beta_4) + \epsilon_4(P\chi_2 - U\beta_2)^\top - H_2 - H_4^\top\\
\phi_{25} &= -\epsilon_2 P + \epsilon_5(P\chi_2 - U\beta_2)^\top - H_5^\top,\\
\phi_{26} &= +\epsilon_6(P\chi_2 - U\beta_2)^\top - H_6^\top
\end{aligned}
\tag{3.57}
$$

$$\phi_{27} = -(1-d_2)U_2 + \epsilon_7(P\chi_2 - U\beta_2)^{\top} - H_7^{\top},$$
$$\phi_{28} = +\epsilon_8(P\chi_2 - U\beta_2)^{\top} - H_8^{\top},$$
$$\phi_{29} = +\epsilon_9(P\chi_2 - U\beta_2)^{\top} - H_9^{\top},$$
$$\phi_{33} = -(1-d_3)Q_3 + \epsilon_3 \operatorname{sym}(P\chi_3 - U\beta_3) - \operatorname{sym}(H_3)$$
$$\phi_{34} = +\epsilon_3(P\chi_4 - U\beta_4) + \epsilon_4(P\chi_3 - U\beta_3)^{\top} - H_3 - H_4^{\top},$$
$$\phi_{35} = -\epsilon_3 P + \epsilon_5(P\chi_3 - U\beta_3)^{\top} - H_5^{\top}$$
$$\phi_{36} = +\epsilon_6(P\chi_3 - U\beta_3)^{\top} - H_6^{\top}, \quad \phi_{37} = +\epsilon_7(P\chi_3 - U\beta_3)^{\top} - H_7^{\top}$$
$$\phi_{38} = +\epsilon_8(P\chi_3 - U\beta_3)^{\top} - (1-d_3)U_3 - H_8^{\top},$$
$$\phi_{39} = +\epsilon_9(P\chi_3 - U\beta_3)^{\top} - H_9^{\top}$$
$$\phi_{44} = -(1-d_4)Q_4 + \epsilon_4 \operatorname{sym}(P\chi_4 - U\beta_4)^{\top} - \operatorname{sym}(H_4)$$
$$\phi_{45} = -\epsilon_4 P + \epsilon_5(P\chi_4 - U\beta_4)^{\top} - H_5^{\top},$$
$$\phi_{46} = +\epsilon_6(P\chi_4 - U\beta_4)^{\top} - H_6^{\top}$$
$$\phi_{47} = +\epsilon_7(P\chi_4 - U\beta_4)^{\top} - H_7^{\top},$$
$$\phi_{48} = +\epsilon_8(P\chi_4 - U\beta_4)^{\top} - H_8^{\top},$$
$$\phi_{49} = -(1-d_4)U_4 + \epsilon_9(P\chi_4 - U\beta_4)^{\top} - H_9^{\top}$$
$$\phi_{55} = \sum_{i=1}^{4}(R_i + h_i Z_i) - \epsilon_5 \operatorname{sym}(P)$$
$$\phi_{56} = -\epsilon_6 P^{\top},$$
$$\phi_{57} = -\epsilon_7 P^{\top},$$
$$\phi_{58} = -\epsilon_8 P^{\top},$$
$$\phi_{59} = -\epsilon_9 P^{\top},$$
$$\phi_{66} = -(1-d_1)R_1,$$
$$\phi_{67} = 0,$$
$$\phi_{68} = 0,$$
$$\phi_{69} = 0,$$
$$\phi_{77} = -(1-d_2)R_2,$$
$$\phi_{78} = 0,$$
$$\phi_{79} = 0,$$
$$\phi_{88} = -(1-d_3)R_3,$$
$$\phi_{89} = 0,$$
$$\phi_{99} = -(1-d_4)R_4.$$

# Chapter 4

# Stability of Flow Networks under Switching Attacks

## 4.1 Introduction

This chapter considers the initial-boundary value problem governed by systems of linear hyperbolic partial differential equations in the canonical diagonal form, and studies conditions for exponential stability when the system discontinuously switches between a finite set of modes. The main motivation of this chapter is to model the random failures and malicious attacks on automated water distribution networks as a switching system. The switching system considered here is fairly general in that the system matrix functions as well as the boundary conditions may switch in time. It is shown how the stability mechanism developed for classical solutions of hyperbolic initial boundary value problems can be generalized to the case in which weaker solutions become necessary due to arbitrary switching. An explicit dwell-time bound for guaranteeing exponential stability of the switching system is provided when, for each mode, the system is exponentially stable. The stability conditions presented here only depend on the system parameters and boundary data. These conditions easily generalize to switching systems in the non-diagonal form under a simple commutativity assumption. Finally, some tutorial examples are presented to illustrate the instabilities that can result from switching.

### 4.1.1 Switching Infinite-Dimensional Systems

Switched systems are a convenient modeling paradigm for a variety of control applications in which evolution processes involve logical decisions. However, in contrast to their simplicity on modeling grounds, the stability analysis of switched systems is often nontrivial. An extensive body of literature now exists for the case of switched (linear and non-linear) ordinary differential equations (ODEs) and more generally for differential algebraic equations (DAEs) in finite dimensional spaces. As surveyed in Shorten et al. [2007] and Lin and Antsaklis [2009], two different approaches have been mainly considered in the literature: Either one designs switching signals such that solutions of the switched system

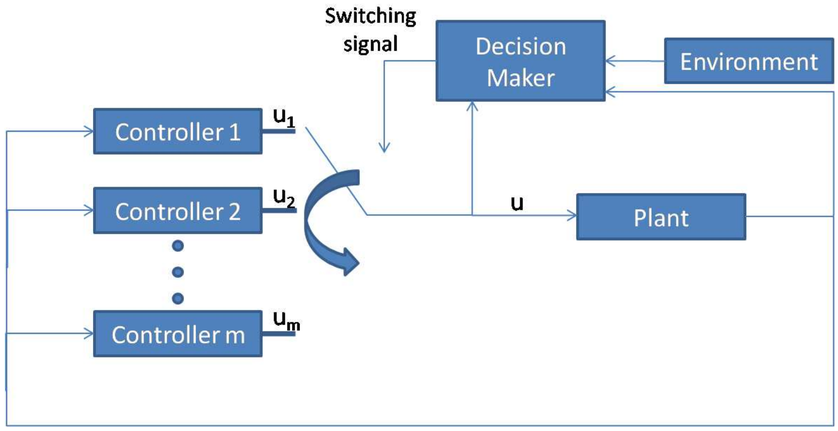


Figure 4.1: Switching triggered by the controller.

decay exponentially (or otherwise behave ‘optimally’), or one tries to identify conditions which guarantee exponential stability of the switched system for arbitrary switching signals. The later approach is of particular interest when the switching mechanism is either unknown or too complicated for a more careful stability analysis D. Liberzon [2003], Morse [1996]. The traditional motivation to study stability under arbitrary switching comes from operational scenarios in which certain parameters of the system may exhibit switching in time triggered by external factors, or the controller based on externally specified logical rules may switch between one of several possible control actions; see Fig. 4.1. Stability under arbitrary switching is mainly achieved by constructing common Lyapunov functions or, more directly, by identifying algebraic/geometric conditions on the involved parameters.

During the past years, several attempts have been made to also consider switched systems in the context of infinite dimensional control theory. Mostly, the problem of designing (optimal or stabilizing) switching control is considered for problems in which the state equation is fixed and just the controller is switched. For example, in El-Farra and Christofides [2004], model reduction together with control synthesis for the reduced finite dimensional model is used to construct switching control for quasi-linear parabolic equations. The design of boundary switching control actions for semi-linear hyperbolic balance equations using switching time sensitivities is considered in Hante and Leugering [2009]. An algorithm to construct optimal switching control for abstract linear systems on Hilbert spaces with switching control operator at fixed switching times is proposed in Iftime and Demetriou [2009]. Moreover, for the heat equation, a systematic way of building switching control based on variational methods is described in Zuazua [2011] and, in a similar context, Gugat [2008] gives conditions under which such switching controls exist for the one dimensional wave equation.

Despite the aforementioned developments, much less is known for problems when not only the controller, but also the state equation is switched. Some general ideas are sketched in Seidman [2009] and, for semi-linear hyperbolic equations with application to transport networks, optimal open-loop and closed-loop switching control is addressed in Hante et al.

[2009] and Hante et al. [2010]. For problems concerning the stability of switched infinite dimensional systems, the construction of common Lyapunov functions gets very difficult when the state equation is switched, even for abstract switched linear systems on Hilbert spaces. The only available result appears to be Sasane [2005], in which a common quadratic Lyapunov function is provided for the case when the semigroup generators commute. This condition is, however, too restrictive for some applications. Nevertheless, it is interesting to note that without further restrictions on the generators, common (not necessarily quadratic) Lyapunov functions exist, even more generally for switched linear systems on Banach spaces Hante and Sigalotti [2011]. Under constrained switching, some algebraic conditions for stability of switched non-linear systems on Banach spaces utilizing Lyapunov functions in each mode are provided in Michel et al. [2005].

### 4.1.2 Stability under Switching Attacks

In this chapter we are interested in the stability properties of solutions to switched linear hyperbolic systems with reflecting boundary conditions when the boundary conditions and the state equation are switched arbitrarily. Our motivation here is to model the random failures and malicious attacks on automated water distribution networks as a switched system. Consider the following example:

*Example* 4.1.1. Consider, for example, the networked control setting for a reservoir-canal system in Fig 4.2. The state variables are the water level and the water flow. The downstream gate is equipped with a water level sensor and the upstream gate can be controlled to move vertically, thereby controlling the water inflow into the system. The level sensor transmits measurements to a remote controller that computes the control signal to be transmitted to the upstream gate actuator. The remote controller designs a control signal that will regulate the water inflow at the upstream gate such that the downstream water level remains within prescribed safety-bounds at all times, and under the effect of downstream demand perturbations. The framework presented in this chapter can be used to model attack scenarios in which the demand perturbations, the level sensor transmits measurements, and the control signal can be modeled as piecewise-constant switching signals.

Let us first introduce the following (unswitched) system of $n$ linear hyperbolic partial differential equations (PDEs) defined for some interval $[a, b] \subset \mathbb{R}$:

$$\frac{\partial \xi}{\partial t} + \Lambda(s)\frac{\partial \xi}{\partial s} + B(s)\xi = 0, \quad s \in (a, b),\ t > 0, \tag{4.1}$$

where $\Lambda(s) = \operatorname{diag}(\lambda_1(s), \dots, \lambda_n(s))$ is a diagonal real matrix function and $B(s)$ is a $n \times n$ real matrix function on $[a, b]$. Assuming appropriate regularity of the matrix functions $\Lambda(\cdot)$ and $B(\cdot)$ and under the hyperbolicity assumption that for some $1 < m < n$

$$\lambda_1(s), \dots, \lambda_m(s) < 0 \text{ and } \lambda_{m+1}(s), \dots, \lambda_n(s) > 0 \tag{4.2}$$

uniformly in $s \in [a, b]$, a $n$-dimensional vector solution $\xi(t, s)$ of the system (4.1) with components $\xi_i(t, s)$ for $i = 1, \dots, n$, arrayed as

$$\xi_{\mathrm{I}}(t, s) = (\xi_1(t, s), \dots, \xi_m(t, s))^\top \text{ and } \xi_{\mathrm{II}}(t, s) = (\xi_{m+1}(t, s), \dots, \xi_n(t, s))^\top,$$

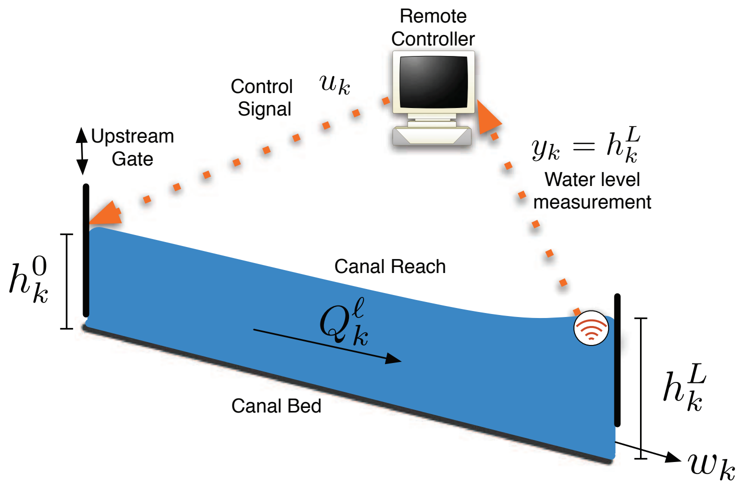


Figure 4.2: Network controlled reservoir-canal system.

is uniquely determined on the time-space strip $\mathbb{R}_+ \times (a, b)$ with the initial condition

$$\xi(0, s) = \bar{\xi}(s), \quad s \in (a, b), \tag{4.3}$$

for specified $\mathbb{R}^n$-valued initial data $\bar{\xi}(s)$ and boundary conditions

$$\xi_{\mathrm{II}}(t, a) = G_{\mathrm{L}} \xi_{\mathrm{I}}(t, a), \;\; \xi_{\mathrm{I}}(t, b) = G_{\mathrm{R}} \xi_{\mathrm{II}}(t, b), \quad t \geqslant 0 \tag{4.4}$$

where $G_{\mathrm{L}}$, $G_{\mathrm{R}}$ are constant matrices of dimensions $(n - m) \times m$ and $m \times (n - m)$, respectively. A common class of problems studied for initial-boundary value problems (4.1)–(4.4) is the stability and stabilization under boundary control actions specified by the matrices $G_{\mathrm{L}}$ and $G_{\mathrm{R}}$. These problems are of interest because hyperbolic PDE systems can model flows in networks that are monitored and controlled at the boundary nodes Leugering and Schmidt [2002]. Examples include transportation systems Haut and Bastin [2007], Bayen et al. [2006], canal systems Litrico et al. [2008], and gas distribution systems Banda et al. [2006]. The available results for this class of problems for linear hyperbolic systems can be found in Rauch and Taylor [1974], Besson et al. [2006], and more generally for quasilinear hyperbolic systems in Li [1994], deHalleux et al. [2003], Coron et al. [2008] and Prieur et al. [2008].

Here we are interested in the stability properties of the hyperbolic initial boundary value problem (4.1)–(4.4) when $\Lambda(\cdot)$, $B(\cdot)$, $G_{\mathrm{L}}$ and $G_{\mathrm{R}}$ are not fixed, but are known to satisfy

$$(\Lambda(\cdot), B(\cdot), G_{\mathrm{L}}, G_{\mathrm{R}}) \in \{(\Lambda^j(\cdot), B^j(\cdot), G_{\mathrm{L}}^j, G_{\mathrm{R}}^j) : j \in Q\}$$

at any time $t > 0$, where $Q = \{1, \ldots, N\}$ is a finite set of modes and, for all $j \in Q$, the data $\Lambda^j(\cdot), B^j(\cdot), G_{\mathrm{L}}^j, G_{\mathrm{R}}^j$ is given. This is equivalent to studying the stability of the switching system

$$\begin{cases} \dfrac{\partial \xi}{\partial t} + \Lambda^{\sigma(t)}(s) \dfrac{\partial \xi}{\partial s} + B^{\sigma(t)}(s) \xi = 0, \\ \xi_{\mathrm{II}}(t, a) = G_{\mathrm{L}}^{\sigma(t)} \xi_{\mathrm{I}}(t, a), \;\; \xi_{\mathrm{I}}(t, b) = G_{\mathrm{R}}^{\sigma(t)} \xi_{\mathrm{II}}(t, b), \\ \xi(0, s) = \bar{\xi}(s), \end{cases} \tag{4.5}$$

for the time-space strip $[0,\infty) \times [a,b]$ where switching occurs according to a piecewise-constant switching signal $\sigma(\cdot)\colon \mathbb{R}_+ \to Q$. Preliminaries and wellposedness of the switched system (4.5) will be discussed in Section 4.2. Then, recalling the classical observation in the finite dimensional control theory of switched systems that exponential stability of all subsystems does not necessarily guarantee an exponential decay of the solution when the system is switched D. Liberzon [2003], we study, motivated by a simple PDE counterpart to this observation, the following two specific problems for the switched system (4.5) in Section 4.3:

(A) Find conditions on the matrix functions $\Lambda^j(\cdot)$, $B^j(\cdot)$ and the matrices $G_{\mathrm{L}}^j$ and $G_{\mathrm{R}}^j$ that guarantee exponential stability for arbitrary switching signals.

(B) Alternatively, characterize a (preferably large) class of switching signals for which exponential stability of all subsystems is sufficient for exponential stability of the switched system.

There are two main contributions of this chapter. Firstly, we show how the techniques mainly developed for classical solutions (with $C^1$ data) can be used for weaker solutions (with $L^\infty$ data) based on the geometric picture of propagation along characteristics. This is necessary because switching boundary conditions may introduce discontinuities into the solution. Secondly, we show how the switching enters the known stability mechanism such that the decay rate obtained in this way is independent of the switching signal (Theorem 1). Following from our analysis, we also obtain an explicit dwell-time bound guaranteeing exponential stability of the system under constrained switching when all subsystems satisfy the known stability condition individually (Corollary 1). In Section 4.4, we discuss how our results for switched diagonal system (4.5) generalize to switched hyperbolic systems in non-diagonal form under a commutativity assumption (Proposition 1). In Sections 4.3 and 4.4, we also provide illustrative examples of instabilities which can result from switching. In Section 4.5, we discuss the stability of canal cascade under attack scenarios that can be modeled as a switched system. Some remarks are mentioned in Section 4.6.

## 4.2 Preliminaries

For an interval $(a,b) \subset \mathbb{R}$ and a measurable function $f\colon (a,b) \to \mathbb{R}^n$, let

$$\|f\|_\infty := \operatorname*{ess\,sup}_{\substack{s\in(a,b)\\ i=1,\dots,n}} |f_i(s)|.$$

We call $L^\infty((a,b);\mathbb{R}^n)$ the space of all measurable functions $f\colon (a,b) \to \mathbb{R}^n$ for which $\|f\|_\infty < \infty$. For an $n \times n$ real matrix $M = (m_{ij})$, we define

$$\|M\|_\infty := \max_{1\leqslant i\leqslant n} \sum_{j=1}^{n} |m_{ij}|.$$

Also define the non-negative matrix of $M$ as $|M| := (|m_{ij}|)$ and for eigenvalues $\lambda_1, \dots, \lambda_n$ of $|M|$ define the spectral radius of $|M|$ as $\rho(|M|) = \max_{1 \leqslant i \leqslant n} |\lambda_i|$.

A *switching signal* $\sigma(\cdot)$ is a piecewise-constant function $\sigma(\cdot)\colon \mathbb{R}_+ \to Q$. Here, we restrict admissible piecewise-constant signals to those for which during each finite time interval of $\mathbb{R}_+$, there are only finitely many switches $j \curvearrowright j'$ to avoid *Zeno behavior*. This assumption anticipated with the accumulation of switching times is commonly made in the field of switched and hybrid systems to obtain global existence results; see for e. g. Zhang et al. [2001]. Thus, necessarily, $\sigma(\cdot)$ has *switching times* $\tau_k \in \mathbb{R}_+$ $(k \in \mathbb{N})$ at which $\sigma(\cdot)$ switches discontinuously from one mode $j_{k-1} \in Q$ to another mode $j_k \in Q$. We denote $\mathcal{S}(\mathbb{R}_+, Q)$ for the set of all such switching signals $\sigma(\cdot)$.

We say that for a given $\sigma(\cdot) \in \mathcal{S}(\mathbb{R}_+, Q)$ the system (4.5) is *exponentially stable* (with respect to the norm $\|\cdot\|_\infty$) if there exist constants $c \geqslant 1$ and $\beta > 0$ such that the solution $\xi(t, \cdot)$ satisfies

$$\|\xi(t,\cdot)\|_\infty \leqslant c\exp(-\beta t)\|\xi(0,\cdot)\|_\infty, \ t \geqslant 0. \tag{4.6}$$

In view of problem (A), we say that the switched system (4.5) is *absolutely exponentially stable* (with respect to a norm $\|\cdot\|_\infty$) if (4.6) holds for all $\sigma(\cdot) \in \mathcal{S}(\mathbb{R}_+, Q)$ with constants $c \geqslant 1$ and $\beta > 0$ independently of $\sigma(\cdot)$. In view of problem (B), we say that a value $\tau > 0$ is a *dwell-time* of a switching signal $\sigma(\cdot)$, if the intervals between consecutive switches are no shorter than $\tau$, that is, $\tau_{k+1} - \tau_k \geqslant \tau$ for all $k > 0$ and we let $\mathcal{S}_\tau(\mathbb{R}_+; Q) \subset \mathcal{S}(\mathbb{R}_+; Q)$ denote the subset of switching signals with dwell-time $\tau$.

## 4.3 Diagonal Switching System

For each $j \in Q$, we have the diagonal subsystem

$$\begin{cases} \dfrac{\partial \xi^j}{\partial t} + \Lambda^j(s)\dfrac{\partial \xi^j}{\partial s} + B^j(s)\xi^j = 0, \ \ s \in (a,b), \ t > 0 \\ \xi^j_{\mathrm{II}}(t,a) = G^j_{\mathrm{L}}\xi^j_{\mathrm{I}}(t,a), \ \xi^j_{\mathrm{I}}(t,b) = G^j_{\mathrm{R}}\xi^j_{\mathrm{II}}(t,b), \ t \geqslant 0 \end{cases} \tag{4.7}$$

for which we impose the following assumptions:

(A1) The matrix function $\Lambda^j(s) = \operatorname{diag}(\lambda^j_1(s), \dots, \lambda^j_n(s))$ is such that the characteristic speeds $\lambda^j_i(\cdot)$ are uniformly bounded, Lipschitz-continuous functions of $s \in [a,b]$ for $i = 1, \dots, n$, and there exists $m_j$ such that for some $0 < m_j < n$, $\lambda^j_r(s) < 0$ $(r = 1, \dots, m_j)$ and $\lambda^j_l(s) > 0$ $(l = m_j + 1, \dots, n)$; the matrix function $B^j(s)$ is such that $B^j(\cdot) : [a,b] \mapsto \mathbb{R}^{n \times n}$ is bounded measurable with respect to $s$.

(A2) For all $j, j' \in Q$, $m_j = m_{j'} =: m$.

It is well-known that under the hyperbolicity assumption (A1) for any $j \in Q$, $T > 0$, and initial data $\xi^j(0,\cdot) = \bar{\xi}^j(\cdot)$ where $\bar{\xi}^j : (a,b) \mapsto \mathbb{R}^n$ is bounded measurable with respect to $s$, a solution $\xi^j_i$ of (4.7) in the *broad sense* can be defined by the method of characteristics

Courant and Hilbert [1962], Kreiss [1970]. In this method, for each $i$ and each point $(t^*, s^*)$, one uses that the ODE

$$\frac{d}{dt} z_i^j(t) = \lambda_i^j(z_i^j(t)), \quad z_i^j(t^*) = s^* \tag{4.8}$$

has a unique Carathéodory solution, defined for all $t$. As usual, we say that this solution $t \mapsto z_i^j(t; t^*, s^*)$ passing through $(t^*, s^*)$ is the $i$-th *characteristic curve* for the $j$-th subsystem. The broad solution $\xi^j(\cdot,\cdot)$ is then defined as a vector function with components $\xi_i^j$, $i = 1, \dots, n$, that are absolutely continuous and satisfy

$$\frac{d}{dt} \xi_i^j(t, z_i^j(t; t^*, s^*)) = -\sum_{k=1}^{n} b_{ik}^j(z_i^j(t; t^*, s^*)) \xi_k^j(t, z_i^j(t; t^*, s^*)) \tag{4.9}$$

along almost every characteristic curve $z_i^j(t; t^*, s^*)$. Here $b_{ik}^j(\cdot)$ corresponds to the $i$-th row and $k$-th column of $B^j(\cdot)$.

Existence and uniqueness of such broad solutions $\xi^j(\cdot,\cdot)$ with initial data and boundary conditions for the subsystems (4.7) with $\xi^j(t,\cdot) \in L^\infty((a,b); \mathbb{R}^n)$ for all $t$ can be obtained on arbitrary finite time horizons using Banach's fixed point theorem. Uniqueness then has to be understood within the usual Lebesgue almost everywhere equivalence class. For further details on the existence and uniqueness of broad solutions, we refer to the iteration method of Courant and Hilbert [1962], pages 470–475, and to the text of Bressan Bressan [2000], pages 46–50, though noting that the latter does not treat boundary conditions. For treatment of the boundary conditions see, instead, Kreiss [1970].

We now justify the existence and uniqueness of solutions for the switching system (4.5), which we need in deriving the main stability result in Section 4.3. Any switching signal $\sigma(\cdot) \in \mathcal{S}(\mathbb{R}_+, Q)$ defines a mode $j_k \in Q$ for each interval $[\tau_k, \tau_{k+1})$. For an initial condition, $\bar{\boldsymbol{\xi}} := \bar{\xi}(\cdot) \in L^\infty((a,b); \mathbb{R}^n)$, we define $\boldsymbol{\xi}(t) = \xi(t,\cdot)$ where

$$\xi(t,\cdot) = \xi^{j_k}(t,\cdot), \quad \text{for } t \in [\tau_k, \tau_{k+1}]$$

and $\xi^{j_k}(t,\cdot)$ is a solution of the subsystem corresponding to mode $j = j_k$ in (4.7) with the initial condition

$$\xi^{j_k}(\tau_k,\cdot) = \begin{cases} \xi^{j_{k-1}}(\tau_k,\cdot) & \text{if } k > 0, \\ \bar{\xi}(\cdot) & \text{if } k = 0. \end{cases}$$

Thus, under Hypothesis (A1), for every $\sigma(\cdot) \in \mathcal{S}(\mathbb{R}_+, Q)$, by construction there exists a unique broad solution $\boldsymbol{\xi}(\cdot)$ with data $\boldsymbol{\xi}(t) \in L^\infty((a,b); \mathbb{R}^n))$ for all $t \in \mathbb{R}_+$ of the switching system (4.5). Again, uniqueness then has to be understood within the usual Lebesgue almost everywhere equivalence class.

In the following, we denote by $z_i^{\sigma(t)}(t; t^*, s^*)$ the $i$-th characteristic path that passes through a point $(t^*, s^*) \in [0,\infty) \times [a,b]$ and is the concatenation of the characteristic curves $z_i^j(t)$ through switching times defined by the switching signal $\sigma(\cdot)$. When needed, we omit the dependence of $z_i^{\sigma(t)}(t; t^*, s^*)$ on $\sigma(t)$ for notational convenience and simply write $z_i(t; t^*, s^*)$.

Observe that, if (A2) holds in addition to (A1), each characteristic path can be classified into left- and right-going depending on the sign of the corresponding characteristic speeds

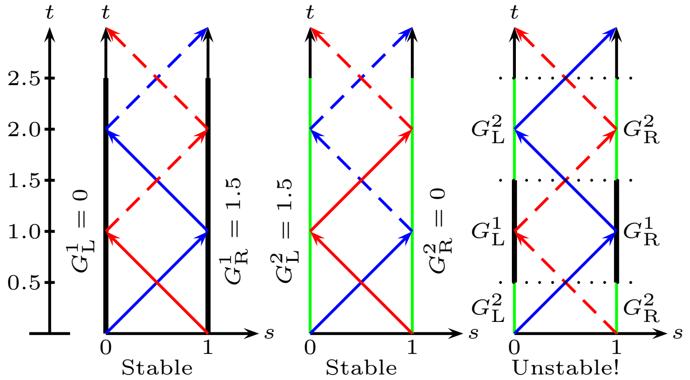


Figure 4.3: Instability by switching.

$\lambda^j_i(s)$, independently of the switching signal $\sigma(\cdot)$. Although (A2) is not required for the existence and uniqueness of the solution, it is crucial for the kind of stabilizing mechanisms that we consider here. This is further discussed in Example 4.3.5.

Furthermore, for the switching system (4.5) we define

$$\bar{\tau} := \frac{b-a}{\min\limits_{\substack{i=1,\ldots,m_j \\ s\in[a,b], j\in Q}} |\lambda^j_i(s)|} + \frac{b-a}{\min\limits_{\substack{i=m_j+1,\ldots,n \\ s\in[a,b], j\in Q}} |\lambda^j_i(s)|} \tag{4.10}$$

Geometrically, $\bar{\tau}$ is an upper bound of the time in which the slowest of all possible characteristic paths will have undergone reflections at both boundaries. Our motivation to study the stability of the diagonal switching system (4.5) is inspired a simple PDE counterpart to the classical ODE observation D. Liberzon [2003] that exponential stability of all subsystems is *not* sufficient for the exponential stability of the switching system.

*Example* 4.3.1. Let $Q = \{1, 2\}$, $[a, b] = [0, 1]$, $\Lambda^j = \mathrm{diag}(-1, 1)$, $B^j = \mathrm{diag}(0, 0)$, $G^j_L = 1.5(j-1)$, $G^j_R = 1.5(2-j)$, and consider $\bar{\xi}(s) = \begin{bmatrix} 1 & 1 \end{bmatrix}^\top$ for $s \in (0, 1)$. For the case of no switching, that is when $\sigma(t) = 1$ or $\sigma(t) = 2$ for all $t \in \mathbb{R}_+$, the solution $\boldsymbol{\xi}(\cdot)$ of the system (4.5) is zero after $t > 2$, but the solution of the system with a switching signal $\sigma(t)$ that is defined over the switching times $\tau_k = 0.5, 1.5, 2.5, \ldots$ and alternates between modes in $Q$ starting with $\sigma(0) = 2$ is not exponentially stable. Indeed, $\|\boldsymbol{\xi}(t)\|_\infty$ is not bounded as $t \to \infty$, because the values on the right-going characteristic emerging from $s \in (0, 0.5)$ always increase by reflection of the characteristics along the boundary; see Figure 4.3. Thus, we can conclude that the instability due to switching can occur for certain combinations between the characteristic speeds and the switching times. (Note, however, that with a switching signal $\sigma(t)$ that is defined over the switching times $\tau_k = 0.5, 1.0, 1.5, 2.0, \ldots$ the system is exponentially stable.) □

We now focus on conditions on the matrix functions $\Lambda(s)^j$, $B^j(s)$ and the boundary data $G^j_L$, $G^j_R$ under which the switching system is absolutely exponentially stable. Our main result, presented next, shows that if a spectral radius condition is jointly satisfied for the left and right boundary data and all pairs of modes $j, j' \in Q$ then a sufficiently small bound on $\|B^j(s)\|_\infty$ exists such that the switching system is absolutely exponentially stable with respect to the norm $\|\cdot\|_\infty$.

**Theorem 4.3.2.** *Assume Hypotheses (A1) and (A2) and suppose that for $j, j' \in Q$ the following condition holds:*

$$\rho\left(\begin{bmatrix} 0 & |G_{\mathrm{R}}^{j'}| \\ |G_{\mathrm{L}}^{j}| & 0 \end{bmatrix}\right) < 1. \tag{4.11}$$

*Then there exists an $\epsilon > 0$ such that if $\|B^j(s)\|_\infty \leqslant \epsilon$ for all $s \in [a, b]$ and $j \in Q$, the switching system (4.5) is absolutely exponentially stable with respect to the norm $\|\cdot\|_\infty$.*

*Proof.* We define the following constants in terms of boundary data

$$K_1 := \max\{1, \tilde{K}_1\}, \quad K_2 := \max\{1, \tilde{K}_2\}, \quad K := \max\{K_1, K_2\} \tag{4.12}$$

where $\tilde{K}_1 = \max\limits_{\substack{r=1,\dots,m \\ j\in Q}} \sum_{l=m+1}^{n} |g_{rl}^{R,j}|$ and $\tilde{K}_2 = \max\limits_{\substack{l=m+1,\dots,n \\ j\in Q}} \left\{\sum_{p=1}^{m} |g_{lp}^{L,j}|\right\}$. From the Lemma 2.1 of Li Li [1994], we note that the condition (4.11) implies

$$\begin{aligned} \theta &:= \max_{j,j'\in Q}\{\||G_{\mathrm{L}}^{j}||G_{\mathrm{R}}^{j'}|\|_\infty, \||G_{\mathrm{R}}^{j'}||G_{\mathrm{L}}^{j}|\|_\infty\} \\ &= \max_{\substack{r=1,\dots,m \\ l=m+1,\dots,n \\ j,j'\in Q}} \left\{\sum_{p=1}^{m}\sum_{k=m+1}^{n} |g_{rk}^{R,j'}||g_{kp}^{L,j}|, \sum_{k=m+1}^{n}\sum_{p=1}^{m} |g_{lp}^{L,j}||g_{pk}^{R,j'}|\right\} < 1, \end{aligned} \tag{4.13}$$

where $G_{\mathrm{L}}^{j} = (g_{pq}^{L,j})$ and $G_{\mathrm{R}}^{j'} = (g_{pq}^{R,j'})$. Let us define

$$T_{\min} := \frac{b-a}{\max\limits_{\substack{a\leqslant s\leqslant b \\ i=1,\dots,n \\ j=1,\dots,N}} |\lambda_i^j(s)|}, \quad T_{\max} := \frac{b-a}{\min\limits_{\substack{a\leqslant s\leqslant b \\ i=1,\dots,n \\ j=1,\dots,N}} |\lambda_i^j(s)|}.$$

Thus, $T_{\min}$ (resp. $T_{\max}$) is the time in which the fastest (resp. slowest) of all possible characteristic paths will have traveled the domain $(a, b)$.

Under the assumption of the theorem, we choose a $c \geqslant 1$ such that

$$c := \frac{K}{\theta}, \tag{4.14}$$

and we choose an $\omega$ such that $\theta < \omega < 1$, and select a $\beta > 0$ such that

$$\beta := \frac{1}{2T_{\max}} \ln\left(\frac{\omega}{\theta}\right). \tag{4.15}$$

We also choose an $\eta > 0$ such that $\theta < \omega < \eta < 1$, and select an $\epsilon > 0$ such that

$$\epsilon := \min\{\epsilon_1, \epsilon_2\} \tag{4.16}$$

where

$$\epsilon_1 = \frac{\theta}{T_{\min} K\omega} \ln\left(\frac{\eta}{\omega}\right), \quad \epsilon_2 = \frac{\theta(1-\eta)}{2T_{\max} K\eta} \ln\left(\frac{\omega}{\theta}\right).$$

We will show that under the aforementioned assumptions and the choice of constants, if the bound

$$\|B^j(s)\|_\infty \leqslant \epsilon \tag{4.17}$$

holds for $s \in [a,b]$ and for all $j \in Q$, then

$$\|\boldsymbol{\xi}(t)\|_\infty \leqslant c\exp(-\beta t)\|\bar{\xi}\|_\infty, \; t \geqslant 0 \tag{4.18}$$

uniformly for all switching signals $\sigma(\cdot) \in \mathcal{S}(\mathbb{R}_+, Q)$. Note that the chosen $c$, $\beta$, and $\epsilon$ are independent of $\sigma(\cdot)$ and only depend on the boundary data and system parameters. We will prove (4.18) using the method of characteristics and induction. To this end, we will first prove the induction basis in **Part A** and the induction step in **Part B**. We define

$$\widehat{\|\boldsymbol{\xi}(t)\|}_\infty := \exp(\beta t)\|\boldsymbol{\xi}(t)\|_\infty. \tag{4.19}$$

**Part A.** Proof of the induction basis. We show that under the chosen constants $\beta > 0, c \geqslant 1, \epsilon > 0$, (4.18) holds on the domain $[0,\delta] \times (a,b)$ when $\delta$ satisfies $0 \leqslant \delta < T_{\min}$. For any $\sigma(\cdot)$, let $z_i(t;t^*,s^*)$ denote the $i$-th characteristic path passing through the point $(t^*,s^*) \in [0,\delta] \times (a,b)$, $(i = 1,\ldots,n)$. Then, we have

$$\begin{aligned} \frac{dz_i(t;t^*,s^*)}{dt} &= \lambda_i^{\sigma(t)}(z_i(t;t^*,s^*)) \\ z_i(t^*;t^*,s^*) &= s^*. \end{aligned}$$

For any fixed $r = 1,\ldots,m$, consider the $r$-th characteristic path $z_r(t;t^*,s^*)$ passing through $(t^*,s^*)$. Under the assumptions (A1) and (A2), backwards in time, $z_r(t;t^*,s^*)$ either intersects $t = 0$ within the interval $[a,b]$ before hitting any boundary (case $A.1$) or it intersects the line $s = b$ (case $A.2$). See Figure 4.4 for an illustration of both possible cases. The point of intersection of the characteristic path with the boundary of the domain is denoted by $(0, z_r(0;t^*,s^*))$ for case $A.1$ and $(t_r(t^*,s^*), b)$ for case $A.2$ with $z_r(t_r(t^*,s^*);t^*,s^*) = b$. Furthermore, let $z_l(t;t_r(t^*,s^*),b)$ denote the $l$-th characteristic path passing through $(t_r(t^*,s^*),b)$ $(l = m+1,\ldots,n)$. Then, since $\delta < T_{\min}$, $z_l(t;t_r(t^*,s^*),b)$ intersects the line $t = 0$ before hitting the line $s = a$. We denote the point of intersection by $(0, z_l(0;t_r(t^*,s^*),b))$. For the ease of notation, we will use $t_r$ for $t_r(t^*,s^*)$.

Estimate for paths with negative slope. We first obtain an estimate of $e^{\beta t^*}|\xi_r(t^*,s^*)|$ for any $(t^*,s^*) \in [0,\delta] \times (a,b)$ by considering cases $A.1$ and $A.2$ for the $r$-th characteristic path $z_r(t;t^*,s^*)$ passing through $(t^*,s^*)$ $(r = 1,\ldots,m)$.

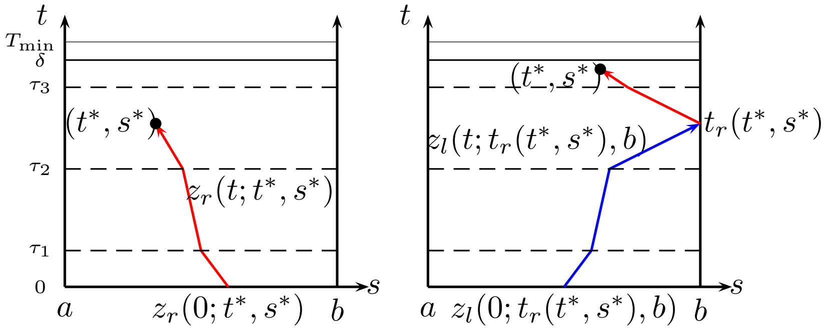


Figure 4.4: Illustration of cases for the proof of induction basis.

Case A.1: Using $j = \sigma(t)$ in (4.9), and integrating the $r$-th equation from 0 to $t^*$ for any $r = 1, \dots, m$ we get

$$\xi_r(t^*, s^*) = \xi_r(0, \tilde{s}_1) - \int_0^{t^*} \sum_{k=1}^{n} b_{rk}^{\sigma(t)}(z_r(t))\xi_k(t, z_r(t))dt$$

where we use the notation $\tilde{s}_1$ for $z_r(0; t^*, s^*)$ and $z_r(t)$ for $z_r(t; t^*, s^*)$. Using the bound (4.17), we obtain

$$|\xi_r(t^*, s^*)| \leqslant \|\bar{\xi}\|_\infty + \epsilon \int_0^{t^*} \|\boldsymbol{\xi}(t)\|_\infty dt.$$

Multiplying both sides by $e^{\beta t^*}$, and noting that $t^* \leqslant \delta < T_{\min}$, we obtain

$$\begin{aligned} e^{\beta t^*}|\xi_r(t^*, s^*)| &\leqslant e^{\beta t^*}\|\bar{\xi}\|_\infty + \epsilon \int_0^{t^*} e^{\beta(t^*-t)}\widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt \\ &\leqslant C_1\|\bar{\xi}\|_\infty + C_2 \int_0^{t^*} \widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt \end{aligned} \tag{4.20}$$

where $C_1 = e^{\beta T_{\min}}$ and $C_2 = e^{\beta T_{\min}}\epsilon$.

Case A.2: Integrating the $r$-th equation from $t_r$ to $t^*$ we get

$$\xi_r(t^*, s^*) = \xi_r(t_r, b) - \int_{t_r}^{t^*} \sum_{k=1}^{n} b_{rk}^{\sigma(t)}(z_r(t))\xi_k(t, z_t(t))dt.$$

Using $\xi_r(t_r, b) = \sum_{l=m+1}^{n} g_{rl}^{R,j}\xi_l(t_r, b)$ with $j = \sigma(t_r)$,

$$|\xi_r(t^*, s^*)| \leqslant \sum_{l=m+1}^{n} |g_{rl}^{R,j}||\xi_l(t_r, b)| + \epsilon \int_{t_r}^{t^*} \|\boldsymbol{\xi}(t)\|_\infty dt. \tag{4.21}$$

Integrating $l$-th equation from 0 to $t_r$ we get

$$\xi_l(t_r, b) = \xi_l(0, \tilde{s}_2) - \int_0^{t_r} \sum_{k=1}^{n} b_{lk}^{\sigma(t)}(z_l(t))\xi_k(t, z_l(t))dt,$$

where we use the notation $\tilde{s}_2$ for $z_l(0; t_r(t^*, s^*), b)$ and $z_l(t)$ for $z_l(t; t_r, b)$. Again using the bound (4.17),

$$|\xi_l(t_r, b)| \leqslant \|\bar{\xi}\|_\infty + \epsilon \int_0^{t_r} \|\boldsymbol{\xi}(t)\|_\infty dt$$

Substituting this bound in equation (4.21), we obtain

$$\begin{aligned} |\xi_r(t^*, s^*)| &\leqslant \tilde{K}_1\|\bar{\xi}\|_\infty + \epsilon\tilde{K}_1 \int_0^{t_r} \|\boldsymbol{\xi}(t)\|_\infty dt + \epsilon \int_{t_r}^{t^*} \|\boldsymbol{\xi}(t)\|_\infty dt \\ &\leqslant K_1\|\bar{\xi}\|_\infty + K_1\epsilon \int_0^{t^*} \|\boldsymbol{\xi}(t)\|_\infty dt \end{aligned}$$

where $\tilde{K}_1$ and $K_1$ are defined in (4.12). Multiplying by $e^{\beta t^*}$ and noting again that since $t^* \leqslant \delta < T_{\min}$, we have

$$e^{\beta t^*}|\xi_r(t^*, s^*)| \leqslant C_3\|\bar{\xi}\|_\infty + C_4 \int_0^{t^*} \widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt \tag{4.22}$$

with $C_3 = K_1 e^{\beta T_{\min}}$ and $C_4 = K_1\epsilon e^{\beta T_{\min}}$.

Combination of cases $A.1$, $A.2$. From inequalities (4.20) and (4.22) we obtain a combined estimate

$$e^{\beta t^*}|\xi_r(t^*, s^*)| \leqslant C_5\|\bar{\xi}\|_\infty + C_6 \int_0^{t^*} \widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt \tag{4.23}$$

with $C_5 = K_1 e^{\beta T_{\min}}$ and $C_6 = K_1\epsilon e^{\beta T_{\min}}$.

Estimate for paths with positive slope. Similarly, we can estimate $e^{\beta t^*}|\xi_l(t^*, s^*)|$ $(l = m+1, \ldots, n)$ for $(t^*, s^*) \in [0, \delta] \times (a, b)$ by considering the corresponding cases for $l$-th characteristic path $z_l(t; t^*, s^*)$ passing through $(t^*, s^*)$ $(l = m+1, \ldots, n)$. We have

$$e^{\beta t^*}|\xi_l(t^*, s^*)| \leqslant C_7\|\bar{\xi}\|_\infty + C_8 \int_0^{t^*} \widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt \tag{4.24}$$

with $C_7 = K_2 e^{\beta T_{\min}}$ and $C_8 = K_2\epsilon e^{\beta T_{\min}}$, where $K_2$ is defined in (4.12).

Estimate for all paths. From (4.23) and (4.24), by taking the maximum over $r$-th and $l$-th characteristic paths $(r = 1, \ldots, m$ and $l = m+1, \ldots, n)$ respectively, and taking the essential supremum over $s^* \in (a, b)$ we obtain the estimate

$$\| \widehat{\boldsymbol{\xi}(t^*)} \|_\infty \leqslant C_9\|\bar{\xi}\|_\infty + C_{10} \int_0^{t^*} \widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt \tag{4.25}$$

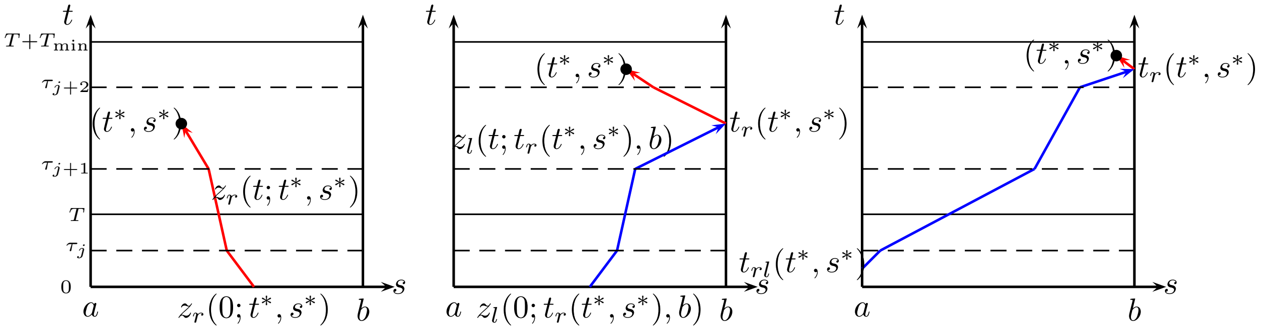


Figure 4.5: Illustration of cases for the proof of induction step.

with $C_9 = Ke^{\beta T_{\min}}$ and $C_{10} = K\epsilon e^{\beta T_{\min}}$, where $K$ is defined in (4.12).

Now, by using $c$ as defined in (4.14) and noting that $T_{\min} < 2T_{\max}$, we can write

$$\| \widehat{\boldsymbol{\xi}(t^*)} \|_\infty \leqslant C_{11}\|\bar{\xi}\|_\infty + C_{12}\int_0^{t^*} \widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt \tag{4.26}$$

with $C_{11} = c\theta e^{2\beta T_{\max}}$ and $C_{12} = K\epsilon e^{\beta T_{\min}}$. By applying Gronwall's lemma, we obtain the inequality for any $(t^*, s^*) \in [0, \delta] \times (a, b)$

$$\widehat{\|\boldsymbol{\xi}(t^*)\|}_\infty \leqslant C_{11}\exp(C_{12}T_{\min})\|\bar{\xi}\|_\infty, \tag{4.27}$$

for all $\sigma(\cdot)$. With the $\beta > 0$ and $\epsilon > 0$ chosen according to (4.15) and (4.16) respectively, we note that

$$\theta\exp(2\beta T_{\max})\exp\left(K\epsilon T_{\min}\exp(\beta T_{\min})\right) \leqslant \omega\exp\left(KT_{\min}\epsilon_1\frac{\omega}{\theta}\right) = \eta < 1.$$

Then by expanding the right-hand-side of inequality (4.27) we obtain,

$$\widehat{\|\boldsymbol{\xi}(t^*)\|}_\infty \leqslant c\left[\theta\exp(2\beta T_{\max})\exp\left(K\epsilon T_{\min}\exp(\beta T_{\min})\right)\right]\|\bar{\xi}\|_\infty < c\|\bar{\xi}\|_\infty$$

holds on $(t^*, s^*) \in [0, \delta] \times (a, b)$ for all switching signals $\sigma(\cdot) \in \mathcal{S}(\mathbb{R}_+, Q)$. Finally, using the definition (4.19), we obtain that

$$\|\boldsymbol{\xi}(t)\|_\infty \leqslant c\exp\left(-\beta t\right)\|\bar{\xi}\|_\infty, \quad 0 \leqslant t \leqslant \delta < T_{\min}.$$

This completes the proof of the induction basis.

**Part B.** Proof of the Induction Step. We will now show that under the chosen constants $\beta > 0, c \geqslant 1, \epsilon > 0$, if (4.18) holds on the domain $[0, T] \times (a, b)$, then it still holds on domain $[0, T + T_{\min}] \times (a, b)$. Let $T > 0$ and assume that (4.18) holds on $[0, T] \times (a, b)$. In this case we have to distinguish three cases as illustrated in Figure 4.5.

Proceeding as before, for any fixed $r = 1, \dots, m$, the $r$-th characteristic path $z_r(t; t^*, s^*)$ passing through $(t^*, s^*)$ considered backward in time, either intersects $t = 0$ within the interval $[a, b]$ before hitting any boundary (case $B.1$) or it intersects the line $s = b$ (case $B.2$); the points of intersection with the boundary of the domain are denoted by $(0, z_r(0; t^*, s^*))$ and $(t_r(t^*, s^*), b)$ respectively, where $z_r(t_r(t^*, s^*); t^*, s^*) = b$. Furthermore, the $l$-th characteristic path $z_l(t; t_r(t^*, s^*), b)$ passing through $(t_r(t^*, s^*), b)$ $(l = m+1, \dots, n)$ either $z_l(t; t_r(t^*, s^*), b)$ intersects the line $t = 0$ before hitting the line $s = a$ (case $B.2(i)$) or it hits $s = a$ (case $B.2(ii)$). The point of intersection is denoted by $(0, z_l(0; t_r(t^*, s^*), b))$ for case $B.2(i)$ and $(t_{rl}(t^*, s^*), a)$ for case $B.2(ii)$. We will again use $t_r$ for $t_r(t^*, s^*)$ and $t_{rl}$ for $t_{rl}(t^*, s^*)$.

Estimate for paths with negative slope. We first obtain an estimate of $e^{\beta t^*}|\xi_r(t^*, s^*)|$ for any $(t^*, s^*) \in [T + T_{\min}] \times (a, b)$ by considering the above three cases for the $r$-th characteristic path $z_r(t; t^*, s^*)$ passing through $(t^*, s^*)$ $(r = 1, \dots, m)$.

For case $B.1$: Using $j = \sigma(t)$ in (4.9), and integrating the $r$-th equation from 0 to $t^*$ for any $r = 1, \dots, m$, and using the bound (4.17),

$$\begin{aligned} |\xi_r(t^*, s^*)| &\leqslant \|\bar{\xi}\|_\infty + \epsilon \int_0^{t^*} \|\boldsymbol{\xi}(t)\|_\infty dt \\ &\leqslant \left(1 + \frac{\epsilon c}{\beta}\right) \|\bar{\xi}\|_\infty + \epsilon \int_T^{t^*} \|\boldsymbol{\xi}(t)\|_\infty dt \end{aligned}$$

where the second inequality is obtained using the assumption that (4.18) holds on $[0, T] \times (a, b)$. Multiplying both sides by $e^{\beta t^*}$, using definition (4.19); and noting that for the present situation (case $B.1$), we have $t^* \leqslant T_{\max}$, then $T \leqslant T_{\max}$, and for $t^* \in [T, T + T_{\min}]$, $t \in (T, t^*)$ then $(t^* - t) \leqslant T_{\min}$, we obtain

$$e^{\beta t^*}|\xi_r(t^*, s^*)| \leqslant C_{13}\|\bar{\xi}\|_\infty + C_{14} \int_T^{t^*} \widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt \tag{4.28}$$

with $C_{13} = (1 + \frac{\epsilon c}{\beta})e^{\beta T_{\max}}$ and $C_{14} = \epsilon e^{\beta T_{\min}}$.

For case $B.2$: Again integrating the $r$-th equation from $t_r$ to $t^*$, and using that $\xi_r(t_r, b) = \sum_{l=m+1}^{n} g_{rl}^{R,j} \xi_l(t_r, b)$ with $j = \sigma(t_r)$,

$$|\xi_r(t^*, s^*)| \leqslant \sum_{l=m+1}^{n} |g_{rl}^{R,j}||\xi_l(t_r, b)| + \epsilon \int_{t_r}^{t^*} \|\boldsymbol{\xi}(t)\|_\infty dt. \tag{4.29}$$

For case $B.2(i)$ Integrating $l$-th equation from 0 to $t_r$ and using the bound (4.17) we have

$$|\xi_l(t_r, b)| \leqslant \|\bar{\xi}\|_\infty + \epsilon \int_0^{t_r} \|\boldsymbol{\xi}(t)\|_\infty dt$$

Substituting this bound in equation (4.29), we obtain

$$
\begin{aligned}
|\xi_r(t^*,s^*)| &\leqslant \tilde{K}_1\|\bar{\xi}\|_\infty + \epsilon\tilde{K}_1\int_0^{t_r}\|\boldsymbol{\xi}(t)\|_\infty dt + \epsilon\int_{t_r}^{t^*}\|\boldsymbol{\xi}(t)\|_\infty dt\\
&\leqslant K_1\|\bar{\xi}\|_\infty + K_1\epsilon\int_0^{t^*}\|\boldsymbol{\xi}(t)\|_\infty dt\\
&\leqslant K_1\left(1+\frac{\epsilon c}{\beta}\right)\|\bar{\xi}\|_\infty + K_1\epsilon\int_T^{t^*}\|\boldsymbol{\xi}(t)\|_\infty dt,
\end{aligned}
$$

where the last inequality is obtained using the assumption that (4.18) holds on $[0,T]\times(a,b)$. Noting that for the present situation (case $B.2(i)$), $t^*\leqslant 2T_{\max}$ then $T\leqslant 2T_{\max}$, and for $t^*\in[T,T+T_{\min}]$, $t\in(T,t^*)$ then $(t^*-t)\leqslant T_{\min}$ we obtain

$$
e^{\beta t^*}|\xi_r(t^*,s^*)| \leqslant C_{15}\|\bar{\xi}\|_\infty + C_{16}\int_T^{t^*}\widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt \tag{4.30}
$$

with $C_{15}=K_1\left(1+\frac{\epsilon c}{\beta}\right)e^{2\beta T_{\max}}$ and $C_{16}=K_1\epsilon e^{\beta T_{\min}}$.

For case $B.2(ii)$, we have

$$
\xi_l(t_r,b)=\xi_l(t_{rl},a)-\int_{t_{rl}}^{t_r}\sum_{k=1}^{n}b_{lk}^{\sigma(t)}(z_l(t))\xi_k(t,z_l(t))dt
$$

Using $\xi_l(t_{rl},a)=\sum_{p=1}^{m}g_{lp}^{L,j'}\xi_p(t_{rl},a)$ with $j'=\sigma(t_{rl})$, we have

$$
|\xi_l(t_r,b)|\leqslant\sum_{p=1}^{m}|g_{lp}^{L,j'}||\xi_p(t_{rl},a)|+\epsilon\int_{t_{rl}}^{t_r}\|\boldsymbol{\xi}(t)\|_\infty dt
$$

Substituting this bound in equation (4.29), and using the induction hypothesis, we obtain

$$
\begin{aligned}
|\xi_r(t^*,s^*)| &\leqslant \theta c e^{-\beta t_{rl}}\|\bar{\xi}\|_\infty + K_1\epsilon\int_{t_{rl}}^{t^*}\|\boldsymbol{\xi}(t)\|_\infty dt\\
&\leqslant \left(\theta+\frac{K_1\epsilon}{\beta}\right)ce^{-\beta t_{rl}}\|\bar{\xi}\|_\infty + K_1\epsilon\int_T^{t^*}\|\boldsymbol{\xi}(t)\|_\infty dt
\end{aligned}
$$

with $\theta$ as in (4.13). Again, noting that for $t^*\in[T,T+T_{\min}]$, in the present situation (case $B.2(ii)$), $0\leqslant t_{rl}\leqslant T$ and $T-t_{rl}\leqslant 2T_{\max}$, we obtain

$$
e^{\beta t^*}|\xi_r(t^*,s^*)| \leqslant C_{17}\|\bar{\xi}\|_\infty + C_{18}\int_T^{t^*}\widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt \tag{4.31}
$$

with $C_{17}=c\left(\theta+\frac{K_1\epsilon}{\beta}\right)e^{2\beta T_{\max}}$ and $C_{18}=K_1\epsilon e^{\beta T_{\min}}$.

Combination of cases $B.1$, $B.2(i)$ and $B.2(ii)$. From inequalities (4.28), (4.30), (4.31) and the $K$ defined in (4.12), we obtain

$$e^{\beta t^*}|\xi_r(t^*,s^*)| \leqslant C_{19}\|\bar{\xi}\|_\infty + C_{20}\int_T^{t^*} \widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt, \tag{4.32}$$

with $C_{19} = c\left(\theta + \frac{K\epsilon}{\beta}\right)e^{2\beta T_{\max}}$ and $C_{20} = K\epsilon e^{\beta T_{\min}}$.

Estimate for paths with positive slope. By using similar arguments, we also obtain an estimate of $e^{\beta t^*}|\xi_l(t^*,s^*)|$ for any $(t^*,s^*) \in [T+T_{\min}]\times(a,b)$ by considering the corresponding cases for the $l$-th characteristic path $z_l(t;t^*,s^*)$ passing through $(t^*,s^*)$ $(l = m+1,\ldots,n)$, for $c$ chosen according to (4.14), and $K$ defined in (4.12)

$$e^{\beta t^*}|\xi_l(t^*,s^*)| \leqslant C_{21}\|\bar{\xi}\|_\infty + C_{22}\int_T^{t^*} \widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt, \tag{4.33}$$

with $C_{21} = c\left(\theta + \frac{K\epsilon}{\beta}\right)e^{2\beta T_{\max}}$, $C_{22} = K\epsilon e^{\beta T_{\min}}$.

Estimate for all paths. We now combine (4.32) and (4.33) by taking the maximum over $r$-th and $l$-th characteristic paths $(r = 1,\ldots,m$ and $l = m+1,\ldots,n)$ respectively, taking the essential supremum over $s^* \in (a,b)$ to obtain the estimate

$$\widehat{\|\boldsymbol{\xi}(t^*)\|}_\infty \leqslant C_{23}\|\bar{\xi}\|_\infty + C_{24}\int_T^{t^*} \widehat{\|\boldsymbol{\xi}(t)\|}_\infty dt \tag{4.34}$$

where $C_{23} = c\left(\theta + \frac{K\epsilon}{\beta}\right)e^{2\beta T_{\max}}$, $C_{24} = K\epsilon e^{\beta T_{\min}}$. By applying Gronwall's lemma, we obtain the inequality

$$\widehat{\|\boldsymbol{\xi}(t^*)\|}_\infty \leqslant C_{23}\exp(C_{24}(t^*-T))\|\bar{\xi}\|_\infty \tag{4.35}$$

for any $(s^*,t^*) \in [T, T+T_{\min}]\times(a,b)$ and thus

$$\widehat{\|\boldsymbol{\xi}(t^*)\|}_\infty \leqslant C_{23}\exp(C_{24}T_{\min}))\|\bar{\xi}\|_\infty$$

for all $\sigma(\cdot)$. Using (4.34), plugging in the expressions for $C_{23}$ and $C_{24}$, given the expression of $\beta$ in (4.15) we obtain the inequality

$$\widehat{\|\boldsymbol{\xi}(t^*)\|}_\infty \leqslant c\left(\theta + \frac{K\epsilon}{\beta}\right)\frac{\omega}{\theta}\exp\left(K\epsilon\frac{\omega}{\theta}T_{\min}\right)\|\bar{\xi}\|_\infty. \tag{4.36}$$

With $\beta$ and $\epsilon$ given by (4.15) and (4.16) respectively, we have

$$\left(\theta + \frac{K\epsilon}{\beta}\right)\frac{\omega}{\theta}\exp\left(K\epsilon\frac{\omega}{\theta}T_{\min}\right) \leqslant \left(\theta + \frac{K\epsilon_2}{\beta}\right)\frac{\omega}{\theta}\exp\left(K\epsilon_1\frac{\omega}{\theta}T_{\min}\right) = 1,$$

and using this in the right hand side of (4.36) we obtain

$$\|\widehat{\boldsymbol{\xi}(t^*)}\|_\infty \leqslant c\|\bar{\xi}\|_\infty$$

holds on $(t^*, s^*) \in [T, T + T_{\min}] \times (a, b)$ for all switching signals $\sigma(\cdot) \in \mathcal{S}(\mathbb{R}_+, Q)$. Finally, from (4.19) we obtain that

$$\|\boldsymbol{\xi}(t)\|_\infty \leqslant c \exp(-\beta t)\|\bar{\xi}\|_\infty, \quad 0 \leqslant t \leqslant T + T_{\min}.$$

This completes the proof of the induction step. □

*Remark* 4.3.3. From the proof of Theorem 4.3.2 we see that with $K$ and $\theta$ given by (4.12) and (4.13) respectively, and the constants $\omega$ and $\eta$ chosen such that $\theta < \omega < \eta < 1$, equation (4.16) gives a concrete value of $\epsilon$ for which the conditions of Theorem 4.3.2 guarantee exponential stability for all switching signals. That is, (4.18) holds uniformly for all switching signals $\sigma(\cdot) \in \mathcal{S}(\mathbb{R}_+, Q)$ with $c$ and $\beta$ given by (4.14) and (4.15) respectively. We then see that the so obtained bound on $\|B^j(s)\|_\infty$ satisfies $\epsilon \to 0$ as $\theta \to 1$. Similar conditions are known for the unswitched case, where such systems with sufficiently small inhomogeneities are called 'almost conservative' Bastin et al. [2008].

For an illustration of the decay estimate and the size of $\epsilon$ obtained by Theorem 4.3.2 and Remark 4.3.3 we provide the following example.

*Example* 4.3.4. Consider a switched system of the form (4.5) with two modes ($Q = \{1, 2\}$) and $[a, b] = [0, 1]$. The parameters and boundary data are specified as

$$\begin{aligned} \Lambda^1 &= \begin{bmatrix} -1.2 & 0 \\ 0 & 1.8 \end{bmatrix}, \quad B^1 = \begin{bmatrix} -0.005 & 0 \\ 0 & -0.005 \end{bmatrix}, \\ \Lambda^2 &= \begin{bmatrix} -0.8 & 0 \\ 0 & 1.4 \end{bmatrix}, \quad B^2 = \begin{bmatrix} 0 & 0.005 \\ 0.005 & 0 \end{bmatrix}, \\ G^1_{\mathrm{L}} &= 0.61,\ G^1_{\mathrm{R}} = 1.15,\ G^2_{\mathrm{L}} = 0.42,\ G^2_{\mathrm{R}} = 1.21. \end{aligned} \tag{4.37}$$

In this example the hypotheses (A1) and (A2) of Theorem 4.3.2 are clearly satisfied. We have $K = 1.21$ and

$$\theta = \max_{j,j' \in Q} \rho\left(\begin{bmatrix} 0 & |G^{j'}_{\mathrm{R}}| \\ |G^j_{\mathrm{L}}| & 0 \end{bmatrix}\right) = 0.7381 < 1. \tag{4.38}$$

Following Remark 4.3.3, we choose $\omega = 0.87$ and $\eta = 0.88$ to obtain that $\|B^{1,2}\|_\infty = 0.0050 < \epsilon = 0.0054$. Therefore, according to Theorem 4.3.2, the switched system is absolutely exponentially stable. Moreover, for equation (4.18), we obtain $c = 1.6393$ and $\beta = 0.0658$ from equations (4.14) and (4.15) respectively. For initial data $\bar{\xi}(s) = \begin{bmatrix} 1 & 1 \end{bmatrix}^\top$ on $s \in (0, 1)$, the exponential bound in (4.18) is plotted together with the observed decay of $\|\boldsymbol{\xi}(t)\|_\infty$ for three different switching signals $\sigma(\cdot)$ in Figure 4.6. The solution approximations are computed using the two-step Lax-Friedrichs finite difference scheme from Shampine [2005]. □

In general, assumption (A2) is necessary for exponential stability under arbitrary switching as evident from the following example.

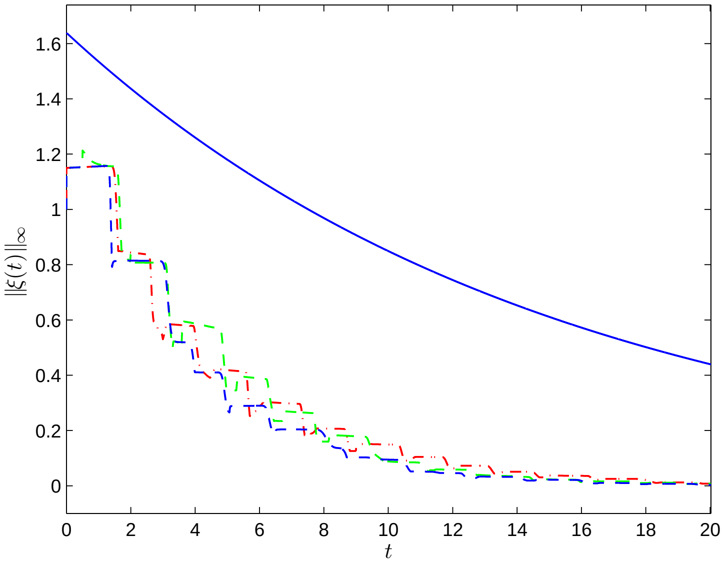


Figure 4.6: Bound from Thm. 4.3.2 (solid) & example switching signals (dashed).

*Example* 4.3.5. Let $Q = \{1, 2\}$, $[a, b] = [0, 1]$, $\Lambda^1 = \text{diag}(-1, 1, 1)$, $\Lambda^2 = \text{diag}(-1, -1, 1)$, $B^j = \text{diag}(0, 0, 0)$ and let $G^1_L$, $G^2_R$, $G^2_L$, and $G^1_R$ be any boundary data of appropriate dimensions. It is clear that this example satisfies assumption (A1) but does not satisfy (A2). Now consider initial data $\bar{\xi}(s) = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top$ on $s \in (0, 1)$, and a switching signal $\sigma(t)$ defined over the switching times $\tau_k = 0.5k$, where $k = 0, 1, 2, \ldots$ and $\sigma(\tau_0) = 1$, $\sigma(\tau_1) = 2$, $\sigma(\tau_2) = 1$ and so on. For the second component of the solution $\boldsymbol{\xi}(t)$, we then have $\xi_2(t, s) = 1$ for $s$ almost everywhere on the interval $(0, 0.5)$ and $t = 1, 2, 3, \ldots$. Hence, the solution $\|\boldsymbol{\xi}(t)\|_\infty$ cannot decay exponentially irrespective of the decay that might be imposed on $\xi_1(t, s)$, $\xi_2(t, s)$ and $\xi_3(t, s)$ by the boundary data. □

A consequence of our results is that, when the only stabilizing mechanism is at the boundary and arbitrary changes of sign of the eigenvalues of $\Lambda$ cannot be ruled out a-priori, the decay of the solution can in general not be concluded from the rate of decay at the boundary (for e. g., in terms of condition (4.11) of Theorem 4.3.2).

*Remark* 4.3.6. The condition (4.11) implies the following spectral radius condition to hold for the subsystems (4.7) with $j \in Q$ fixed:

$$\rho\left(\begin{bmatrix} 0 & |G^j_{\mathrm{R}}| \\ |G^j_{\mathrm{L}}| & 0 \end{bmatrix}\right) < 1. \tag{4.39}$$

Under this assumption, classical solutions of (4.7) are known to be exponentially stable Li [1994]. However, assumption (4.39) for all $j \in Q$ is not sufficient for the switching system to be exponentially stable. Note that $G^j_{\mathrm{L}}, G^j_{\mathrm{R}}$ in Example 4.3.1 satisfy (4.39) but not (4.11) for $j = 1, 2$, i. e.,

$$\rho\left(\begin{bmatrix} 0 & 1.5 \\ 0 & 0 \end{bmatrix}\right) = \rho\left(\begin{bmatrix} 0 & 0 \\ 1.5 & 0 \end{bmatrix}\right) = \rho\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\right) = 0,$$

but

$$\rho\left(\begin{bmatrix} 0 & 1.5 \\ 1.5 & 0 \end{bmatrix}\right) = 1.5.$$

Nevertheless, as shown next in Corollary 4.3.7, the switched system satisfying (4.39) in every mode $j$ can be stabilized by switching slow enough. Note that Corollary 4.3.7 does not require assumption (A2) to hold.

**Corollary 4.3.7.** *(Dwell-Time) Under the hypotheses (A1), there exists an $\epsilon > 0$ such that if $\|B^j(s)\|_\infty < \epsilon$ for all $s \in [a, b]$ and $j \in Q$, the switching system in diagonal form (4.5) is exponentially stable with respect to the norm $\|\cdot\|_\infty$ for all switching signals in $\mathcal{S}_\tau(\mathbb{R}_+; Q)$ for which the dwell-time $\tau > \bar{\tau}$ ($\bar{\tau}$ given by (4.10)) if the condition (4.39) holds for all $j \in Q$.*

*Proof.* From the definition of $\bar{\tau}$ in (4.10) it is easy to see that if $\tau > \bar{\tau}$, then in case $B.2(ii)$, $t_r$ and $t_{rl}$ lie in the same inter switching interval and all the required estimates can be made using a $\tilde{\theta}$ defined similar to (4.13) but where the maximum is only taken over $j \in Q$. □

## 4.4 Non-diagonal Switching System

We now focus on non-diagonal systems. Suppose that the system switches among non-diagonal subsystems

$$\begin{cases} \dfrac{\partial u^j}{\partial t} + A^j(s)\dfrac{\partial u^j}{\partial s} + \tilde{B}^j(s)u^j = 0, \;\; s \in (a, b), \;\; t > 0 \\ D_L^j u^j(t, a) = 0, \quad D_R^j u^j(t, b) = 0, \;\; t \geqslant 0 \end{cases} \tag{4.40}$$

where, for each $j \in Q$, $A^j(s)$, $B^j(s)$ are $n \times n$ dimensional matrix functions on $(a, b)$ and $D_L^j$, $D_R^j$ are constant matrices of appropriate dimensions. Each subsystem can be written in the diagonal form (4.7) under certain assumptions. For instance, if we impose that for each $j \in Q$,

(A1)* The matrix function $A^j(\cdot)\colon [a, b] \mapsto \mathbb{R}^{n\times n}$ is Lipschitz-continuous such that for all $s \in [a, b]$, there exists $m_j$ such that $0 < m_j < n$ and $A^j(s)$ has $m_j$ negative and $(n - m_j)$ positive eigenvalues $\lambda_i^j(s)$ with $n$ corresponding linearly independent left (resp. right) eigenvectors $l_i^j(s)$ (resp. $r_i^j(s)$), $i = 1, \ldots, n$ all Lipschitz-continuous functions of $s$. The matrix function $\tilde{B}^j(\cdot)\colon [a, b] \mapsto \mathbb{R}^{n\times n}$ is bounded measurable with respect to $s$. Furthermore, the following two rank conditions hold for $D_L^j \in \mathbb{R}^{(n-m_j)\times n}$ and $D_R^j \in \mathbb{R}^{m_j\times n}$

$$\operatorname{rank}\left[(D_L^j)^\top \middle| l_1^j(a)\middle| \cdots \middle| l_{m_j}^j(a)\right] = n$$
$$\operatorname{rank}\left[(D_R^j)^\top \middle| l_{m_j+1}^j(b)\middle| \cdots \middle| l_n^j(b)\right] = n.$$

Under the assumption (A1$^*$) the matrix functions $S_j(\cdot) = [l_1^j(\cdot)\big| \dots \big| l_n^j(\cdot)]^\top$ and $S_j^{-1}(\cdot) = [r_1^j(\cdot)\big| \dots \big| r_n^j(\cdot)]^\top$ are Lipschitz-continuous functions with partial derivatives defined a. e. We refer the reader to the text by Bressan Bressan [2000], pages $46 - 50$, for the details about assumption (A1)$^*$.

For all $s \in [a,b]$, we have

$$S_j(s)A^j(s)S_j^{-1}(s) = \Lambda^j(s). \tag{4.41}$$

with $\Lambda^j(s)$ as in (A1). By applying a transformation $u^j(t,s) = S_j^{-1}(s)\xi^j(t,s)$, $\tilde{D}_L^j = D_L^j S_j^{-1}(a)$ and $\tilde{D}_R^j = D_R^j S_j^{-1}(b)$ and using the representation

$$\begin{aligned}
&B^j(s) = S_j(s)\left(A^j(s)\frac{\partial}{\partial s}S_j^{-1}(s) + \tilde{B}^j(s)S_j^{-1}(s)\right),\\
&\tilde{D}_L^j = [\tilde{D}_{L,I}^j \big| \tilde{D}_{L,II}^j],\ \tilde{D}_R^j = [\tilde{D}_{R,I}^j \big| \tilde{D}_{R,II}^j],\\
&G_{\mathrm{L}}^j = -(\tilde{D}_{L,II}^j)^{-1}\tilde{D}_{L,I}^j,\ G_{\mathrm{R}}^j = -(\tilde{D}_{R,I}^j)^{-1}\tilde{D}_{R,II}^j,
\end{aligned} \tag{4.42}$$

with $\tilde{D}_{L,I}^j \in \mathbb{R}^{(n-m_j)\times m_j}$, $\tilde{D}_{L,II}^j \in \mathbb{R}^{(n-m_j)\times(n-m_j)}$, $\tilde{D}_{R,I}^j \in \mathbb{R}^{m_j\times m_j}$, $\tilde{D}_{R,II}^j \in \mathbb{R}^{m_j\times(n-m_j)}$, $G_{\mathrm{L}}^j \in \mathbb{R}^{(n-m_j)\times m_j}$ and $G_{\mathrm{R}}^j \in \mathbb{R}^{m_j\times(n-m_j)}$, the system corresponding to (4.40) and initial data $\bar{u}(s)$ corresponding to mode $j$ becomes (4.7) with initial data $\bar{\xi}^j(s) = S_j(s)\bar{u}(s)$.

Now observing that the switching system in the non-diagonal form for a switching signal $\sigma(\cdot) \in \mathcal{S}(\mathbb{R}_+, Q)$

$$\begin{cases}
\dfrac{\partial u}{\partial t} + A^{\sigma(t)}(s)\dfrac{\partial u}{\partial s} + \tilde{B}^{\sigma(t)}(s)u = 0,\ s \in (a,b),\ t > 0\\
D_L^{\sigma(t)}u(t,a) = 0, \quad D_R^{\sigma(t)}u(t,b) = 0,\ t \geqslant 0\\
u(0,s) = \bar{u}(s), \quad s \in (a,b)
\end{cases} \tag{4.43}$$

can be written as a switching system in the diagonal form with discontinuous resets at the switching times $\tau_k$ for $k = 1, 2, \dots$ and $\tau_0 = 0$, i. e.,

$$\begin{cases}
\dfrac{\partial \xi}{\partial t} + \Lambda^{\sigma(t)}(s)\dfrac{\partial \xi}{\partial s} + B^{\sigma(t)}(s)\xi = 0,\ t \in [\tau_k, \tau_{k+1}]\\
\xi_{\mathrm{II}}(t,a) = G_{\mathrm{L}}^{\sigma(t)}\xi_{\mathrm{I}}(t,a),\ \xi_{\mathrm{I}}(t,b) = G_{\mathrm{R}}^{\sigma(t)}\xi_{\mathrm{II}}(t,b),\\
\xi(0,\cdot) = \bar{\xi}(\cdot) = S_{\sigma(\tau_0)}(\cdot)\bar{u}(\cdot),\\
\xi(\tau_k,\cdot) = S_{j_k}(\cdot)S_{j_{k-1}}^{-1}(\cdot)\lim\limits_{t\to\tau_k, t<\tau_k}\xi(\tau_k,\cdot),\ k > 0,
\end{cases} \tag{4.44}$$

Our next proposition is a very simple consequence of simultaneous diagonalization.

**Proposition 4.4.1.** *Under hypotheses (A1$^*$)-(A2) and under the pairwise commutativity assumption that for all $s \in [a,b]$ and for all $j, j' \in Q$*

$$A^j(s)A^{j'}(s) = A^{j'}(s)A^j(s), \tag{4.45}$$

*and let $G_{\mathrm{L}}^j$, $G_{\mathrm{R}}^j$ and $B^j(s)$ are given by (4.42). Then, if condition (4.11) holds for all $j, j \in Q$, there exists an $\epsilon > 0$ such that if $\|B^j(s)\|_\infty < \epsilon$ for all $s \in [a,b]$ and $j \in Q$, the switching system in non-diagonal form (4.43) is absolutely exponentially stable in $\|\cdot\|_\infty$.*

*Proof.* Recall that a set of diagonalizable matrices are simultaneously diagonalizable if (and only if) they commute. Thus, system (4.43) can be transformed into a switching system in diagonal form (4.44) with a common diagonalizing matrix function $S^j(\cdot) \equiv S(\cdot)$. The assertion then follows. □

Though the commutativity assumption in Proposition 4.4.1 seems very strong, we include an example showing that it is in general necessary for conditions such as in Section 4.3 to be sufficient for absolutely exponential stability.

*Example* 4.4.2. Consider a non-diagonal switching system of form (4.43) with two modes ($Q = \{1,2\}$) and initial data $\bar{u}(s) = \begin{bmatrix}1 & 1\end{bmatrix}^\top$ on $s \in (a,b)$, for an alternating switching signal $\sigma(\cdot)$ with switching times $\tau_k = 0.5k$ where $k = 0,1,2,3,\dots$ and $\sigma(\tau_0) = 1$, $\sigma(\tau_1) = 2$, $\sigma(\tau_2) = 1$ and so on. The parameters and boundary data are specified as

$$A^1 = \begin{bmatrix}-1 & 0\\ 0 & +1\end{bmatrix}, \quad A^2 = \begin{bmatrix}-1 & -4\\ 0 & +1\end{bmatrix}, \quad B^{1,2} = \begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$$
$$D_L^1 = \begin{bmatrix}-\frac{3}{2} & 1\end{bmatrix}, \quad D_L^2 = \begin{bmatrix}-\frac{3}{4} & -1\end{bmatrix},$$
$$D_R^1 = \begin{bmatrix}1 & -\frac{1}{4}\end{bmatrix}, \quad D_R^2 = \begin{bmatrix}1 & \frac{7}{4}\end{bmatrix}$$

The non-diagonal system so specified satisfies (A1*)-(A2) but does not satisfy the commutativity condition (4.45) ($A_1A_2 \neq A_2A_1$). With

$$S_1 = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \quad S_2 = \begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix},$$

and doing a change of variables by this transformation, both the constituting subsystems of the non-diagonal switching system reduce to the same diagonal subsystem

$$\begin{gathered}\frac{\partial \xi}{\partial t} + \begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}\frac{\partial \xi}{\partial s} = 0,\\ \xi^2(t,0) = \frac{3}{2}\xi^1(t,0), \quad \xi^1(t,1) = \frac{1}{4}\xi^2(t,1),\end{gathered} \tag{4.46}$$

which satisfied the spectral radius condition

$$\rho\left(\begin{bmatrix}0 & \frac{1}{4}\\ \frac{3}{2} & 0\end{bmatrix}\right) = 0.6124 < 1,$$

implying that the solution of the subsystem (4.46) starting with initial condition for $\bar{\xi}(s) = \begin{bmatrix}1 & 1\end{bmatrix}^\top$, $s \in (0,1)$ decays exponentially for $t \to \infty$. However, following the representation (4.44), we observe that for the non-diagonal switching system $\|\mathbf{u}(t)\|_\infty$ is not bounded as $t \to \infty$. See Figure 4.7 for the growth of $\|\mathbf{u}(t)\|_\infty$ where the solution $\mathbf{u}$ is again obtained by using a two-step Lax-Friedrichs scheme as in Example 4.3.4. □

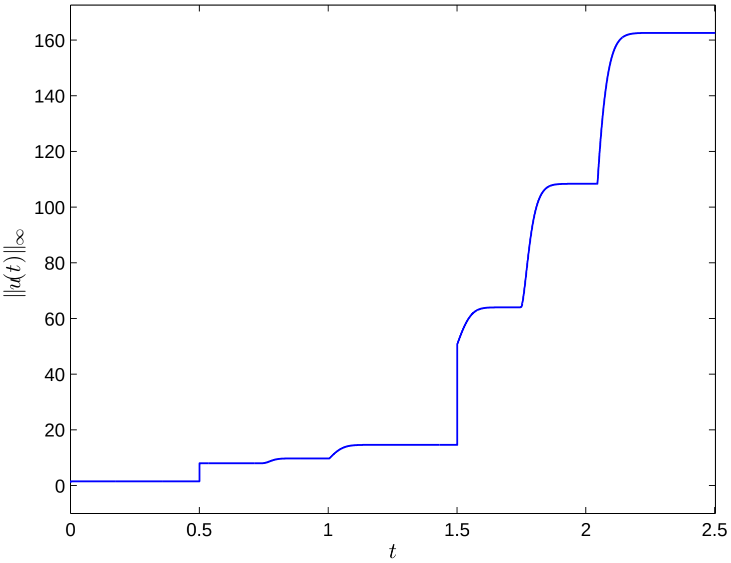


Figure 4.7: Blowup for the system considered in Example 4.4.2.

## 4.5 Application to Stability of Canal Cascade under DoS attacks

We apply the stability results to water flow in a cascade of $m$ canal reaches as depicted in Figure 4.8 (a). Consider a setting in which DoS attacks trigger switching of the boundary control actions. The control actions are applied at the underflow sluice gates, and affect the corresponding gate openings $w_i^j$ for reach $i$ in mode $j$. Theorem 4.3.2 can be applied to investigate the stability of linearized dynamics to a steady-state flow in such a $m-$reach cascade of open channels for attack scenarios in which boundary control actions switch between a number of modes.

The flow of water in reach $i$ is characterized by velocity $V_i(t,s)$ and elevation $H_i(t,s)$. For horizontal, prismatic canals with rectangular cross-section and frictionless walls, the flow under gravity $g$ satisfies the Saint-Venant equations Leugering and Schmidt [2002]

$$\frac{\partial}{\partial t}\begin{pmatrix} H_i \\ V_i \end{pmatrix} + \begin{pmatrix} V_i & H_i \\ g & V_i \end{pmatrix}\frac{\partial}{\partial s}\begin{pmatrix} H_i \\ V_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{4.47}$$

for $i = 1,\ldots,m$, each defined on the domain $\{(t,s) : 0 \leqslant t < \infty, 0 \leqslant s \leqslant 1\}$. Following deHalleux et al. [2003], let the initial data be given by $H_i(0,s)$, $V_i(0,s)$ and the boundary conditions modeling decentralized feedback control actions in mode $j$ together

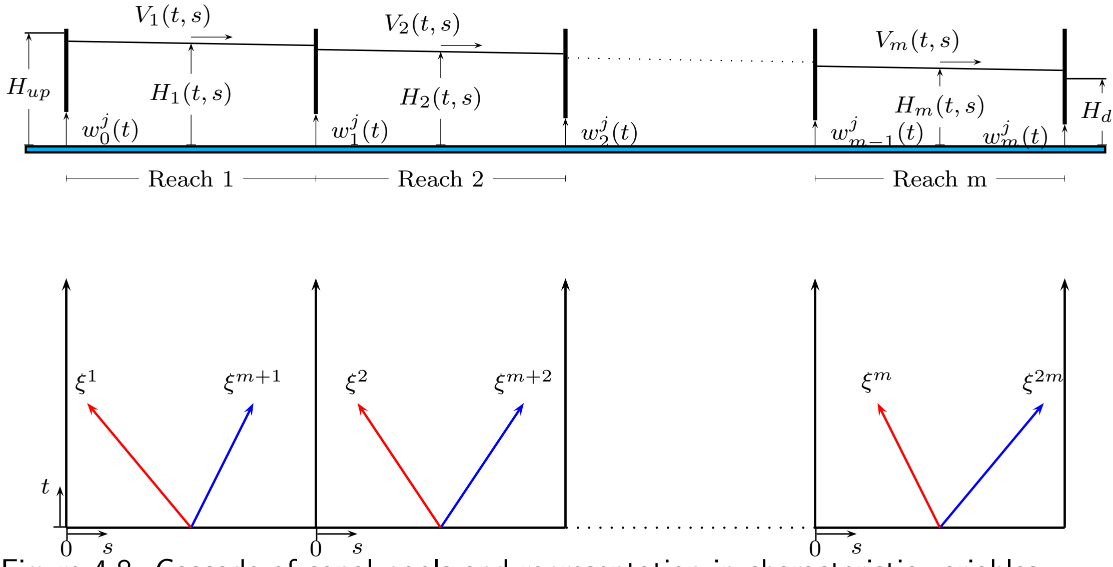


Figure 4.8: Cascade of canal pools and representation in characteristic variables.

with flow conservation for each reach $i$ be given by

$$\begin{aligned}
&f_1^j(w_0^j(t), H_{\text{up}}, H_1(t,0), V_1(t,0)) = 0\\
&f_i^j(w_i^j(t), H_i(t,1), H_{i+1}(t,0), V_i(t,1)) = 0\\
&f_m^j(w_m^j(t), H_m(t,1), H_{\text{do}}, V_m(t,1)) = 0\\
&H_i(t,1)V_i(t,1) - H_{i+1}(t,0)V_{i+1}(t,0) = 0
\end{aligned}$$

where $H_{\text{up}}$, $H_{\text{do}}$ are the (known) up and down stream water levels.

Assume that under constant gate openings $\bar{w}_i$ and constant $H_{\text{up}}$, $H_{\text{do}}$, each reach attains a uniform steady state $(\bar{H}_i, \bar{V}_i)$ such that $H_{\text{do}} < \bar{H}_m < \ldots < \bar{H}_1 < H_{\text{up}}$ and $\bar{H}_1\bar{V}_1 > 0$. Using $v_i(x,t) = V_i(x,t) - \bar{V}_i$ and $h_i(x,t) = H_i(x,t) - \bar{H}_i$, the linearized model can be written as

$$\frac{\partial}{\partial t}\begin{pmatrix} h_i \\ v_i \end{pmatrix} + \begin{pmatrix} \bar{V}_i & \bar{H}_i \\ g & \bar{V}_i \end{pmatrix} \frac{\partial}{\partial s}\begin{pmatrix} h_i \\ v_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{4.48}$$

with initial conditions $h_i(0,\cdot)$, $v_i(0,\cdot)$ for $i = 1,\ldots,m$. The traditional Riemann coordinate change Leugering and Schmidt [2002] $\xi_i(t,s) = h_i(t,s) + v_i\sqrt{\bar{H}_i/g}$, $\xi_{m+i}(t,s) = h_i(t,s) - v_i\sqrt{\bar{H}_i/g}$ leads to a diagonal system:

$$\frac{\partial}{\partial t}\begin{pmatrix} \xi_i \\ \xi_{m+i} \end{pmatrix} + \begin{pmatrix} \lambda_i & 0 \\ 0 & \lambda_{m+i} \end{pmatrix} \frac{\partial}{\partial s}\begin{pmatrix} \xi_i \\ \xi_{m+i} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{4.49}$$

with $\lambda_i = (\sqrt{g\bar{H}_i} - \bar{V}_i)$ and $\lambda_{m+i} = (\sqrt{g\bar{H}_i} + \bar{V}_i)$.

Under sub-critical flow, the eigenvalues satisfy $\lambda_i < 0 < \lambda_{m+i}$. For the system of $m-$canal reaches, equation (4.49) can be written in the form

$$\partial_t\xi + \Lambda\partial_s\xi = 0, \tag{4.50}$$

where $\xi = (\xi_1,\ldots,\xi_m,\xi_{m+1},\ldots,\xi_{2m})^\top$ and $\Lambda = \text{diag}(\lambda_1,\ldots,\lambda_{2m})$ (see Figure 4.8 (b)). Moreover, setting $\xi_{\text{I}} = (\xi_1,\ldots,\xi_m)$, $\xi_{\text{II}} = (\xi_{m+1},\ldots,\xi_{2m})$ and taking into account the coordinate transformation while assuming sufficient regularity of $f_i^j$, the boundary conditions in linearized form for each $j$ can be rewritten as

$$\xi_{\mathrm{II}}(t,0) = G_L^j \xi_{\mathrm{I}}(t,0) \quad \xi_{\mathrm{I}}(t,1) = G_R^j \xi_{\mathrm{II}}(t,1) \tag{4.51}$$

with appropriately defined jacobians $G_L^j$, $G_R^j$ (for details on the derivation for an explicit control law $f_i^j$ see deHalleux et al. [2003]).

Our results from Section 4.3 provide a set of sufficient conditions for solutions of (4.50)-(4.51) to decay for any admissible attack scenarios.

## 4.6 Discussion

We presented a generalization of a well-known mechanism for stability of hyperbolic PDE systems Li [1994] to the case in which the switching occurs among a set of systems that may differ in the system matrix function and/or boundary conditions. When constituent PDEs are in the canonical diagonal form, we derived a sufficient condition for exponential stability under arbitrary switching signals. For the case in which the system matrix functions are not diagonal, the result holds when they are jointly diagonalizable. This results in a commutativity condition that has a counterpart in the switched ODE literature D. Liberzon [2003].

It is also clear that, although the switching signal represents joint switching of the boundary conditions and system matrices, the results apply for switching the boundary conditions or system matrices individually by introducing appropriate auxiliary modes, which is just a matter of notational convenience. Thus, the treatment presented in this chapter might be of interest in control settings under abruptly changing boundary conditions and operating regimes such as the opening and closing of gates in a cascade of open-canal pools, the dynamics of which are classically modeled by the linearized Saint-Venant equations Bastin et al. [2008].

A limitation of the results obtained here is that they are valid only for almost conservative systems (see Remark 4.3.3). Thus, it will be interesting to investigate if, possibly by using different methods, other conditions can be found that guarantee absolute exponential stability for less conservative systems. In particular, our results motivate a Lyapunov theory for switching infinite dimensional systems.

Finally, we also argued that the results presented here can be used to study stability of a cascade of canal pools under attack scenarios which result in arbitrary switching of boundary control actions.

# Chapter 5

# Detection of Deception Attacks on Process Control Systems

## 5.1 Introduction

In the previous chapters we have described the cyber-security threat assessment, attack diagnosis, and resilient control methods for water SCADA systems used to operate cascaded canal networks. The developments presented earlier use the tools from robust and switched control system theory to provide control system guarantees for water SCADA systems against a class of deception and DoS attacks. In this chapter, we extend the ideas presented earlier to develop cyber-security tools for process control systems.

In the last years there has been an increasing interest in the security of process control and SCADA systems. Furthermore, recent computer attacks such as the Stuxnet worm, have shown there are parties with the motivation and resources to effectively attack process control systems. While a significant body of research work exists on the security mechanisms of process control systems, few researchers have explored practically implementable solutions for securing these systems. In particular, the sophistication of new malware attacking control systems–malware including zero-days attacks, rootkits created for control systems, and software signed by trusted certificate authorities–has shown that it is very difficult to prevent and detect these attacks based solely on IT system information.

In this chapter it is shown how, by incorporating knowledge of the physical system under control, one can detect computer attacks that change the behavior of the targeted control system. By using knowledge of the physical system we are able to focus on the final objective of the attack, and not on the particular mechanisms of how vulnerabilities are exploited, and how the attack is hidden. We also analyze the safety of our mechanisms by exploring the effects of stealthy attacks, and by ensuring that automatic attack-response mechanisms will not drive the system to an unsafe state.

An accurate assessment of potential losses under cyber-attacks is a pre-requisite for any risk management program. Risk management is the process of shifting the odds in your favor by finding among all possible alternatives, the one that minimizes the impact of uncertain events. Probably the best well known risk metric is the average loss $R_\mu = \mathbb{E}[L] \approx \sum_i L_i p_i$, where $L_i$ is the loss if event $i$ occurs, and $p_i$ is the probability that event $i$ occurs. Other risk metrics try to get more information about the probability distribution of the losses, and not only its mean value ($R_\mu$). For example the variance of the losses $R_\chi = \mathbb{E}[L^2] - R_\mu$ is very useful in finance since it gives more information to risk averse individuals. This is particularly important if the average loss is computed for a large period of time (e.g. annually). If the loss is considered every time there is a computer event then we believe the average loss by itself provides enough risk information to make a rational decision.

In this chapter, we focus on attacks on sensor networks and the effects they have on the process control system. Therefore $p_i$ denotes the likelihood that an attacker will compromise sensor $i$, and $L_i$ denotes the losses associated with that particular compromise. To simplify our presentation we assume that $p_i$ is the same for all sensors, therefore our focus in the remaining of this section is to estimate the potential losses $L_i$. The results can then be used to identify high priority sensors and to invest a given security budget in the most cost-effective way.

### 5.1.1 Attack models

We consider the case when the state of the system is measured by a sensor network of $p$ sensors with measurement vector $y(k) = \{y_1(k), \ldots, y_p(k)\}$, where $y_i(k)$ denotes the measurement by sensor $i$ at time $k$. All sensors have a dynamic range that defines the domain of $y_i$ for all $k$. That is, all sensors have defined minimum and maximum values $\forall k, y_i(k) \in [y_i^{\min}, y_i^{\max}]$. Let $\mathcal{Y}_i = [y_i^{\min}, y_i^{\max}]$. We assume each sensor has a unique identity protected by a cryptographic key.

Let $\tilde{y}(k) \in \mathbb{R}^p$ denote the received measurements by the controller at time $k$. Based on these measurements the control system defines control actions to maintain certain operational goals. If some of the sensors are under attack, $\tilde{y}(k)$ may be different from the real measurement $y(k)$; however, we assume that the attacked signals $\tilde{y}_i(k)$ also lie within $\mathcal{Y}_i$ (signals outside this range can be easily detected by fault-tolerant algorithms).

Let $\mathcal{K}_a = \{k_s, \ldots, k_e\}$ represent the attack duration; between the start time $k_s$ and stop time $k_e$ of an attack. A general model for the observed signal is the following:

$$\tilde{y}_i(k) = \begin{cases} y_i(k) & \text{for } k \notin \mathcal{K}_a \\ a_i(k) & \text{for } k \in \mathcal{K}_a,\ a_i(k) \in \mathcal{Y}_i \end{cases}$$

where $a_i(k)$ is the attack signal. This general sensor attack model can be used to represent integrity attacks and DoS attacks. In an integrity attack we assume that if attackers have compromised a sensor, then they can inject any arbitrary value, therefore in this case, $a_i(k)$ is some arbitrary non-zero value.

In a DoS attack, the controller will notice the lack of new measurements and will react accordingly. An intuitive response for a controller to implement against a DoS attack is to use the last signal received: $a_i(k) = y_i(k_s)$, where $y_i(k_s)$ is the last measurement received before the DoS attack starts.

### 5.1.2 Experiments

To test our attacks, we use the Tennessee-Eastman process control system (TE-PCS) model and the associated multi-loop PI control law as proposed by Ricker Ricker [1993]. We briefly describe the process architecture and the control loops in Figure 5.1. The original process model is implemented in Fortran and the PI control law is implemented in Matlab. We use this code for our study.

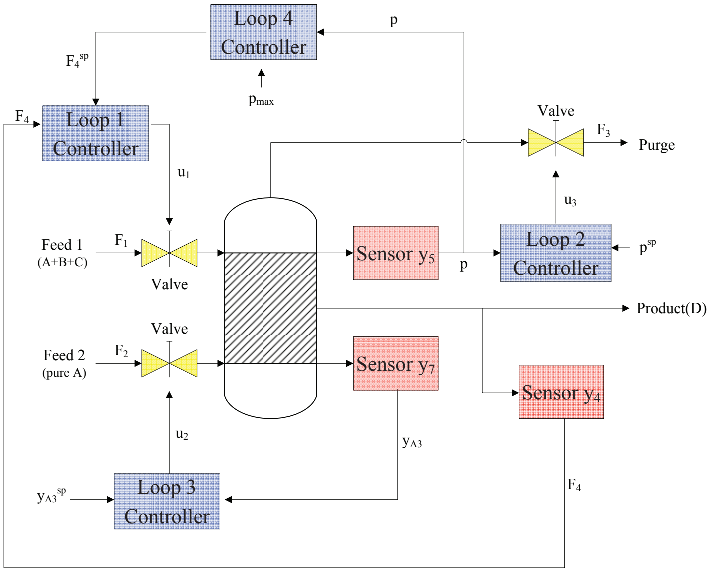


Figure 5.1: Architecture of the Simplified TE Plant.

The chemical process consists of an irreversible reaction which occurs in the vapor phase inside a reactor of fixed volume $V$ of 122 ($m^3$). Two non-condensible reactants $A$ and $C$ react in the presence of an inert $B$ to form a non-volatile liquid product $D$:

$$A + C \xrightarrow{B} D.$$

The feed stream 1 contains $A$, $C$ and trace of $B$; feed stream 2 is pure $A$; stream 3 is the purge containing vapors of $A$, $B$, $C$; and stream 4 is the exit for liquid product $D$. The measured flow rates of stream $i$ is denoted by $F_i$ (kmol $h^{-1}$). The *control objectives* are

- *Regulate* $F_4$, the rate of production of the product $D$, at a set-point $F_4^{sp}$ (kmol $h^{-1}$),
- Maintain $P$, the operating pressure of the reactor, below the shut-down limit of 3000 kPa as dictated *safety* considerations,
- Minimize $C$, the *operating cost* measured in (kmol-of-product). The cost depends linearly on the purge loss of $A$ and $C$ relative to the production rate of $D$. The cost considerations dictate that the pressure be maintained as close as possible to 3000 kPa.

The production rate of $D$, denoted by $r_D$ (kmol h$^{-1}$) is

$$r_D = k_0 y_{A3}^{v1} y_{C3}^{v2} P^{v3},$$

where $y_{A3}$ and $y_{C3}$ denote the respective fractions of $A$ and $C$ in the purge and $v_1$, $v_2$, $v_3$ are given constants.

There are four *input variables* (or command signals) available to achieve the above control objectives. The first three input variables, denoted as $u_1$, $u_2$ and $u_3$, trigger the actuators that can change the positions of the respective valves. The fourth input variable, denoted as $u_4$, is the set point for the proportional controller for the liquid inventory. The input variables as used by the controller in the following way:

- Production rate $y_4 = F_4$ is controlled using Feed 1 ($u_1$) by loop$-1$ controller,
- Pressure $y_5 = P$ is controlled using the purge rate ($u_3$) by loop$-2$ controller,
- Partial pressure of product $A$ in the purge $y_7 = y_{A3}$ is controlled using Feed 2 ($u_3$) by loop$-3$ controller,

When $u_3$ saturates, the loop$-4$ controller uses $u_1$ to control the pressure $P$. The controllers for all four loops in figure 5.1 are *proportional integral* (PI) controllers.

In steady-state operation, the production rate $F_4$ is 100 kmol h$^{-1}$, the pressure $P$ is 2700 KPa and the fraction of $A$ in the purge is 47 mol%.

We study the security issues of control systems by experimenting and simulating cyber attacks on sensor signals in the TE-PCS model. Because operating the chemical reactor with a pressure larger than 3000 kPa is unsafe (it may lead to an explosion or damage of the equipment) We.assume that that the goal of the attacker is to raise the pressure level of the tank to a value larger than 3000 kPa. We model an attacker that only has access to a single sensor at a given time. We also assume $L_i > L_j$, when an attack $i$ can drive the system to an unsafe state and an attack $j$ cannot, and $L_i = L_j$ if both attacks $i$ and $j$ either do not drive the system to an unsafe state, or both can compromise the safety of the system.

From the experimental results, we found that the most effective of these attacks were max/min attacks (i.e., when $a_i(k) = y_i^{\min}$ or $a_i(k) = y_j^{\max}$). However, not all of the max/min attacks were able to drive the pressure to unsafe levels. We now summarize some of the results.

- By attacking the sensors, a controller is expected to respond with incorrect control signals since it receives wrong information from the compromised sensors. For example, by forging $y_7$ as $y_7^{\max}$ from $t = 0$ to 30, the controller believes there is a large amount of component $A$ in the tank.

  From the experiments, we found that the plant system can go back to the steady state after the attack finishes, as illustrated in Fig 5.2. Furthermore, the pressure in the main tank never reaches 3000 kPa. In general we found that the plant is very resilient to attacks on $y_7$ and $y_4$. Attacks in the limit of the sensing range ($y^{\min}$ and $y^{\max}$) were the more damaging, but they did not force the system into an unsafe state.

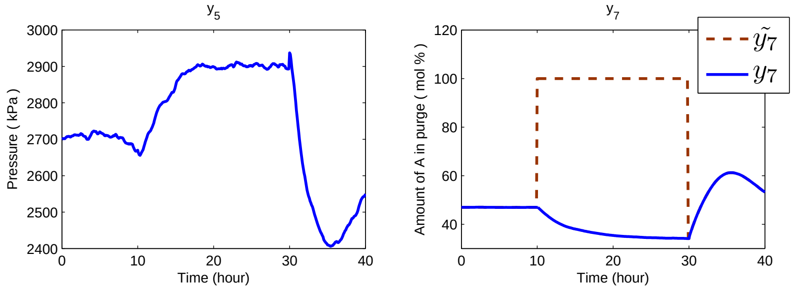
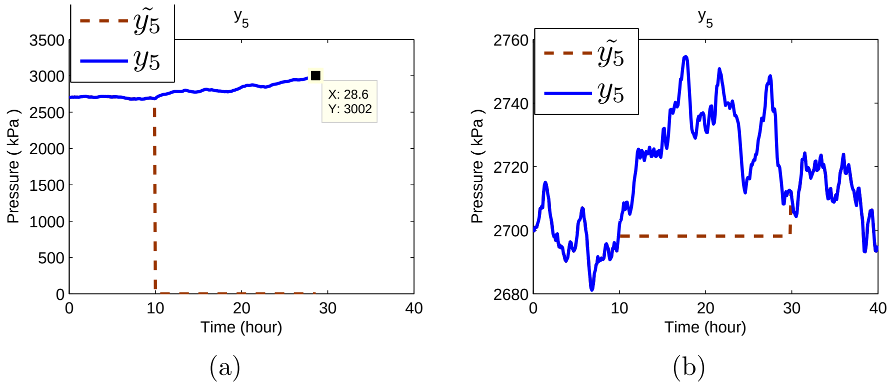


Figure 5.2: Integrity attack $y_7^{\max}$ from $t = 0$ to 30. Safety preserved for attacks on $y_7$.

- By launching attack $y_5^{\min}$ the controller turns down the purge valve to increase the pressure and prevent the liquid products from accumulating. We can see that the real pressure of the tank ($y_5$ in Fig 5.3(a)) keeps increasing past 3000 kPa and the system operates in an unsafe state. In this experiment, it takes about 20 hours ($t = 10$ to $t = 30$) to shut down (or cause an explosion to) the plant. This long delay in causing an effective attack may give defenders the advantage: for physical processes with *slow-dynamics*, it is possible that human system operators may have enough time to observe unusual phenomenon and take proper actions against the attack.

- We found out that in general DoS attacks do not affect the plant. We ran the plant 20 times for 40 hours each and for a DoS attack lasting 20 hours the pressure in the tank never exceeded 2900kPa.



(a)

(b)

Figure 5.3: Safety violated by compromising $y_5$. DoS attacks do not cause damage.

We conclude that if the plant operator wants to prevent an attack from making the system operate in an unsafe state, it should prioritize defenses against integrity attacks rather than on DoS attacks. If the plant operator only has enough budget to deploy advanced security mechanisms for one sensor (e.g., tamper resistance, or TPM chips), $y_5$ should be the priority.

## 5.2 Detection of Attacks

Detecting attacks to control systems can be formulated as an anomaly-based intrusion detection problem Denning [1987]. One big difference in control systems compared to traditional IT systems, is that instead of creating models of network traffic or software behavior, we can use a representative model of the physical system.

The intuition behind this approach is the following: if we know how the output sequence of the physical system, $y(k)$, should react to the control input sequence, $u(k)$, then any attack to the sensor data can be potentially detected by comparing the expected output $\hat{y}(k)$ with the received (and possibly compromised) signal $\tilde{y}(k)$. Depending on the quality of our estimate $\hat{y}(k)$ we may have some false alarms. We revisit this problem in the next section.

To formalize the anomaly detection problem, we need (1) a model of the behavior of the physical system, and (2) an anomaly detection algorithm. In section 5.2.1 we discuss our choice of linear models as an approximation of the behavior of the physical system. In section 5.2.2, we describe change detection theory and the detection algorithm we use–a nonparametric cumulative sum (CUSUM) statistic.

### 5.2.1 Linear Model

To develop accurate control algorithms, control engineers often construct a representative model that captures the behavior of the physical system in order to predict how the system will react to a given control signal. A process model can be derived from first principles (a model based on the fundamental laws of physics) or from empirical input and output data (a model obtained by simulating the process inputs with a carefully designed test sequence). It is also very common to use a combination of these two models; for example, first-principle models are typically calibrated by using process test data to estimate key parameters. Likewise, empirical models are often adjusted to account for known process physics Quin and Badgwell [2003]; Rawlings [2000].

For highly safety-critical applications, such as the aerospace industry, it is technically and economically feasible to develop accurate models from first principles Quin and Badgwell [2003]. However, for the majority of process control systems, the development of process models from fundamental physics is difficult.

In many cases such detailed models are difficult to justify economically, and even impossible to obtain in reasonable time due to the complex nature of many systems and processes. (The TE-PCS system used in our experiments is one of the few cases available in the literature of a detailed nonlinear model of an industrial control problem; this is the reason why the TE-PCS system has been used as a standard testbed in many industrial control papers.)

To facilitate the creation of physical models, most industrial control vendors provide tools (called identification packages) to develop models of physical systems from training data. The most common models are *linear* systems. Linear systems can be used to model dynamics that are linear in state $x(k)$ and control input $u(k)$

$$x(k+1) = Ax(k) + Bu(k) \tag{5.1}$$

where time is represented by $k \in \mathbb{Z}^+$, $x(k) = (x_1(k), \ldots, x_n(k)) \in \mathbb{R}^n$ is the state of the system, and $u(k) = (u_1(k), \ldots, u_m(k)) \in \mathbb{R}^m$ is the control input. The matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ models the physical dependence of state $i$ on state $j$, and $B = (b_{ij}) \in \mathbb{R}^{n \times m}$ is the input matrix for state $i$ from control input $j$.

Assume the system (5.1) is monitored by a *sensor network* with $p$ sensors. We obtain the measurement sequence from the observation equations

$$\hat{y}(k) = Cx(k), \tag{5.2}$$

where $\hat{y}(k) = (\hat{y}_1(k), \ldots, \hat{y}_p(k)) \in \mathbb{R}^p$, and $\hat{y}_l(k) \in \mathbb{R}$ is the estimated measurement collected by sensor $l$ at time $k$. The matrix $C \in \mathbb{R}^{p \times n}$ is called output matrix.

### 5.2.2 Detection Methods

The physical-model-based attack detection method presented in this chapter can be viewed as complementary to intrusion detection methods based on network and computer systems models.

Because we need to detect anomalies in real time, we can use results from sequential detection theory to give a sound foundation to our approach. Sequential detection theory considers the problem where the measurement time is not fixed, but can be chosen online as and when the measurements are obtained. Such problem formulations are called *optimal stopping problems*. Two such problem formulations are: sequential detection (also known as sequential hypothesis testing), and quickest detection (also known as change detection). A good survey of these problems is given by Kailath and Poor Kailath and Poor [1998].

In optimal stopping problems, we are given a time series sequence $z(1), z(2), \ldots, z(N)$, and the goal is to determine the minimum number of samples, $N$, the anomaly detection scheme should observe before making a decision $d_N$ between two hypotheses: $H_0$ (normal behavior) and $H_1$ (attack).

The difference between sequential detection and change detection is that the former assumes the sequence $z(i)$ is generated either by the normal hypothesis ($H_0$), or by the attack hypothesis ($H_1$). The goal is to decide which hypothesis is true in minimum time. On the other hand, change detection assumes the observation $z(i)$ starts under $H_0$ and then, at a given $k_s$ it changes to hypothesis $H_1$. Here the goal is to detect this change as soon as possible.

Both problem formulations are very popular, but security researchers have used sequential detection more frequently. However, for our attack detection method, the change detection formulation is more intuitive. To facilitate this intuition, we now briefly describe the two formulations.

#### Sequential Detection

Given a fixed probability of false alarm and a fixed probability of detection, the goal of sequential detection is to minimize the number of observations required to make a decision between two hypotheses. The solution is the classic sequential probability ratio test (SPRT) of Wald Wald [1947] (also referred as the threshold random walk (TRW) by some security papers). SPRT has been widely used in various problems in information security such as detecting portscans Jung et al. [2004], worms Schechter and Berger [2004], proxies used by spammers Xie et al. [2006], and botnets Gu et al. [2008].

Assuming that the observations $z(k)$ under $H_j$ are generated with a probability distribution $p_j$, the SPRT algorithm can be described by the following equations:

$$\begin{aligned} S(k+1) &= \log \frac{p_1(z(k))}{p_0(z(k))} + S(k) \\ N &= \inf_n \{n : S(n) \notin [L, U]\}, \end{aligned}$$

starting with $S(0) = 0$. The SPRT decision rule $d_N$ is defined as:

$$d_N = \begin{cases} H_1 & \text{if } S(N) \geqslant U \\ H_0 & \text{if } S(N) \leqslant L, \end{cases} \tag{5.3}$$

where $L \approx \ln \frac{b}{1-a}$ and $U \approx \ln \frac{1-b}{a}$, and where $a$ is the desired probability of false alarm and $b$ is the desired probability of missed detection (usually chosen as small values).

### Change Detection

The goal of the change detection problem is to detect a possible change, at an unknown change point $k_s$.Cumulative sum (CUSUM) and Shiryaev-Roberts statistics are the two most commonly used algorithms for change detection problems. In this chapter, we use the CUSUM statistic because it is very similar to the SPRT.

Given a fixed false alarm rate, the CUSUM algorithm attempts to minimize the time $N$ (where $N \geqslant k_s$) for which the test stops and decides that a change has occurred. Let $S(0) = 0$. The CUSUM statistic is updated according to

$$S(k+1) = \left( \log \frac{p_1(z(k))}{p_0(z(k))} + S(k) \right)^+ \tag{5.4}$$

where $(a)^+ = a$ if $a \geqslant 0$ and zero otherwise. The stopping time is:

$$N = \inf_n \{n : S(n) \geqslant \tau\} \tag{5.5}$$

for a given threshold $\tau$ selected based on the false alarm constraint.

We can see that the CUSUM algorithm is an SPRT test with $L = 0$, $U = \tau$, and whenever the statistic reaches the lower threshold $L$, it re-starts.

We now describe how to adapt the results of change detection theory to the particular problem of detecting compromised sensors. In the following, we use the subscript $i$ to denote the sequence corresponding to sensor $i$.

One problem that we have in our case is that we do not know the probability distribution for an attack $p_1$. In general, an adaptive adversary can select any arbitrary (and possibly) non-stationary sequence $z_i(k)$. Assuming a fixed $p_1$ will thus limit our ability to detect a wide range of attacks.

To avoid making assumptions about the probability distribution of an attacker, we use ideas from nonparametric statistics. We do not assume a parametric distribution for $p_1$ and $p_0$; instead, only place mild constraints on the observation sequence. One of the simplest constraints is to assume the expected value of the random process $Z_i(k)$ that generates the sequence $z_i(k)$ under $H_0$ is less than zero ($\mathbb{E}_0[Z_i] < 0$) and the expected value of $Z_i(k)$ under $H_1$ is greater than zero ($\mathbb{E}_1[Z_i] > 0$).

To achieve these conditions let us define

$$z_i(k) := \|\tilde{y}_i(k) - \hat{y}_i(k)\| - b_i \tag{5.6}$$

where $b_i$ is a small positive constant chosen such that

$$\mathbb{E}_0[\|\tilde{y}_i(k) - \hat{y}_i(k)\| - b_i] < 0. \tag{5.7}$$

The nonparametric CUSUM statistic for sensor $i$ is then:

$$S_i(k) = (S_i(k-1) + z_i(k))^+,\ S_i(0) = 0 \tag{5.8}$$

and the corresponding decision rule is

$$d_{N,i} \equiv d_\tau(S_i(k)) = \begin{cases} H_1 & \text{if } S_i(k) > \tau_i \\ H_0 & \text{otherwise.} \end{cases} \tag{5.9}$$

where $\tau_i$ is the threshold selected based on the false alarm rate for sensor $i$.

Following Brodsky and Darkhovsky [1993], we state the following two important results for Eq. (5.8)-(5.9):

- The probability of false alarm decreases exponentially as the threshold $\tau_i$ increases,
- The time to detect an attack, $(N_i - k_{s,i})^+$, is inversely proportional to $b_i$.

### 5.2.3 Stealthy Attacks

A fundamental problem in intrusion detection is the existence of adaptive adversaries that will attempt to evade the detection scheme; therefore, we now consider an adversary that knows about our anomaly detection scheme. We take a conservative approach in our models by assuming a very powerful attacker with knowledge of: (1) the exact linear model that we use (i.e., matrices $A$,$B$, and $C$), the parameters ($\tau_i$ and $b_i$), and (3) the control command signals. Such a powerful attacker may be unrealistic in some scenarios, but we want to test the resiliency of our system to such an attacker to guarantee safety for a wide range of attack scenarios.

The goal of the attacker is to raise the pressure in the tank without being detected (i.e., raise the pressure while keeping the statistic he controls below the corresponding threshold $\tau_i$).

We model three types of attacks: surge attacks, bias attacks and geometric attacks. Surge attacks model attackers that want to achieve maximum damage as soon as they get access to the system. A bias attack models attackers that try to modify the system discretely by adding small perturbations over a large period of time. Finally, geometric attacks model attackers that try to shift the behavior of the system very discretely at the beginning of the attack and then maximize the damage after the system has been moved to a more vulnerable state.

### 5.2.4 Surge Attacks

In a surge attack the adversary tries to maximize the damage as soon as possible, but when the statistic reaches the threshold, it then stays at the threshold level: $S_i(k) = \tau$ for the remaining time of the attack. To stay at the threshold, the attacker needs to solve the following quadratic equation:

$$S_i(k) + \sqrt{(\hat{y}_i(k) - \tilde{y}_i(k))^2} - b_i = \tau_i$$

The resulting attack (for $y_5$ and $y_4$) is:

$$\tilde{y}_i(k) = \begin{cases} y_i^{min} & \text{if } S_i(k+1) \leqslant \tau_i \\ \hat{y}_i(k) - |\tau_i + b_i - S_i(k)| & \text{if } S_i(k+1) > \tau_i \end{cases}$$

For $y_7$ we use

$$\tilde{y}_7(k) = \begin{cases} y_7^{max} & \text{if } S_{y_7}(k) \leqslant \tau_7 \\ \hat{y}_7 + |\tau_7 + b_7 - S_{y_7}(k)| & \text{if } S_{y_7}(k) > \tau_7 \end{cases}$$

### 5.2.5 Bias Attacks

In a bias attack the attacker adds a small constant $c_i$ at each time step.

$$\tilde{y}_{i,k} = \hat{y}_{i,k} - c_i \in \mathcal{Y}_i$$

In this case, the nonparametric CUSUM statistic can be written as:

$$S_i(n) = \sum_{k=0}^{n-1} |\hat{y}_i(k) - \tilde{y}_i(k)| - nb_i$$

Assuming the attack starts at time $k = 0$ and assuming the attacker wants to be undetected for $n$ time steps the attacker needs to solve the following equation:

$$\sum_{k=0}^{n-1} c_i = \tau_i + nb_i$$

Therefore $c_i = \tau_i/n + b$. This attack creates a bias of $\tau_i/n + b_i$ for each attacked signal.

This equation shows the limitations of the attacker. If an attacker wants to maximize the damage (maximize the bias of a signal), the attacker needs to select the smallest $n$ it can find. Because $\tilde{y}_i \in \mathcal{Y}_i$ this attack reduces to an impulse attack.

If an attacker wants to attack for a long time, then $n$ will be very large. If $n$ is very large then the bias will be smaller.

### 5.2.6 Geometric Attacks

In a geometric attack, the attacker wants to drift the value very slowly at the beginning and maximize the damage at the end. This attack combines the slow initial drift of the bias attack with a surge attack at the end to cause maximum damage.

Let $\alpha \in (0, 1)$. The attack is:

$$\tilde{y}_i(k) = \hat{y}_i(k) - \beta_i \alpha_i^{n-k}.$$

Now we need to find $\alpha$ and $\beta$ such that $S_i(n) = \tau_i$.

Assume the attack starts at time $k = 0$ and the attacker wants to be undetected for $n$ time steps. The attacker then needs to solve the following equation.

$$\sum_{k=0}^{n-1} \beta_i \alpha_i^{n-k} - nb_i = \tau_i$$

This addition is a geometric progression.

$$\sum_{k=0}^{n-1} \beta_i \alpha_i^{n-k} = \beta_i \alpha_i^n \sum_{k=0}^{n-1} (\alpha_i^{-1})^k = \beta_i \frac{1 - \alpha_i^n}{\alpha_i^{-1} - 1}$$

By fixing $\alpha$ the attacker can select the appropriate $\beta$ to satisfy the above equation.

### 5.2.7 Experiments

We continue our use of the TE-PCS model. In this section we first describe our selection criteria for matrices $A$, $B$, and $C$ for the linear model, and the parameters $b_i$ and $\tau_i$ for the CUSUM statistic. We then describe the tradeoffs between false alarm rates and the delay for detecting attacks. The section ends with the study of stealthy attacks.

#### Linear Model

In this chapter, we use the linear system characterized by the matrices $A$, $B$, and $C$, obtained by linearizing the non-linear TE-PCS model about the steady-state operating conditions. (See Ricker Ricker [1993].) The linear model is a good representative of the actual TE-PCS model when the operating conditions are reasonably close to the steady-state.

### Nonparametric CUSUM parameters

In order to select $b_i$ for each sensor $i$, we need to estimate the expected value of the distance $|\hat{y}_i(k) - y_i(k)|$ between the linear model estimate $\hat{y}_i(k)$ and the sensor measurement $y_i(k)$ (i.e., the sensor signal without attacks).

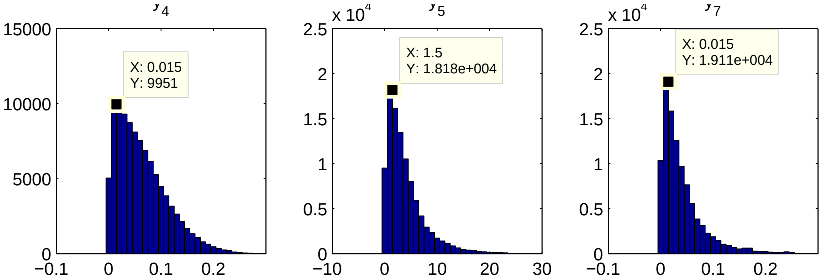


Figure 5.4: Anomaly detection module (ADM) parameter $b$.

We run experiments for ten thousand times (and for 40 hours each time) without any attacks to gather statistics. Fig 5.4 shows the estimated probability distributions (without normalization).

To obtain $b_i$, we compute the empirical expected value for each distance and then round up to the two most significant units. We obtain $b_{y_4} = 0.065$, $b_{y_5} = 4.1$, $b_{y_7} = 0.042$.

Once we have $b_i$ for each sensor, we need to find a threshold $\tau_i$ to balance the tradeoff between false alarms and detection time.

**False Alarm Rate** We run simulations for twenty times without attacks and compute the total number of false alarms for different values of $\tau$ (and for each sensor). Fig 5.5 shows the results. Taking $y_4$ as an example, we notice that $S_{y_4}$ alerts frequently if we set $\tau_{y_4} < 6$.

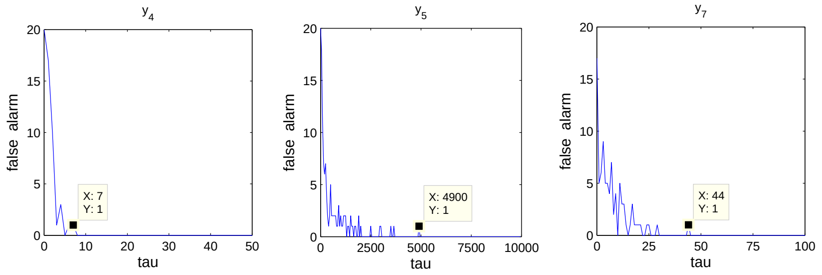


Figure 5.5: The number of false alarms decreases exponentially with increasing $\tau$.

In general, we would like to select $\tau$ as high as possible for each sensor to avoid any false alarm; however, increasing $\tau$ increases the time to detect attacks.

**Detection Time** To measure the time to detect attacks, we run simulations by launching scaling attacks ($a_i(k) = \lambda_m y_i(k)$) on sensors $y_4$, $y_5$ and $y_7$. Figs 5.6 and 5.7 shows the experimental results.

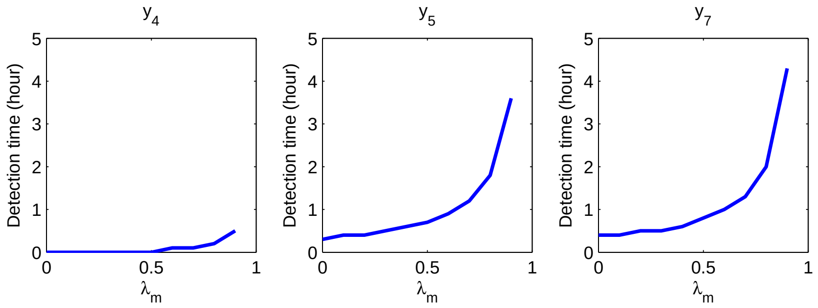


Figure 5.6: Detection time v.s. scaling attack. Note that for $\lambda_i^m = 1$ there is no alarm.

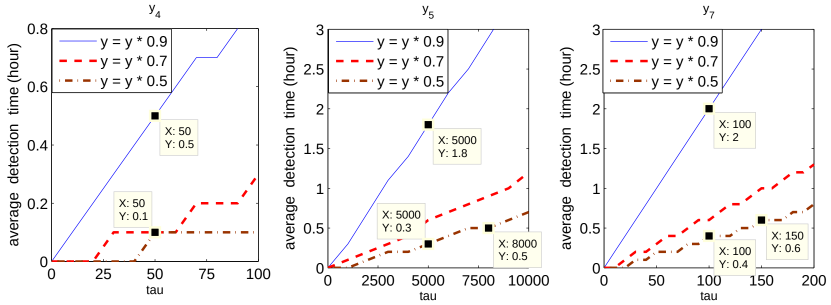


Figure 5.7: The time for detection increases linearly with increasing $\tau$.

The selection of $\tau$ is a trade-off between detection time and the number of false alarms. The appropriate value differs from system to system. Because the large number of false alarms is one of the main problems for anomaly detection systems, and because the TE-PCS process takes at least 10 hours to reach the unsafe state (based on our risk assessment section), we choose the conservative set of parameters $\tau_{y_4} = 50$, $\tau_{y_5} = 10000$, $\tau_{y_7} = 200$. These parameters allow us to detect attacks within a couple of hours, while not raising any false alarms.

## Stealthy Attacks

To test if our selected values for $\tau$ are resilient to stealthy attacks, we decided to investigate the effect of stealthy attacks as a function of $\tau$. To test how the attacks change for all thresholds we parameterize each threshold by a parameter $p$: $\tau_i^{test} = p\tau_i$. Fig. 5.8 shows the percentage of times that geometric stealthy attacks (assuming the attacker controls all three sensor readings) were able to drive the pressure above 3000kPa while remaining undetected (as a function of $p$).

We implemented all stealth attacks starting at time $T = 10$ (hrs). We assume the goal of the attacker is to be undetected until $T = 30$ (hrs). For example, Fig. 5.9 shows the results of attacking all three sensors with a geometric attack. The nonparametric CUSUM statistic shown in Fig. 5.10 shows how the attacker remains undetected until time $T = 30$ (hrs).

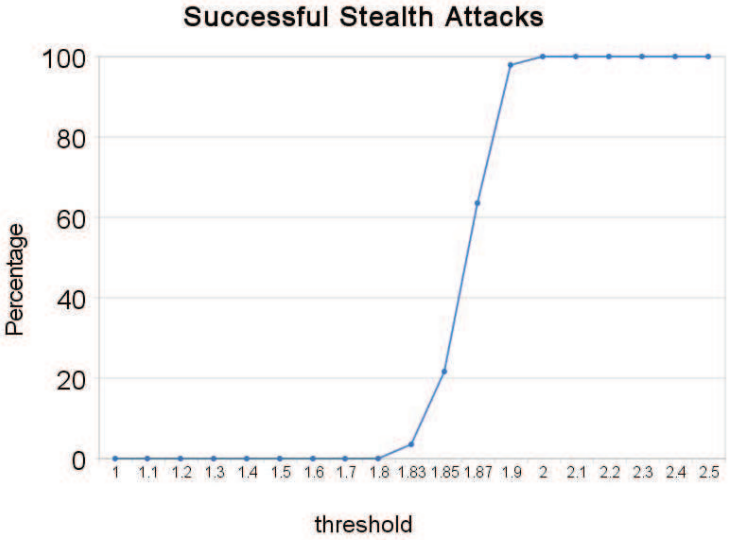


Figure 5.8: Percentage of unsafe stealthy attack percentage vs. scaling parameter $p$.

We found that a surge attack does not cause significant damages because of the inertia of the chemical reactor: by the time the statistic reaches the threshold $\tau$, the chemical reactor is only starting to respond to the attack. However, since the attacker can only add very small variations to the signal once it is close to the threshold, the attack ceases to produce any effect and the plant continues operating normally.

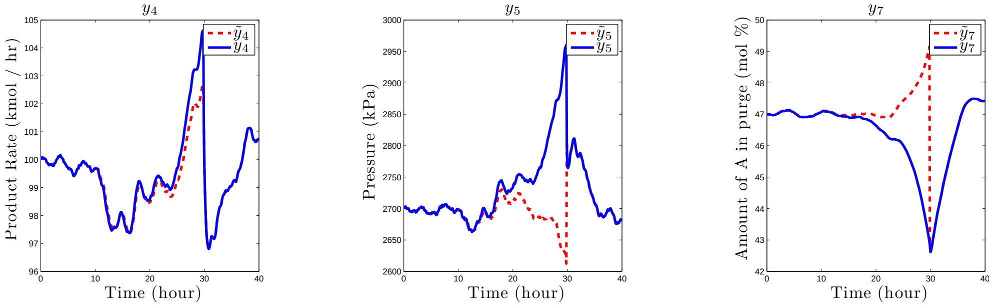


Figure 5.9: Geometric attacks to sensors: real state (solid), false data (dotted).

Finally, we assume two types of attackers. An attacker that has compromised $y_5$ (but who does not know the values of the other sensors, and therefore can only control $S_{y_5}(k)$), and an attacker that has compromised all three sensors (and therefore can control the statistic $S(k)$ for all sensors). We launched each attack 20 times. The results are summarized in Figure 5.11.

Our results show that even though our detection algorithm fails to detect stealthy attacks, we can keep the the plant in safe conditions. We also find that the most successful attack strategy are geometric attacks.

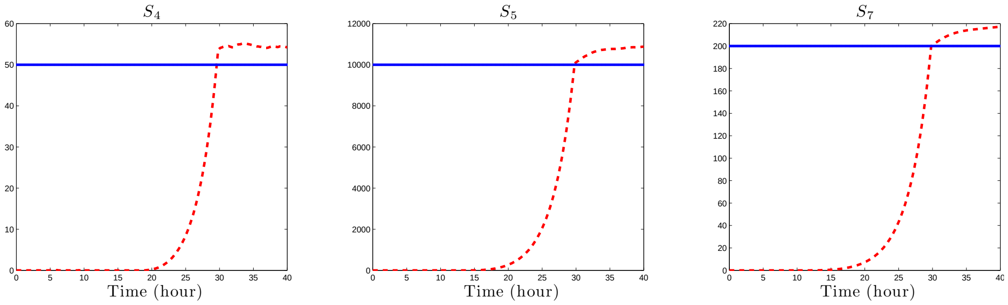


Figure 5.10: Statistics of geometric attacks with sensors compromised.

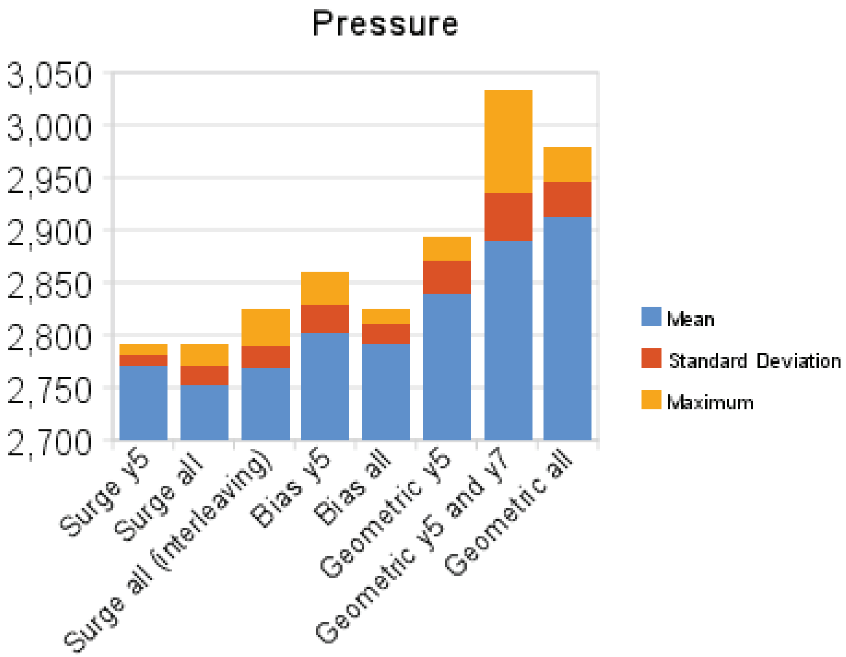


Figure 5.11: Effect of stealthy attacks. Each attack last 20 hours.

## 5.3 Response to Attacks

A comprehensive security posture for any system should include mechanisms for prevention, detection, and response to attacks. Automatic response to computer attacks is one of the fundamental problems in information assurance. While most of the research efforts found in the literature focus on prevention (authentication, access controls, cryptography etc.) or detection (intrusion detection systems), in practice there are quite a few response mechanisms. For example, many web servers send CAPTCHAs to the client whenever they find that connections resemble bot connections, firewalls drop connections that conform to their rules, the execution of anomalous processes can be slowed down by intrusion detection systems, etc.

Given that we already have an estimate for the state of the system (given by a linear model), a natural response strategy for control systems is to use this estimate when the anomaly detection statistic fires an alarm. Fig 5.12 shows our proposed architecture. Specifically: for sensor $i$, if $S_i(k) > \tau_i$, the ADM replaces the sensor measurements $\tilde{y}_i(k)$ with measurements generated by the linear model $\hat{y}_i(k)$ (that is the controller will receive as input $\hat{y}_i(k)$ instead of $\tilde{y}_i(k)$). Otherwise, it treats $\tilde{y}_i(k)$ as the correct sensor signal.

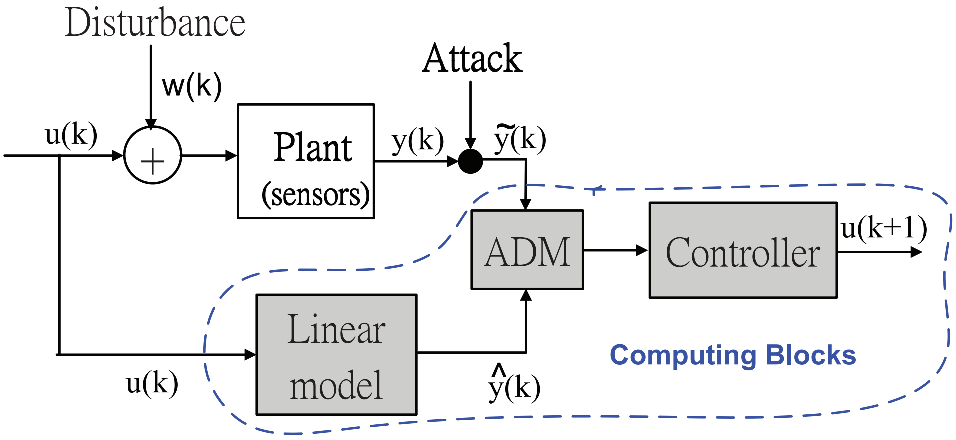


Figure 5.12: An Anomaly Detection Module (ADM).

Introducing automatic response mechanisms is, however, not an easy solution. Every time systems introduce an automatic response to an alarm, they have to consider the cost of dealing with false alarms. In our proposed detection and response architecture (Fig. 5.12), we have to make sure that if there is a false alarm, controlling the system by using the estimated values from the linear system will not cause any safety concerns.

### 5.3.1 Experiments

The automatic response mechanism works well when we are under attack. For example, Fig. (5.13) shows that when an attack is detected, the response algorithm manages to keep

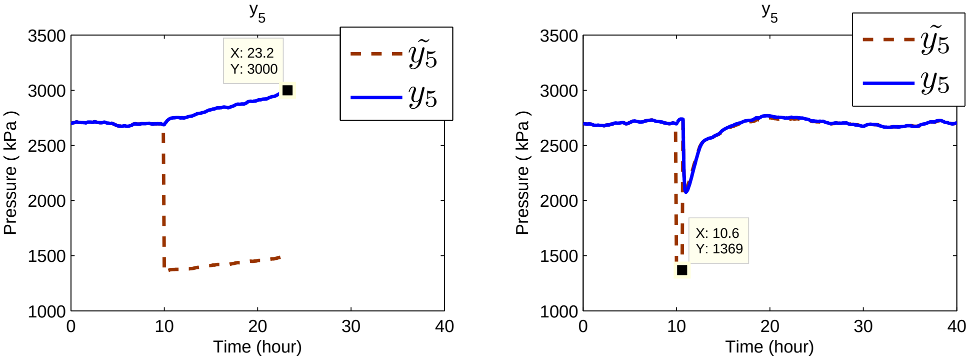
| Alarms | Avg $y_5$ | Std Dev | Max $y_5$ |
|---|---|---|---|
| 0 | 2700.4 | 14.73 | 2757 |

Table 5.1: For Thresholds $\tau_{y_4} = 50, \tau_{y_5} = 10000, \tau_{y_7} = 200$ we obtain no false alarm. Therefore we only report the expected pressure, the standard deviation of the pressure, and the maximum pressure reached under no false alarm.

| | Alarms | Avg $y_5$ | Std Dev | Max $y_5$ |
|---|---|---|---|---|
| $y_4$ | 61 | 2710 | 30.36 | 2779 |
| $y_5$ | 106 | 2705 | 18.72 | 2794 |
| $y_7$ | 53 | 2706 | 20.89 | 2776 |

Table 5.2: Behavior of the plant after response to a false alarm with thresholds $\tau_{y_4} = 5, \tau_{y_5} = 1000, \tau_{y_7} = 20$.

the system in a safe state. Similar results were obtained for all detectable attacks.



(a) Without ADM

(b) ADM detects and responds to the attack at $T = 10.7$ (hr)

Figure 5.13: Safety of response mechanism under deception attack $\tilde{y_5} = y_5 * 0.5$.

While our attack response mechanism is a good solution when the alarms are indeed an indication of attacks, Our main concern in this section is the cost of false alarms. To address these concerns we ran the simulation scenario without any attacks 1000 times; each time the experiment ran for 40 hours. As expected, with the parameter set $\tau_{y_4} = 50$, $\tau_{y_5} = 10000$, $\tau_{y_7} = 200$ our system did not detect any false alarm (see Table 5.1); therefore we decided to reduce the detection threshold to $\tau_{y_4} = 5$, $\tau_{y_5} = 1000$, $\tau_{y_7} = 20$ and run the same experiments again. Table 5.2 shows the behavior of the pressure *after the response to a false alarm*. We can see that while a false response mechanism increases the pressure of the tank, it never reaches unsafe levels. The maximum pressure obtained while controlling the system based on the linear model was 2779 kPa, which is in the same order of magnitude than the normal variation of the pressure without any false alarm (2757 kPa).

In our case, even if the system is kept in a safe state by the automated response, our response strategy is meant as a temporary solution before a human operator responds to the alarm. Based on our results we believe that the time for a human response can be very large (a couple of hours).

## 5.4 Discussion

In this work we identified three new research challenges for securing control systems. We showed that by incorporating a physical model of the system we were able to identify the most critical sensors and attacks. We also studied the use of physical models for anomaly detection and proposed three generic types of stealthy attacks. Finally, we proposed the use of automatic response mechanisms based on estimates of the state of the system. Automatic responses may be problematic in some cases (especially if the response to a false alarm is costly); therefore, we would like to emphasize that the automatic response mechanism should be considered as a temporary solution before a human investigates the alarm. A full deployment of any automatic response mechanism should take into consideration the amount of time in which it is reasonable for a human operator to respond, and the potential side effects of responding to a false alarm.

In our experiments with the TE-PCS process we found several interesting results. (1) Protecting against integrity attacks is more important than protecting against DoS attacks. In fact, we believe that DoS attacks have negligible impact to the TE-PCS process. (2) The chemical reactor process is a well-behaved system, in the sense that even under perturbations, the response of the system follows very closely our linear models. In addition, the slow dynamics of this process allows us to be able to detect attacks even with large delays with the benefit of not raising any false alarms. (3) Even when we configure the system to have false alarms, we saw that the automatic response mechanism was able to control the system in a safe mode.

One of our main conclusions regarding the TE-PCS plant, is that it is a very resiliently-designed process control system. Design of resilient process control systems takes control system design experience and expertise. The design process is based on iteratively evaluating the performance on a set of bad situations that can arise during the operation of the plant and modifying control loop structures to build in resilience. In particular, Ricker's paper discusses the set of random faults that the four loop PI control is able to withstand.

We would like to make two points in this regard: (1). The PI control loop structure is distributed, in the sense that no PI control loop controls all actuators and no PI loop has access to all sensor measurements, and (2). The set of bad situations to which this control structure is able to withstand may itself result from the one or more cyber attacks. However, even though the resilience of TE-PCS plant is ensured by expert design, we find it interesting to directly test this resilience within the framework of assessment, detection and response that we present in this article.

However, as a word of caution, large scale control system designs are often not to resilient by design and may become prey to such stealth attacks if sufficient resilience is not built by design in the first place. Thus, our ideas become all the more relevant for operational security until there is a principled way of designing fully attack resilient control structures and algorithms (which by itself is a very challenging research endeavor and may not offer a cost effective design solution).

Even though we have focused on the analysis of a chemical reactor system, our principles and techniques can be applied to many other physical processes. An automatic detection and response module may not be a practical solution for all control system processes; however, we believe that many processes with similar characteristics to the TE-PCS can benefit from this kind of response.

# Chapter 6

# Deception Attacks on Power System State Estimators

## 6.1 Introduction

In Chapter 5 we described the design of an attack detection module based on an approximate model of system dynamics and used it for detection and response under cyber threats to a benchmark process control system. The aim of this chapter is to analyze the cyber security of state estimators in Supervisory Control and Data Acquisition (SCADA) systems operating in power grids. Safe and reliable operation of these critical infrastructure systems is a major concern in our society. The power system state estimation algorithms which are currently in use also employ bad data detection (BDD) schemes to detect random outliers in the measurement data. Such schemes are based on high measurement redundancy. Although such methods may detect a set of basic cyber attacks, they may fail in the presence of a more intelligent attacker. We explore the latter by considering scenarios in which deception attacks are performed, sending false information to the control center. Similar attacks have been studied before for linear state estimators, assuming the attacker has perfect model knowledge. Here, we instead assume the attacker only possesses a perturbed model. Such a model may correspond to a partial model of the true system, or even an out-dated model. We characterize the attacker by a set of objectives, and propose policies to synthesize stealthy deceptions attacks, both in the case of linear and nonlinear estimators. We show that the more accurate model the attacker has access to, the larger deception attack he can perform undetected. Specifically, we quantify trade-offs between model accuracy and possible attack impact for different BDD schemes. The developed tools can be used to further strengthen and protect the critical state-estimation component in SCADA systems.

Power networks are operated through supervisory control and data acquisition (SCADA) systems complemented by a set of application specific software, usually called energy management systems (EMS). Modern EMS provide information support for a variety of applications related to power network monitoring and control. The power system state estimator (PSSE) is an on-line application which uses redundant measurements and a network model to provide the EMS with an accurate state estimate at all times. The PSSE has become an integral tool for EMS for instance for contingency-constrained optimal power flow. The PSSE also provides important information to pricing algorithms. Monitoring and control of power systems is done through SCADA systems, which collect data from remote terminal units (RTUs) installed in various substations, and relay aggregated measurements to the central master station located at the control center. Several cyber attacks on SCADA systems operating power networks have been reported. Major blackouts, as the August 2003 Northeast blackout, may be caused due to the misuse of the SCADA systems. The 2003 blackout also highlighted the need of robust state estimators that converge accurately and rapidly in such extreme situations, so that necessary preventive actions can be taken in a timely manner. As discussed in Giani et al. [2009], there are several vulnerabilities in the SCADA system architecture, including the direct tampering of RTUs, communication links from RTUs to the control center, and the IT software and databases in the control center. For instance, the RTUs could be targets of denial-of-service (DoS) or deceptions attacks injecting false data Liu et al. [2009].

Power networks, being systems for which control loops are closed over communication networks, represent an important class of networked control systems (NCS). Unlike other IT systems where cyber security mainly involves encryption and protection of data, here cyber attacks may influence the physical processes through the digital controllers. Therefore encryption may not be enough to guarantee security. In order to increase the resilience of these systems, one needs appropriate tools to first understand and then to protect NCS against cyber attacks. Some of the literature has already tackled these problems such as false data injection in power system state estimation Liu et al. [2009], security constrained control Amin et al. [2009*b*], and replay attacks Mo and Sinopoli [2009].

In this chapter, we analyze the cyber security of the PSSE in the SCADA system. In current implementations of PSSE algorithms there are bad data detection (BDD) schemes designed to detect random outliers in the measurement data. Such schemes are based on high measurement redundancy and are performed at the end of the state estimation process. Although such methods may detect basic attacks, they may fail in the presence of more intelligent attackers that wish to stay undetected. We explore the latter by considering scenarios where deception attacks are performed by sending false information to the control center. A related study was performed in Liu et al. [2009] for linear state estimators, assuming the attacker has perfect model knowledge. Here we instead assume the attacker only possesses a perturbed model. Such a model may correspond to a partial model of the true system, or an out-dated model. We characterize the attacker by defining a set of objectives, and propose policies to synthesize stealthy deceptions attacks, both for linear and nonlinear estimators. We show that the more accurate model the attacker has access to, the larger deception attack he can perform undetected. Specifically, we quantify trade-offs between model accuracy and possible attack impact for different BDD schemes.

The outline of this chapter is as follows. We present the main concepts behind state estimation in power systems, the attacker model, and problem formulation in Section 6.2. The properties of estimation algorithm which are deployed in practice are discussed in Section 6.3. In Section 6.4, two common BDD methods are reviewed. The analysis of stealthy deception attacks with partial knowledge is performed in Section 6.5. An example that illustrates the results is presented in Section 6.6, followed by the conclusions in Section 6.7.

## 6.2 Stealthy Deception Attacks

We focus on additive deception attacks aimed toward manipulating the measurements to be processed by the PSSE in such a manner that the resulting systematic errors introduced by the adversary are either undetected or only partially detected by a BDD method. We call such attacks *stealthy deception attacks* on the PSSE. We are also interested in find the class of stealthy deception attacks that do not pose significant convergence issues for the estimator.

### 6.2.1 Power System State Estimation (PSSE)

The basic PSSE problem is to find the best $n$-dimensional state $x$ for the measurement model

$$z = h(x) + \epsilon, \tag{6.1}$$

in a weighted least square (WLS) sense. Here $z$ is the $m$-dimensional vector of measurements, $h$ is a nonlinear function modeling the power network, and $\epsilon \sim \mathcal{N}(0, R)$ is a vector of independent zero-mean Gaussian variables with covariance matrix $R = \text{diag}(\sigma_1^2, \ldots, \sigma_m^2)$. For an electric power network with $N$ buses, the state vector $x = \theta^\top, V^\top)^\top$, where $V = (V_1, \ldots, V_N)^\top$ is the vector of bus voltage magnitudes and $\theta = (\theta_2, \ldots, \theta_N)^\top$ the vector of phase angles. Without loss of generality, bus 1 is considered as the reference bus with $\theta_1 = 0$, so the state dimension is $n = N - 1$. The measurements $z$ can be grouped into two categories: (1) $z_P$, the active power flow measurements $P_{ij}$ from bus $i$ to $j$ and active power injection measurement $P_i$ at bus $i$, and (2) $z_Q$, the reactive power flow measurements $Q_{ij}$ from bus $i$ to $j$, reactive power injection measurement $Q_i$ and $V_i$ voltage magnitude measurement at bust $i$.

Defining the residual vector $r(x) = z - h(x)$, we can write the WLS problem as

$$\min_{x \in \mathbb{R}^n} J(x) = \frac{1}{2} r(x)^\top R^{-1} r(x).$$

The PSSE yields a *state estimate* $\hat{x}$ as a minimizer to this minimization problem. The *measurement estimates* are defined as $\hat{z} := h(\hat{x})$. The WLS estimate $\hat{x}$ satisfies the following first order necessary condition for optimality

$$F(\hat{x}) := \nabla J(\hat{x}) = -H^\top(\hat{x}) R^{-1} r(\hat{x}) = 0, \tag{6.2}$$

where $H = dh/dx$ is the $m \times n$ dimensional measurement Jacobian matrix. The solution $\hat{x}$ of the nonlinear equation $F(\hat{x}) = 0$ may be obtained by the *Newton method* in which a linear equation is solved at each iteration to compute the correction $\Delta x^k := x^{k+1} - x^k$:

$$[F'(x^k)](\Delta x^k) = -F(x^k), \quad k = 0, 1, \ldots, \tag{6.3}$$

where the Hessian matrix $[F'(x^k)] = \nabla^2 J(x^k)$ is given by

$$[F'(x^k)] = H^\top(x^k)R^{-1}H(x^k) + \sum_{i=1}^{m} \frac{r_i(x^k)}{\sigma_i^2}\nabla^2 r_i(x^k).$$

The iterates (6.3) guarantee the convergence to a local minimum as long as the generated sequence $\{x^k\}$ converges and the matrices $[F'(x^k)]$ remain non-singular during the iteration process. A nearly singular Hessian matrix $[F'(x^k)]$ can result in a convergence failure.

The second order information in $[F'(x^k)]$ is computationally expensive, and its effect often negligible when applied to PSSE. Thus, the symmetric approximation is used in practice

$$[F'(x^k)] \approx H^\top(x^k)R^{-1}H(x^k) =: K^k$$

where $K^k$ is called the *gain* (or information) matrix. This approximation leads to the *Gauss-Newton* steps obtained by solving the so called *normal equations*:

$$\left(H^\top(x^k)R^{-1}H(x^k)\right)(\Delta x^k) = H^\top(x^k)R^{-1}r(x^k), \tag{6.4}$$

for $k = 0, 1, \ldots$. For an observable power network, the measurement Jacobian matrix $H(x^k)$ is full column rank. Consequently, the gain matrix $K^k = \sum_{i=1}^m \frac{H_i^\top(x^k)H_i(x^k)}{\sigma_i^2}$ in (6.4) is positive definite and the Gauss-Newton step generates a descent direction, i.e, for the direction $\Delta x^k = x^{k+1} - x^k$ the condition $\nabla J(x^k)^\top \Delta x^k < 0$ is satisfied. . One can observe that $K^k$ is also relatively sparse, i.e., a $H_i(x^k)$ with $p$ non-zero elements introduces $p^2$ non-zero elements in $K^k$. We now present the attacker model.

### 6.2.2 Attacker Model

The main goal of the stealthy attacker is to deceive the PSSE and introduce a desired bias in a set of targeted measurements, known as the *target set*, while remaining undetected by the BDD scheme. More precisely, the goal of a stealthy deception attacker is to compromise the telemetered measurements available to the PSSE such that: 1) The PSSE algorithm converges; 2) For the target set, the estimated measurements at convergence are close to the compromised measurements introduced by the attacker; and 3) The attack remains fully undetected by the BDD scheme.

As a consequence of the attacker's stealthy action, the incorrect state estimates generated by the PSSE can have different affects on other power management functions. In fact, as depicted in Figure 6.1, the state estimate is used as an input to other software applications, in particular the contingency analysis and optimal power flow. These components analyze the state of the grid based on the estimates generated by the PSSE, and compute the optimal control action which reduces the costs while maintaining the grid in a safe state.

Let the corrupted measurement be denoted $z^a$. We assume the following additive attack model

$$z^a = z + a, \tag{6.5}$$

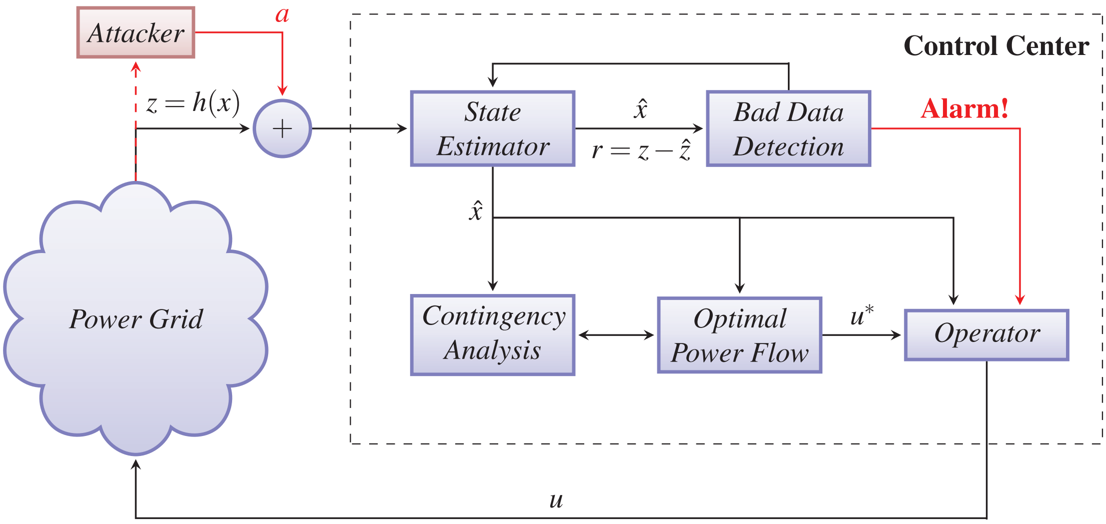


Figure 6.1: The state estimator under a cyber attack.

where $a \in \mathbb{R}^m$ is the attack vector introduced by the attacker. The vector $a$ has zero entries for uncompromised measurements. Under attack, the normal equations (6.4), give the estimates

$$\tilde{x}^{k+1} = \tilde{x}^k + \left(H^\top(\tilde{x}^k)R^{-1}H(\tilde{x}^k)\right)^{-1} H^\top(\tilde{x}^k)R^{-1}r^a(\tilde{x}^k),$$

for $k = 0, 1, \ldots,$ where $\tilde{x}^k$ is the *biased* estimate at iterate $i$, and $r^a(\tilde{x}^k) := z^a - h(\tilde{x}^k)$. If the local convergence conditions hold, then these iterations converge to $\hat{x}^a$, which is the biased state estimate resulting from the use of $z^a$. Thus, the convergence behavior can be expressed as the following statement:

1) The sequence $\{\tilde{x}^0, \tilde{x}^1, \ldots\}$ generated by the mapping

$$G(x) = x + (H^\top(x)R^{-1}H(x))^{-1}H^\top(x)R^{-1}r^a(x),$$

converges to a fixed point $\hat{x}^a$ of $G$ in a region $\mathcal{S}^a_\vartheta$,

where $\mathcal{S}^a_\vartheta$ is a closed ball in $\mathbb{R}^n$ of radius $\vartheta$ governed by the conditions required for the local convergence to hold. We will occasionally use the notation $\hat{x}^a(z^a)$ to emphasize the dependence on $z^a$.

The BDD schemes for PSEE are based on checking if the weighted $p$-norm of the measurement residual is below some threshold $\tau$, which is selected based on permissible false-alarm rate. Thus, the attackers action will be undected by the BDD scheme provided that the following condition holds:

2) The measurement residual under attack $r^a := r(\hat{x}^a) = z^a - h(\hat{x}^a)$, satisfies the condition $\|Wr(\hat{x}^a)\|_p < \tau$.

Finally, let the target set be represented by $\mathcal{I}_{tgrt}$ containing indices of the measurements which are targeted by the attacker. For each $i \in \mathcal{I}_{tgrt}$, the attacker would like the estimated measurement $\hat{z}_i^a := h_i(\hat{x}^a(z^a))$ to be equal to the actual corrupted measurement $z_i^a$. However, such a condition may not be satisfied since corrupted measurements may not be consistent with the model, and can result in violation of conditions 1), and 2) mentioned above. Therefore, we arrive at the following condition which will additionally govern the synthesis of attack vector $a$:

3) The attack vector $a$ is chosen such that $|z_i^a - \hat{z}_i^a| < \eta$ for $i \in \mathcal{I}_{tgrt}$, where $\eta$ is a small positive constant.

The aim of a stealthy deception attacker is then to find and apply an attack $a$ that satisfies conditions 1), 2), and 3). In Section V, we take a similar approach as in Liu et al. [2009] to synthesize stealthy attack policies of the form of $a = \tilde{H}c$, where $\tilde{H}$ is the impefect model known by the attacker. Unlike in Liu et al. [2009], we do not assume the attacker has the exact model of the system and we consider both linear and nonlinear estimators.

## 6.3 PSSE Iterates as Linear WLS Problems

### 6.3.1 Normal Equations as Linear Least Squares

The normal equation can be interpreted as the solution of a linear least squares problem. In particular, writing $H(x^k)$ as $H$, and $\Delta x^k$ as $\Delta x$, and $r(x^k) = z - h(x^k)$ as $\Delta z$ for notational convenience, and defining $\Delta\bar{z} = R^{-1/2}\Delta z$ and $\bar{H} = R^{-1/2}H$, the $k-$th iteration as given by equation (6.4) is the solution of the linear least squares problem

$$\min_{\Delta x}(\Delta\bar{z} - \bar{H}\Delta x)^\top(\Delta\bar{z} - \bar{H}\Delta x).$$

It can be obtained as a solution of the overdetermined system of equations

$$\bar{H}\Delta x \cong \Delta\bar{z}. \tag{6.6}$$

Given that $\bar{H}$ has full column rank and using the notation of the pseudo-inverse $\bar{H}^\dagger := (\bar{H}^\top\bar{H})^{-1}\bar{H}^\top$,

$$\Delta x = \bar{H}^\dagger\Delta\bar{z} = (\bar{H}^\top\bar{H})^{-1}\bar{H}^\top\Delta\bar{z}.$$

For the approximate (linear) model

$$\Delta\bar{z} = \bar{H}\Delta\bar{x} + \bar{\epsilon}$$

where $\bar{\epsilon} = R^{-1/2}\epsilon$, the measurement residual can be expressed as

$$\bar{r} = \bar{S}\bar{\epsilon}, \tag{6.7}$$

where $\bar{S} = (I - \bar{H}(\bar{H}^\top\bar{H})^{-1}\bar{H}^\top)$ is called the weighted sensitivity matrix. Since the matrix $\bar{T} = \bar{H}(\bar{H}^\top\bar{H})^{-1}\bar{H}^\top$ is symmetric and orthogonal with range space $\mathrm{Im}(\bar{H}(\bar{H}^\top\bar{H})^{-1}\bar{H}^\top))$ same as $\mathrm{Im}(\bar{H})$, we call it the *orthogonal projector* on to $\mathrm{Im}(\bar{H})$ and denote it by $\mathcal{P}_{\mathrm{Im}(\bar{H})}$. Such matrix is known as the *hat matrix* in the power system literature. Consequentially, we see that $\bar{S}$ in (6.7) is the orthogonal projector onto the null-space (kernel) of $\bar{H}^\top$, *i.e.* $\bar{S} = (I - \mathcal{P}_{\mathrm{Im}(\bar{H})}) = \mathcal{P}_{\mathrm{Ker}(\bar{H}^\top)}$. Since $\bar{\epsilon} \sim \mathcal{N}(0, \mathcal{I})$, we note that in the absence of gross measurement errors we have $\bar{r} \sim \mathcal{N}(0, \bar{S})$.

### 6.3.2 Decoupled State Estimation

A useful observation in electric power systems is that of active-reactive decoupling, i.e., the active measurements $z_P$ (resp. reactive measurement $z_Q$) predominantly affect the phase angles $\theta$ (resp. the voltage magnitudes $V$). In the decoupled state estimation, the approximate values of the corrections $\Delta\theta$ and $\Delta V$ are then not computed simultaneously, but independently Wu [1990].

Following (6.6), the correction to state estimate $\Delta x = (\Delta\theta^\top, \Delta V^\top)^\top$ at each iteration can be obtained as the solution to the overdetermined system

$$\begin{pmatrix} \bar{H}_{P\theta} & \bar{H}_{PV} \\ \bar{H}_{Q\theta} & \bar{H}_{QV} \end{pmatrix} \begin{pmatrix} \Delta\theta \\ \Delta V \end{pmatrix} = \begin{pmatrix} \Delta\bar{z}_P \\ \Delta\bar{z}_Q \end{pmatrix}, \tag{6.8}$$

where the submatrices $\bar{H}_{P\theta}$ and $\bar{H}_{PV}$ correspond to active measurements and $\bar{H}_{Q\theta}$ and $\bar{H}_{QV}$ correspond to reactive measurements. The mismatches are $\Delta\bar{z}_P$ and $\Delta\bar{z}_Q$. The submatrices and mismatches depend on $\theta$ and $V$, and hence vary from iteration to iteration. The traditional version of fast decoupled state estimation is based on the following decoupled normal equations, where the coupling submatrices $\bar{H}_{PV}$ and $\bar{H}_{Q\theta}$ have been set to zero:

$$\begin{aligned} \Delta\theta^k &= \bar{H}^\dagger_{P\theta}\Delta\bar{z}_P(\theta^k, V^k), \\ \Delta V^k &= \bar{H}^\dagger_{QV}\Delta\bar{z}_Q(\theta^k, V^k). \end{aligned} \tag{6.9}$$

Equations (6.9) are alternately solved for $\Delta\theta^k$ and $\Delta V^k$, where the mismatches $\Delta\bar{z}_P$ and $\Delta\bar{z}_Q$ are evaluated at the latest estimates. The submatrices $\bar{H}_{P\theta}$ and $\bar{H}_{QV}$ are evaluated at flat start and branch series resistances are ignored in forming $\bar{H}_{P\theta}$.

In the new version of fast decoupled state estimator, the matrices $\bar{H}_{PV}$ and $\bar{H}_{Q\theta}$ are not ignored. Using simple matrix operations, (6.8) can be transformed in to the following decoupled form

$$\begin{pmatrix} \bar{H}_{P\theta} & 0 \\ 0 & \tilde{H}_{QV} \end{pmatrix} \begin{pmatrix} \Delta\theta \\ \Delta V \end{pmatrix} = \begin{pmatrix} \Delta\tilde{z}_P \\ \Delta\tilde{z}_Q \end{pmatrix}, \tag{6.10}$$

where $\tilde{H}_{QV} = \bar{H}_{QV} - \bar{H}_{Q\theta}\bar{H}^\dagger_{P\theta}\bar{H}_{PV}$, $\Delta\tilde{z}_Q = \Delta\bar{z}_Q - \bar{H}_{Q\theta}\bar{H}^\dagger_{P\theta}\Delta\bar{z}_P$, and $\Delta\tilde{z}_P = \Delta\bar{z}_P - \bar{H}_{PV}\bar{H}^\dagger_{QV}\Delta\tilde{z}_Q$. The basic (primal) decoupled algorithm which solves (6.10) is presented as follows.

i Compute intermediate angle corrections

$$\Delta\theta^k_{int} = \bar{H}^\dagger_{P\theta}\Delta\bar{z}_P(\theta^k, V^k),$$

ii Compute voltage corrections

$$\begin{aligned} &\Delta\tilde{z}_Q(\theta^k, V^k) = \Delta\bar{z}_Q(\theta^k, V^k) - \bar{H}_{Q\theta}\Delta\theta^k_{int} \\ &\Delta V^k = \tilde{H}^\dagger_{QV}\Delta\tilde{z}_Q(\theta^k, V^k), \end{aligned}$$

iii Compute complementary angle corrections

$$\Delta\theta^k_{com} = -H^\dagger_{P\theta} H_{PV} \Delta V^k$$
$$\Delta\theta^k = \Delta\theta^k_{int} + \Delta\theta^k_{com}$$

For the above algorithm, it can be shown that if measurements are normalized, i.e., measurements are replaced by the normalized measurements $P_{i,m}/V_{i,m}$, $P_{ij,m}/V_{i,m}$, $Q_{i,m}/V_{i,m}$, and $Q_{ij,m}/V_{i,m}$, the matrix $\bar{H}$ can be approximated by a constant matrix evaluated at a flat voltage profile ($V = 1$ and $\theta = 0$). Also, for the $QV$ iteration, it is observed that $\tilde{H}_{QV}$ can be directly obtained from the network topology and element impedances as in the case of $\bar{H}_{QV}$; however, branch susceptances $b_{km}$ are replaced by corresponding reactances $1/x_{km}$.

## 6.4 Bad Data Detection

The measurements used in PSSE may be corrupted by random errors and so a necessary security capability of the PSSE is bad data detection (BDD). Traditionally, the bad data is understood as a result of parameter errors which corrupt the values of modeled circuit elements, incorrect network topology descriptions, and gross measurement errors due to device failures and incorrect meter scans. However, in view of new security threats, bad data can be deliberately introduced by an active adversary which manipulates the communication between remote RTUs and the SCADA system.

Through BDD the PSSE detects gross errors in the measurements, meaning it detects measurements corrupted by errors whose statistical properties exceed the presumed standard deviation or mean. This is achieved by hypothesis tests using the statistical properties of the weighted measurement residual (6.7). We now introduce two of the BDD hypothesis tests widely used in practice, the *performance index test* and the *largest normalized residual test*. These indices are used to model the BDD objective in Section 6.2.2.

### Performance index test

For the measurement error $\bar{\epsilon} \sim \mathcal{N}(0, I)$, the random variable $y := \sum_{i=1}^m \bar{\epsilon}_i^{\,2}$ has a chi-square distribution with $m$ degrees of freedom ($\chi^2_m$) with $\mathbb{E}\{y\} = m$. Consider the quadratic cost function evaluated at the optimal estimate $\hat{x}$

$$J(\hat{x}) = \bar{r}^\top \bar{r} = \bar{\epsilon}^\top \bar{S} \bar{\epsilon}. \tag{6.11}$$

Recalling that $\text{rank}(\bar{H}) = n$, $\text{Im}(\bar{H}) \oplus \text{Ker}(\bar{H}^\top) = \mathbb{R}^m$, and using the definition of orthogonal projector, we note that $\bar{S} = \mathcal{P}_{\text{Ker}(\bar{H}^\top)}$, and we have $\text{rank}(\bar{S}) = m - n$. Therefore, in the absence of bad data, the quadratic form $\bar{\epsilon}^\top \bar{S} \bar{\epsilon}$ has a chi-squares distribution with $m - n$ degrees of freedom, *i.e.* $J(\hat{x}) \sim \chi^2_{m-n}$ with $\mathbb{E}\{J(\hat{x})\} = m - n$. The main idea behind the performance index test is to use $J(\hat{x})$ as an approximation of $y$ and check if $J(\hat{x})$ follows the distribution $\chi^2_{m-n}$. This can be posed as a hypothesis test with a null hypothesis $H_0$, which if accepted means there is no bad data, and an alternative bad data hypothesis $H_1$ where

$$H_0 : \mathbb{E}\{J(\hat{x})\} = m - n, \quad H_1 : \mathbb{E}\{J(\hat{x})\} > m - n$$

Defining $\alpha \in [0,\ 1]$ as the significance level of the test corresponding to the false alarm rate, and $\tau_{\chi}(\alpha)$ such that

$$\int_{0}^{\tau_{\chi}(\alpha)} g^{\chi}(u)du = 1 - \alpha, \tag{6.12}$$

where $g^{\chi}(u)$ is the probability distribution function (pdf) of $\chi^2_{m-n}$, and noting that $J(\hat{x}) = \|R^{-1/2}r(\hat{x})\|_2$ the result of the test is

$$\begin{aligned} \text{reject } H_0 \text{ if } \|R^{-1/2}r\|_2 &> \sqrt{\tau_{\chi}(\alpha)}, \\ \text{accept } H_0 \text{ if } \|R^{-1/2}r\|_2 &\leqslant \sqrt{\tau_{\chi}(\alpha)}. \end{aligned}$$

### Largest normalized residual test

From (6.7), we note that $\bar{r} \sim \mathbb{N}(0, \bar{S})$ and equivalently $r \sim \mathbb{N}(0, \Omega)$ with $\Omega = R^{1/2}\bar{S}R^{1/2}$. Now consider the normalized residual vector

$$r^N = D^{-1/2}r, \tag{6.13}$$

with $D \in \mathbb{R}^{m\times m}$ being a diagonal matrix defined as $D = \operatorname{diag}(\Omega)$. In the absence of bad date each element $r_i^N,\ i = 1, \ldots, m$ of the normalized residual vector then follows a normal distribution with zero mean and unit variance, *i.e.* $r_i^N \sim \mathbb{N}(0, 1),\ \forall i = 1, \ldots, m$. Thus, bad data could be detected by checking if $r_i^N$ follows $\mathbb{N}(0, 1)$. Posing this as hypothesis test for each element $r_i^N$

$$H_0 : \mathbb{E}\left\{r_i^N\right\} = 0, \quad H_1 : \mathbb{E}\left\{|r_i^N|)\right\} > 0$$

Again defining $\alpha \in [0,\ 1]$ as the significance level of the test and $\tau_{\mathbb{N}}$ such that

$$\int_{-\tau_{\mathbb{N}}(\alpha)}^{\tau_{\mathbb{N}}(\alpha)} g^{\mathbb{N}}(u)du = 1 - \alpha, \tag{6.14}$$

where $g^{\mathbb{N}}(u)$ is the pdf of $\mathbb{N}(0, 1)$, and noting (6.13), the result of the test is

$$\begin{aligned} \text{reject } H_0 \text{ if } \|D^{-1/2}r\|_\infty &> \tau_{\mathbb{N}}(\alpha) \\ \text{accept } H_0 \text{ if } \|D^{-1/2}r\|_\infty &\leqslant \tau_{\mathbb{N}}(\alpha) \end{aligned}$$

We observe that for the case of single measurement with bad data, the largest normalized residual element $|r_i^N|$ corresponds to the corrupted measurement. It is clear that both tests may be written as $\|Wr(\hat{x})\|_p < \tau$, for suitable $W$, $p$ and $\tau$.

## 6.5 Deception Attacks on Linear State Estimator

Several scenarios of stealthy deception attacks on PSSE for the DC case have been analyzed in Liu et al. [2009]. The authors of Liu et al. [2009] considered linear models, which was fully known by the attacker, and focused on additive attack policies that would guarantee the measurement residual to remain unchanged for the linear least squares algorithm. The feasibility of such attack policies was then analyzed for several IEEE benchmarks under different resource constraints of the attacker (for e.g., number of sensors the attacker could corrupt) and attacker objectives (for e.g., random attack, targeted attack). The main result related to attack policies was that if the attack vector $a$ was in the range space of $H$, then the measurement residual $r^a = (z+a) - H\hat{x}$ would be the same as the residual $r$ when there was no attack. Thus, such attack vectors would not increase the residual. Such undetectable errors have been analyzed previously within the power system's community, see Wu and Liu [1989].

In this section we analyze how the attacker may fulfill the objective Section 6.2.2, and thereby remain undetected.

### 6.5.1 Attack Synthesis

As indicated in Section 6.2.2, the attacker aims at injecting false data in a few targeted measurements without being detected. In general a stealthy attack requires the corruption of more measurements than the targeted ones, see Liu et al. [2009]; Sandberg et al. [2010]. This relates to the fact that a stealthy attack must have the attack vector $a$ fitting the measurement model, which for the weighted linear case is equivalent to have $a \in \operatorname{Im}(\bar{H})$.

We now present a general methodology for synthesizing stealthy attacks for the linear case with specific target constraints. Suppose the attacker wishes to compute an attacker vector $a$ such that $\bar{z}^a = \bar{z} + a$ satisfies a set of goals, encoded by $a \in \mathcal{G}$, and the attack is stealthy, i.e. $a \in \operatorname{Im}(\bar{H})$. Assuming the attacker knows the weighted measurement model $\bar{H}$, such attack could be computed by solving the optimization problem

$$\begin{aligned} \min_{a} \; & \|a\|_p \\ \text{s.t.} \; & a \in \mathcal{G}, \;\; a \in \operatorname{Im}(\bar{H}) \;, \end{aligned} \tag{6.15}$$

corresponding to the "least-effort“ attack in the $p$-norm sense. An interesting case is that of $p = 0$, which means the attacker is computing the attack with minimum cardinality, e.g., minimizing the number of sensors to corrupt. Another particular formulation is the 2-norm case with a single attack target, $z_a^i = z_i + 1$ or $a_i = 1$. By recalling that $a \in \operatorname{Im}(\bar{H})$ means that $a = \bar{H}c$ for a given $c$, the optimization problem may be recast as

$$\begin{aligned} \min_{c} \; & \|\bar{H}c\|_2^2 \\ \text{s.t.} \; & e_i^\top \bar{H}c = 1 \end{aligned} \;, \tag{6.16}$$

where $e_i$ is a unitary vector with 1 in the $i$-th component. Recall $\bar{T} = \mathcal{P}_{\operatorname{Im}(\bar{H})} = \bar{H}\bar{H}^\dagger$.

**Proposition 6.5.1.** *The optimal solution $a^*$ to the optimization problem (6.16) is given by $a^* = \frac{\bar{T}}{\bar{T}_{ii}} e_i$*

*Proof.* The Lagrangian of this optimization problem is $L(c,\nu) = c\bar{H}^\top \bar{H}c + \nu(e_i^\top \bar{H}c - 1)$ and the KKT conditions for an optimal solution $(c^*, \nu^*)$ are

$$\begin{cases} \bar{H}^\top \bar{H}c^* + \nu^* \bar{H}^\top e_i = 0 \\ \qquad\quad e_i^\top \bar{H}c^* - 1 = 0 \end{cases}. \tag{6.17}$$

Since it is assumed the power network is observable, the solution for the first equation is $c^* = \nu^* \bar{H}^\dagger e_i$. Including this in the second equation results in $\nu^* e_i^\top \bar{T} e_i = 1$ which is equivalent to $\nu^* = \frac{1}{\bar{T}_{ii}}$ with $\bar{T}_{ii}$ being the $i$-th diagonal element of $\bar{T}$. We then have that $a^* = \bar{H}c^* = \frac{\bar{T}}{\bar{T}_{ii}} e_i$. □

In the power system's literature, the hat matrix $\bar{T}$ is known to have information regarding measurement redundancy and correlation. This result highlights a new meaning: each column of $\bar{T}$ actually corresponds to an optimal attack vector yielding a zero residual.

### 6.5.2 Relaxing the Assumptions on Adversarial Knowledge

Here we consider the scenario where the attacker has only a partial or corrupted knowledge of the measurement model. Such knowledge may be obtained, for instance, by recording and analyzing data sent from the RTUs to the control center using suitable statistical methods. The corrupted measurement model may also correspond to an out-dated model or an estimated model using the power network topology, usual parameter values and uncertain operating point. We further assume that the covariance matrix $R$ is known.

In the following analysis we provide bounds on the measurement residual under this kind of attack scenario. These bounds give some insights on what attacks may go undetected, given the model uncertainty. For the moment we assume there are no random errors in the measurements and so we consider the weighted measurements $\bar{z} = \bar{H}x$.

Let the perturbed measurement model known by the attacker be denoted by $\tilde{H}$, such that

$$\tilde{H} = \bar{H} + \Delta\bar{H}, \tag{6.18}$$

and consider the linear policy to compute attacks on the measurements to be $a = \tilde{H}c$, resulting in the corrupted set of measurements $\bar{z}^a = \bar{z} + a$. Recall the objectives of the attacker as defined in Section 6.2.2.

The third objective, being undetected, depends both on the desired bias on the flow measurements $a$ and on the model uncertainty $\Delta\bar{H}$. The measurement residual under attack, $r^a := \bar{r}(\bar{z}^a)$, can be written as

$$\bar{r}(\bar{z}^a) = \bar{S}(\bar{z} + \tilde{H}c) = \bar{S}\bar{z} + \bar{r}_a 6) \tag{6.19}$$

Using (6.18) and the fact that $\bar{S} = \mathsf{P}_{\mathrm{Ker}(\bar{H}^\top)}$, we can rewrite it as

$$\bar{r}(\bar{z}^a) = \bar{S}(\bar{z} + \bar{H}c) + \bar{S}\Delta\bar{H}c = \bar{S}\Delta\bar{H}c. \tag{6.20}$$

We denote $\bar{r}_a = \bar{S}\Delta\bar{H}c$ as the residual due to the attack, since it only depends on $c$ and $\Delta\bar{H}$. Furthermore, we see that $\|\bar{r}_a\| \leqslant \|\bar{S}\|\|\Delta\bar{H}\|\|c\| = \|\Delta\bar{H}\|\|c\|$, since $\bar{S}$ is an orthogonal projector, showing that the residual norm is linear in terms of the model uncertainty. However, this bound does not capture an important property of the sensitivity matrix $\bar{S}$, *i.e.*, $\bar{S}$ is the orthogonal projector on to $\mathrm{Ker}(\bar{H}^\top)$. To show this, assume $\tilde{H} = \delta\bar{H}$ for some nonzero $\delta$, yielding $\Delta\bar{H} = (1-\delta)\bar{H}$. From the previous result we have $\|\bar{r}_a\| \leqslant \|(1-\delta)\bar{H}\|\|c\|$. However, since $\bar{S}$ is the orthogonal projector onto $\mathrm{Ker}(\bar{H}^\top)$ and this subspace is the orthogonal complement of $\mathrm{Im}(\bar{H})$ we know that $\bar{r}_a = \bar{S}\Delta\bar{H}c = 0$. Therefore, although there is model uncertainty, the residual is still zero. This reasoning indicates that there is a geometrical meaning in the residual, since all the model perturbations $\Delta\bar{H}$ spanning $\mathrm{Im}(\bar{H})$ will yield a zero residual. To further explore this property, we will make use of the so-called principal angles and projection theory described in Galántai [2006]. The main results and definitions used in this work are now given.

**Definition 6.5.2** (Galántai [2006])**.** Let $M_1$ and $M_2$ be subspaces of $\mathbb{C}^m$. The smallest principal angle $\gamma_1 \in [0, \pi/2]$ between $M_1$ and $M_2$ is defined by

$$\begin{aligned} \cos(\gamma_1) &= \max_{u\in M_1}\max_{v\in M_2} |u^H v| \\ &\text{subject to } \|u\| = \|v\| = 1 \end{aligned} \tag{6.21}$$

**Lemma 6.5.3** (Galántai [2006])**.** *Let $\mathcal{P}_1, \mathcal{P}_2 \in \mathbb{R}^{m\times m}$ be orthogonal projectors of $M_1$ and $M_2$, respectively. Then the following holds*

$$\|\mathcal{P}_1\mathcal{P}_2\|_2 = \cos(\gamma_1) \tag{6.22}$$

**Proposition 6.5.4.** *Let $\gamma_1$ be the smallest principal angle between $\mathrm{Ker}(\bar{H}^\top)$ and $\mathrm{Im}(\tilde{H})$. The residual increment due to a deception attack following the policy $a = \tilde{H}c$ satisfies*

$$\|\bar{r}_a\|_2 \leqslant \cos\gamma_1 \|a\|_2. \tag{6.23}$$

*Proof.* Recall the so-called hat matrix defined by $T = \bar{H}\bar{H}^\dagger$, which is the orthogonal projector onto $\mathrm{Im}(\bar{H})$ and define $\tilde{T} = \mathcal{P}_{\mathrm{Im}(\tilde{H})} = \tilde{H}\tilde{H}^\dagger$. The residual under attack in Eq. (6.19) may be rewritten as

$$\bar{r}_a = \bar{S}\tilde{T}\tilde{H}c, \tag{6.24}$$

since $\tilde{T}\tilde{H} = \tilde{H}$. The residual norm can be upper bounded as

$$\|\bar{r}_a\|_2 \leqslant \|\bar{S}\tilde{T}\|_2\|\tilde{H}c\|_2 = \cos\gamma_1\|a\|_2, \tag{6.25}$$

where $\gamma_1$ is the smallest principal angle between $\mathrm{Ker}(\bar{H}^\top)$ and $\mathrm{Im}(\tilde{H})$. □

Analyzing the example where $\tilde{H} = \delta\bar{H}$, we see that $\mathrm{Im}(\tilde{H}) = \mathrm{Im}(\bar{H})$ is orthogonal to $\mathrm{Ker}(\bar{H}^\top)$. Hence the smallest principal angle between these subspaces is $\gamma_1 = \frac{\pi}{2}$, yielding $\|\bar{r}_a\|_2 \leqslant \cos(\gamma_1)\|a\|_2 = 0$.

Thus we achieved a tighter bound that explores the geometrical properties of the residual subspace. In brief, $\gamma_1$ measures how close the subspaces $\mathrm{Ker}(\bar{H}^\top)$ and $\mathrm{Im}(\tilde{H})$ are from each other. In order for the model uncertainty not to affect the residual, it is desired that $\mathrm{Ker}(\bar{H}^\top)$ and $\mathrm{Im}(\tilde{H})$ are as close to orthogonal as possible.

### 6.5.3 Stealthy Attacks

Consider the measurement residual under attack in (6.19). Taking into account the random error vector $\bar{\epsilon}$ we can rewrite the residual as

$$\bar{r}(\bar{z}^a) = \bar{S}\bar{\epsilon} + \bar{S}a. \tag{6.26}$$

The residual then has the following distribution $\bar{r}(\bar{z}^a) \sim \mathcal{N}(\bar{r}_a, \bar{S})$. Note that due to the model uncertainties the residual has a non-zero mean, which increases the chances of triggering an alarm in the BDD. Recall that one of the attacker's objective is to keep such probability as low as possible, *i.e.* $\|Wr(\hat{x}^a)\|_p < \tau$. We now provide insights on how such objective may be fulfilled for the two BDD schemes presented in Section 6.4.

#### Performance index test

Recall that without any attack on the measurements we have $J(\hat{x}) \sim \chi^2_{m-n}$. Under attack the cost function $J_a(\hat{x}) = \bar{r}(\bar{z}^a)^\top \bar{r}(\bar{z}^a)$ will have the so-called *non-central chi-squares* distribution Muirhead [1982], due to the non-zero mean which affects all the statistical moments of the $\chi^2_{m-n}$ distribution. We denote $J_a(\hat{x}) \sim \chi^2_{m-n}(\lambda)$ where $\lambda = \|\bar{S}a\|_2^2$. Recalling the relationship between the false alarm probability $\alpha$ and the detection threshold $\tau_\chi(\alpha)$ in (6.12), in the presence of attacks we have

$$\int_{\tau_\chi(\alpha)}^{\infty} g_\lambda(u)du = \alpha + \delta_\lambda(\lambda), \tag{6.27}$$

with $g_\lambda(u)$ being the pdf of $\chi^2_{m-n}(\lambda)$. We call $\delta_\lambda(\lambda)$ the increase in the alarm probability that the attacker must minimize to remain undetected. It is not possible to attack the PSSE and guarantee that no alarm is triggered, due to the presence of random measurement errors. Therefore we assume the attacker has an upper limit on $\delta_\lambda(\lambda)$ which is considered acceptable, $\bar{\delta}_\lambda$. Given reasonable values of $\alpha$, the attacker is able to compute feasible values of $\lambda$ by solving

$$\int_{\tau_\chi(\alpha)}^{\infty} g_\lambda(u)du \leqslant \alpha + \bar{\delta}_\lambda. \tag{6.28}$$

Under the reasonable assumption that $\delta_\lambda(\lambda)$ increases with $\lambda$, since the mean of $\chi^2_{m-n}(\lambda)$ is shifted along the positive direction and its variance increases as $\lambda$ increases, we provide the following result.

**Proposition 6.5.5.** *Given $\alpha$ and $\bar{\delta}_\lambda$ an attack is stealthy regarding the performance index test if the following holds*

$$\cos\gamma_1 \|a\|_2 \leqslant \sqrt{\bar{\lambda}(\alpha, \bar{\delta}_\lambda)} \tag{6.29}$$

*where $\bar{\lambda}(\alpha, \bar{\delta}_\lambda)6$ is the maximum value of $\lambda$ for which (6.28) is satisfied.*

*Proof.* First note that from our assumption $\delta_\lambda(\lambda)$ increases with $\lambda$. Therefore stealthy attack vectors satisfy $\|\bar{r}_a\|_2 \leqslant \sqrt{\bar{\lambda}}$, as this implies by definition that $\lambda \leqslant \bar{\lambda}$ and $\delta_\lambda(\lambda) \leqslant \bar{\delta}_\lambda$. The rest of the proof follows from Prop. 6.5.4. □

### Largest normalized residual test

Recall that the residuals without attack follow a normal distribution $\bar{r} \sim \mathcal{N}(0, \bar{S})$, whereas under attack we have $\bar{r}_a \sim \mathcal{N}(d, \bar{S})$ with $d = \bar{S}a$. Each element of the normalized residual vector then has distribution $r^N_{a_i} \sim \mathcal{N}(d^N_i, 1)$ with $d^N_i = D^{-1/2}_{ii} d_i$ being the bias introduced by the attack vector. Similarly as before, defining $\bar{\delta}_d$ as the maximum admissible increase in the alarm probability and given $\alpha$, the biases $d^N_i$ providing the required level of stealthiness satisfy the inequality

$$\int_{-\tau_{\mathbb{N}}(\alpha)}^{\tau_{\mathbb{N}}(\alpha)} g^{\mathbb{N}}_{d^N_i}(u)du \geqslant 1 - \alpha - \bar{\delta}_d, \tag{6.30}$$

with $g^{\mathbb{N}}_{d^N_i}(u)$ being the pdf of $r^N_{a_i}$.

**Proposition 6.5.6.** *Given $\alpha$ and $\bar{\delta}_d$ an attack is stealthy regarding the largest normalized residual test if the following holds*

$$\|D^{-1/2}\|_2 \cos\gamma_1 \|a\|_2 \leqslant \bar{d}^N(\alpha, \bar{\delta}_d) \ , \tag{6.31}$$

*where $\bar{d}^N(\alpha, \bar{\delta}_d)$ is the maximum value of $\|d^N\|_\infty$ for which (6.30) is satisfied with $d^N_i = \|d^N\|_\infty$.*

*Proof.* Clearly it is sufficient to require (6.30) to hold for $|d^N_i| = \|d^N\|_\infty$, as this corresponds to the worst-case bias. Note that the increase in alarm probability $\delta_d$ increases with $|d^N_i|$ due to the symmetrical nature of $g^{\mathbb{N}}_{d^N_i}(u)$. Thus (6.30) reaches equality for $\|d^N\|_\infty = \bar{d}^N$ and a sufficient condition for (6.30) to hold is to have $\|d^N\|_\infty \leqslant \bar{d}^N$. Recalling $d^N = D^{-1/2}\bar{S}a$ and $\|\cdot\|_\infty \leqslant \|\cdot\|_2$, we conclude the attack is stealthy if $\|D^{-1/2}\bar{S}a\|_2 \leqslant \bar{d}^N$, which is satisfied by $\|D^{-1/2}\|_2\|\bar{S}a\|_2 \leqslant \bar{d}^N$. The rest follows from Prop. 6.5.4. □

The main result of this section is as follows:

**Theorem 6.5.7.** *Given the perturbed model $\tilde{H}$, the false-alarm probability $\alpha$ and the maximum admissible increase in alarm probability $\bar{\delta}$, an attack following the policy $a = \tilde{H}c$ is stealthy if*

$$\|a\|_2 \leqslant \beta(\alpha, \bar{\delta}) \ , \tag{6.32}$$

*where $\beta(\alpha, \bar{\delta})$ is given by the BDD scheme of the SCADA system.*

*Proof.* Assuming the BDD method is the performance index and taking $\beta(\alpha, \bar{\delta}) = \frac{\sqrt{\bar{\lambda}(\alpha, \bar{\delta}_\lambda)}}{\cos\gamma_1}$, the proof directly follows from Prop. 6.5.5. For the largest normalized residual, defining $\beta(\alpha, \bar{\delta}) = \frac{\bar{d}^N(\alpha, \bar{\delta}_d)}{\|D^{-1/2}\|_2 \cos\gamma_1}$ the proof follows from Prop. 6.5.6. □

Note that in the scenario analyzed here the designer of the BDD scheme chooses both the detection method as well as the false-alarm probability $\alpha$. These elements are fixed and usually unknown to the attacker, who defines the maximum risk $\bar{\delta}$ he is willing to take and has some knowledge of the power network $\tilde{H}$, that used to compute the attack vector $a$. However $\alpha$ can be estimated by reasonable values and the same happens for the degrees of freedom of the chi-squares distribution.

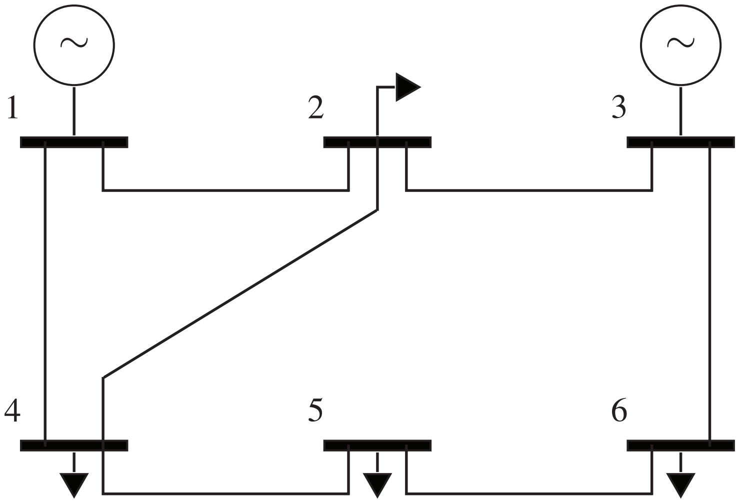


Figure 6.2: Power network with 6 buses.

## 6.6 Case study

### 6.6.1 Worst-Case Model Uncertainty

An interesting analysis is to understand what is the worst-case uncertainty $\Delta\bar{H}$ maximizing the orthogonality between $\mathrm{Im}(\tilde{H})$ and $\mathrm{Ker}(\bar{H}^\top)$. This corresponds to maximize the effect of the attack vector $a$ on the measurement residual. From the attacker's view, this could lead to a set of robust attack policies. As for the control center this could be useful to implement security measures based on decoys, for instance. It is known that the network model used in the PSSE can be kept in the databases of the SCADA system with little protection. Thus a possible defensive strategy would be to replace that model by a perturbed one which, if used by an attacker, would increase the residuals and increase the detection of intelligent attacks.

The first observation at this point is that it is of little interest to consider cases when only the maximum magnitude of the model perturbation is considered, *i.e.* $\|\Delta\bar{H}\| \leqslant \omega$. Note that this formulation only tells us that the uncertainty is within a ball of radius $\omega$ from the nominal model $\bar{H}$. Thus one can always choose a worst-case perturbation satisfying $\|\Delta\bar{H}\| = \omega$ which is orthogonal to $\bar{H}$, yielding $\|\bar{S}\bar{T}_\Delta\| = 1$. Hence scenarios where the uncertainty is more structured are of greater interest.

We now apply the previous results to the scenario where the attacker knows the exact topology of the network but has an error on the transmission line's parameters of ±20%. The detectability of attacks in this scenario is intimately related to the detectability of parameter or topology errors. Consider the power network in Fig. 6.2 with the data in Tab. 6.1 and linear measurement model $z = Hx$.[1] The parameter errors in Tab. 6.1 were computed so that $\cos(\gamma_1) = \|\bar{S}\tilde{T}\|_2$ is maximized for errors up to ±20%, corresponding to the worst-

[1] The author is grateful to Andre Teixeira for conducting this case study.

Table 6.1: Data of the 6 buses network.

| Branch | From bus | To bus | Reactance (pu) | Parameter Error |
|---|---|---|---|---|
| $b1$ | 1 | 4 | 0.370 | -20% |
| $b2$ | 1 | 2 | 0.518 | +20% |
| $b3$ | 6 | 5 | 1.05 | -20% |
| $b4$ | 6 | 3 | 0.640 | -20% |
| $b5$ | 5 | 4 | 0.133 | -20% |
| $b6$ | 4 | 2 | 0.407 | -20% |
| $b7$ | 3 | 2 | 0.300 | +20% |

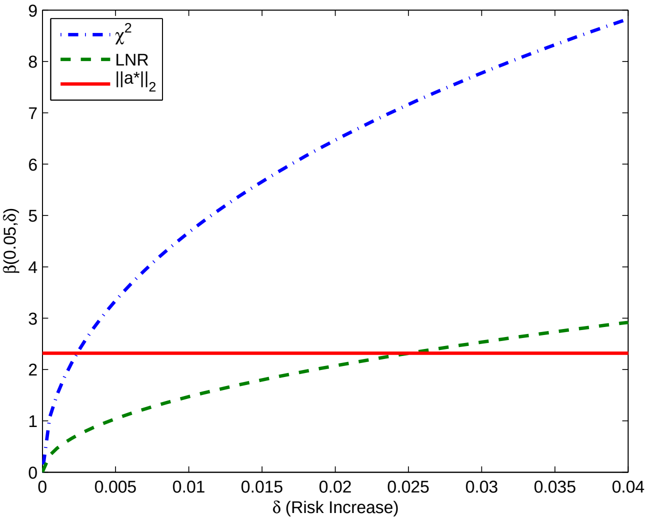


Figure 6.3: Attack stealthiness as a function of the detection risk.

case uncertainty. Note that this actually corresponds to the constrained maximization of a convex function, which was solved using the numerical solvers available in Matlab.

In Fig. 6.3 we show how the maximum 2-norm of a stealthy attack vector $\beta(\alpha, \delta)$ in terms of Thm. 6.5.7 varies with respect to the detection risk $\delta$, for $\alpha = 0.05$. The solid line represents the 2-norm of the optimal attack vector $a^*$ constrained by $a_{b_1} = 1$, where $a_{b_1}$ is the power flow in branch $b_1$. The curves denoted as $\chi^2$ and $LNR$ represent the value of $\beta(0.05, \delta)$ for the performance index test and largest normalized residual test, respectively. As it is seen, the performance index test allows for larger attacks than the largest normalized residual test. Since attacks following $a = \tilde{H}c$ have a similar meaning to multiple interacting bad data, this validates the known fact that largest normalized residual test is more robust to such bad data than the performance index test. Note that the norm of the optimal attack vector in the sense of (6.16) when targeting the power flow between buses 1 and 4 is also shown. We see that such attack would have a small risk, even for the largest normalized residual.

## 6.7 Discussion

In this chapter we provided methods to analyze cyber-security of PSSE in scenarios where the attacker has a limited knowledge of the network and unlimited resources. In particular we proposed a framework to model such attackers, which is capable of taking into account resource constraints. We also considered two BBD methods widely used and showed that such tools do not guarantee security against cyber-attacks.

# Chapter 7

# Security Constrained Networked Control

## 7.1 Introduction

In this chapter, we consider the problem of security constrained optimal control for discrete-time, linear dynamical systems in which control and measurement packets are transmitted over a communication network. The packets may be jammed or compromised by a malicious adversary. For a class of denial-of-service (DoS) attack models, the goal is to find an (optimal) causal feedback controller that minimizes a given objective function subject to safety and power constraints. We present a semi-definite programming based solution for solving this problem. Our analysis also presents insights on the effect of attack models on solution of the optimal control problem.

As discussed in the previous chapters, attacks to computer networks have become prevalent over the last decade. While most control networks have been safe in the past, they are currently more vulnerable to malicious attacks Cárdenas et al. [2008]; Turk [2005]. The consequences of a successful attack on control networks can be more damaging than attacks on other networks because control systems are at the core of many critical infrastructures. Therefore, analyzing the security of control systems is a growing concern Cárdenas et al. [2008]; Nguyen et al. [2008]; Pinar et al. [2010]; Salmeron et al. [2004]; Turk [2005].

There is a significant body of work on networked control Schenato et al. [2007], stochastic verification Amin et al. [2006]; Chatterjee, de Alfaro and Henzinger [2009], robust control Amin et al. [2007]; Ben-tal et al. [2005]; D.Q. [2000]; Goulart et al. [2006], and fault-tolerant control Yu et al. [1994]. We argue that several major security concerns for control systems are not addressed by the current literature. For example, fault analysis of control systems usually assumes independent modes of failure, while during an attack, the modes of failure will be highly correlated. The existing body of work in networked control systems assumes that the failure modes follow a given class of probability distributions; however, a real attacker has no incentives to follow this assumed distribution, and may attack in a non-deterministic manner. Finally, the work in stochastic system verification has addressed safety and reachability problems for fairly general systems; however, the potential applicability of these results for securing control systems has not been studied.

In this chapter, we formulate and analyze the problem of secure control for discrete-time linear dynamical systems. Our work is based on two ideas: (1) the introduction of safety-constraints as one of the top security requirements of a control system, and (2) the introduction of new adversary models—we generalize traditional uncertainty classes for control systems to incorporate more realistic attacks. The goal in our model is to minimize a performance function such that a safety specification is satisfied with high probability and power limitations are obeyed in expectation when the sensor and control packets can be dropped by a random or a resource-constrained attacker. Our analysis uses tools from optimal control theory such as dynamic and convex programming.

### 7.1.1 Attacks on control systems

As discussed in the previous chapters, malicious cyber attacks to networked control systems can be classified as either *deception* attacks or *denial-of-service* DoS attacks.
In the context of control systems, integrity refers to the trustworthiness of sensor and control data packets. A lack of integrity results in deception: when a component receives false data and believes it to be true. Figure 7.1 adopts a simplistic viewpoint, where A1 and A3 represent deception attacks, and the adversary sends false information $\tilde{y} \neq y$ or $\tilde{u} \neq u$ from (one or more) sensors or controllers. The false information can include: an incorrect measurement, the incorrect time stamp, or the incorrect sender identity. The adversary can launch these attacks by compromising some sensors (A1) or controllers (A3).
On the other hand, availability of a control system refers to the ability of all components of being accessible. Lack of availability results in a DoS of sensor and control data. A2 and A4 represent *DoS attacks* in Figure 7.1, where the adversary prevents two entities from communicating. To launch a DoS the adversary can jam the communication channels, compromise devices and prevent them from sending data, attack the routing protocols, flood with network traffic some devices, etc.
Lastly, A5 represents a direct attack against the actuators or the plant. Solutions to these attacks, fall in the realm of detecting such attacks and improving the physical security of the system.

As shown by the analysis of a database that tracked cyber-incidents affecting industrial control systems from 1982 to 2003 Byres and Lowe [2004], DoS is the most likely threat to control systems; therefore in this chapter we focus on DoS attacks.

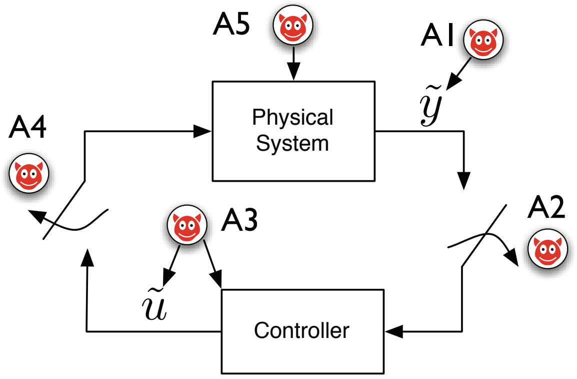


Figure 7.1: Attacks on a control system.

## 7.2 Problem Setting

### 7.2.1 System Model

We consider a linear time invariant stochastic system over a time horizon $k = 0, \ldots, N-1$ with measurement and control packets subject to DoS attacks $(\gamma_k, \nu_k)$:

$$x_{k+1} = Ax_k + Bu_k^a + w_k \qquad k = 0, \ldots, N-1, \tag{7.1}$$

$$u_k^a = \nu_k u_k \qquad \nu_k \in \{0, 1\}, \tag{7.2}$$

$$x_k^a = \gamma_k x_k \qquad \gamma_k \in \{0, 1\}, \tag{7.3}$$

where $x_k \in \mathbb{R}^n$ and $u_k \in \mathbb{R}^m$ denote the state and the control input respectively, $w_k \in \mathbb{R}^n$ is independent, Gaussian distributed noise with mean 0 and covariance $W$ (denoted as $w_k \sim \mathcal{N}(0, W)$), $x_0 \sim \mathcal{N}(\bar{x}, P_0)$ is the initial state, and $\{\gamma_k\}$ (resp. $\{\nu_k\}$) is the sensor (resp. actuator) attack sequence. Also, $x_0$ and $w_k$ are uncorrelated. The available state (resp. available control input) is denoted by $x_k^a$ (resp. $u_k^a$) after a DoS attack on the measurement (resp. control) packet. Following Schenato et al. [2007], for an acknowledgment based communication protocol such as TCP, the information set available at time $k$ is $\mathcal{I}_k = \{x_0^a, \ldots, x_k^a, \gamma_0^k, \nu_0^{k-1}\}$ where $\gamma_i^j = (\gamma_i, \ldots, \gamma_j)$ and $\nu_i^j = (\nu_i, \ldots, \nu_j)$. Define $u_0^{N-1} = (u_0, \ldots, u_{N-1})$.

We note that due to (7.3), the controller receives perfect state information $x_k$ when $\gamma_k = 1$ and 0 when $\gamma_k = 0$. However, our analysis presented can also be extended for the case of measurement equation $y_k^a = \gamma_k C_s x_k + v_k$.

### 7.2.2 Goals and Requirements

At this stage, we have not specified any restrictions on the DoS attack actions except that $(\gamma_k, \nu_k) \in \{0, 1\}^2$ for $k = 0, \ldots, N-1$. We will impose constraints on the attacker actions in Section 7.3.1. Given such constraints, our goal is to synthesize a causal feedback control law $u_k = \mu_k(\mathcal{I}_k)$ such that for the system (7.1), (7.2), and (7.3), the following finite-horizon objective function is minimized

$$J_N(\bar{x}, P_0, u_0^{N-1}) = \mathsf{E}\left[x_N^\top Q^{xx} x_N + \sum_{k=0}^{N-1} \begin{pmatrix} x_k \\ u_k \end{pmatrix}^\top \begin{pmatrix} I_n & 0 \\ 0 & \nu_k I_m \end{pmatrix} Q \begin{pmatrix} x_k \\ u_k \end{pmatrix} |u_0^{N-1}, \bar{x}, P_0\right] \tag{7.4}$$

where $Q^{xx} \succ 0$, and $Q \succeq 0$ is partitioned as

$$Q = \begin{pmatrix} Q^{xx} & 0 \\ 0 & Q^{uu} \end{pmatrix} \in \mathbb{R}^{(n+m)\times(n+m)},$$

and constraints on *both* the state and the input in an expected sense

$$\mathsf{E}\left[\begin{pmatrix} x_k \\ u_k \end{pmatrix}^\top \begin{pmatrix} I_n & 0 \\ 0 & \nu_k I_m \end{pmatrix} H_i \begin{pmatrix} x_k \\ u_k \end{pmatrix}\right] \leqslant \beta_i \quad \text{for } i = 1, \ldots, L, \text{ and } k = 0, \ldots, N-1 \tag{7.5}$$

with $H_i \succeq 0$ and scalar constraints on the state and the input in a probabilistic sense

$$\mathsf{P}\left[t_i^\top \begin{pmatrix} I_n & 0 \\ 0 & \nu_k I_m \end{pmatrix} \begin{pmatrix} x_k \\ u_k \end{pmatrix} \leqslant \alpha_i\right] \geqslant (1-\varepsilon) \quad \text{for } i = 1, \ldots, T, \text{ and } k = 0, \ldots, N-1 \tag{7.6}$$

with $t_i \in \mathbb{R}^{n+m}$ are satisfied. The constraints (7.5) can be viewed as *power constraints* that limit the energy of state and control inputs at each time step. The constraint (7.6) can be interpreted as a *safety specification* stipulating that the state and the input remain within the hyperplanes specified by $t_i$ and $\alpha_i$ with a sufficiently high probability, $(1-\varepsilon)$, for $k = 0, \ldots, N-1$. Equations (7.5) and (7.6) are to be interpreted as conditioned on the initial state, i.e., $\mathsf{E}[\cdot] := \mathsf{E}[\cdot|x_0]$ and $\mathsf{P}[\cdot] := \mathsf{P}[\cdot|x_0]$.

## 7.3 Optimal control with constraints and random attacks

### 7.3.1 A random DoS attack model

Networked control formulations have previously considered the loss of sensor or control packets and their impact on the system. While previous results model packet drops caused by random events (and not by an attacker) we believe these packet drop models can be used as a first-step towards understanding the impact of DoS attacks to our objective and constraints.

One of these models is the Bernoulli packet drop model, in which at each time, the attacker randomly jams a measurement (resp. control) packet according to independent Bernoulli trials with success probability $\bar{\gamma}$ (resp. $\bar{\nu}$). This attack model, referred as the $\mathrm{Ber}(\bar{\gamma}, \bar{\nu})$ adversary, has the following admissible attack actions

$$\mathcal{A}_{\mathrm{Ber}(\bar{\gamma},\bar{\nu})} = \{(\gamma_0^{N-1}, \nu_0^{N-1}) | \mathsf{P}(\gamma_k = 1) = \bar{\gamma}, \mathsf{P}(\nu_k = 1) = \bar{\nu},\ k = 0, \ldots, N-1\}. \tag{7.7}$$

For the $\mathcal{A}_{\text{Ber}(\bar{\gamma},\bar{\nu})}$ model, we can write the Kalman filter equations for the state estimate $\hat{x}_{k|k} := \mathsf{E}[x_k|\mathcal{I}_k]$ and the state estimation error $e_{k|k} := (x_k - \hat{x}_{k|k})$. For the update step we have

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + \nu_k B u_k \text{ and, } e_{k+1|k} = Ae_{k|k} + w_k$$

and for the correction step

$$\hat{x}_{k+1|k+1} = \gamma_{k+1}x_{k+1} + (1-\gamma_{k+1})\hat{x}_{k+1|k} \text{ and, } e_{k+1|k+1} = (1-\gamma_{k+1})e_{k+1|k},$$

starting with $\hat{x}_{0|-1} = \bar{x}$ and $e_{0|-1} \sim \mathcal{N}(0, P_0)$. It follows that the error covariance matrices $\Sigma_{k+1|k} := \mathsf{E}[e_{k+1|k}e_{k+1|k}^\top|\mathcal{I}_k]$ and $\Sigma_{k|k} := \mathsf{E}[e_{k|k}e_{k|k}^\top|\mathcal{I}_k]$ do not depend on the control input $u_k$. Thus, the separation principle holds for TCP-like communication Schenato et al. [2007]. Furthermore, it is easy to see that

$$\mathsf{E}[e_{k|k}x_{k|k}^\top] = 0. \tag{7.8}$$

Taking expectations w.r.t. $\{\gamma_k\}$, the expected error covariances follow

$$\mathsf{E}_\gamma[\Sigma_{k+1|k}] = A\mathsf{E}_\gamma[\Sigma_{k|k}]A^\top + W \text{ and, } \mathsf{E}_\gamma[\Sigma_{k+1|k+1}] = (1-\bar{\gamma})\mathsf{E}_\gamma[\Sigma_{k+1|k}],$$

for $k = 0, \ldots, N-1$ starting with the initial condition $\Sigma_{0|-1} = P_0$. For the ease of notation, we denote $\hat{x}_{k+1} := \hat{x}_{k+1|k}$, $e_{k+1} := e_{k+1|k}$, and $\Sigma_{k+1} := \Sigma_{k+1|k}$. Using the Kalman filter equations we obtain for $k = 0, \ldots, N-1$

$$\hat{x}_{k+1} = A\hat{x}_k + \nu_k B u_k + \gamma_k A e_k \tag{7.9}$$

$$e_{k+1} = (1-\gamma_k)Ae_k + w_k \tag{7.10}$$

$$\mathsf{E}_\gamma[\Sigma_{k+1}] = (1-\bar{\gamma})A\mathsf{E}_\gamma[\Sigma_k]A^\top + W. \tag{7.11}$$

**Definition 7.3.1.** For Bernoulli attacks, $(\gamma_0^{N-1}, \nu_0^{N-1}) \in \mathcal{A}_{\text{Ber}(\bar{\gamma},\bar{\nu})}$ over systems controlled over TCP-like communication protocols, the safety-constrained robust optimal control problem is equivalent to minimizing (7.4) subject to (7.9), (7.11), (7.5) and (7.6).

### 7.3.2 Controller parameterization

In this section, we deal with the safety-constrained optimal control problem as defined in Definition 7.3.1. Naive implementation of the control law $u_k^* = -L_k\hat{x}_{k|k}$ may not guarantee constraint satisfaction for any initial state. Recent research has shown that for the optimal control problems involving state and input constraints, more general causal feedback controllers can guarantee a larger set of initial states for which the constrained optimal control problem admits a feasible solution Ben-tal et al. [2005]; Goulart et al. [2006]; Primbs and Sung [2009]; Skaf and Boyd [2010]; van Hessem and Bosgra [2003]. Specifically, these approaches consider the problem of designing causal controllers that are affine in all previous measurements such that a convex objective function is minimized subject to constraints imposed by the system dynamics, and the state and inputs constraints are satisfied.

When considering a system under DoS attacks, (7.1), (7.2), and (7.3), the class of causal feedback controllers can be defined as an affine function of the available measurements, i.e.,

$$u_k = \bar{u}_k + \sum_{j=0}^{k} \gamma_j M_{k,j} x_j, \qquad k = 0, \dots, N-1 \tag{7.12}$$

where $\bar{u}_k \in \mathbb{R}^m$ is the open-loop part of the control, and $M_{k,j} \in \mathbb{R}^{m \times n}$ is the feedback gain or the recourse at time $k$ from sensor measurement $x_j$. For a lost measurement packet, say $x_{j'}$ for $\gamma_{j'} = 0$, the corresponding feedback gain $M_{k,j'}$ has no contribution toward the control policy. We note that the above parameterization can be re-expressed as an affine function of innovations $v_{k|k-1} := \gamma_k(x_k - \hat{x}_{k|k-1}) = \gamma_k e_k$ as

$$u_k = u_k^\circ + \sum_{j=0}^{k} \gamma_j M_{k,j} e_j, \qquad k = 0, \dots, N-1 \tag{7.13}$$

where $u_k^\circ := \bar{u}_k + \sum_{j=0}^{k} \gamma_j M_{i,j} \hat{x}_{j|j-1}$.

*Remark* 7.3.2. When only the current available measurement is used for computing the feedback policy, the mapping $\mu_k$ can be expressed as

$$u_k = \bar{u}_k + \gamma_k M_{k,k} x_k = u_k^\circ + \gamma_k M_k e_k, \qquad k = 0, \dots, N-1, \tag{7.14}$$

where $M_k := M_{k,k}$ for ease of notation and $u_k^\circ := \bar{u}_k + \gamma_k M_k \hat{x}_{k|k-1}$. □

### 7.3.3 Convex characterization

In this section, we will show that unlike (7.12), the use of control parameterization (7.13) yields an affine representation of state and control trajectories in terms of the control parameters $\bar{u}_k$ (or $u_k^\circ$) and $M_{k,j}$. We use $\mathbf{x}$, $\hat{\mathbf{x}}$, $\mathbf{u}$, $\mathbf{e}$ and $\mathbf{w}$ to denote the respective trajectories over the time horizon $0, \dots, N$. That is, $\mathbf{x} = (x_0^\top, \dots, x_N^\top)^\top \in \mathbb{R}^{n(N+1)}$ and similarly for $\hat{\mathbf{x}} \in \mathbb{R}^{n(N+1)}$ and $\mathbf{e} \in \mathbb{R}^{n(N+1)}$; $\mathbf{u} = (u_0^\top, \dots, u_{N-1}^\top)^\top \in \mathbb{R}^{mN}$ and similarly for $\mathbf{w} \in \mathbb{R}^{nN}$. Using this representation, the system (7.1) and the control parameterization (7.12) can be written as

$$\mathbf{x} = \mathbf{A}\mathbf{w} + \mathbf{B}\mathbf{N}\mathbf{u} + \mathbf{x_0}, \tag{7.15}$$

$$\mathbf{u} = \bar{\mathbf{u}} + \mathbf{M}\boldsymbol{\Gamma}\mathbf{x}, \tag{7.16}$$

where $\mathbf{x_0}$, $\mathbf{A}$, $\mathbf{B}$, $\boldsymbol{\Gamma}$, $\mathbf{N}$ are given by:

$$\mathbf{x_0} := \begin{pmatrix} I_n \\ A \\ A^2 \\ \vdots \\ A^N \end{pmatrix} x_0 \in \mathbb{R}^{n(N+1)}, \quad \mathbf{A} := \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ I_n & 0 & 0 & \dots & 0 \\ A & I_n & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ A^{N-1} & A^{N-2} & A^{N-3} & \dots & I_n \end{pmatrix} \in \mathbb{R}^{n(N+1) \times nN},$$

$$\mathbf{B} := \mathbf{A}(I_N \otimes B) = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ B & 0 & 0 & \dots & 0 \\ AB & B & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ A^{N-1}B & A^{N-2}B & A^{N-3}B & \dots & B \end{pmatrix} \in \mathbb{R}^{n(N+1)\times mN},$$

$$\mathbf{\Gamma} = \mathrm{diag}(\gamma_0^{N-1}) \otimes I_n = \begin{pmatrix} \gamma_0 I_n & & \\ & \ddots & \\ & & \gamma_{N-1} I_n \end{pmatrix} \in \mathbb{R}^{nN\times nN},$$

$$\mathbf{N} = \mathrm{diag}(\nu_0^{N-1}) \otimes I_m = \begin{pmatrix} \nu_0 I_m & & \\ & \ddots & \\ & & \nu_{N-1} I_m \end{pmatrix} \in \mathbb{R}^{mN\times mN},$$

and

$$\mathbf{e_0} = \begin{pmatrix} I_n \\ (1-\gamma_0)A \\ (1-\gamma_0)(1-\gamma_1)A^2 \\ \vdots \\ \prod_{j=0}^{N-1}(1-\gamma_j)A^N \end{pmatrix} e_0 \in \mathbb{R}^{n(N+1)}$$

$$\mathbf{H} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ I_n & 0 & \dots & 0 \\ (1-\gamma_1)A & I_n & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \prod_{j=1}^{N-1}(1-\gamma_j)A^{N-1} & \prod_{j=2}^{N-1}(1-\gamma_j)A^{N-2} & \dots & I_n \end{pmatrix} \in \mathbb{R}^{n(N+1)\times nN},$$

and

$$\mathbf{M} = \begin{pmatrix} M_{0,0} & 0 & \dots & 0 \\ M_{1,0} & M_{1,1} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ M_{N-1,0} & \dots & M_{N-1,N-1} & 0 \end{pmatrix} \in \mathbb{R}^{mN\times n(N+1)}, \quad \bar{\mathbf{u}} = \begin{pmatrix} \bar{u}_0 \\ \vdots \\ \bar{u}_{N-1} \end{pmatrix} \in \mathbb{R}^{mN} \tag{7.17}$$

Using (7.15) and (7.16), we can show that the closed-loop system response can be written as

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{G}}_{\mathbf{xw}} \\ \tilde{\mathbf{G}}_{\mathbf{uw}} \end{pmatrix} \mathbf{w} + \begin{pmatrix} \tilde{\mathbf{x}} \\ \tilde{\mathbf{u}} \end{pmatrix} \tag{7.18}$$

where

$$\begin{aligned} \tilde{\mathbf{G}}_{\mathbf{xw}} &= \big(\mathbf{A} + \mathbf{BNM\Gamma}(I - \mathbf{BNM\Gamma})^{-1}\mathbf{A}\big) \\ \tilde{\mathbf{G}}_{\mathbf{uw}} &= \big(\mathbf{M\Gamma}(I - \mathbf{BNM\Gamma})^{-1}\mathbf{A}\big) \\ \tilde{\mathbf{x}} &= \mathbf{x_0} + \mathbf{BN\bar{u}} + \mathbf{BNM\Gamma}(I - \mathbf{BNM\Gamma})^{-1}(\mathbf{x_0} + \mathbf{BN\bar{u}}) \\ \tilde{\mathbf{u}} &= \mathbf{M\Gamma}(I - \mathbf{BNM\Gamma})^{-1}(\mathbf{x_0} + \mathbf{BN\bar{u}}) + \mathbf{\bar{u}} \end{aligned}$$

Equation (7.18) is nonlinear in the control parameters $(\mathbf{\bar{u}}, \mathbf{M})$ and hence, parameterization (7.12) cannot be directly used for solving constrained stochastic optimal control problems. On the other hand, using (7.10), the error trajectory can be written as

$$\mathbf{e} = \mathbf{e}_0 + \mathbf{H}\mathbf{w} \tag{7.19}$$

where $\mathbf{e_0}$ and $\mathbf{H}$ are also given in the Appendix. Using (7.19), (7.15) and the control parameterization (7.13) we can re-express the closed-loop system response as

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{\hat{G}_{xw}} \\ \mathbf{\hat{G}_{uw}} \end{pmatrix} \mathbf{w} + \begin{pmatrix} \mathbf{\hat{x}} \\ \mathbf{\hat{u}} \end{pmatrix} \tag{7.20}$$

where

$$\begin{aligned} \mathbf{\hat{G}_{xw}} &= (\mathbf{A} + \mathbf{BNM\Gamma H}), & \mathbf{\hat{G}_{uw}} &= \mathbf{M\Gamma H} \\ \mathbf{\hat{x}} &= \mathbf{BNM\Gamma e_0} + \mathbf{x_0} + \mathbf{BNu}^\circ, & \mathbf{\hat{u}} &= \mathbf{M\Gamma e_0} + \mathbf{u}^\circ \end{aligned}$$

Thus, we arrive at the following result

**Theorem 7.3.3.** *Under the error feedback parameterization* (7.13)*, the closed loop system response* (7.20) *is affine in the control parameters* $(\mathbf{u}^\circ, \mathbf{M})$*.* □

We will now use the error feedback parameterization (7.13) for our analysis. Alternatively, we also note the following result:

*Remark* 7.3.4. Using the transformation

$$\mathbf{Q} := \mathbf{M\Gamma}(I - \mathbf{BNM\Gamma})^{-1}, \qquad \mathbf{r} := (I + \mathbf{QBN})\mathbf{\bar{u}} \tag{7.21}$$

where $\mathbf{Q} \in \mathbb{R}^{mN \times n(N+1)}$ and $\mathbf{r} \in \mathbb{R}^{mn}$, the terms in equation (7.18) can be written as: $\mathbf{G_{xw}} = (I + \mathbf{BNQ})\,\mathbf{A}$, $\mathbf{G_{uw}} = \mathbf{QA}$, $\mathbf{\tilde{x}} = (I + \mathbf{BNQ})\,\mathbf{\bar{x}} + \mathbf{BNr}$, and $\mathbf{\tilde{u}} = \mathbf{Q\bar{x}} + \mathbf{r}$. Using simple matrix operations, the relations in (7.21) can be inverted as $\mathbf{M\Gamma} = (I + \mathbf{QBN})^{-1}\mathbf{Q}$ and $\mathbf{\bar{u}} = (I - \mathbf{M\Gamma HN})\mathbf{r}$. Thus, under parameterization (7.21), the closed-loop system response also becomes affine in the control parameters $(\mathbf{r}, \mathbf{Q})$. □

### 7.3.4 Safety-constrained optimal control for Bernoulli attacks

For the control parameterization (7.12), and for the Bernoulli attack model, $\mathcal{A}_{\text{Ber}(\bar{\gamma},\bar{\nu})}$ we will now solve the safety-constrained optimal control problem as stated in Lemma 7.3.1, i.e., minimize (7.4) subject to (7.9), (7.11), (7.5), and (7.6). We state the following useful lemma

**Lemma 7.3.5** (Schur Complements)**.** *For all* $X \in \mathbb{S}^n$, $Y \in \mathbb{R}^{m\times n}$, $Z \in \mathbb{S}^m$, *the following statements are equivalent:*

$$\text{a)} Z \succ 0, X - Y^\top Z^{-1} Y \succeq 0,$$

$$\text{b)} Z \succ 0, \begin{pmatrix} X & Y^\top \\ Y & Z \end{pmatrix} \succeq 0$$

For the sake of simplicity we will consider the parameterization (7.14). However, our results can be re-derived for the parameterization (7.12). First, we will derive the expression for

$$V_k = \mathsf{E}\left[\begin{pmatrix}\hat{x}_k\\ u_k^\circ\end{pmatrix}\begin{pmatrix}\hat{x}_k\\ u_k^\circ\end{pmatrix}^\top\right]$$

Using (7.14), the update equation for the state estimate (7.9) becomes

$$\hat{x}_{k+1} = A\hat{x}_k + \nu_k B u_k^\circ + \gamma_k(A + \nu_k B M_k)e_k, \tag{7.22}$$

and further defining $F = [I_n, 0] \in \mathbb{R}^{n\times(n+m)}$ we have,

$$\begin{aligned}
FV_{k+1}F^\top &= V_{k+1}^{\hat{x}\hat{x}} = \mathsf{E}\left[\hat{x}_{k+1}\hat{x}_{k+1}^\top\right]\\
&= \mathsf{E}\left[(A\hat{x}_k + \nu_k B u_k^\circ + \gamma_k(A+\nu_k BM_k)e_k)(A\hat{x}_k + \nu_k B u_k^\circ + \gamma_k(A+\nu_k BM_k)e_k)^\top\right]\\
&= \left[A \quad \middle|\sqrt{\bar{\nu}}B\right] \mathsf{E}\left[\begin{pmatrix}\hat{x}_k\\ u_k^\circ\end{pmatrix}\begin{pmatrix}\hat{x}_k\\ u_k^\circ\end{pmatrix}^\top\right]\left[A \quad \middle|\sqrt{\bar{\nu}}B\right]^\top\\
&+ \sqrt{\bar{\gamma}}(A + \sqrt{\bar{\nu}}BM_k)\mathsf{E}_\gamma[\Sigma_k](A+\sqrt{\bar{\nu}}BM_k)^\top\sqrt{\bar{\gamma}}\\
&= \left[AV_k \quad \middle|\sqrt{\bar{\nu}}BV_k\right](V_k)^{-1}\left[AV_k \quad \middle|\sqrt{\bar{\nu}}BV_k\right]^\top\\
&+ \sqrt{\bar{\gamma}}(A\mathsf{E}_\gamma[\Sigma_k] + \sqrt{\bar{\nu}}BU_k)(\mathsf{E}_\gamma[\Sigma_k])^{-1}(A\mathsf{E}_\gamma[\Sigma_k] + \sqrt{\bar{\nu}}BU_k)^\top\sqrt{\bar{\gamma}}
\end{aligned}$$

where we have used $U_k = M_k\mathsf{E}_\gamma[\Sigma_k]$. An upper bound on $V$ can be obtained in the form of the following LMI by replacing the equality by $\succeq$ and using Schur complements for $k = 0,\ldots,N-1$:

$$\begin{bmatrix}
(FV_{k+1}F^\top) & * & * & *\\
\left[AV_k \quad \sqrt{\bar{\nu}}BV_k\right]^\top & 0 & V_k & *\\
\sqrt{\bar{\gamma}}(A\mathsf{E}_\gamma[\Sigma_k] + \sqrt{\bar{\nu}}BU_k)^\top & 0 & 0 & \mathsf{E}_\gamma[\Sigma_k]
\end{bmatrix} \succeq 0 \tag{7.23}$$

The objective function (7.4) can be expressed as

$$\begin{aligned}
&\mathsf{E}\left[\mathbf{Tr}\left\{Q^{xx}x_N x_N^\top\right\}\right] + \sum_{k=0}^{N-1}\mathsf{E}\left[\mathbf{Tr}\left\{\begin{pmatrix}Q^{xx} & 0\\ 0 & \nu_k Q^{uu}\end{pmatrix}\right\}\begin{pmatrix}x_k\\ u_k\end{pmatrix}\begin{pmatrix}x_k\\ u_k\end{pmatrix}^\top\right]\\
&= \mathbf{Tr}\left\{Q^{xx}\mathsf{E}\left[x_N x_N^\top\right]\right\} + \sum_{k=0}^{N-1}\mathbf{Tr}\left\{\begin{pmatrix}Q^{xx} & 0\\ 0 & \mathsf{E}[\nu_k]Q^{uu}\end{pmatrix}\mathsf{E}\left[\begin{pmatrix}x_k\\ u_k\end{pmatrix}\begin{pmatrix}x_k\\ u_k\end{pmatrix}^\top\right]\right\}\\
&= \mathbf{Tr}\left\{Q^{xx}\mathsf{E}\left[\hat{x}_N \hat{x}_N^\top\right]\right\} + \sum_{k=0}^{N-1}\mathbf{Tr}\left\{\begin{pmatrix}Q^{xx} & 0\\ 0 & \bar{\nu}Q^{uu}\end{pmatrix}\mathsf{E}\left[\begin{pmatrix}\hat{x}_k\\ u_k\end{pmatrix}\begin{pmatrix}\hat{x}_k\\ u_k\end{pmatrix}^\top\right]\right\}\\
&+ \sum_{k=0}^{N}\mathbf{Tr}\left\{Q^{xx}\mathsf{E}_\gamma[\Sigma_k]\right\}
\end{aligned}$$

Since $\Sigma_k$ does not depend on the control input (refer to eq. (7.11)), $\sum_{k=0}^{N}\mathbf{Tr}\left\{Q^{xx}\mathsf{E}_\gamma[\Sigma_k]\right\}$ is a constant and minimizing $J_N(\bar{x}, P_0, u_0^{N-1})$ is the same as minimizing

$$\mathbf{Tr}\left\{Q^{xx}V_N^{\hat{x}\hat{x}}\right\} + \sum_{k=0}^{N-1}\mathbf{Tr}\left\{\begin{pmatrix}Q^{xx} & 0\\ 0 & \bar{\nu}Q^{uu}\end{pmatrix}P_k\right\} \tag{7.24}$$

where $V_N^{\hat{x}\hat{x}}$ is equal to $\mathsf{E}\left[\hat{x}_N \hat{x}_N^\top\right]$ and the upper bound $P_k$ is defined as

$$P_k \succeq \mathsf{E}\left[\begin{pmatrix}\hat{x}_k\\ u_k\end{pmatrix}\begin{pmatrix}\hat{x}_k\\ u_k\end{pmatrix}^\top\right] = \mathsf{E}\left[\begin{pmatrix}\hat{x}_k\\ u_k^\circ + \gamma_k M_k e_k\end{pmatrix}\begin{pmatrix}\hat{x}_k\\ u_k^\circ + \gamma_k M_k e_k\end{pmatrix}^\top\right]$$
$$= \mathsf{E}\left[\begin{pmatrix}\hat{x}_k\\ u_k^\circ\end{pmatrix}\begin{pmatrix}\hat{x}_k\\ u_k^\circ\end{pmatrix}^\top\right] + \begin{bmatrix}0 & 0\\ 0 & \bar{\gamma} U_k (\mathsf{E}_\gamma[\Sigma_k])^{-1} U_k^\top\end{bmatrix}$$

Again using Schur complement, we obtain for $k = 0, \ldots, N-1$

$$\begin{bmatrix} P_k & * & * \\ V_k & V_k & * \\ \begin{bmatrix} 0 \\ \sqrt{\bar{\gamma}} U_k \end{bmatrix}^\top & 0 & \mathsf{E}_\gamma[\Sigma_k] \end{bmatrix} \succeq 0 \tag{7.25}$$

The power constraints (7.5) can be written as

$$\mathbf{Tr}\left\{H_i \begin{bmatrix} I_n & 0 \\ 0 & \mathsf{E}[\nu_k] I_m \end{bmatrix} \mathsf{E}\left[\begin{pmatrix}x_k\\ u_k\end{pmatrix}\begin{pmatrix}x_k\\ u_k\end{pmatrix}^\top\right]\right\}$$
$$= \mathbf{Tr}\left\{H_i \begin{bmatrix} I_n & 0 \\ 0 & \bar{\nu} I_m \end{bmatrix} \mathsf{E}\left[\begin{pmatrix}\hat{x}_k\\ u_k\end{pmatrix}\begin{pmatrix}\hat{x}_k\\ u_k\end{pmatrix}^\top\right]\right\} + \mathbf{Tr}\left\{H_i^{xx} \mathsf{E}_\gamma[\Sigma_k]\right\}$$

Therefore the power constraints (7.5) become for $i = 1, \ldots, L, k = 0, \ldots, N-1$

$$\mathbf{Tr}\left\{H_i \begin{bmatrix} I_n & 0 \\ 0 & \bar{\nu} I_m \end{bmatrix} P_k\right\} \leqslant \beta_i - \mathbf{Tr}\left\{H_i^{xx} \mathsf{E}_\gamma[\Sigma_k]\right\}. \tag{7.26}$$

Thus, we can now state the following theorem

**Theorem 7.3.6.** *For the* $(\gamma_0^{N-1}, \nu_0^{N-1}) \in \mathcal{A}_{\mathrm{Ber}(\bar{\gamma},\bar{\nu})}$ *attack model the optimal causal controller of the form* (7.14) *for the system* (7.1), (7.2), (7.3) *that minimizes the objective function* (7.4) *subject to power constraints* (7.5) *is equivalent to solving the following semidefinite program (SDP):*

$$\mathcal{P}(\bar{x}, P_0, N) : \begin{cases} \min_{V_i, P_i, U_i} (7.24) \\ \textit{subject to } (7.23), (7.25), (7.26). \end{cases} \tag{7.27}$$

□

To address safety specification (7.6), we refer to Theorem 3.1 in Calafiore and El Ghaoui [2007] which says that for any $\epsilon \in (0, 1)$, the chance constraint of the form

$$\inf_{d \sim \mathcal{D}} \mathsf{P}\left[d^\top \tilde{x} \leqslant 0\right] \geqslant 1 - \epsilon$$

is equivalent to the second order cone constraint (SOCP)

$$\sqrt{\frac{1-\epsilon}{\epsilon}\tilde{x}^\top \Gamma \tilde{x}} + \hat{d}^\top \tilde{x} \leqslant 0$$

where $\mathcal{D}$ is the set of all probability distributions with mean $\hat{d}$ and covariance $\Gamma$, $d$ is the uncertain data with distributions in the set of distributions $\mathcal{D}$, and $\tilde{x}$ is the decision variable. We claim without proof that safety specifications of type (7.6) can be converted to SOCP constraints following Calafiore and El Ghaoui [2007],van Hessem and Bosgra [2003].

## 7.4 Modeling general DoS attacks

From the security viewpoint, it might be difficult to justify the incentive for the attacker to follow a $\mathcal{A}_{\text{Ber}(\bar{\gamma},\bar{\nu})}$ model. Therefore, in this section we introduce more general attack models that impose constraints on the DoS attack actions $(\gamma_k, \nu_k)$.

First, note that if we know in advance the strategy of the attacker—for any arbitrary sequence $(\gamma_0^{N-1}, \nu_0^{N-1})$—we can use the results from the previous theorem.

**Corollary 7.4.1.** *The results of Theorem 7.3.6 be specialized to any given attack signature* $(\gamma_0^{N-1}, \nu_0^{N-1}) \in \{0,1\}^{2N}$. □

However, in practice we do not know the strategy of the attacker, thus we need to prepare for all possible attacks. Our model constrains the attacker action in time by restricting the DoS attacks on the measurement (resp. control) packet for *at most* $p < N$ (resp. $q < N$) time steps anywhere in the time interval $i = 0, \ldots, N-1$. This attack model is motivated by limitations on the resources of the adversary—such as its battery power, or the response time of the defenders—which in turn limits the number of times it can block a transmission. We refer this attack model as the $(p,q)$ adversary and it has the following admissible attack actions

$$\mathcal{A}_{pq} = \{(\gamma_0^{N-1}, \nu_0^{N-1}) \in \{0,1\}^{2N} | \parallel \gamma_0^{N-1} \parallel_1 \geqslant N-p, \parallel \nu_0^{N-1} \parallel_1 \geqslant N-q\}, \tag{7.28}$$

where $\parallel \cdot \parallel_1$ denotes the 1−norm. The size of $\mathcal{A}_{pq}$ is $\sum_{i=0}^{p}\binom{N}{N-i} \cdot \sum_{j=0}^{q}\binom{N}{N-j}$.

An interesting sub-class of $\mathcal{A}_{pq}$ attack actions is the class of block attack strategies

$$\mathcal{A}_{pq}^{\tau_x\tau_u} = \{(\gamma_0^{N-1}, \nu_0^{N-1}) \in \{0,1\}^{2N} | \gamma_{\tau_x}^{\tau_x+p-1} = 0, \nu_{\tau_u}^{\tau_u+q-1} = 0\} \tag{7.29}$$

where $\tau_x \in \{0, \ldots, N-p\}$ and $\tau_u \in \{0, \ldots, N-q\}$ are the times at which the attacker starts jamming the measurement and control packets respectively. The size of $\mathcal{A}_{pq}^{\tau_x\tau_u}$ is $(N-p+1)\cdot(N-q+1)$. The intuition behind this attack sub-class is that an attacker will consume all of its resources continuously in order to maximize the damage done to the system. In this attack sub-class, $p$ and $q$ can represent the response time of defensive mechanisms. For example, a packet-flooding attack may be useful until network administrators implement filters or replicate the node under attack; similarly a jamming attack may be useful only until the control operators find the jamming source and neutralize it. We note that $\mathcal{A}_{pq}$ and $\mathcal{A}_{pq}^{\tau_x\tau_u}$ are *non-deterministic attack models* in that the attacker can choose its action non-deterministically as long as the constraints defined by the attack model are satisfied.

### 7.4.1 DoS attacks against the safety constraint

One possible objective of the attacker can be to violate safety constraints:

**Definition 7.4.2.** [Most unsafe attack] For a given attack model $\mathcal{A}$ and control strategy $\mu_k(\mathcal{I}_k)$, the best attack plan to violate safety specification that a output vector $z_k := (Cx_k + \nu_k D u_k)$ remains within safe set $\mathcal{S}$ is

$$\max_{\mathcal{A}} \mathsf{P}[(Cx_k + \nu_k D\mu(I_k)) \in \mathcal{S}^c] \text{ for } k = 0, \dots, N-1 \tag{7.30}$$

where $S^c$ denotes the unsafe set.

We will now show that for control parameterization (7.12), the block $pq$ attacks, $\mathcal{A}_{pq}^{\tau_x\tau_u}$ can be viewed as the best attack plan for violating the safety constraint (refer to Definition 7.4.2). We can write the system equation (7.1) as

$$x_{k+1} = Ax_k + \nu_k B\bar{u}_k + \nu_k \sum_{j=0}^{k} \gamma_j M_{k,j} x_j + w_k$$

and for the attack strategy $\mathcal{A}_{pq}^{\tau_x\tau_u}$:

$$x_{k+1} = \begin{cases} Ax_k + w_k \text{ for } k = \tau_u, \dots, \tau_u + q - 1 \\ Ax_k + B\bar{u}_k + B\sum_{j=0}^{\min(\tau_x - 1, k)} M_{k,j} x_j \\ +\mathbf{1}(k \geqslant \tau_x + p) B \sum_{j=0}^{k} M_{k,j} x_j + w_k \text{ for } k = \begin{cases} 0, \dots, \tau_u - 1 \\ \tau_u + q, \dots, N-1. \end{cases} \end{cases} \tag{7.31}$$

Now, if we ignore $\bar{u}_k$ and substitute $\tau_x = 0$, $\tau_u = p$ in (7.31) we obtain

$$x_{k+1} = \begin{cases} Ax_k + w_k \text{ for } k = 0, \dots, p + q - 1 \\ Ax_k + B\sum_{j=p}^{k} M_{k,j} x_j \text{ for } k = p + q, \dots, N-1 \end{cases} \tag{7.32}$$

Thus, using the attack strategy $\mathcal{A}_{pq}^{0p}$, the first $p + q - 1$ time steps evolve as open-loop and beyond time step $p + q$, the system evolves as closed using available measurements since time $p$. With this strategy output vector $z_k$ is expected to violate the safety constraint in the shortest time.

## 7.5 Discussion

From the controller's viewpoint, it is of interest to design control laws that are robust against all attacker actions, i.e.:

**Definition 7.5.1.** [Minimax (robust) control] For a given attack model $\mathcal{A}$, the security constrained robust optimal control problem is to synthesize a control law that minimizes the maximum cost over all $(\gamma_0^{N-1}, \nu_0^{N-1}) \in \mathcal{A}$, subject to the power and safety constraints. This can be written as the minimax problem

$$\min_{\mu_k(\mathcal{I}_k)} \max_{\mathcal{A}} [(7.4) \text{ subject to } (7.1), (7.2), (7.3), (7.5) \text{ and}, (7.6)]. \tag{7.33}$$

In general, we note that the problem (7.33) may not always be feasible. When $\mathcal{A}$ is probabilistic, Definition 7.5.1 can be treated in sense of expectation or almost-surely.

On the other hand, from the attacker's viewpoint, it is of interest to determine the optimal *attack plan* that degrades performance, i.e.,

**Definition 7.5.2.** [Maximin (worst-case) attack] For a given attack model $\mathcal{A}$, the optimal attack plan is the attacker action that maximizes the minimum operating costs. This can be written as the maximin problem

$$\max_{\mathcal{A}} \min_{\mu_k(\mathcal{I}_k)} \left[ (7.4) \text{ subject to } \; (7.1), (7.2), \; (7.3) \right]. \tag{7.34}$$

To analyze these goals, we consider the classical linear quadratic control problem, and analyze the cost function for the case of (1) no attacks, (2) $\mathcal{A}_{\text{Ber}(\bar{\gamma},\bar{\nu})}$attacks, and (3) $\mathcal{A}_{pq}$ attacks.

The problem is to find the optimal control policy $u_k = \mu_k(\mathcal{I}_k)$ that minimizes the objective (7.4) for the system (7.1), (7.2), and (7.3). The solution of this problem can be obtained in closed form using dynamic programming (DP) recursions Gattami [2007]; Schenato et al. [2007].

We recall that for the case of no-attack, i.e., $(\gamma_k, \nu_k) = (1, 1)$ for all $k$, the optimal control law is given by $u_k^* = -L_k x_k$ where $L_k := (B^\top S_{k+1} B + Q^{uu})^{-1} B^\top S_{k+1} A$ and the matrices $S_k$ are chosen such that $S_N = Q^{xx}$ and for $k = N-1, \ldots, 0$,

$$S_k = A^\top S_{k+1} A + Q^{xx} - R_k$$

with $R_k = L_k^\top (B^\top S_{k+1} B + Q^{uu}) L_k$. The optimal cost is given by

$$J_N^* = \bar{x}^\top S_0 \bar{x} + \mathbf{Tr}\{S_0 P_0\} + \sum_{k=0}^{N-1} \mathbf{Tr}\{S_{k+1} W\}. \tag{7.35}$$

Following Schenato et al. [2007], the optimal control law for the case of $\mathcal{A}_{\text{Ber}(\bar{\gamma},\bar{\nu})}$ attack model is given by $u_k^* = -L_k \hat{x}_{k|k}$ where $\hat{x}_{k|k}$ is given by the Kalman filter equations; the expressions for $L_k$, $R_k$, $S_N$ are same as those for the no-attack case, and for $k = N-1, \ldots, 0$,

$$S_k = A^\top S_{k+1} A + Q^{xx} - \bar{\nu} R_k.$$

The optimal cost in this case is given by

$$J_{N,\mathcal{A}_{\text{Ber}(\bar{\gamma},\bar{\nu})}}^* = \bar{x}^\top S_0 \bar{x} + \mathbf{Tr}\{S_0 P_0\} + \sum_{k=0}^{N-1} \mathbf{Tr}\{S_{k+1} W\} + \sum_{k=0}^{N-1} \mathbf{Tr}\{\bar{\nu} R_k \mathsf{E}_\gamma[\Sigma_{k|k}]\} \tag{7.36}$$

**Lemma 7.5.3.** *$J_{N,\mathcal{A}_{\text{Ber}(\bar{\gamma},\bar{\nu})}}^* \geqslant J_N^*$ for all $(\bar{\gamma}, \bar{\nu}) \in [0, 1]$.* □

We now consider the case of $\mathcal{A}_{pq}$ attacks. We can solve the problem of optimal attack plan for the $\mathcal{A}_{pq}$ attack class (refer to Definition 7.5.2):

For any *given* attack signature, $(\gamma_0^{N-1}, \nu_0^{N-1}) \in \{0,1\}^{2N}$, the update equations of error covariance are $\Sigma_{k+1|k} = A\Sigma_{k|k}A^\top + W$ and $\Sigma_{k+1|k+1} = (1-\gamma_{k+1})\Sigma_{k+1|k}$ and the optimal cost is given by

$$\begin{aligned} J_{N,\mathcal{A}_{pq}} =& \bar{x}^\top S_0 \bar{x} + \mathbf{Tr}\{S_0 P_0\} \\ &+ \sum_{k=0}^{N-1} \mathbf{Tr}\{S_{k+1}Q\} + \sum_{k=0}^{N-1} \mathbf{Tr}\{(A^\top S_{k+1}A + Q^{xx} - S_k)\Sigma_{k|k}\} \end{aligned} \tag{7.37}$$

where $S_N = Q^{xx}$ and for $k = N-1, \ldots, 0$,

$$S_k = A^\top S_{k+1} A + Q^{xx} - \nu_k A^\top S_{k+1} B(B^\top S_{k+1} B + Q^{uu})^{-1} B^\top S_{k+1} A. \tag{7.38}$$

and for $k = 1, \ldots, N-1$,

$$\Sigma_{k|k} = \prod_{j=1}^{k}(1-\gamma_j)A^k P_0 {A^k}^\top + \sum_{i=0}^{k-1} \prod_{j=(k-i)}^{k} (1-\gamma_j) A^i W {A^i}^\top. \tag{7.39}$$

**Proposition 7.5.4.** *An optimal attack plan for $\mathcal{A}_{pq}$ attack model is a solution of the following optimization problem:*

$$\begin{aligned} &\max_{\mathcal{A}_{pq}} (7.37) \ \textit{subject to} \ \ (7.38), \ (7.39), \\ &\| \gamma_0^{N-1} \|_1 \geqslant (N-p), \ \textit{and} \ \ \| \nu_0^{N-1} \|_1 \geqslant (N-q). \end{aligned}$$

We note that while $\Sigma_{k|k}$ is affected by the *past* measurement attack sequence $\{\gamma_0^k\}$, $S_k$ is affected by the *future* control attack sequence $\{\nu_k^{N-1}\}$.

# Chapter 8

# Stabilization of Networked Control Systems using Bounded Inputs

## 8.1 Introduction

In Chapter 7, the problem of controlling stochastic linear systems for networked control settings was considered when the sensor-control data is prone to packet loss and jamming. For a class of packet drop models, a synthesis procedure was presented to compute feedback control policies which minimize a given objective function subject to safety constraints. In this chapter, we consider the problem of controlling marginally stable linear systems using bounded control inputs for networked control settings in which the communication channel between the remote controller and the system is unreliable. We assume that the states are perfectly observed, but the control inputs are transmitted over a noisy communication channel. Under mild hypotheses on the noise introduced by the control communication channel and large enough control authority, we construct a control policy that renders the state of the closed-loop system mean-square bounded. The noise introduced by the control channel is assumed to be independent and identically distributed and hence, this chapter is only concerned with stabilization using an unreliable control channel. However, the complexity of the problem considered here arises from the fact that hard bounds on the control inputs must be satisfied. Moreover, the analysis developed in this chapter only requires mild assumptions on the distribution of control channel noise. The relationship between unreliable channel communication and the insecurity introduced due to interdependent network risks is explored in Chapter 9.

In applications such as remotely operated robotic systems Hokayem and Spong [2006], the measurement and control signals are exchanged via a lossy and noisy communication channels, which makes the system a *networked control system* (NCS). The research in NCS has branched into many different directions that deal with the effects of delays, limited information exchanged, and information losses on the stability of the plant, see, e.g., Nair et al. [2007] and the references therein. Control under information loss in the communication channel has been extensively studied within the Linear Quadratic Gaussian (LQG) framework Imer et al. [2006]. Typically, the communication channel(s) are modeled by an independent and identically distributed (i.i.d) Bernoulli process, which assign probabilities to the successful transmission of packets. Perhaps the most well known result in this setting is: When the transmission of sensor and control data packets happens over a network with TCP-like protocols, the closed-loop system under LQG controller can be mean-square stabilized provided that the probabilities of successful transmission are above a certain threshold. Since the TCP-like protocols enable the receiver to obtain an acknowledgment of whether or not the packets were successfully transmitted, the separation principle holds and the optimal LQG controller is linear in the estimated state. Thus, this result is a proper generalization of the classical LQG control problem to the networked control setting.

Within the LQG setting, control inputs are not assumed bounded and therefore linear state feedback is a permissible and optimal strategy. However, guaranteeing hard bounds on the control inputs is of paramount importance in applications. Consequently, many researchers have pursued the problem of optimal control and stabilization for linear systems with bounded control inputs Bernstein and Michel [1995]; Saberi et al. [1999]; Toivonen [1983]; Wonham and Cashman [1969]. This problem has also received a renewed interest in recent years Chatterjee, Hokayem and Lygeros [2009]; Digaĭlova and Kurzhanskiĭ [2004]; Hokayem et al. [2009]; Wang and Boyd [2009]. In the deterministic setting, it is well-known Yang et al. [1997] that global asymptotic stabilization of a linear system $x_{t+1} = Ax_t + Bu_t$ is possible if and only if the pair $(A, B)$ is stabilizable under unbounded controls and the spectral radius of the system matrix $A$ is at most 1. In the stochastic setting, it was argued in Nair and Evans [2004] that ensuring a mean-square bound for every initial condition is not possible for linear systems with bounded control elements if the system matrix $A$ is unstable. The article Ramponi et al. [2010] establishes the existence of a policy with sufficiently large control authority that ensures mean-square boundedness of the states of the system under the assumption that $A$ is Lyapunov stable. Although Lyapunov stability of $A$ is a stronger requirement than the spectral radius of $A$ being at most 1, to the best of our knowledge, this is the current state of the art.

In this chapter we generalize the results of Ramponi et al. [2010] to incorporate noisy control channels. We consider mean-square boundedness of stochastic linear systems under the following specification:

1. the communication channel between the controller and the system actuators is noisy whereas the communication channel between sensors and controller is noiseless, and

2. hard constraints on the control inputs must be satisfied.

We are thus concerned with a networked setting as proposed in Elia [2005]; Schenato et al. [2007] when generalized to incorporate bounded control inputs Ramponi et al. [2010]. The control input $u^{(i)}$ for the $i$-th plant is communicated to the corresponding plant actuator via a lossy communication channel, which is characterized by the noise $\nu^{(i)}$ affecting the control input multiplicatively as shown in Figure 8.1. We assume that the states are perfectly observed and are transmitted to the controller without any loss.

The remainder of this chapter is organized as follows. In Section 8.2 we formalize the main problem with all the underlying assumptions. In Section 8.3 we state the main results and defer the proofs until Section 8.4. Section 8.5 mentions some future work.

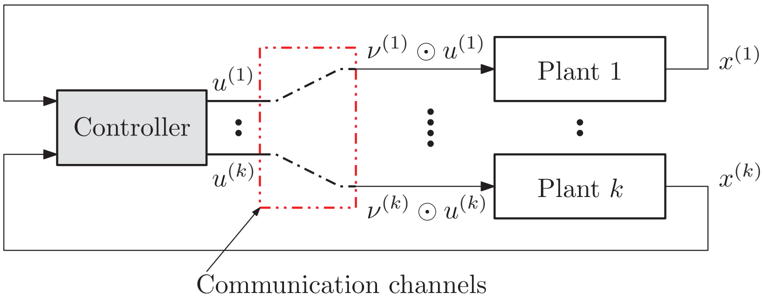


Figure 8.1: Topology of the control system.

### Notation

For any random vector $\nu$ let $\mu_\nu := \mathsf{E}[\nu]$ denote its mean and $\sigma_\nu := \operatorname{var}(\nu) := \mathsf{E}\left[\|\nu - \mu_\nu\|^2\right]$ denote its second moment. For a matrix $M$ we let $\|M\|$ denote the induced Euclidean norm of $M$. We shall employ the standard notation $\operatorname{diam}(S) := \sup_{x,y\in S} \|x - y\|$ to denote the diameter of a subset $S$ of Euclidean space. For $n \in \mathbb{N}$, by $\mathbf{1}_n$ we denote a vector of length $n$ with all entries equal to 1. For $r > 0$ and $n \in \mathbb{N}$, define the saturation function $\operatorname{sat}_r : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, as

$$\operatorname{sat}_r(z) = \begin{cases} z & \text{if } \|z\| < r \\ rz/\|z\| & \text{otherwise} \end{cases} \tag{8.1}$$

## 8.2 Problem Setup

Consider the following discrete-time stochastic linear system subjected to packet drops in the control communication channel

$$\begin{aligned} x_{t+1} &= Ax_t + B\tilde{u}_t + w_t, \\ \tilde{u}_t &:= \nu_t \odot u_t, \end{aligned} \qquad t \in \mathbb{N}_0, \tag{8.2}$$

where $x_t \in \mathbb{R}^d$ is the state, $u_t \in \mathbb{R}^m$ is the control input, $A \in \mathbb{R}^{d\times d}$ is the dynamics matrix, $B \in \mathbb{R}^{d\times m}$ is the input matrix, $(w_t)_{t\in\mathbb{N}_0}$ is an $\mathbb{R}^d$-valued random process noise, and $(\nu_t)_{t\in\mathbb{N}_0}$ is an $\mathbb{R}^m$-valued random process modelling the uncertainty in the control communication channel, and $\odot$ denotes the Schur or Hadamard product of matrices.[1] The initial condition $x_0 = \bar{x}$ is given and the state $x_t$ is perfectly observed by the controller.

The controller determines the control input $u_t$ based on the history of $k$ states $\zeta_{t,k} : = (x_{t-k+1}, \ldots, x_{t-1}, x_t)$. (For $t = 0, \ldots, k-2$, $\zeta_{t,k} := (\underbrace{x_0, \ldots, x_0}_{(k-1-t)\text{-times}}, x_0, x_1, \ldots, x_t)$.) The controller synthesizes a deterministic control policy $\pi = (\pi_t)_{t\in\mathbb{N}_0}$ which maps the states vector $\zeta$ into a control set $\mathcal{U}$. To wit,

$$u_t = \pi_t(\zeta_{t,k}), \quad t \in \mathbb{N}_0,$$

[1] Recall [Bernstein, 2009, p. 444] that if $M', M''$ are $n_1 \times n_2$ matrices with real entries, then $M' \odot M''$ is the $n_1 \times n_2$ matrix defined by $(M' \odot M'')_{i,j} := (M')_{i,j}(M'')_{i,j}$.

where the maps $\pi_t : \mathbb{R}^{kd} \longrightarrow \mathcal{U} \subseteq \mathbb{R}^m$, $t \in \mathbb{N}_0$, are Borel measurable. Such a control policy $\pi$ is known as a *$k$-history dependent policy.* The control set $\mathcal{U}$ is assumed to be nonempty, compact, and containing the origin. Any control policy $\pi = (\pi_t)_{t\in\mathbb{N}_0}$ which guarantees that the control input sequence $(u_t)_{t\in\mathbb{N}_0}$ satisfies

$$u_t \in \mathcal{U}, \qquad t \in \mathbb{N}_0, \tag{8.3}$$

is called an *admissible $k$-history dependent policy.* In many practical situations involving saturating actuators and hard bounds on control inputs, $\mathcal{U}$ is chosen to be a ball, i.e.,

$$\mathcal{U} := \big\{ z \in \mathbb{R}^m \,\big|\, \|z\| \leqslant U_{\max} \big\}, \tag{8.4}$$

where $U_{\max} > 0$ is called the *control authority* available to the controller.

Our control objective is to synthesize an admissible $k$-history dependent policy which ensures that the second moment of the closed-loop system, for any initial condition $\bar{x} \in \mathbb{R}^n$,

$$x_{t+1} = Ax_t + B\nu_t \odot \pi_t(\zeta_{t,k}) + w_t, \qquad t \in \mathbb{N}_0, \tag{8.5}$$

remains bounded for all $t \in \mathbb{N}_0$. We shall focus on the following problem:

**Problem 8.2.1.** Find, if possible, a control authority $U_{\max}$ and an admissible policy $\pi = (\pi_t)_{t\in\mathbb{N}_0}$ with control authority $U_{\max}$, such that the following condition holds: for every initial condition $\bar{x} \in \mathbb{R}^d$ there exists a constant $\zeta > 0$ such that the closed-loop system (8.5) satisfies

$$\mathsf{E}_{\bar{x}}[\|x_t\|^2] \leqslant \zeta, \qquad \forall t \in \mathbb{N}_0.$$

In practice, a performance index that accounts for the average sum of cost-per-stage functions (involving the state and control inputs) of the system is often required to be minimized; however, in this chapter we are only concerned with the stability property defined in Problem 8.2.1.

We shall make the following standing hypotheses:

**Assumption 8.2.2.**

(i) The matrix $A$ is Lyapunov stable, i.e., all the eigenvalues of $A$ lie in the closed unit circle, and all eigenvalues $\lambda$ satisfying $|\lambda| = 1$ have equal algebraic and geometric multiplicities.

(ii) The pair $(A, B)$ is stabilizable.

(iii) The process noise $(w_t)_{t\in\mathbb{N}_0}$ is an independent sequence, and has bounded fourth moment, i.e., $C_4 := \sup_{t\in\mathbb{N}_0} \mathsf{E}[\|w_t\|^4] < \infty$.

(iv) The control input sequence $(u_t)_{t\in\mathbb{N}_0}$ satisfies (8.3) and (8.4).

(v) The control channel noise $(\nu_t)_{t\in\mathbb{N}_0}$ is i.i.d. ◇

It follows from Assumption 8.2.2-(iii) that there exists $C_1 > 0$ such that $\mathsf{E}[\|w_t\|] \leqslant C_1$ for all $t \in \mathbb{N}_0$. (For instance, Jensen's inequality shows that $C_1 \leqslant \sqrt[4]{C_4}$.) Note also that Assumption 8.2.2-(iii) does not require that the process noise vectors $(w_t)_{t\in\mathbb{N}_0}$ be identically distributed. The assumption of mutual independence of $(w_t)_{t\in\mathbb{N}_0}$ can also be relaxed, but we shall not pursue this line of generalization here.

Without any loss of generality, we also assume that $A$ is in real Jordan canonical form (cf. Nair and Evans [2004]). Indeed, given a linear system described by system matrices $(\tilde{A}, \tilde{B})$, there exists a coordinate transformation in the state-space that brings the pair $(\tilde{A}, \tilde{B})$ to the pair $(A, B)$, where $A$ is in real Jordan form [Horn and Johnson, 1990, p. 150]. In particular, choosing a suitable ordering of the Jordan blocks, we can ensure that the pair $(A, B)$ has the form $\left(\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}, \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}\right)$, where $A_1 \in \mathbb{R}^{d_1\times d_1}$ is Schur stable, and $A_2 \in \mathbb{R}^{d_2\times d_2}$ has its eigenvalues on the unit circle. By Assumption 8.2.2-(i), $A_2$ is therefore block-diagonal with elements on the diagonal being either ±1 or 2 × 2 rotation matrices. As a consequence, $A_2$ is orthogonal. Moreover, since $(A, B)$ is stabilizable by Assumption 8.2.2-(ii), the pair $(A_2, B_2)$ must be reachable in a number of steps $\kappa \leqslant d_2$ that depends on the dimension of $A_2$ and the structure of $(A_2, B_2)$, i.e., $\operatorname{rank}(\mathfrak{R}_\kappa(A_2, B_2)) = d_2$, where

$$\mathfrak{R}_\kappa(A_2, B_2) := \begin{bmatrix} A_2^{\kappa-1}B_2 & \cdots & A_2B_2 & B_2 \end{bmatrix}.$$

The smallest such $\kappa$ is called the *controllability index* of $(A_2, B_2)$ and is fixed throughout the rest of this chapter. Summing up, we can start by considering that the state equation (8.2) has the form

$$\begin{bmatrix} x_{t+1}^{(1)} \\ x_{t+1}^{(2)} \end{bmatrix} = \begin{bmatrix} A_1 x_t^{(1)} \\ A_2 x_t^{(2)} \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \tilde{u}_t + \begin{bmatrix} w_t^{(1)} \\ w_t^{(2)} \end{bmatrix}, \tag{8.6}$$

where $A_1$ is Schur stable, $A_2$ is orthogonal, and the subsystem $(A_2, B_2)$ is reachable in $\kappa$ steps. Since the matrix $\mathfrak{R}_\kappa(A_2, B_2)$ has rank full rank, its Moore-Penrose pseudoinverse exists and is given by

$$\mathfrak{R}_\kappa(A_2, B_2)^+ := \mathfrak{R}_\kappa(A_2, B_2)^\mathsf{T}\left(\mathfrak{R}_\kappa(A_2, B_2)\mathfrak{R}_\kappa(A_2, B_2)^\mathsf{T}\right)^{-1}.$$

## 8.3 Admissible Policy of Bounded Control Authority

We now state the main result pertaining to the existence of a policy of bounded authority that renders the state of the system (8.2) mean-square bounded. Let us define the normalized measure of dispersion or the noise-to-signal ratio of the channel $\varPsi := \sqrt{\sigma_\nu} \cdot \max_{i=1,\ldots,m} |(\mu_\nu)_i|^{-1}$. We impose the following additional requirements:

**Assumption 8.3.1.** In addition to Assumption 8.2.2 we stipulate that:

(vi) The control channel noise has bounded range, i.e., $\nu_0 \in \mathrm{T}$, where $\mathrm{T}$ is a bounded subset of $\mathbb{R}^m$, and that $\mu_\nu$ has nonzero entries.

(vii) The following two technical conditions hold:

(vii.a) $\kappa \Psi \sqrt{\sigma_\nu} \left\| \mathfrak{R}_\kappa(A_2, B_2)^+ \right\| \left\| \mathfrak{R}_\kappa(A_2, B_2) \right\| < 1$.

(vii.b) $U_{\max} > \big(\sqrt{\kappa} C_1 \max_{i=1,\ldots,m} \left|(\mu_\nu)_i\right|^{-1} \left\| \mathfrak{R}_\kappa(A_2, B_2)^+ \right\| \left\| \mathfrak{R}_\kappa(A_2, I_2) \right\|\big) \div \big(1 - \kappa \Psi$ $\left\| \mathfrak{R}_\kappa(A_2, B_2)^+ \right\| \left\| \mathfrak{R}_\kappa(A_2, B_2) \right\|\big)$. ◇

**Proposition 8.3.2.** *Consider the system* (8.2)*, and suppose that Assumption* 8.3.1 *holds. Then there exists a* $\kappa$*-history dependent policy* $\pi := (\pi_t)_{t \in \mathbb{N}_0}$ *with control authority at most* $U_{\max}$*, such that for every initial condition* $\bar{x}$ *there exists a constant* $\zeta = \zeta(\bar{x}, \kappa, \mu_\nu, \Psi, C_1) > 0$ *with*

$$\mathsf{E}_{\bar{x}}[\|x_t\|^2] \leqslant \zeta \qquad \textit{for all } t \in \mathbb{N}_0$$

*in closed-loop.*

*Remark* 8.3.3. Proposition 8.3.2 assumes minimal structure from the set in which the control channel noise takes its values. In particular, we do not assume that the control channel noise takes values in a finite set—in fact, T may be uncountable. While the standard choice of modelling uncertainty in the control channels has focussed on a multiplicative Bernoulli $\{0, 1\}$ random variable multiplying the entire control vector, there are cases in which the uncertainty model considered in (8.2) (i.e., different random variables multiplying the components of the controller,) makes sense. For instance, the standard processes of control quantization or "binning" can be viewed as introducing uncertainty to the controller—components of the controller being multiplied by bounded but not necessarily identically distributed random variables; the set T has the natural interpretation of the "largest bin." In view of this, Assumption 8.3.1-(vii.a) is a technical condition stipulated as a trade-off for the absence of any further structure in the set T. ◁

In Section 8.4 we prove Proposition 8.3.2 by a constructive method. It turns out that our policy (see (8.12) below) is derived from the $\kappa$-subsampled system $(x_{\kappa t})_{t \in \mathbb{N}_0}$, and is $\kappa$-history dependent. To wit, for each $n \in \mathbb{N}_0$, at time $\kappa n$, based on the state $x_{\kappa n}$, the policy synthesizes a $\kappa$-long sequence of control values for time steps $\kappa n, \kappa n + 1, \ldots, \kappa(n+1) - 1$.

Let us assume that the same uncertainty enters all the control channels, i.e., $\tilde{u}_t = \nu_t u_t$, where $\nu_t \in \mathbb{R}$. The structure of our control policy permits us to transmit the control data packets in a single burst each $\kappa$ steps. This however, necessitates the presence of a buffer at the actuator to store the $\kappa$ control values $\{\nu_{\kappa n} u_{\kappa n}, \nu_{\kappa n} u_{\kappa n+1}, \ldots, \nu_{\kappa n} u_{\kappa(n+1)-1}\}$ transferred in a burst at time $\kappa n$, such that at each time $t \in \{\kappa n, \ldots, \kappa(n+1) - 1\}$, the control $\nu_{\kappa n} u_t$ can be applied.

**Assumption 8.3.4.** In addition to Assumption 8.2.2, we require that:

(vi′) Control signals are sent to the actuator every $\kappa$ steps, and for each $t \in \mathbb{N}_0$, the control channel noise is of the form $\nu_{\kappa t} \mathbf{1}_{\kappa m}$, with $\nu_{\kappa n} \in \{0, 1\}$ and $\mathsf{P}(\nu_{\kappa n} = 1) = p \in \,]0, 1[$ for each $n \in \mathbb{N}_0$.

(vii′) $U_{\max} > \sqrt{\kappa} C_1 \left\| \mathfrak{R}_\kappa(A_2, B_2)^+ \right\| \left\| \mathfrak{R}_\kappa(A_2, I_2) \right\| / p$. ◇

**Proposition 8.3.5.** *Consider the system* (8.2)*, and suppose that Assumption* 8.3.4 *holds. Then there exists a $\kappa$-history dependent policy $\pi \coloneqq (\pi_t)_{t\in\mathbb{N}_0}$ with control authority at most $U_{\max}$, such that for every initial condition $\bar{x}$ there exists a constant $\zeta' = \zeta'(\bar{x}, \kappa, p, C_1) > 0$ with*

$$\mathsf{E}_{\bar{x}}[\|x_t\|^2] \leqslant \zeta' \qquad \textit{for all } t \in \mathbb{N}_0$$

*in closed-loop.*

*Remark* 8.3.6. We noted in Remark 8.3.3 that Proposition 8.3.2 assumes minimal structure from the bounded set T. In contrast, Proposition 8.3.5 assumes a rather specific structure of the set T—that it consists of two elements (note that $\nu_{\kappa n} \in \{0, 1\}$ for each $n \in \mathbb{N}_0$ in Assumption 8.3.4-(vi′)). The i.i.d Bernoulli assumption on $(\nu_{\kappa t})_{t\in\mathbb{N}_0}$ leads to a simpler description of the control authority $U_{\max}$ in Assumption 8.3.4-(vii′) compared to Assumption 8.3.1-(vii.b), and the analog of Assumption 8.3.1-(vii.a) is not required here. ◁

*Example* 8.3.7. Consider the scalar system $x_{t+1} = x_t + \tilde{u}_t + w_t$, $t \geqslant 0$, with initial condition $x_0 = \bar{x}$, $\tilde{u}_t = \nu_t u_t$. Suppose that $(\nu_t)_{t\in\mathbb{N}_0}$ is i.i.d Bernoulli $\{0, 1\}$ with $\mathsf{P}(\nu_t = 1) = p > 0$, and let $(w_t)_{t\in\mathbb{N}_0}$ be i.i.d, and satisfy $\sup_{t\in\mathbb{N}_0} \mathsf{E}[|w_t|^4] = C_4' < \infty$. This implies, in particular, that $\sup_{t\in\mathbb{N}_0} \mathsf{E}[|w_t|] \leqslant C_1' \leqslant \sqrt[4]{C_4'}$. Suppose that $U_{\max} > C_1'/p$, where $u_t \in [-U_{\max}, U_{\max}]$ for all $t$. With this much data it is easy to verify the conditions of Assumption 8.3.4. We conclude by Proposition 8.3.5 that there exists a policy with control authority at most $U_{\max}$ such that the system is mean-square bounded. In fact, we see that for every nonzero probability $p$ of transmission of the control signal, there exists a control authority $U_{\max} > 0$ and a policy with control authority at most $U_{\max}$, under which the state of the system is mean-square bounded. △

*Remark* 8.3.8. Notice that Proposition 8.3.5 does not contradict the classical results of NCS under packet losses. For e.g., it was proved in [Schenato et al., 2007, Lemma 5.4] that there exists a threshold probability of i.i.d. Bernoulli packet drops such that a stabilizing linear feedback for unstable linear systems can be found provided the drop probability is less than that threshold. Indeed, in Assumption 8.2.2 we have specifically ruled out unstable $A$. ◁

## 8.4 Proofs for the Existence of Admissible Policies

For our proofs of Propositions 8.3.2 and 8.3.5 we shall employ the following immediate adaptation of [Pemantle and Rosenthal, 1999, Theorem 1] on $\mathrm{L}_2$ bounds of nonnegative random variables:

**Proposition 8.4.1** (Pemantle and Rosenthal [1999])**.** *Let $(\Omega, \mathfrak{F}, \mathsf{P})$ be a probability space, and let $(\mathfrak{F}_t)_{t\in\mathbb{N}_0}$ be a filtration on $(\Omega, \mathfrak{F}, \mathsf{P})$. Suppose that $(\xi_t)_{t\in\mathbb{N}_0}$ is a family of nonnegative random variables adapted to $(\mathfrak{F}_t)_{t\in\mathbb{N}_0}$, such that there exist constants $a, M, J > 0$ such that $\xi_0 < J$, and for all $t \in \mathbb{N}_0$,*

$$\mathsf{E}^{\mathfrak{F}_t}[\xi_{t+1} - \xi_t] \leqslant -a \ \ \textit{on the set } \{\xi_t > J\}, \tag{8.7}$$

$$\mathsf{E}[|\xi_{t+1} - \xi_t|^4 \mid \xi_0, \ldots, \xi_t] \leqslant M. \tag{8.8}$$

*Then there exists* $c = c(a, J, M) > 0$ *such that* $\sup_{t\in\mathbb{N}_0} \mathsf{E}\big[\xi_t^2\big] \leqslant c$.

In what follows we let $I_2$ denote the $d_2 \times d_2$ identity matrix.

**Lemma 8.4.2.** *Given the system* (8.2), *suppose that Assumption* 8.2.2 *holds, and consider the decomposition* (8.6). *Let* $\mathfrak{F}_t$ *be the* $\sigma$*-algebra generated by* $(x_t)_{t\in\mathbb{N}_0}$. *Then there exists* $a, J > 0$ *such that*

$$\mathsf{E}^{\mathfrak{F}_{\kappa t}}\left[\left\|x^{(2)}_{\kappa(t+1)}\right\| - \left\|x^{(2)}_{\kappa t}\right\|\right] \leqslant -a \quad \text{on the set } \left\{\left\|x^{(2)}_{\kappa t}\right\| > J\right\}$$

*for all* $t \in \mathbb{N}_0$.

*Proof.* To simplify notation we write compactly

$$\boldsymbol{\nu}_{\kappa t} := \begin{bmatrix} \nu_{\kappa t} \\ \nu_{\kappa t+1} \\ \vdots \\ \nu_{\kappa(t+1)-1} \end{bmatrix} \quad \text{and} \quad \mathbf{u}_{\kappa t} := \begin{bmatrix} u_{\kappa t} \\ u_{\kappa t+1} \\ \vdots \\ u_{\kappa(t+1)-1} \end{bmatrix}. \tag{8.9}$$

It follows from the system dynamics that

$$\begin{aligned} x^{(2)}_{\kappa(t+1)} = A_2^{\kappa} x^{(2)}_{\kappa t} &+ \mathfrak{R}_\kappa(A_2, B_2)\mathbf{u}_{\kappa t} \\ &+ \mathfrak{R}_\kappa(A_2, I_2) w^{(2)}_{\kappa t:\kappa(t+1)-1}, \qquad t \in \mathbb{N}_0, \end{aligned}$$

where $w^{(2)}_{\kappa t:\kappa(t+1)-1} := \left[(w^{(2)}_{\kappa t})^{\mathsf{T}} \quad \cdots \quad (w^{(2)}_{\kappa(t+1)-1})^{\mathsf{T}}\right]^{\mathsf{T}}$. Therefore,

$$\begin{aligned} &\mathsf{E}^{\mathfrak{F}_{\kappa t}}\left[\left\|x^{(2)}_{\kappa(t+1)}\right\| - \left\|x^{(2)}_{\kappa t}\right\|\right] \\ &= \mathsf{E}^{\mathfrak{F}_{\kappa t}}\Big[\big\|A_2^{\kappa} x^{(2)}_{\kappa t} + \mathfrak{R}_\kappa(A_2, B_2)\mathbf{u}_{\kappa t} \\ &\quad + \mathfrak{R}_\kappa(A_2, I_2) w^{(2)}_{\kappa t:\kappa(t+1)-1}\big\| - \left\|x^{(2)}_{\kappa t}\right\|\Big] \\ &\leqslant \mathsf{E}^{\mathfrak{F}_{\kappa t}}\left[\left\|A_2^{\kappa} x^{(2)}_{\kappa t} + \mathfrak{R}_\kappa(A_2, B_2)\mathbf{u}_{\kappa t}\right\| - \left\|x^{(2)}_{\kappa t}\right\|\right] \\ &\quad + \left\|\mathfrak{R}_\kappa(A_2, I_2)\right\| \mathsf{E}\left[\left\|w^{(2)}_{\kappa t:\kappa(t+1)-1}\right\|\right] \\ &\leqslant \mathsf{E}^{\mathfrak{F}_{\kappa t}}\left[\left\|A_2^{\kappa} x^{(2)}_{\kappa t} + \mathfrak{R}_\kappa(A_2, B_2)\mathbf{u}_{\kappa t}\right\| - \left\|x^{(2)}_{\kappa t}\right\|\right] \\ &\quad + \sqrt{\kappa}\left\|\mathfrak{R}_\kappa(A_2, I_2)\right\| C_1. \end{aligned}$$

Since $A_2$ is orthogonal, we have $\left\|A_2^{\kappa} x^{(2)}_{\kappa t}\right\| = \left\|x^{(2)}_{\kappa t}\right\|$. We require $\mathbf{u}_{\kappa t}$ be $\mathfrak{F}_{\kappa t}$-measurable.

Employing Jensen's inequality and sublinearity of the square-root function, we get

$$\begin{aligned}
&\mathsf{E}^{\mathfrak{F}_{\kappa t}}\Big[\Big\|A_2^\kappa x_{\kappa t}^{(2)} + \mathfrak{R}_\kappa(A_2, B_2)\mathbf{u}_{\kappa t:\kappa(t+1)-1}\Big\|\Big] \\
&\leqslant \sqrt{\mathsf{E}^{\mathfrak{F}_{\kappa t}}\Big[\Big\|A_2^\kappa x_{\kappa t}^{(2)} + \mathfrak{R}_\kappa(A_2, B_2)\tilde{u}_{\kappa t:\kappa(t+1)-1}\Big\|^2\Big]} \\
&= \Big(\Big\|x_{\kappa t}^{(2)}\Big\|^2 + 2\big(x_{\kappa t}^{(2)}\big)^\mathsf{T}(A_2^\kappa)^\mathsf{T}\mathfrak{R}_k(A_2, B_2)(\mathsf{E}^{\mathfrak{F}_{\kappa t}}[\boldsymbol{\nu}_{\kappa t}] \odot \mathbf{u}_{\kappa t}) \\
&\quad + \mathsf{E}^{\mathfrak{F}_{\kappa t}}\big[\big(\mathfrak{R}_\kappa(A_2, B_2)(\boldsymbol{\nu}_{\kappa t} \odot \mathbf{u}_{\kappa t})\big)^\mathsf{T} \\
&\qquad\qquad \cdot \big(\mathfrak{R}_\kappa(A_2, B_2)(\boldsymbol{\nu}_{\kappa t} \odot \mathbf{u}_{\kappa t})\big)\big]\Big)^{1/2} \\
&\leqslant \Big\|A_2^\kappa x_{\kappa t}^{(2)} + \mathfrak{R}_\kappa(A_2, B_2)(\mathsf{E}^{\mathfrak{F}_{\kappa t}}[\boldsymbol{\nu}_{\kappa t}] \odot \mathbf{u}_{\kappa t})\Big\| \\
&\quad + \Big(\mathsf{E}^{\mathfrak{F}_{\kappa t}}\big[\|\mathfrak{R}_\kappa(A_2, B_2)(\boldsymbol{\nu}_{\kappa t} \odot \mathbf{u}_{\kappa t})\|^2\big] \\
&\qquad\qquad - \big\|\mathfrak{R}_\kappa(A_2, B_2)(\mathsf{E}^{\mathfrak{F}_{\kappa t}}[\boldsymbol{\nu}_{\kappa t}] \odot \mathbf{u}_{\kappa t})\big\|^2\Big)^{1/2},
\end{aligned}$$

The last term under the square-root is the conditional variance of vector $\mathfrak{R}_\kappa(A_2, B_2)(\boldsymbol{\nu}_{\kappa t} \odot \mathbf{u}_{\kappa t})$ given $\mathfrak{F}_{\kappa t}$. Since $(\nu_n)_{n=\kappa t}^{\kappa(t+1)-1}$ is independent of $\mathfrak{F}_{\kappa t}$, $\bar{\boldsymbol{\nu}} := \mathsf{E}^{\mathfrak{F}_{\kappa t}}[\boldsymbol{\nu}_{\kappa t}]$ is a constant, and equals $\operatorname{vec}\{\underbrace{\mu_\nu, \dots, \mu_\nu}_{\kappa\text{-times}}\}$.) Thus, we see that

$$\begin{aligned}
&\mathsf{E}^{\mathfrak{F}_{\kappa t}}\Big[\Big\|A_2^\kappa x_{\kappa t}^{(2)} + \mathfrak{R}_\kappa(A_2, B_2)\tilde{u}_{\kappa t:\kappa(t+1)-1}\Big\|\Big] \\
&\leqslant \Big\|A_2^\kappa x_{\kappa t}^{(2)} + \mathfrak{R}_\kappa(A_2, B_2)(\bar{\boldsymbol{\nu}} \odot \mathbf{u}_{\kappa t})\Big\| \\
&\qquad + \sqrt{\mathsf{E}^{\mathfrak{F}_{\kappa t}}\big[\|\mathfrak{R}_\kappa(A_2, B_2)((\boldsymbol{\nu}_{\kappa t} - \bar{\boldsymbol{\nu}}) \odot \mathbf{u}_{\kappa t})\|^2\big]} \\
&\leqslant \Big\|A_2^\kappa x_{\kappa t}^{(2)} + \mathfrak{R}_\kappa(A_2, B_2)(\bar{\boldsymbol{\nu}} \odot \mathbf{u}_{\kappa t})\Big\| \\
&\qquad + \sqrt{\kappa}\,\|\mathfrak{R}_\kappa(A_2, B_2)\|\, U_{\max}\sqrt{\mathsf{E}^{\mathfrak{F}_{\kappa t}}\big[\|\boldsymbol{\nu}_{\kappa t} - \bar{\boldsymbol{\nu}}\|^2\big]} \\
&\leqslant \Big\|A_2^\kappa x_{\kappa t}^{(2)} + \mathfrak{R}_\kappa(A_2, B_2)(\bar{\boldsymbol{\nu}} \odot \mathbf{u}_{\kappa t})\Big\| \\
&\qquad + \kappa\,\|\mathfrak{R}_\kappa(A_2, B_2)\|\, U_{\max}\sqrt{\sigma_\nu}.
\end{aligned}$$

Collecting the inequalities above, we see that

$$\begin{aligned}
&\mathsf{E}^{\mathfrak{F}_{\kappa t}}\Big[\Big\|x_{\kappa(t+1)}^{(2)}\Big\| - \Big\|x_{\kappa t}^{(2)}\Big\|\Big] \\
&\leqslant \Big\|A_2^\kappa x_{\kappa t}^{(2)} + \mathfrak{R}_\kappa(A_2, B_2)(\bar{\boldsymbol{\nu}} \odot \mathbf{u}_{\kappa t})\Big\| - \Big\|A_2^\kappa x_{\kappa t}^{(2)}\Big\| \\
&\quad + \kappa\big(\|\mathfrak{R}_\kappa(A_2, B_2)\|\, U_{\max}\sqrt{\sigma_\nu} + \tfrac{\|\mathfrak{R}_\kappa(A_2, I_2)\| C_1}{\sqrt{\kappa}}\big).
\end{aligned} \tag{8.10}$$

In view of Assumption 8.3.1-(vii.a) we see that there exists

$$\begin{aligned}0 < a &:= U_{\max}\big(1 - \kappa\varPsi \left\|\mathfrak{R}_\kappa(A_2, B_2)^+\right\| \left\|\mathfrak{R}_\kappa(A_2, B_2)\right\|\big)\\ &\quad - \sqrt{\kappa}C_1\left(\max_{i=1,\ldots,m} |(\mu_\nu)_i|^{-1}\right)\left\|\mathfrak{R}_\kappa(A_2, B_2)^+\right\| \left\|\mathfrak{R}_\kappa(A_2, I_2)\right\|.\end{aligned}$$

We now define

$$\begin{aligned} r &:= a + \kappa\big(\left\|\mathfrak{R}_\kappa(A_2, B_2)\right\| U_{\max}\sqrt{\sigma_\nu} + \tfrac{\left\|\mathfrak{R}_\kappa(A_2, I_2)\right\| C_1}{\sqrt{\kappa}}\big)\\ &\leqslant U_{\max}. \end{aligned} \tag{8.11}$$

By Assumption 8.3.1-(vi), every entry of $\bar{\boldsymbol{\nu}}$ is nonzero; we let $\bar{\boldsymbol{\nu}}^{(-1)}$ be the vector of reciprocals of each entry of $\bar{\boldsymbol{\nu}}$ (i.e., $(\bar{\boldsymbol{\nu}}^{(-1)})_i = (\bar{\boldsymbol{\nu}}_i)^{-1}$ for each $i$). We define our control policy[2]

$$\mathbf{u}_{\kappa t} := \mathbf{u}_{\kappa t}\big(x^{(2)}_{\kappa t}\big) := -\mathfrak{R}_\kappa(A_2, B_2)^+ \operatorname{sat}_r\big(A_2^\kappa x^{(2)}_{\kappa t}\big) \odot \bar{\boldsymbol{\nu}}^{(-1)}, \tag{8.12}$$

where $\operatorname{sat}_r$ is the function defined in (8.1). Clearly, $\mathbf{u}_{\kappa t}$ is $\mathfrak{F}_{\kappa t}$-measurable. Substituting into (8.10) we see that

$$\begin{aligned} &\mathsf{E}^{\mathfrak{F}_{\kappa t}}\left[\left\|x^{(2)}_{\kappa(t+1)}\right\| - \left\|x^{(2)}_{\kappa t}\right\|\right]\\ &\leqslant -r + \kappa\big(\left\|\mathfrak{R}_\kappa(A_2, B_2)\right\| U_{\max}\sqrt{\sigma_\nu} + \tfrac{\left\|\mathfrak{R}_\kappa(A_2, I_2)\right\| C_1}{\sqrt{\kappa}}\big)\\ &\qquad\qquad \text{on the set } \left\{\left\|x^{(2)}_{\kappa t}\right\| \geqslant r\right\}\\ &\leqslant -a \qquad \text{on the set } \left\{\left\|x^{(2)}_{\kappa t}\right\| \geqslant r\right\}, \end{aligned}$$

where the last inequality follows from the definition of $r$ above.

Thus, it only remains to see that the control policy defined in (8.12) satisfies the bound $\|u_t\| \leqslant U_{\max}$ for each $t$. But in view of the definition of $r$ in (8.11) and our policy (8.12), we see that $\|\mathbf{u}_{\kappa t}\| \leqslant U_{\max}$, and the assertion follows. □

**Lemma 8.4.3.** *Given the system (8.2), suppose that Assumption 8.2.2 holds, and consider the decomposition (8.6). Then there exists $M > 0$ such that*

$$\mathsf{E}\left[\left|\left\|x^{(2)}_{\kappa(t+1)}\right\| - \left\|x^{(2)}_{\kappa t}\right\|\right|^4 \,\middle|\, \left\|x^{(2)}_0\right\|, \ldots, \left\|x^{(2)}_{\kappa t}\right\|\right] \leqslant M$$

*for all $t \in \mathbb{N}_0$.*

[2] This controller resembles in part the Ackermann's formula in linear control theory [Franklin et al., 2006, p. 477] employed in unconstrained deadbeat controllers.

*Proof.* We retain the notation $w^{(2)}_{\kappa t:\kappa(t+1)-1}$ from the proof of Lemma 8.4.2. Fix $t \in \mathbb{N}_0$. Observe that since $A_2$ is orthogonal, $\left\|x^{(2)}_{\kappa t}\right\| = \left\|A_2^\kappa x^{(2)}_{\kappa t}\right\|$, and therefore,

$$\begin{aligned}
&\mathsf{E}\Big[\Big|\left\|x^{(2)}_{\kappa(t+1)}\right\| - \left\|x^{(2)}_{\kappa t}\right\|\Big|^4 \;\Big|\; \big\{\|x^{(2)}_{\kappa n}\|\big\}_{n=0}^{t}\Big] \\
&= \mathsf{E}\Big[\Big|\Big\|A_2^\kappa x^{(2)}_{\kappa t} + \mathfrak{R}_\kappa(A_2, B_2)\tilde{u}_{\kappa t:\kappa(t+1)-1} + \\
&\quad \mathfrak{R}_\kappa(A_2, I_2) w^{(2)}_{\kappa t:\kappa(t+1)-1}\Big\| - \left\|A_2^\kappa x^{(2)}_{\kappa t}\right\|\Big|^4 \;\Big|\; \big\{\|x^{(2)}_{\kappa n}\|\big\}_{n=0}^{t}\Big] \\
&\leqslant \mathsf{E}\Big[\Big\|A_2^\kappa x^{(2)}_{\kappa t} + \mathfrak{R}_\kappa(A_2, B_2)\tilde{u}_{\kappa t:\kappa(t+1)-1} + \\
&\quad \mathfrak{R}_\kappa(A_2, I_2) w^{(2)}_{\kappa t:\kappa(t+1)-1} - A_2^\kappa x^{(2)}_{\kappa t}\Big\|^4 \;\Big|\; \big\{\|x^{(2)}_{\kappa n}\|\big\}_{n=0}^{t}\Big] \\
&= \mathsf{E}\Big[\Big\|\mathfrak{R}_\kappa(A_2, B_2)\tilde{u}_{\kappa t:\kappa(t+1)-1} + \mathfrak{R}_\kappa(A_2, I_2) w^{(2)}_{\kappa t:\kappa(t+1)-1}\Big\|^4 \\
&\qquad\qquad \Big|\; \big\{\|x^{(2)}_{\kappa n}\|\big\}_{n=0}^{t}\Big].
\end{aligned}$$

By Assumption 8.2.2-(iv), $\|\tilde{u}_t\| \leqslant \sqrt{m} U_{\max} \operatorname{diam}(\mathrm{T})$, which implies that

$$\begin{aligned}
&\mathsf{E}\Big[\Big\|\mathfrak{R}_\kappa(A_2, B_2)\tilde{u}_{\kappa t:\kappa(t+1)-1} + \mathfrak{R}_\kappa(A_2, I_2) w^{(2)}_{\kappa t:\kappa(t+1)-1}\Big\|^4 \\
&\qquad\qquad \Big|\; \big\{\|x^{(2)}_{\kappa n}\|\big\}_{n=0}^{t}\Big] \\
&\leqslant \mathsf{E}\Big[\Big(\kappa\sqrt{m} U_{\max} \operatorname{diam}(\mathrm{T}) \left\|\mathfrak{R}_\kappa(A_2, B_2)\right\| + \\
&\quad \left\|\mathfrak{R}_\kappa(A_2, I_2)\right\| \left\|w^{(2)}_{\kappa t:\kappa(t+1)-1}\right\|\Big)^4 \;\Big|\; \big\{\|x^{(2)}_{\kappa n}\|\big\}_{n=0}^{t}\Big].
\end{aligned}$$

Noting that $w^{(2)}_{\kappa t:\kappa(t+1)-1}$ is independent of $\left\|x^{(2)}_0\right\|, \ldots, \left\|x^{(2)}_{\kappa t}\right\|$ in view of Assumption 8.2.2-(iii), applying Jensen's inequality to the right-hand side above yields

$$\begin{aligned}
&\mathsf{E}\Big[\Big(\kappa\sqrt{m} U_{\max} \operatorname{diam}(\mathrm{T}) \left\|\mathfrak{R}_\kappa(A_2, B_2)\right\| + \\
&\quad \left\|\mathfrak{R}_\kappa(A_2, I_2)\right\| \left\|w^{(2)}_{\kappa t:\kappa(t+1)-1}\right\|\Big)^4\Big] \\
&= \mathsf{E}\Big[\Big(\kappa\sqrt{m} U_{\max} \operatorname{diam}(\mathrm{T}) \left\|\mathfrak{R}_\kappa(A_2, B_2)\right\| + \\
&\quad \kappa \left\|\mathfrak{R}_\kappa(A_2, I_2)\right\| \left\|w^{(2)}_{\kappa t}\right\|\Big)^4\Big] \\
&\leqslant \kappa^4\Big(\sqrt{m} U_{\max} \operatorname{diam}(\mathrm{T}) \left\|\mathfrak{R}_\kappa(A_2, B_2)\right\| + \\
&\quad \left\|\mathfrak{R}_\kappa(A_2, I_2)\right\| C_1\Big)^4.
\end{aligned}$$

The assertion follows at once with $M$ equal to the right-hand side of the last inequality. □

*Proof of Proposition 8.3.2:* From (8.6) we see that the system splits into two parts, $x^{(1)}$ and $x^{(2)}$, with the sequence $(x_t^{(1)})_{t\in\mathbb{N}_0}$ describing the evolution of the Schur stable component of the state, and $(x_t^{(2)})_{t\in\mathbb{N}_0}$ describing the evolution of the orthogonal component of the state. It is well-known that $(x_t^{(1)})_{t\in\mathbb{N}_0}$ is mean-square bounded so long as the control is bounded, which by Assumption 8.3.1-(vi) clearly holds (i.e., there exists $\zeta^{(1)} > 0$ such that $\mathsf{E}_{\bar{x}}\big[\big\|x_t^{(1)}\big\|^2\big] \leqslant \zeta^{(1)}$ for all $t \in \mathbb{N}_0$). It thus suffices to concentrate on $(x_t^{(2)})_{t\in\mathbb{N}_0}$. We let $\xi_t := \big\|x_{\kappa t}^{(2)}\big\|$ for each $t \in \mathbb{N}_0$. We see that:

- the condition (8.7) of Proposition 8.4.1 holds with $J = r$, where $r$ is as defined in (8.11), by Lemma 8.4.2, and
- the condition (8.8) of Proposition 8.4.1 holds by Lemma 8.4.3.

Defining $J := \max\big\{\big\|x_0^{(2)}\big\|, r\big\}$, we see that by Proposition 8.4.1 there exists a $\widetilde{\zeta}^{(2)} = \widetilde{\zeta}^{(2)}(a, M, J)$ such that $\mathsf{E}_{\bar{x}}\big[\big\|x_{\kappa t}^{(2)}\big\|^2\big] \leqslant \widetilde{\zeta}^{(2)}$ for all $t \in \mathbb{N}_0$. Since the subsampled process $(x_{\kappa t}^{(2)})_{t\in\mathbb{N}_0}$ is mean-square bounded, and $x^{(2)}$ is generated by a linear dynamical system, we conclude that there exists $\zeta^{(2)} > 0$ such that $\mathsf{E}_{\bar{x}}\big[\big\|x_t^{(2)}\big\|^2\big] \leqslant \zeta^{(2)}$ for all $t \in \mathbb{N}_0$. The assertion of Proposition 8.3.2 follows with $\zeta := \zeta^{(1)} + \zeta^{(2)}$ and noticing that $a$ and $J$ depend on $\bar{x}$, $\kappa$, $\mu_\nu$, $\Psi$ and $C_1$. ■

*Proof of Proposition 8.3.5* Let us consider the $\kappa$-subsampled system

$$\begin{aligned} x_{\kappa(t+1)}^{(2)} = A_2^\kappa x_{\kappa t}^{(2)} &+ \mathfrak{R}_\kappa(A_2, B_2)\nu_{\kappa t} u_{\kappa t:\kappa(t+1)-1} \\ &+ \mathfrak{R}_\kappa(A_2, I_2) w_{\kappa t:\kappa(t+1)-1}, \qquad t \in \mathbb{N}_0, \end{aligned}$$

where $u_{\kappa t:\kappa(t+1)-1} := \begin{bmatrix} u_{\kappa t}^{\mathsf{T}} & \ldots & u_{\kappa(t+1)-1}^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}}$. For this subsampled system we propose the control policy:

$$u_{\kappa t:\kappa(t+1)-1} = -\mathfrak{R}_\kappa(A_2, B_2)^+ \operatorname{sat}_r\big(A_2^\kappa x_{\kappa t}^{(2)}\big), \tag{8.13}$$

for some $r > 0$ to be defined shortly. Let us verify the conditions of Proposition 8.4.1 for the process $\left(\left\|x^{(2)}_{\kappa t}\right\|\right)_{t\in\mathbb{N}_0}$ under the control policy proposed above. We see immediately that

$$\begin{aligned}
&\mathsf{E}^{\mathfrak{F}_{\kappa t}}\left[\left\|x^{(2)}_{\kappa(t+1)}\right\| - \left\|x^{(2)}_{\kappa t}\right\|\right]\\
&\leqslant \mathsf{E}^{\mathfrak{F}_{\kappa t}}\left[\left\|A_2^{\kappa}x^{(2)}_{t\kappa} + \mathfrak{R}_{\kappa}(A_2, B_2)\nu_{\kappa t}u_{\kappa t:\kappa(t+1)-1}\right\|\right]\\
&\qquad - \left\|x^{(2)}_{\kappa t}\right\| + \mathsf{E}\left[\left\|\mathfrak{R}_{\kappa}(A_2, I_2)w_{\kappa t:\kappa(t+1)-1}\right\|\right]\\
&= p\left\|A_2^{\kappa}x^{(2)}_{\kappa t} + \mathfrak{R}_{\kappa}(A_2, B_2)u_{\kappa t:\kappa(t+1)-1}\right\|\\
&\qquad + (1-p)\left\|A_2^{\kappa}x^{(2)}_{\kappa t}\right\| - \left\|x^{(2)}_{\kappa t}\right\|\\
&\qquad + \mathsf{E}\left[\left\|\mathfrak{R}_{\kappa}(A_2, I_2)w_{\kappa t:\kappa(t+1)-1}\right\|\right]\\
&= p\left(\left\|A_2^{\kappa}x^{(2)}_{\kappa t} + \mathfrak{R}_{\kappa}(A_2, B_2)u_{\kappa t:\kappa(t+1)-1}\right\| - \left\|x^{(2)}_{\kappa t}\right\|\right)\\
&\qquad + \sqrt{\kappa}\left\|\mathfrak{R}_{\kappa}(A_2, I_2)\right\| C_1\\
&= -pr + \sqrt{\kappa}\left\|\mathfrak{R}_{\kappa}(A_2, I_2)\right\| C_1,
\end{aligned}$$

where we have employed orthogonality of $A_2$ to arrive at the second equality above. By Assumption 8.3.4-(vii′) we see that there exists $a > 0$ such that

$$\left\|\mathfrak{R}_{\kappa}(A_2, B_2)^{+}\right\| \left(a + \sqrt{\kappa}C_1\left\|\mathfrak{R}_{\kappa}(A_2, I_2)\right\|/p\right) \leqslant U_{\max}.$$

Letting $r := a + \sqrt{\kappa}C_1\left\|\mathfrak{R}_{\kappa}(A_2, I_2)\right\|/p$, we see that the condition (8.7) is verified with $J = r$. The condition (8.8) follows readily from Lemma 8.4.3, since the elements of the control input are uniformly bounded. Letting $J := \max\left\{r, \left\|x^{(2)}_0\right\|\right\}$, we see that by Proposition 8.4.1 there exists a constant $\zeta^{(2)} > 0$ such that $\mathsf{E}_{\bar{x}}\left[\left\|x^{(2)}_t\right\|^2\right] \leqslant \zeta^{(2)}$ for all $t \in \mathbb{N}_0$. By the same argument involving the Schur stable part $x^{(1)}$ as in the proof of Proposition 8.3.2, we see that there exists a constant $\zeta' > 0$ such that $\mathsf{E}_{\bar{x}}[\|x_t\|^2] \leqslant \zeta'$ for all $t \in \mathbb{N}_0$. In view of the fact that $a$ and $J$ depend on $\bar{x}$, $\kappa$, $p$, and $C_1$, this concludes the proof. ■

## 8.5 Discussion

We considered a networked systems setup in which the control inputs are transmitted via a noisy communication channel. Under mild assumptions on the statistics of the channel noise and control authority, we constructed a causal $\kappa$-steps control strategy that renders the state of the closed-loop networked system mean-square bounded. Future work will focus on extending the current results to encompass the case of imperfect and incomplete state measurements. In addition, the noise model presented in this chapter will be extended to adversarial stochastic noise models. The next chapter investigates the effect of network-induced risks on the failure probabilities affecting the sensor and control communication channels.

# Chapter 9

# Security Interdependencies for Networked Control Systems

## 9.1 Introduction

In this chapter, we study the security choices of identical networked controlled systems (NCS), when their security is interdependent due to the exposure to network induced risks. Each NCS is modeled by a discrete-time stochastic linear system, which is sensed and controlled over a communication network. Today, NCS already exhibit substantial interdependence. An imminent wider deployment of smart devices is only likely to result in a higher degree of interdependence Anderson and Fuloria [2010]; Weiss [2010].

The current state-of-the-art literature on NCS assumes independent and identically distributed (IID) packet losses for systems, even when the systems use the same communication network for their operation Amin et al. [2009*a*]; Garone et al. [2010]; Hespanha et al. [2007]; Imer et al. [2006]; Schenato et al. [2007]. In such settings, attacks on the availability of sensor and control data packets for one system do not affect the availability of data packets for other systems.

The analysis based on the IID packet loss models does not capture the environments in which an attack on the availability of data packets of one system can propagate to other systems due to fact that they share the same communication network. An important example of such attacks are the so-called distributed-denial-of-service (DDOS) attacks Amin et al. [2009*a*]; Cárdenas et al. [2008]; Weiss [2010]. Since the DDOS attacks affect the availability of sensor and control data packets of multiple systems, any security choice of one system is also likely to influence the security of other systems. This chapter contributes to the existing literature by considering a setting where security of one system affects security of other systems.

Several factors exacerbate the severity of the losses which may be caused by security interdependencies. First, only a small number of vendors provide embedded controller devices Weiss [2010], causing a danger of highly correlated software–hardware malfunctions. Due to the prevalence of identical devices, a single glitch could bring major disruption of NCS functioning. Second, since the NCS will soon govern the operation of critical infrastructure systems, the NCS interdependencies could be exploited by nation states. So far, no such occurrences have been recorded, but presence of aforementioned cyber attack capabilities is well documented W.A. Owens and Lin [2009], and cannot be ignored Weiss [2010]. The risks of such rare (but extremely disruptive) events are similar to risks of terrorist attacks Bier et al. [2007], and it is established that private mitigation of such risks fails, thus likely requiring governments to step in Heal and Kunreuther [2004].

We model the problem of operator's security choice as a non-cooperative two-stage game between m plant-controller systems (or players). Each of these players is modeled in a standard NCS setting (e.g., Imer et al. [2006]; Schenato et al. [2007]). In the first stage, each player has a binary choice of investing versus non-investing into enhanced security measures at his plant. In the second stage, players choose optimal control inputs for their respective plants. Each players' objective is to minimize the average long-term cost, which is comprised of the plant operating costs and the cost of security measures. We compare the individually optimal choices with that of the social planner, whose objective is to minimize the sum of aggregate operating costs of all the players (which include costs of security measures). The approach in this chapter compliments the existing and growing literature on investment efficient security strategies for critical systems systems Alpcan and Başar [2011]; Başar and Olsder [1999]; Cavusoglu et al. [2005]. By imposing penalties on the players not investing in security, we induce individually optimal player choices that coincide with the socially optimal ones. Such correction of individual incentives is frequently referred as internalizing the externalities Alpcan and Başar [2011].

The importance of network externalities for incentives to invest in security have been noted and modeled by many researches (e.g., Anderson et al. [2008]; Böhme and Schwartz [2010] and the references therein). The relevance of these effects for critical infrastructures, and in particular, the provision of electricity was raised in Anderson and Fuloria [2009, 2010], but to the best of our knowledge, so far nobody attempted formal modeling of security interdependencies in NCS. The closest models to ours are the application of security interdependencies to Internet security such as Lelarge and Bolot [2008], where the authors apply Heal and Kunreuther [2004], and present an analytical model, which permits them to study the deployment of security features and protocols in the sub-nets with different network topologies. Also, Lelarge [2009] expands on Heal and Kunreuther [2004] to study economics of malware (propagation of viruses and worms).

Our modeling of security choices builds on the Heal and Kunreuther's interdependent security model (see Heal and Kunreuther [2003, 2004]; Kunreuther and Heal [2002]). We refer the reader to Mounzer et al. [2010], Grossklags et al. [2008]; Hofmann [2007] for similar approaches.

In our setting, player actions differ from social optimum ones; this reflects the presence of externalities. Indeed, in general, when player costs are affected by other player's choices, players impose externalities on each other. The externalities manifest by the gap between the individually and socially optimal security choices Alpcan and Başar [2011]. In the case of negative externalities, players tend to under-invest in security. This is commonly referred to free-riding in economics. To internalize the externalities, an instrument (e.g., penalty) is commonly suggested which alters individually optimal security choices and makes them

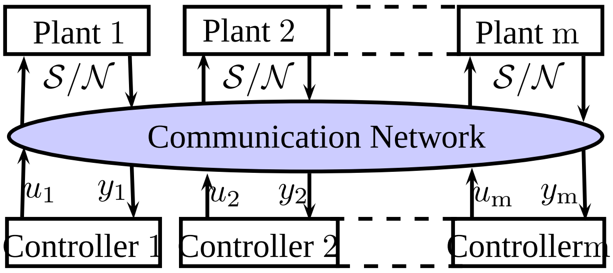


Figure 9.1: Networked Control System (NCS).

coincide with the socially optimum ones.

This chapter is organized as follows: In Section 9.2, we formulate the game between NCS when interdependencies are present. In Sections 9.3 and 9.4, we present the analysis of the game of two and m players respectively. For the two player case, we also derive the penalties to be imposed on players for not investing in security under which the individually and socially optimal choices coincide. Section 9.5 discusses some consequences of the results presented in this chapter.

## 9.2 Problem Setup

### 9.2.1 The Game

We consider an m−player stochastic two-stage game. The players are denoted by $\mathbf{P}1, \mathbf{P}2, \ldots, \mathbf{P}\mathrm{m}$, and the index set $\{1, \ldots, \mathrm{m}\}$ is denoted by $\boldsymbol{M}$. We model each player as a NCS (e.g., Schenato et al. [2007]) in which each $\mathbf{P}i$'s plant and controller communicate over a network; see Fig. 9.1. In the *first stage*, each $\mathbf{P}i$ $(i \in \boldsymbol{M})$ chooses to make a security investment $(\mathcal{S})$ or not $(\mathcal{N})$. Let $\mathcal{V}^i$ denote the security choice of $\mathbf{P}i$, i.e.,

$$\mathcal{V}^i := \begin{cases} \mathcal{S}, & \mathbf{P}i \text{ invests in security,} \\ \mathcal{N}, & \mathbf{P}i \text{ does not invest in security,} \end{cases}$$

and let $\mathcal{V}$ denote the set of player security choices, i.e.,

$$\mathcal{V} := \{\mathcal{V}^1, \ldots, \mathcal{V}^{\mathrm{m}}\}.$$

Once player security choices are made, they are irreversible and observable by all the players. The $\mathbf{P}i$'s first stage investment is given by

$$J_{\mathrm{I}}^i(\mathcal{V}) := (1 - \mathcal{I}^i)\ell, \quad i \in \boldsymbol{M}, \tag{9.1}$$

where $\mathcal{I}^i$ is the indicator function:

$$\mathcal{I}^i := \begin{cases} 0, & \mathcal{V}^i = \mathcal{S}, \\ 1, & \mathcal{V}^i = \mathcal{N}, \end{cases} \tag{9.2}$$

and $\ell > 0$ is the security investment incurred by $\mathbf{P}i$ only if it (or he) has chosen $\mathcal{S}$, i.e., $\mathcal{V}^i = \mathcal{S}$.

The plant of $\mathbf{P}i$ is modeled as the discrete-time stochastic linear system:

$$\begin{aligned} x_{t+1}^i &= Ax_t^i + \nu_t^i Bu_t^i + w_t^i \\ y_t^i &= \gamma_t^i Cx_t^i + v_t^i \end{aligned} \qquad t \in \mathbb{N}_0, \quad i \in \boldsymbol{M}, \tag{9.3}$$

where $x_t^i \in \mathbb{R}^d$ denotes the system state, $u_t^i \in \mathbb{R}^m$ the control input, $w_t^i \in \mathbb{R}^d$ the process noise, $y_t^i \in \mathbb{R}^p$ the measured output, $v_t^i \in \mathbb{R}^p$ the measurement noise, for $\mathbf{P}i$ at the $t-$th time step. The matrices $A \in \mathbb{R}^{d\times d}$, $B \in \mathbb{R}^{d\times m}$, $C \in \mathbb{R}^{p\times d}$ are given. We assume that $w_t^i$ (resp. $v_t^i$), for any $i \in \boldsymbol{M}$ and $t \in \mathbb{N}_0$, are independent and identically distributed (i.i.d.) Gaussian random vectors with mean 0 and covariance $Q \in \mathbb{R}^{d\times d}$ (resp. $R \in \mathbb{R}^{p\times p}$). The initial state $x_0^i$ is also Gaussian with mean $\bar{x} \in \mathbb{R}^d$ and covariance $\bar{P} \in \mathbb{R}^{d\times d}$. We assume uncorrelated $x_0^i$, $w_t^i$, and $v_t^i$. For a fixed $i \in \boldsymbol{M}$ and any $t \in \mathbb{N}_0$, the random variables $\gamma_t^i$ (resp. $\nu_t^i$) are i.i.d. Bernoulli with the failure probability $\tilde{\gamma}^i$ (resp. $\tilde{\nu}^i$), and model the packet loss in the sensor (resp. control) communication channel.

In contrast to the existing NCS literature (e.g., Imer et al. [2006],Schenato et al. [2007]), we assume that the failure probabilities $\tilde{\gamma}^i$ and $\tilde{\nu}^i$ are interdependent between the players due to the exposure to network induced insecurities. In order to reflect security interdependencies, in our model, the failure probabilities $\tilde{\gamma}^i$ and $\tilde{\nu}^i$ depend on the $\mathbf{P}i$'s own security choice $\mathcal{V}^i$ *and* on the other players' security choices $\{\mathcal{V}^j, j \neq i\}$ (chosen in the first stage), i.e.,

$$\mathsf{P}[\gamma_t^i = 0 \mid \mathcal{V}] = \tilde{\gamma}^i(\mathcal{V}), \quad \mathsf{P}[\nu_t^i = 0 \mid \mathcal{V}] = \tilde{\nu}^i(\mathcal{V}), \quad t \in \mathbb{N}_0,$$

where the failure probabilities $\tilde{\gamma}_i(\mathcal{V})$ and $\tilde{\nu}_i(\mathcal{V})$ for $\mathbf{P}i$ are introduced below by (9.9) and (9.10) in Sec. 9.2.2 for the case of $\mathrm{m} = 2$ and $\mathrm{m} > 2$ players, respectively.

In the *second stage*, each $\mathbf{P}i$ ($i \in \boldsymbol{M}$) chooses a control input sequence $\mathcal{U}^i := \{u_t^i,\, t \in \mathbb{N}_0\}$ for its plant based on the available information defined as[1]:

$$\zeta_t^i = \zeta_{t-1}^i \cup \left\{y_t^i, \nu_{t-1}^i, \gamma_t^i\right\}, \quad t \in \mathbb{N}, \tag{9.4}$$

with $\zeta_0^i = \{\mathcal{V}, y_0^i, \gamma_0^i\}$. The class of control policies considered here consist of the sequence of functions $\mu_0^i, \mu_1^i, \ldots$ such that each $\mu_t^i$ maps $\zeta_t^i$ into $\mathbb{R}^m$, i.e.,

$$u_t^i = \mu_t^i(\zeta_t^i), \quad t \in \mathbb{N}_0, \quad i = 1 \ldots \mathrm{m}. \tag{9.5}$$

Let $\mathcal{U}$ denote the set of player control input sequences:

$$\mathcal{U} := \{\mathcal{U}^1 \cup \cdots \cup \mathcal{U}^{\mathrm{m}}\}.$$

For given $\mathcal{V}$ and $\mathcal{U}$, the $\mathbf{P}i$'s second stage cost is given by the average Linear Quadratic Gaussian (LQG) cost:

$$J_{\mathrm{II}}^i(\mathcal{V},\mathcal{U}) := \limsup_{T\longrightarrow\infty} \frac{1}{T}\mathsf{E}\left[\sum_{t=0}^{T-1} {x_t^i}^\top G x_t^i + \nu_t^i {u_t^i}^\top H u_t^i\right], \tag{9.6}$$

[1] This information set corresponds to the packet acknowledgment behavior of TCP-like protocols (see Imer et al. [2006]).

where $G \geqslant 0$ (resp. $H > 0$) is a known matrix in $\mathbb{R}^{d\times d}$ (resp. $\mathbb{R}^{p\times p}$).

The objective of each $\mathbf{P}i$ is to minimize his total cost:

$$J^i(\mathcal{V},\mathcal{U}) = J^i_{\mathrm{I}}(\mathcal{V}) + J^i_{\mathrm{II}}(\mathcal{V},\mathcal{U}), \quad i \in \boldsymbol{M}, \tag{9.7}$$

where $J^i_{\mathrm{I}}(\mathcal{V})$ (resp. $J^i_{\mathrm{II}}(\mathcal{V},\mathcal{U})$) is given by (9.1) (resp. (9.6)). To summarize, in the first stage, each $\mathbf{P}i$ makes a security choice $\mathcal{V}^i$. In the subgame that starts after the first stage, each $\mathbf{P}i$ chooses the control input sequence $\mathcal{U}^i$ to minimize the average cost (9.6). The solution concept for the game is subgame perfect Nash equilibrium. Next, we introduce the baseline case of a social planner whose objective is to minimize the aggregate cost of all players:

$$J^{\mathrm{SO}}(\mathcal{V},\mathcal{U}) = \sum_{i=1}^{\mathrm{m}} J^i(\mathcal{V},\mathcal{U}). \tag{9.8}$$

### 9.2.2 Security Interdependence

For a two player game ($\mathrm{m} = 2$), we model the failure probabilities for $\mathbf{P}i$ as follows:

$$\begin{aligned} \tilde{\gamma}^i(\mathcal{V}) &= \mathcal{I}^i\bar{\gamma} + (1-\mathcal{I}^i\bar{\gamma})\alpha(\mathcal{I}^i,\mathcal{I}^{-i}),\\ \tilde{\nu}^i(\mathcal{V}) &= \underbrace{\mathcal{I}^i\bar{\nu}}_{\text{reliability}} + \underbrace{(1-\mathcal{I}^i\bar{\nu})\alpha(\mathcal{I}^i,\mathcal{I}^{-i})}_{\text{security}}, \end{aligned} \tag{9.9}$$

where the superscript $-i$ denotes the other player. In (9.9), the first term reflects the probability of a *direct* failure, and the second term reflects the probability of an *indirect* failure. The second term in (9.9) reflects player interdependence due to being networked and subjected to communication losses (for e.g., resulting from distributed denial-of-service (DDOS) attacks). We define the interdependence term $\alpha : \{0,1\}^2 \to ]0,1[$ as follows:

$$0 =: \alpha(0,0) = \alpha(1,0) < \alpha(0,1) := \underline{\alpha} < \alpha(1,1) := \bar{\alpha} < 1,$$

where $\bar{\alpha}$ is such that $\bar{\gamma} + (1-\bar{\gamma})\bar{\alpha} < 1$ and $\bar{\nu} + (1-\bar{\nu})\bar{\alpha} < 1$. Thus, we assume that, due to network interdependence, the probability of indirect failure increases when more players insecure. Here $\bar{\gamma}$ (resp. $\bar{\nu}$) is the failure probability of the sensor (resp. control) communication channel (identical for both players) when $\alpha(\mathcal{I}^i,\mathcal{I}^{-i}) = 0$, i.e., no interdependence. Then, the failure probabilities in our model coincide with the existing NCS literature Imer et al. [2006],Schenato et al. [2007].

We now extend (9.9) to $\mathrm{m} > 2$ players as follows:

$$\begin{aligned} \tilde{\gamma}^i(\mathcal{V}) &= \mathcal{I}^i\bar{\gamma} + (1-\mathcal{I}^i\bar{\gamma})\beta(\eta^i),\\ \tilde{\nu}^i(\mathcal{V}) &= \mathcal{I}^i\bar{\nu} + (1-\mathcal{I}^i\bar{\nu})\beta(\eta^i), \end{aligned} \tag{9.10}$$

where $\eta^i := \sum_{j\neq i}\mathcal{I}^j$ denotes the number of players (excluding $\mathbf{P}i$) who have chosen $\mathcal{N}$. As in the two-player case, we assume that the probability of indirect failure (the second term in (9.10)) increases when more players are insecure. To reflect this, we define the interdependence term $\beta : \{0,1,\ldots,\mathrm{m}-1\} \longrightarrow ]0,1[$ as follows:

$$0 =: \beta\left(0\right) < \cdots < \beta\left(\eta^i\right) < \cdots < \beta\left(\mathrm{m}-1\right) := \bar{\beta} < 1, \tag{9.11}$$

where $\bar{\gamma} + (1-\bar{\gamma})\bar{\beta} < 1$, and $\bar{\nu} + (1-\bar{\nu})\bar{\beta} < 1$. In contrast to (9.9), the interdependence for $\mathbf{P}i$ as defined in (9.10) does not depend his own choice of security investment. Notice that although we do not specifically model the interdependencies *between* the failure probabilities of sensor and control channels, they are still interdependent due to the second terms in (9.9) and (9.10).

### 9.2.3 Second Stage LQG Problem

For any fixed security choices $\mathcal{V}$, the problem of minimizing $\mathbf{P}i$'s expected second stage cost $J_{\mathrm{II}}^i(\mathcal{V},\mathcal{U}^i)$ over $u_t^i = \mu_t^i(\zeta_t^i)$ becomes an infinite horizon LQG problem defined by (9.3)–(9.6). Following Schenato et al. [2007][2], we assume that $(A, B)$ and $(A, Q^{1/2})$ are controllable, $(A, C)$ and $(A, G^{1/2})$ are observable, and the maximum failure probabilities are below "certain" thresholds, i.e., for (9.9):

$$\bar{\gamma} + (1-\bar{\gamma})\bar{\alpha} < \tilde{\gamma}_c, \quad \bar{\nu} + (1-\bar{\nu})\bar{\alpha} < \tilde{\nu}_c,$$

where $\tilde{\gamma}_c$ (resp. $\tilde{\nu}_c$) depends on $A$, $C$, $Q$, and $R$ (resp. $A$, $B$, $G$, and $H$); similarly for (9.10). In general, the minimum second stage cost cannot be analytically expressed; however, Theorem 5.6 of Schenato et al. [2007] provides analytical expressions for the upper and lower bounds of this cost. To simplify the exposition, we restrict our attention to the case of invertible $C$ and $R = 0$, which allows us to analytically express the minimum cost[3]:

$$\begin{aligned} J_{\mathrm{II}}^{i*}(\mathcal{V}) := \min_{\mathcal{U}^i \ni u_t^i = \mu_t^i(\zeta_t^i)} & J_{\mathrm{II}}^i(\mathcal{V},\mathcal{U}) = \mathrm{tr}(S^i(\mathcal{V})Q) \\ & + \tilde{\gamma}^i(\mathcal{V})\, \mathrm{tr}\left((A^\top S^i(\mathcal{V})A + G - S^i(\mathcal{V}))\underline{P}^i(\mathcal{V})\right), \end{aligned} \tag{9.12}$$

where the matrices $S^i(\mathcal{V})$ and $P^i(\mathcal{V})$ are the respective positive definite solutions of the following equations:

$$\begin{aligned} S^i(\mathcal{V}) =& A^\top S^i(\mathcal{V})A + G - (1 - \tilde{\nu}^i(\mathcal{V})) \\ & \times A^\top S^i(\mathcal{V})B(B^\top S^i(\mathcal{V})B + H)^{-1}B^\top S^i(\mathcal{V})A, \\ P^i(\mathcal{V}) =& \tilde{\gamma}^i(\mathcal{V})AP^i(\mathcal{V})A^\top + Q. \end{aligned} \tag{9.13}$$

The following lemma provides that $J_{\mathrm{II}}^{i*}(\mathcal{V})$ decreases in failure probabilities:

**Lemma 9.2.1.** *Let $\tilde{\gamma}^i(\mathcal{V}^1) < \tilde{\gamma}^i(\mathcal{V}^2)$ and $\tilde{\nu}^i(\mathcal{V}^1) < \tilde{\nu}^i(\mathcal{V}^2)$. Then, $J_{\mathrm{II}}^{i*}(\mathcal{V}^1) < J_{\mathrm{II}}^{i*}(\mathcal{V}^2)$.*

*Proof.* From (9.13) $S^i$ and $P^i$, are increasing with $\tilde{\nu}^i$ and $\tilde{\gamma}^i$ respectively. The proof follows from (9.12). □

*Remark* 9.2.2. From (9.9) and (9.10), when $\mathbf{P}i$ invests in security, the probability of direct failure is reduced to 0. However, our results easily extend to cases when $\mathbf{P}i$'s investment in security reduces this probability to a non-zero value.

[2] In Schenato et al. [2007], these expressions are given for the arrival probabilities $1-\tilde{\gamma}_i$ and $1-\tilde{\nu}_i$, while we work with $\tilde{\gamma}_i$ and $\tilde{\nu}_i$.

[3] In a general case, the minimum $J_{\mathrm{II}}^i(\mathcal{V},\mathcal{U}^i)$ can be obtained via Monte-Carlo simulations.

*Example* 9.2.3. Consider (9.3) for the scalar setting with $d = 1$, $B = 1$, $C = 1$. Then $Q, R, G, H$ are scalars. For $|A| > 1$, $\tilde{\gamma}_c = \tilde{\nu}_c = A^{-2}$. Following Schenato et al. [2007], the upper and lower bounds for $J^{i*}_{\mathrm{II}}(\mathcal{V})$ are given by:

$$\begin{aligned} \bar{J}^{i*}_{\mathrm{II}}(\mathcal{V}) &= QS^i(\mathcal{V}) + \frac{\bar{P}^i(\mathcal{V})T^i(\mathcal{V})\left(\bar{P}^i(\mathcal{V})\tilde{\gamma}^i(\mathcal{V}) + R\right)}{\bar{P}^i(\mathcal{V}) + R} \\ \underline{J}^{i*}_{\mathrm{II}}(\mathcal{V}) &= QS^i(\mathcal{V}) + \tilde{\gamma}^i(\mathcal{V})\underline{P}^i(\mathcal{V})T^i(\mathcal{V}) \end{aligned} \tag{9.14}$$

where

$$S^i(\mathcal{V}) = \frac{(A^2H + G - H) + \sqrt{(A^2H + G - H)^2 + 4GH(1 - A^2\tilde{\nu}^i(\mathcal{V}))}}{2(1 - A^2\tilde{\nu}^i(\mathcal{V}))},$$

and

$$\bar{P}^i(\mathcal{V}) = \frac{(A^2R + Q - R) + \sqrt{(A^2R + Q - R)^2 + 4QR(1 - A^2\tilde{\gamma}^i(\mathcal{V}))}}{2(1 - A^2\tilde{\gamma}^i(\mathcal{V}))},$$

$$\underline{P}^i(\mathcal{V}) = \frac{Q}{1 - A^2\tilde{\gamma}^i(\mathcal{V})},$$

and $T^i(\mathcal{V}) = ((A^2 - 1)S^i(\mathcal{V}) + G)$. Notice that $\bar{J}^{i*}_{\mathrm{II}}(\mathcal{V}) = \underline{J}^{i*}_{\mathrm{II}}(\mathcal{V})$ if $R = 0$. $\Delta$

## 9.3 Equilibria for two player game

Consider a 2−player game, where the interdependent failure probabilities are given by (9.9). For any fixed security choices $\mathcal{V}$, each **P**$i$'s minimum expected cost in the second stage $J^{i*}_{\mathrm{II}}(\mathcal{V})$ is given by (9.12)–(9.13). Following (9.7), the player objectives for the second stage subgame are presented in Fig. 9.2(top). Following (9.8), the social planner objectives are presented in Fig. 9.2(bottom). To derive optimal player actions in the first stage (security choices $\mathcal{V}^i$), we will distinguish the following two cases:

| | | **P2** | |
|---|---|---|---|
| | | $\mathcal{S}$ | $\mathcal{N}$ |
| **P1** | $\mathcal{S}$ | $J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{S}\}) + \ell$, $J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{S}\}) + \ell$ | $J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{N}\}) + \ell$, $J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{S}\})$ |
| | $\mathcal{N}$ | $J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{S}\})$, $J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{N}\}) + \ell$ | $J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{N}\})$, $J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{N}\})$ |

| | $\mathcal{S}$ | $\mathcal{N}$ |
|---|---|---|
| $\mathcal{S}$ | $2(J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{S}\}) + \ell)$ | $J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{N}\}) + J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{S}\}) + \ell$ |
| $\mathcal{N}$ | $J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{N}\}) + J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{S}\}) + \ell$ | $2J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{N}\})$ |

Figure 9.2: Objectives: 2−player game (top) & social planner (bottom).

$$J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{N}\}) - J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{N}\}) \leqslant J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{S}\}) - J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{S}\}), \tag{9.15}$$

$$J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{S}\}) - J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{S}\}) \leqslant J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{N}\}) - J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{N}\}). \tag{9.16}$$

If (9.15) holds and a player invests in security, other player gain from investing in security *increases.* However, if (9.16) holds, each player decision to secure *decreases* the other player gain from investing in security. In Sections 9.3.1 and 9.3.2, we present equilibria for different $\ell$, and compare with social optima.

### 9.3.1 Increasing incentives

Let (9.15) hold, and let us define

$$\underline{\ell}_1 := J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{N}\}) - J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{N}\}),\ \bar{\ell}_1 := J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{S}\}) - J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{S}\}).$$

From Fig. 9.2(top), we infer that if $\ell < \underline{\ell}_1$ (resp. $\ell > \bar{\ell}_1$), $\{\mathcal{S},\mathcal{S}\}$ (resp. $\{\mathcal{N},\mathcal{N}\}$) is unique Nash equilibrium. Thus, $\underline{\ell}_1$ (resp. $\bar{\ell}_1$) is the cut-off cost below (resp. above) which both players invest (resp. neither player invests) in security. However, if $\underline{\ell}_1 \leqslant \ell \leqslant \bar{\ell}_1$, both $\{\mathcal{S},\mathcal{S}\}$ and $\{\mathcal{N},\mathcal{N}\}$ are individually optimal. From Fig. 9.2(bottom), if $\ell \leqslant \ell_1^{\mathrm{SO}}$, the socially optimum choices are $\{\mathcal{S},\mathcal{S}\}$ with

$$\ell_1^{\mathrm{SO}} := J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{N}\}) - J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{S}\}). \tag{9.17}$$

For $\ell$ in the range $\bar{\ell}_1 \leqslant \ell \leqslant \ell_1^{\mathrm{SO}}$, individually optimal choices are $\{\mathcal{N},\mathcal{N}\}$, while the socially optimal choices are still $\{\mathcal{S},\mathcal{S}\}$. If $\ell \geqslant \ell_1^{\mathrm{SO}}$, the individually and socially optimal choices coincide at $\{\mathcal{N},\mathcal{N}\}$. Case 1 of Fig. 9.3 summarizes pure strategy equilibria for different $\ell$.

For $\ell$ in the range $\underline{\ell}_1 \leqslant \ell \leqslant \bar{\ell}_1$, a mixed strategy equilibrium exists. Let $\theta_1^i$ (resp. $(1-\theta_1^i)$) denote the mixing probability with which $\mathbf{P}i$ chooses $\mathcal{S}$ (resp. $\mathcal{N}$). Then, $\mathbf{P}1$'s mixing probability $\theta_1^1$ is such that the $\mathbf{P}2$'s expected costs for both choices $\mathcal{S}$ or $\mathcal{N}$ are equal, i.e.,

$$\begin{aligned}\theta_1^1\left[J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{S}\})+\ell\right] &+ (1-\theta_1^1)\left[J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{N}\})+\ell\right]\\ &= \theta_1^1 J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{S}\}) + (1-\theta_1^1)J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{N}\}).\end{aligned}$$

Simplifying the above equation, we obtain

$$\theta_1^1 = \frac{\ell-\underline{\ell}_1}{\bar{\ell}_1-\underline{\ell}_1},\ \text{for } \ell \in (\underline{\ell}_1,\bar{\ell}_1).$$

By writing a similar equation for $\mathbf{P}1$, it is easy to check that $\theta_1^2 = \theta_1^1$. Thus, mixed equilibrium is symmetric.

### 9.3.2 Decreasing incentives

Let (9.16) hold, and let us define

$$\underline{\ell}_2 := J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{S}\}) - J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{S}\}),\ \bar{\ell}_2 := J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{N}\}) - J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{N}\}).$$

Using Fig. 9.2(top), we infer that if $\ell < \underline{\ell}_2$ (resp. $\ell > \bar{\ell}_2$) then $\{\mathcal{S},\mathcal{S}\}$ (resp. $\{\mathcal{N},\mathcal{N}\}$) is unique Nash equilibrium. However, if $\underline{\ell}_2 \leqslant \ell \leqslant \bar{\ell}_2$, both $\{\mathcal{S},\mathcal{N}\}$ and $\{\mathcal{N},\mathcal{S}\}$ are individually

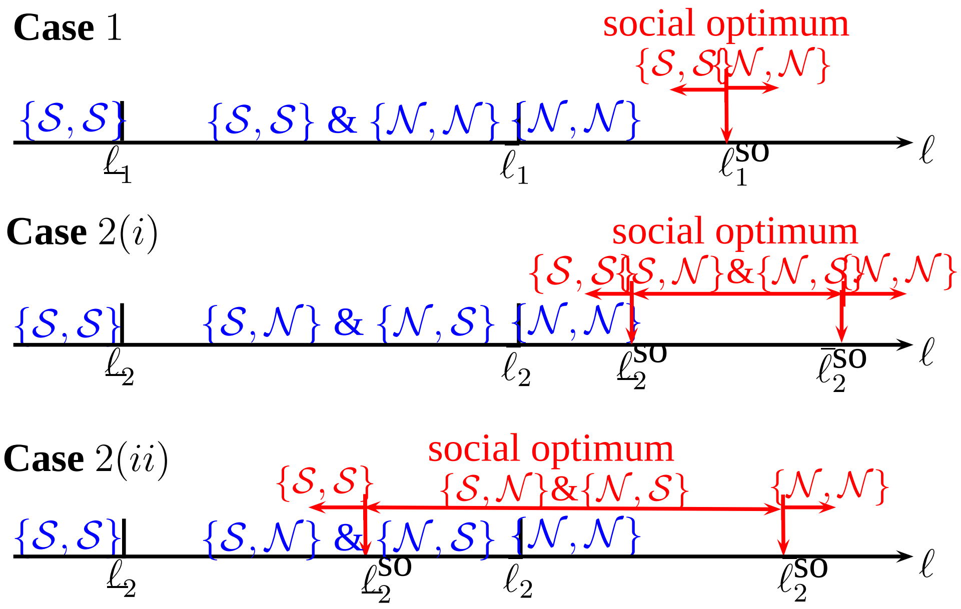

Case 1
social optimum
$\{\mathcal{S},\mathcal{S}\}$ $\{\mathcal{N},\mathcal{N}\}$
$\{\mathcal{S},\mathcal{S}\}$
$\{\mathcal{S},\mathcal{S}\}$ & $\{\mathcal{N},\mathcal{N}\}$
$\{\mathcal{N},\mathcal{N}\}$
$\underline{\ell}_1$
$\bar{\ell}_1$
$\ell_1^{\text{SO}}$
$\ell$
Case $2(i)$
social optimum
$\{\mathcal{S},\mathcal{S}\}$ $\{\mathcal{S},\mathcal{N}\}$&$\{\mathcal{N},\mathcal{S}\}$ $\{\mathcal{N},\mathcal{N}\}$
$\{\mathcal{S},\mathcal{S}\}$
$\{\mathcal{S},\mathcal{N}\}$ & $\{\mathcal{N},\mathcal{S}\}$
$\{\mathcal{N},\mathcal{N}\}$
$\underline{\ell}_2$
$\bar{\ell}_2$
$\underline{\ell}_2^{\text{SO}}$
$\bar{\ell}_2^{\text{SO}}$
$\ell$
Case $2(ii)$
social optimum
$\{\mathcal{S},\mathcal{S}\}$
$\{\mathcal{S},\mathcal{N}\}$&$\{\mathcal{N},\mathcal{S}\}$
$\{\mathcal{N},\mathcal{N}\}$
$\{\mathcal{S},\mathcal{S}\}$
$\{\mathcal{S},\mathcal{N}\}$ & $\{\mathcal{N},\mathcal{S}\}$
$\{\mathcal{N},\mathcal{N}\}$
$\underline{\ell}_2$
$\underline{\ell}_2^{\text{SO}}$
$\bar{\ell}_2$
$\bar{\ell}_2^{\text{SO}}$
$\ell$

Figure 9.3: Nash equilibria and social optima for different $\ell$.

optimal. From Fig. 9.2(bottom), if $\ell < \underline{\ell}_2^{\text{SO}}$ (resp.$\ell > \bar{\ell}_2^{\text{SO}}$), the socially optimum choices are $\{\mathcal{S},\mathcal{S}\}$ (resp. $\{\mathcal{N},\mathcal{N}\}$) with

$$\begin{aligned}\underline{\ell}_2^{\text{SO}} &:= J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{S}\}) + J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{N}\}) - 2J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{S}\}),\\ \bar{\ell}_2^{\text{SO}} &:= 2J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{N}\}) - J^*_{\mathrm{II}}(\{\mathcal{S},\mathcal{N}\}) - J^*_{\mathrm{II}}(\{\mathcal{N},\mathcal{S}\}).\end{aligned} \tag{9.18}$$

Note that $\underline{\ell}_2^{\text{SO}}$ can be either above or below $\bar{\ell}_2$. If $\underline{\ell}_2^{\text{SO}} \leqslant \ell \leqslant \bar{\ell}_2^{\text{SO}}$, both $\{\mathcal{S},\mathcal{N}\}$ and $\{\mathcal{N},\mathcal{S}\}$ are socially optimum choices. Case $2(i)$ (resp. Case $2(ii)$) of Fig. 9.3 summarizes the pure strategy equilibria for different $\ell$ when $\bar{\ell}_2 < \underline{\ell}_2^{\text{SO}}$ (resp. $\bar{\ell}_2 > \underline{\ell}_2^{\text{SO}}$).

Finally, a symmetric mixed strategy equilibrium exists for $\ell$ in the range $\underline{\ell}_2 \leqslant \ell \leqslant \bar{\ell}_2$ where each player invests in security with probability:

$$\theta_2^1 = \theta_2^2 = \frac{\bar{\ell}_2 - \ell}{\bar{\ell}_2 - \underline{\ell}_2}, \text{ for } \ell \in (\underline{\ell}_2, \bar{\ell}_2).$$

We now provide an example system for each case of Fig 9.3.

*Example* 9.3.1. **Case** 1. Let $A = 0.80$, $G = Q = H = R = 1$, and $\bar{\gamma} = \bar{\nu} = \underline{\alpha} = \bar{\alpha} = 0.1$. From (9.14), this system satisfies (9.15). **Case** $2(i)$. Let $A = 1.2$, $G = H = Q = R = 1$, $\bar{\gamma} = \bar{\nu} = 0.1$, $\underline{\alpha} = \bar{\alpha} = 0.25$. This system satisfies (9.16) and $\bar{\ell}_2 < \underline{\ell}_2^{\text{SO}}$. **Case** $2(ii)$. Let $\bar{\gamma} = \bar{\nu} = 0.25$ and all other parameters be as in Case $2(i)$. This system satisfies (9.16) and $\bar{\ell}_2 > \underline{\ell}_2^{\text{SO}}$. $\Delta$

### 9.3.3 Penalties for insecure players

In both increasing and decreasing incentive cases for the 2−player games of Sections 9.3.1 and 9.3.2, the individual and socially optimal security choices differ for a range of security costs. From Fig. 9.3, we observe that players tend to under-invest in security relative to the social planner. This reflects the presence of negative externalities. We suggest an instrument (penalty) to alter individually optimal security choices and make them coincide with the socially optimum ones. Let $\mathscr{F}$ denote the penalty imposed on the players who do not invest in security. In the game with penalties, when $\mathbf{P}i$ chooses $\mathcal{S}$ (resp. $\mathcal{N}$), the cost of $\mathbf{P}-i$ when he chooses $\mathcal{N}$ is $J^*_{\mathrm{I\!I}}(\{\mathcal{N},\mathcal{S}\})+\mathscr{F}$ (resp. $J^*_{\mathrm{I\!I}}(\{\mathcal{N},\mathcal{N}\})+\mathscr{F}$). We now show that a range of penalties can be computed such that the individually optimum choices in the game with penalties coincide with the social optimum ones.

With (9.15) imposed, the individual and socially optimal choices coincide if the penalties $\mathscr{F}_1$ for the corresponding game satisfy:

$$\ell_1^{\mathrm{SO}} + J^*_{\mathrm{I\!I}}(\{\mathcal{S},\mathcal{S}\}) \leqslant \mathscr{F}_1 + J^*_{\mathrm{I\!I}}(\{\mathcal{N},\mathcal{S}\}), \tag{9.19}$$

and

$$J^*_{\mathrm{I\!I}}(\{\mathcal{N},\mathcal{N}\}) + \mathscr{F}_1 \leqslant J^*_{\mathrm{I\!I}}(\{\mathcal{S},\mathcal{N}\}) + \ell_1^{\mathrm{SO}}. \tag{9.20}$$

From (9.19) and (9.20), and using (9.17), we obtain:

$$J^*_{\mathrm{I\!I}}(\{\mathcal{N},\mathcal{N}\}) - J^*_{\mathrm{I\!I}}(\{\mathcal{N},\mathcal{S}\}) \leqslant \mathscr{F}_1 \leqslant J^*_{\mathrm{I\!I}}(\{\mathcal{S},\mathcal{N}\}) - J^*_{\mathrm{I\!I}}(\{\mathcal{S},\mathcal{S}\}).$$

Similarly, with (9.16) imposed, the individual and socially optimal choices coincide if the penalties $\mathscr{F}_2$ for the corresponding game satisfy:

$$\underline{\ell}_2^{\mathrm{SO}} + J^*_{\mathrm{I\!I}}(\{\mathcal{S},\mathcal{S}\}) \leqslant \mathscr{F}_2 + J^*_{\mathrm{I\!I}}(\{\mathcal{N},\mathcal{S}\}), \tag{9.21}$$

and

$$J^*_{\mathrm{I\!I}}(\{\mathcal{N},\mathcal{N}\}) + \mathscr{F}_2 \leqslant J^*_{\mathrm{I\!I}}(\{\mathcal{S},\mathcal{N}\}) + \overline{\ell}_2^{\mathrm{SO}}. \tag{9.22}$$

From (9.21) and (9.22), and using (9.18), we obtain:

$$J^*_{\mathrm{I\!I}}(\{\mathcal{S},\mathcal{N}\}) - J^*_{\mathrm{I\!I}}(\{\mathcal{S},\mathcal{S}\}) \leqslant \mathscr{F}_2 \leqslant J^*_{\mathrm{I\!I}}(\{\mathcal{N},\mathcal{N}\}) - J^*_{\mathrm{I\!I}}(\{\mathcal{N},\mathcal{S}\}).$$

## 9.4 Equilibria for M player game

We now extend the analysis of Section 9.3 to m−player games (m > 2), where the interdependent failure probabilities are given by (9.10). Consider the $\mathbf{P}i$'s security choice of $\mathcal{S}$ or $\mathcal{N}$, and let the security choices of all other players be fixed. Recall from Sec. 9.2.2 that $\eta^i$ denotes the number of players (excluding $\mathbf{P}i$) who have chosen $\mathcal{N}$. To simplify the notation, we will henceforth omit the superscript $i$. Let $\eta$ other players be insecure. Without loss of generality, we assume that $\mathbf{P}1,\dots,\mathbf{P}(i-1)$ (resp. $\mathbf{P}(i+1),\dots,\mathbf{P}\mathrm{m}$) have chosen $\mathcal{S}$ (resp. $\mathcal{N}$), where $i=\mathrm{m}-\eta$. We use the following simplifying notation:

$$\langle\mathcal{S},\eta\rangle := \left\{\mathcal{V}^1,\dots,\mathcal{V}^{\mathrm{m}} \middle| \mathcal{V}^i = \mathcal{S}, \sum_{-i}\mathcal{I}^{-i} = \eta\right\},$$

$$\langle\mathcal{N},\eta\rangle := \left\{\mathcal{V}^1,\dots,\mathcal{V}^{\mathrm{m}} \middle| \mathcal{V}^i = \mathcal{N}, \sum_{-i}\mathcal{I}^{-i} = \eta\right\},$$

where $\mathcal{I}^j$ is the indicator function defined by (9.2). Let $\Delta(\eta)$ denote the gain of a player from investing in security when $\eta$ other players are insecure, i.e.,

$$\Delta(\eta) \coloneqq J^*_{\mathrm{II}}(\langle\mathcal{N},\eta\rangle) - J^*_{\mathrm{II}}(\langle\mathcal{S},\eta\rangle),\ \eta \in \{0,\dots,\mathrm{m}-1\}. \tag{9.23}$$

To derive optimal player security choices $\mathcal{V}^i$, we will distinguish the following two cases (which generalize the increasing and decreasing incentive cases for the 2−player games of Sections 9.3.1 and 9.3.2):

$$\Delta(\eta) \leqslant \Delta(\eta-1), \quad \text{for all } \eta \in \{1,2,\dots,\mathrm{m}-1\}, \tag{9.24}$$

and

$$\Delta(\eta) \geqslant \Delta(\eta-1), \quad \text{for all } \eta \in \{1,2,\dots,\mathrm{m}-1\}. \tag{9.25}$$

Thus, similar to (9.15), (9.24) corresponds to the case when the decision of an extra player to invest in security *increases* other players' gains from investing in security. Also, similar to (9.16), (9.25) corresponds to the case when player gain from investing in security *decreases* as more players invest in security.

### 9.4.1 Increasing incentives

Analogous to Section 9.3.1, we have:

**Theorem 9.4.1.** *In the* m *player game* (m > 2) *with* (9.24) *imposed, a pure strategy equilibrium exists, and is symmetric. Depending on the magnitude of $\ell \in \mathbb{R}_+$, the equilibrium is*

$$\begin{array}{ll} \{\mathcal{S},\dots,\mathcal{S}\} & \textit{if } \ell < \ell_1^{\mathrm{m}-1} \\ \{\mathcal{N},\dots,\mathcal{N}\} & \textit{if } \ell > \ell_1^0 \\ \{\mathcal{S},\dots,\mathcal{S}\} \textit{ or } \{\mathcal{N},\dots,\mathcal{N}\} & \textit{if } \ell_1^{\mathrm{m}-1} \leqslant \ell \leqslant \ell_1^0 \end{array} \tag{9.26}$$

*where $\ell_1^{\mathrm{m}-1} \coloneqq \Delta(\mathrm{m}-1)$ and $\ell_1^0 \coloneqq \Delta(0)$.*

*Proof.* First, with (9.24) imposed, the existence of symmetric pure strategy Nash equilibrium (9.26) follows from adopting the construction of Section 9.3.1. Indeed, if $\ell < \ell_1^{\mathrm{m}-1} \leqslant \Delta(\eta)$ for all $\eta \in \{0,\dots,\mathrm{m}-2\}$ (resp. $\ell > \ell_1^0 \geqslant \Delta(\eta)$ for all $\eta \in \{1,\dots,\mathrm{m}-1\}$), each $\mathbf{P}i$'s dominant strategy is $\mathcal{S}$ (resp. $\mathcal{N}$). Thus, $\{\mathcal{S},\dots,\mathcal{S}\}$ (resp. $\{\mathcal{N},\dots,\mathcal{N}\}$) is unique Nash equilibrium. If $\ell \leqslant \ell_1^0$ (resp. $\ell \geqslant \ell_1^{\mathrm{m}-1}$), $\{\mathcal{S},\dots,\mathcal{S}\}$ (resp. $\{\mathcal{N},\dots,\mathcal{N}\}$) is a Nash equilibrium. Hence, if $\ell$ is in the range $\ell_1^{\mathrm{m}-1} \leqslant \ell \leqslant \ell_1^0$, both $\{\mathcal{S},\dots,\mathcal{S}\}$ and $\{\mathcal{N},\dots,\mathcal{N}\}$ are equilibria.

Second, we show that no asymmetric equilibrium exists. Assume on the contrary that $\{\mathcal{S},\dots,\mathcal{S},\underbrace{\mathcal{N}\dots,\mathcal{N}}_{\mathrm{m}_1 \text{ players}}\}$ is an equilibrium, i.e., when $\mathbf{P}1,\dots,\mathbf{P}(\mathrm{m}-\mathrm{m}_1)$ invest in security and $\mathbf{P}\mathrm{m}_1,\dots,\mathbf{P}\mathrm{m}$ do not. For $\mathbf{P}(\mathrm{m}-\mathrm{m}_1+1)$,

$$\Delta(\mathrm{m}_1-1) < \ell, \tag{9.27}$$

and for $\mathbf{P}(\mathrm{m}-\mathrm{m}_1)$,

$$\ell < \Delta(\mathrm{m}_1). \tag{9.28}$$

Combining inequalities (9.27) and (9.28), we obtain

$$\Delta(\mathrm{m}_1 - 1) < \Delta(\mathrm{m}_1),$$

which contradicts (9.24) for $\eta = \mathrm{m}_1$. The same contradiction can be shown for any other asymmetric equilibrium. Thus, no asymmetric equilibrium exists. □

### 9.4.2 Decreasing incentives

Analogous to Section 9.3.2, we have:

**Theorem 9.4.2.** *In the game of* m *players with* (9.25) *imposed, a pure strategy equilibrium exists. Depending on the magnitude of* $\ell \in \mathbb{R}_+$*, equilibrium is:*

$$\left\{\mathcal{V}^1, \dots, \mathcal{V}^{\mathrm{m}} \middle| \mathcal{V}^i \in \{\mathcal{S}, \mathcal{N}\}, \sum_{i=1}^{\mathrm{m}} \mathcal{I}^i = \eta \right\},$$

*where*

$$\eta = \begin{cases} 0 & \text{if } \ell \leqslant \ell_2^0 \\ \mathrm{m} & \text{if } \ell \geqslant \ell_2^{\mathrm{m}-1} \\ k & \text{if } \ell_2^{k-1} \leqslant \ell \leqslant \ell_2^k, \quad k \in \{1, \dots, \mathrm{m}-1\}, \end{cases} \tag{9.29}$$

*and* $\ell_2^j := \Delta(j),\ j \in \{0, \dots, \mathrm{m}-1\}$.

*Proof.* If $\ell \leqslant \ell_1^0$ (resp. $\ell \geqslant \ell_1^{\mathrm{m}-1}$), all players invest (resp. no player invests) in security, and $\{\mathcal{S}, \dots, \mathcal{S}\}$ (resp. $\{\mathcal{N}, \dots, \mathcal{N}\}$) is an equilibrium. However, if $\ell$ is in the range $\ell_2^{\eta-1} \leqslant \ell \leqslant \ell_2^{\eta}$, an asymmetric equilibrium exists where $(\mathrm{m}-\eta)$ players choose $\mathcal{S}$ and $\eta$ players choose $\mathcal{N}$, i.e., $\{\mathcal{S}, \dots, \mathcal{S}, \underbrace{\mathcal{N} \dots, \mathcal{N}}_{\eta \text{ players}}\}$ is an equilibrium. The uniqueness of the equilibrium follows by construction.

□

Theorem 9.4.1 (resp. Theorem 9.4.2) characterizes the pure strategy Nash equilibria for the case of increasing (resp. decreasing) incentives. Note that m player game in both cases is symmetric, and a symmetric mixed equilibrium can be computed as follows: Any equilibrium mixing probability $\theta$ is such that any $\mathbf{P}i$'s expected costs for both choices $\mathcal{S}$ or $\mathcal{N}$ are equal. Player expected costs for choosing $\mathcal{S}$ is

$$\begin{aligned} &\theta^{\mathrm{m}-1}\left(J_{\mathrm{II}}^*(\langle \mathcal{S}, 0\rangle) + \ell\right) + \binom{\mathrm{m}}{1}\theta^{\mathrm{m}-2}(1-\theta)\left(J_{\mathrm{II}}^*(\langle \mathcal{S}, 1\rangle) + \ell\right) \\ &+ \dots + \binom{\mathrm{m}}{\mathrm{m}-1}\theta(1-\theta)^{\mathrm{m}-2}\left(J_{\mathrm{II}}^*(\langle \mathcal{S}, \mathrm{m}-2\rangle) + \ell\right) \\ &+ (1-\theta)^{\mathrm{m}-1}\left(J_{\mathrm{II}}^*(\langle \mathcal{S}, \mathrm{m}-1\rangle) + \ell\right) \end{aligned}$$

Similarly, player expected cost for choosing $\mathcal{N}$ is

$$\begin{aligned}&\theta^{\mathrm{m}-1}\left(J^*_{\mathrm{II}}(\langle\mathcal{N},0\rangle)\right)+\binom{\mathrm{m}}{1}\theta^{\mathrm{m}-2}(1-\theta)\left(J^*_{\mathrm{II}}(\langle\mathcal{N},1\rangle)\right)\\&+\cdots+\binom{\mathrm{m}}{\mathrm{m}-1}\theta(1-\theta)^{\mathrm{m}-2}\left(J^*_{\mathrm{II}}(\langle\mathcal{N},\mathrm{m}-2\rangle)\right)\\&+(1-\theta)^{\mathrm{m}-1}\left(J^*_{\mathrm{II}}(\langle\mathcal{N},\mathrm{m}-1\rangle)\right)\end{aligned}$$

Equating the above expressions and noting (9.23), we conclude that mixing probability $\theta \in (0,1)$ is a solution of the following polynomial:

$$\sum_{j=0}^{\mathrm{m}-1}\binom{\mathrm{m}-1}{j}(\ell-\Delta(j))\times\theta^{\mathrm{m}-1-j}(1-\theta)^j=0 \qquad (9.30)$$

## 9.5 Discussion and Concluding Remarks

In this paper, we investigated the incentives to invest in security for players which operate interdependent and identical NCS. We presented a new model of interdependent NCS, where the players' failure probabilities are dependent on the security investments of other players. In such cases, the externalities are present.

We hope that our findings are relevant for analyzing the effects of DDoS attacks on NCS governing the critical infrastructures, for e.g., the next generation electric power grid. It is well accepted that in the future grid, a large number of commodity IT solutions will be deployed. A wider deployment of smart devices is likely to result in a higher number of players (higher $M$), a higher degree of interdependence between the players (a higher second terms in (9.11)), and also a higher security cost $\ell$ due to the increased configuration (and overall system) complexity. Thus, we expect that with the NCS becoming increasingly "smarter", the magnitude of negative externalities, and therefore the gap between the individually and socially optimal outcomes will only widen.

Such underinvestment in the presence of interdependencies raises the possibility of major breakdowns, see Bier et al. [2007], which would create losses (due to higher costs) far beyond the NCS losses considered in this paper. Our model does not incorporate these extra loses, which makes our estimates of security investments, including the socially optimal ones, rather conservative.

# Chapter 10

# Conclusion and Future Plans

The extensive use of Information and Communications Technologies (ICT) raises concerns about the vulnerabilities of the nation's critical infrastructures to security attacks. Cyber-security of Distributed Control Systems (DCS) and Supervisory Data Control and Data Acquisition (SCADA) systems is especially important, because they are used for sensing and control of large physical infrastructures. The use of homogeneous commercial off-the-shelf ICT components makes control systems subject to correlated software and hardware malfunctions. Consequently, DCS/SCADA inherit the vulnerabilities of ICT components (for e.g., the Stuxnet worm), making these systems both safety- and security-critical. The main focus of this thesis is on the design of reliable and secure control for infrastructure systems. The work presented in this thesis can be of particular interest to the next generation SCADA and other networked embedded systems for power grid, water and gas distribution, highway and air transportation, cyber-enabled energy management systems (for e.g., hybrid electric vehicles and buildings equipped with smart meters). The existing research framework in robust and fault-tolerant Networked Control Systems (NCS) does not account for cyber-attacks on SCADA communications. In this thesis, we have developed the basis of a comprehensive theoretical framework and proposed several practical tools for building resilient NCS, with the focus on reliability and security in the presence of ICT vulnerabilities. The following issues will need to be addressed in future:

1. Analysis of attacks on NCS, and attacker interactions with the physical dynamics.
2. Development of a framework to jointly analyze security and reliability failures.
3. Design of resilient control methods and incentive mechanisms to reduce global risks.
4. Testing and evaluation of attack diagnosis and control algorithms.

## 10.1 Proposed future technical approach

Security requirements are traditionally evaluated in three dimensions: confidentiality, integrity, and availability. For the purpose of this research, confidentiality is relatively less important. The availability and integrity are profound concerns for many critical infrastructures. The necessity to satisfy the real-time constraints imposes limitations on the possible defenses against denial-of-service (DoS) attacks. Another challenge in this area is dealing with compromises of unattended nodes deployed in the field; a reliable operation of critical infrastructures must be ensured even if the adversary controls a subset of devices. In addition, wider deployment of distributed sensors networks increases the opportunities for privacy abuses. In many cases, the users are not fully aware of the large number of details which could be inferred from their usage of infrastructure resources. Indeed, information might be revealed indirectly (for instance, the increase in power usage indicates the presence of building's occupants). The operating systems of upcoming infrastructures such as smart buildings and smart structures portend immense data collection in places routinely occupied by individuals. Aggregating privacy-sensitive data permits to uncover the individuals' behavioral patterns from their usage and exploit this information. There is a unique opportunity to improve privacy concerns by considering them early in the security design, and by incorporating them into concurrently developed policy and consumer protection tools.

### 10.1.1 Attack Models and Threat Assessment

Attacks on NCS are aimed at causing degraded closed-loop performance, and may lead to other undesirable effects for system's safety and stability (the word *crash* is literal for control systems). To achieve these goals, an adversary can disrupt control systems by carrying out deception attacks: manipulating set-points, tuning control parameters and/or sensor readings. The DoS attacks increase communication latencies, which could disrupt the communications between NCS components. Several countermeasures have been proposed in the past for the cyber-security of wireless sensor networks. However, the key drawback of these counter-measures is that they do not address security vulnerabilities which are unique to closing-the-loop around wireless sensor networks. In this thesis, a taxonomy of attacks on SCADA systems has been developed, which includes attributes such as (i) mode of attack (availability, integrity, confidentiality); (ii) signature (targeted, resource constrained, random); and (iii) time of attack (surge, bias, geometric).

### 10.1.2 Diagnostic Methods for Stealthy Attacks

In this thesis, we have also developed diagnostic methods for distinguishing whether control variables are manipulated by attackers, and studied the consequences of attacks for the underlying physical system dynamics. Successful attacks may change the master node's perception of the environment, modifying the semantics of the information. Traditional intrusion detection systems were not designed for NCS environments with ICT, and thus perform poorly against cyber-attacks. Both sensing and control data use ICT, and therefore are subject to DoS and integrity attacks. Future work will involve work on a class of model-based detection schemes, formulated as an adversarial game between the detection system and the attacker. For a desired false alarm rate, the detection system's objective is to maximize the detection probability; while the attacker's objective is to minimize this probability via intelligent manipulation of the data that the compromised NCS components send. The creation of a reference model which captures the misbehavior of NCS, and can be used as a baseline for evaluating both existing and new control strategies, will be beneficial for future research.

Several implementable secure control tools were developed in this thesis. The thesis presents the cyber-security analysis of the SCADA system of the Gignac water distribution network in Southern France. We conducted a field experiment to demonstrate considerable losses from hacking water level sensors. To improve the system's resilience, an observer-based diagnostic method, in which each observer estimates the state of a reduced-order flow model, was introduced. The European Task Committee on canal automation has recognized this research as an important contribution to risk assessment and attack diagnosis of water SCADA systems. The thesis also presented the design of attack-detection and response mechanisms to maintain system safety and operational performance for the Tennessee-Eastman benchmark chemical process control system. Finally, an analysis of the cyber security of state estimators in SCADA systems operating the Nordic electric power grid was conducted. It has been shown that while the current power system state estimation algorithms are equipped with detection schemes for measurement errors, they can fail to detect deception attacks. A diagnostic method to study the relationship between attacker's information and his ability to evade power grid monitoring systems was developed. In the aforementioned case studies, the proposed diagnostic and response mechanisms maintained closed-loop stability under a wide class of attacks. The work presented in this thesis is still in its infancy; future work will extend it in several conceptual directions, for e.g., using non-parametric estimation and machine learning tools.

### 10.1.3 Scalable Attack Resilient Control Algorithms

The theory of robust control models the controller-disturbance interaction as a game where disturbance is non-strategic. The proviso of a deliberately malicious (strategic) attacker should be considered to increase the robustness of NCS. In the future, a learning-based approach to optimal control, in which probabilistic constraints must be satisfied for a class of DoS attacks, will be investigated. Guaranteed margins obtained from this approach can be used to improve the resilience of regulatory control level (i.e., the systems which regulate process variables). Attacks to supervisory control level (i.e., higher level controls) can cause instantaneous switches of modes due to the changes in system dynamics, control inputs, and sensor measurements. To address the attacks on supervisory control level, the synthesis of control strategies which guarantee safety and stability for mixed-mode (analog plus digital) hybrid cyber-physical systems will be investigated. This work will incorporate stochastic and distributed nature of physical infrastructure systems in the analysis of safety and stability - a significant contribution to the theory of hybrid cyber-physical systems.

### 10.1.4 Management of Interdependent Network Risks

This thesis has also approached the design of resilient NCS from game theoretic perspective (multi-player games). While the usefulness of game theory methods in modeling cyber-security is well established, the novelty of this approach is the integration of game and control theory. The respective optima of (i) strategic security decisions, and (ii) real-time control actions were computed. Future work will involve the computations of Nash equilibria for multi-player games with heterogeneous players, and broad assumptions on the interdependencies between the individual player's payoffs and the other players' security decisions. Such games with externalities are known to have subtle equilibrium properties. In general, they can only be analyzed numerically. The externalities caused by information incompleteness are of special interest, for e.g., information asymmetries due to unknown attacker type. The effects of multiple stake-holders or players (power generators, system operators, and electric utilities) on the NCS/SCADA resilience, when player security decisions are interdependent due to cyber (e.g., Internet Protocol (IP)) and physical (e.g., power line failures due to external factors) risks, will be studied. Due to network-induced externalities, the individual players tend to under-invest in security (relative to a social planner). Another type of interdependency emerges due to high costs of detecting and isolating between security failures (attacks) and reliability failures (faults). Future work will investigate the connection between NCS security and reliability, and will develop a framework for jointly analyzing failures due to interdependent security and reliability failures.

The effects of security decisions (e.g., investments in security) on system resilience will also be investigated. A widely cited review by Prof. Hal Varian establishes that in non-cooperative Nash equilibria underinvestment in security occurs. For public goods, such as security, regulatory impositions (e.g., due care standards) can be used improve social efficiency (in our case, system resilience). Several recent studies have suggested that the alternative approaches such as raffle scheduling also modify player incentives, and reduce the gap between the non-cooperative allocations and the global societal optimum. Such approaches require modest information exchanges. Thus, they are less susceptible to ICT attacks. Finally, in most physical infrastructures, individual consumption data is privacy-sensitive. Hence, future work will involve the development of approaches which have weak requirements on private data, relative to currently considered alternatives of demand response mechanisms (which require real-time communications between the command center and the individual users).

# Bibliography

Aamo, O., Salvesen, J. and Foss, B. [2006], Observer design using boundary injections for pipeline monitoring and leak detection, *in* 'Proc. IFAC Symposium on advanced control of chemical processes', pp. 53–58.

Alpcan, T. and Başar, T. [2011], *Network Security: A Decision and Game Theoretic Approach*, Cambridge University Press, Philadelphia.

Amin, S., Abate, A., Prandini, M., Lygeros, J. and Sastry, S. [2006], Reachability analysis for controlled discrete time stochastic hybrid systems, *in* J. Hespanha and A. Tiwari, eds, 'HSCC', Vol. 3927 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 49–63.

Amin, S., Bayen, A., Ghaoui, L. E. and Sastry, S. S. [2007], Robust feasibility for control of water flow in a canal reservoir system, *in* 'Decision and Control, 2007 46th IEEE Conference on', pp. 1571–1577.

Amin, S., Cárdenas, A. A. and Sastry, S. S. [2009*a*], Safe and secure networked control systems under denial-of-service attacks, *in* R. Majumdar and P. Tabuada, eds, 'HSCC', Vol. 5469 of *Lecture Notes in Computer Science*, Springer, pp. 31–45.

Amin, S., Cárdenas, A. and Sastry, S. [2009*b*], Safe and secure networked control systems under denial-of-service attacks., *in* R. Majumdar and P. Tabuada, eds, 'HSCC', Vol. 5469 of *Lecture Notes in Computer Science*, Springer, pp. 31–45.

Amin, S., Litrico, X., Sastry, S. and Bayen, A. [2010], Stealthy deception attacks on water SCADA systems, *in* 'Proc. 13th ACM International Conference on Hybrid Systems: Computation and Control (HSCC '10)', pp. 161–170.

Anderson, R., Böhme, R., Clayton, R. and Moore, T. [2008], Security economics and European policy, *in* 'Proceedings of the Workshop on the Economics of Information Security WEIS', Hanover, NH, USA.

Anderson, R. and Fuloria, S. [2009], Security economics and critical national infrastructure, *in* 'The Eighth Workshop on the Economics of Information Security'.

Anderson, R. and Fuloria, S. [2010], On the security economics of electricity metering, *in* 'The Ninth Workshop on the Economics of Information Security'.

Attorney, U. [2007], 'Willows man arrested for hacking into Tehama Colusa Canal Authority computer system', http://www.usdoj.gov/usao/cae/press_releases/.

Başar, T. and Olsder, G. [1999], *Dynamic Noncooperative Game Theory*, second edition edn, SIAM Series in Classics in Applied Mathematics, Philadelphia, PA.

Banda, M. K., Herty, M. and Klar, A. [2006], 'Gas flow in pipeline networks', *AIMS Journal on Networks and Heterogeneous Media (NHM)* **1**(1), 41–56.

Basseville, M. and Nikiforov, I. [1993], *Detection of abrupt changes: theory and application*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Bastin, G., Coron, J. M. and d'Andréa-Novel, B. [2008], Using hyperbolic systems of balance laws for modeling, control and stability analysis of physical networks, *in* 'Lecture notes for the Pre-congress workshop on complex embedded and networked control systems, 17th IFAC World Congress', Seoul, Korea.

Bayen, A. M., Raffard, R. L. and Tomlin, C. J. [2006], 'Adjoint-based control of a new Eulerian network model of air traffic flow', *IEEE Transactions on Control Systems Technology* **14**(5), 804–818.

Bedjaoui, N., Weyer, E. and Bastin, G. [2009], 'Methods for the localization of a leak in open water channels', *Networks and Heterogeneous Media* **4**(2), 189–210.

Bedjaoui, N., Litrico, X., Koenig, D. and Malaterre, P. [2006], $H_\infty$ observer for time-delay systems application to FDI for irrigation canals, *in* 'Proc. 45th IEEE Conference on Decision and Control (CDC '06)', pp. 532–537.

Bedjaoui, N., Litrico, X., Koenig, D., Ribot-Bruno, J. and Malaterre, P.-O. [2008], 'Static and dynamic data reconciliation for an irrigation canal', *Journal of Irrigation and Drainage Engineering* **134**(6), 778–787.

Bedjaoui, N. and Weyer, E. [2011], 'Algorithms for leak detection, estimation, isolation and localization in open water channels', *Control Engineering Practice (forthcoming)* .

Bellovin, S. [2010], 'Stuxnet: The first weaponized software?', http://www.cs.columbia.edu/~smb/blog//2010-09-27.html.

Ben-tal, A., Boyd, S. and Nemirovski, A. [2005], Control of uncertainty-affected discrete time linear systems via convex programming, *in* 'SIAM Journal on Control and Optimization; E-print: http://www.optimizationonline.org/DB HTML/2005/10/1232.html'.

Bernstein, D. S. [2009], *Matrix Mathematics: theory, facts, and formulas*, 2 edn, Princeton University Press.

Bernstein, D. S. and Michel, A. N. [1995], 'A chronological bibliography on saturating actuators', *International Journal on Robust Nonlinear Control* **5**, 375–380.

Besson, T., Tchousso, A. and Xu, C.-Z. [2006], Exponential stability of a class of hyperbolic PDE models from chemical engineering, *in* ‘Proceedings of the 45th IEEE Conference on Decision & Control’, San Diego, USA, pp. 3974–3978.

Bier, V., Oliveros, S. and Samuelson, L. [2007], ‘Choosing what to protect: Strategic defensive allocation against an unknown attacker’, *Journal of Public Economic Theory* **9**(4), 563–587.

Böhme, R. and Schwartz, G. [2010], Modeling cyber-insurance: Towards a unifying framework, *in* ‘Proceedings of the Workshop on the Economics of Information Security WEIS’, Harvard University.

Bressan, A. [2000], *Hyperbolic Systems of Conservation Laws: The One-Dimensional Cauchy Problem*, Oxford University Press.

Brodsky, B. and Darkhovsky, B. [1993], *Non-Parametric Methods in Change-Point Problems*, Kluwer Academic Publishers.

Byres, E., Leversage, D. and Kube, N. [2007], ‘Security incidents and trends in SCADA and process industries’, *The Industrial Ethernet Book* **39**(2), 12–20.

Byres, E. and Lowe, J. [2004], The myths and facts behind cyber security risks for industrial control systems, *in* ‘Proceedings of the VDE Congress, VDE Association for Electrical Electronic & Information Technologies’.

Calafiore, G. and El Ghaoui, L. [2007], Linear programming with probability constraints - part 2, *in* ‘American Control Conference, 2007. ACC ’07’, pp. 2642 –2647.

Cantoni, M., Weyer, E., Li, Y., Ooi, S.-K., Mareels, I. and Ryan, M. [2007], ‘Control of large-scale irrigation networks’, *Proceedings of the IEEE* **95**(1), 75–91.

Cárdenas, A. A., Amin, S., Lin, Z.-Y., Huang, Y.-L. and Sastry, S. [2011], Attacks against process control systems: risk assessment, detection, and response, *in* ‘Prof. of the 6th ACM Symposium on InformAtion, Computer and Communications Security (ASIACCS ’11)’.

Cárdenas, A. A., Amin, S. and Sastry, S. S. [2008], Research challenges for the security of control systems, *in* N. Provos, ed., ‘HotSec’, USENIX Association.

Cavusoglu, H., Mishra, B. and Raghunathan, S. [2005], ‘The value of intrusion detection systems in information technology security architecture’, *Info. Sys. Research* **16**(1), 28–46.

CCTV [2002], *Colombian Rebels Continue Attacks on Energy Infrastructure*, English Channel, http://www.cctv.com/english/news/20020119/79688.html.

Chatterjee, D., Hokayem, P. and Lygeros, J. [2009], ‘Stochastic receding horizon control with bounded control inputs: a vector space approach’, http://arxiv.org/abs/0903.5444. Submitted to IEEE Transactions on Automatic Control.

Chatterjee, K., de Alfaro, L. and Henzinger, T. A. [2009], Termination criteria for solving concurrent safety and reachability games, *in* 'Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms', SODA '09, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, pp. 197–206.

Cheung, S., Dutertre, B., Fong, M., Lindqvist, U., Skinner, K. and Valdes, A. [2007], Using model-based intrusion detection for SCADA networks, *in* 'Proceedings of the SCADA Security Scientific Symposium', Miami Beach, FL.

Choy, S. and Weyer, E. [2008], 'Reconfiguration schemes to mitigate faults in automated irrigation channels', *Control Engineering Practice* **16**(10), 1184–1194.

Clemmens, A. [2005], A process-based approach to improving the performance of irrigated agriculture, *in* 'International Congress on Irrigation and Drainage', New Delhi, India.

Conte, G. and Perdon, A. [2006], Unknown input observers and residual generators for linear time delay systems, *in* 'Current Trends in Nonlinear Systems and Control', Systems and Control: Foundations & Applications, Birkhäuser, pp. 15–33.

Coron, J. M., Bastin, G. and D'Andrea-Novel, B. [2008], 'Dissipative boundary conditions for one dimensional nonlinear hyperbolic systems', *SIAM Journal on Control and Optimization* **47**(3), 1460–1498.

Cosman, E. [2006], Patch management at Dow chemical, *in* 'ARC Tenth Annual Forum on Manufacturing'.

Courant, R. and Hilbert, D. [1962], *Methods of Mathematical Physics, Part II: Partial Differential Equations*, Interscience, New York, NY.

Craig, P., Mortensen, J. and Dagle, J. [2008], Metrics for the National SCADA Test Bed Program, Technical report, PNNL-18031, Pacific Northwest National Laboratory (PNNL), Richland, WA.

D. Liberzon, l. . B. [2003], *Switching in Systems and Control*, Volume in series Systems and Control: Foundations and Applications, Birkhauser.

Darouach, M., Zasadzinski, M. and Xu, S. [1994], 'Full-order observers for linear systems with unknown inputs', *IEEE Transactions on Automatic Control* **39**(3), 606–609.

deHalleux, J., Prieur, C., Andrea-Novel, B. and Bastin, G. [2003], 'Boundary feedback control in networks of open channels', *Automatica* **39**(8), 1365–1376.

Denning, D. [1987], 'An intrusion-detection model', *Software Engineering, IEEE Transactions on* **SE-13**(2), 222–232.

Digaĭlova, I. A. and Kurzhanskiĭ, A. B. [2004], 'Attainability problems under stochastic perturbations', *Differential Equations* **40**(11), 1573–1578.

D.Q., M. [2000], 'Constrained model predictive control: Stability and optimality', *Automatica* **36**, 789–814.

Eisenhauer, J., Donnelly, P., Ellis, M. and O'Brien, M. [2006], *Roadmap to Secure Control Systems in the Energy Sector*, Energetics Incorporated. Sponsored by the U.S. Department of Energy and the U.S. Department of Homeland Security.

El-Farra, N. H. and Christofides, P. D. [2004], 'Coordinating feedback and switching for control of spatially distributed processes', *Computers and Chemical Engineering* **28**, 111–128.

Elia, N. [2005], 'Remote stabilization over fading channels', *Systems & Control Letters* **54**(3), 237–249.

Falliere, N., Murchu, L. and Chien, E. [2010], *W32.Stuxnet Dossier*, Symantec.

Franklin, G. F., Powell, J. D. and Emami-Naeini, A. [2006], *Feedback Control of Dynamic Systems*, 5 edn, Pearson Prentice Hall.

Galántai, A. [2006], 'Subspaces, angles and pairs of orthogonal projections', *Linear and Multilinear Algebra* **56**(3), 227–260.

Garone, E., Sinopoli, B. and Casavola, A. [2010], 'LQG control over lossy TCP-like networks with probabilistic packet acknowledgements', *International Journal of Systems, Control and Communications* **2**(1/2/3), 55–81.

Gattami, A. [2007], *Optimal Decisions with Limited Information*, PhD Thesis, Department of Automatic Control, Lund University.

Geer, D. [2006], 'Security of critical control systems sparks concern', *Computer* **39**(1), 20–23.

Giani, A., Sastry, S., Johansson, K. H. and Sandberg, H. [2009], The VIKING project: an initiative on resilient control of power networks, *in* 'Proc. 2nd Int. Symp. on Resilient Control Systems', Idaho Falls, ID, USA, pp. 31–35.

Goulart, P. J., Kerrigan, E. C. and Maciejowski, J. M. [2006], 'Optimization over state feedback policies for robust control with constraints', *Automatica* **42**(4), 523 – 533.

Greenberg, A. [2008], 'Hackers cut cities' power', http://www.forbes.com.

Grossklags, J., Christin, N. and Chuang, J., eds [2008], *Secure or Insure? A Game-Theoretic Analysis of Information Security Games*, Proceedings of the 17th International World Wide Web Conference.

Gu, G., Zhang, J. and Lee, W. [2008], Botsniffer: Detecting botnet command and control channels in network traffic, *in* 'Proceedings of the 15th Annual Network and Distributed System Security Symposium (NDSS'08)', San Diego, CA.

Gugat, M. [2008], ‘Optimal switching boundary control of a string to rest in finite time’, *ZAMM - Journal of Applied Mathematics and Mechanics* **88**, 283–305.

Hale, J. and Lunel, S. [1993], *Introduction to Functional Differential Equations*, Vol. 99 of *Applied mathematical sciences*, Springer-Verlag, New-York, NY.

Hamoud, G., Chen, R. and Bradley, I. [2003], Risk assessment of power systems SCADA, *in* ‘IEEE Power Engineering Society General Meeting, 2003’, Vol. 2.

Hante, F. M. and Leugering, G. [2009], Optimal boundary control of convention-reaction transport systems with binary control functions, *in* R. Majumdar and P. Tabuada, eds, ‘Hybrid Systems: Computation and Control’, Lecture Notes in Computer Science 5469, Springer-Verlag, Berlin, Heidelberg, pp. 209–222.

Hante, F. M., Leugering, G. and Seidman, T. I. [2009], ‘Modeling and analysis of modal switching in networked transport systems’, *Appl. Math. Optim.* **59**(2), 275–292.

Hante, F. M., Leugering, G. and Seidman, T. I. [2010], ‘An augmented BV setting for feedback switching control’, *Journal of Systems Science and Complexity* **23**(3), 456–466.

Hante, F. M. and Sigalotti, M. [2011], ‘Converse Lyapunov theorems for switched systems in Banach and Hilbert spaces’, *SIAM Journal on Control and Optimization* **49**(2), 752–770.

Hart [2007], ‘http://www.hartcomm2.org/frontpage/wirelesshart.html’, *WirelessHart whitepaper* .

Haut, B. and Bastin, G. [2007], ‘A second order model of road junctions in fluid models of traffic networks’, *AIMS Journal on Networks and Heterogeneous Media (NHM)* **2**(2), 227–253.

Heal, G. and Kunreuther, H. [2003], ‘Interdependent security’, *Journal of Risk and Uncertainty* **26**(2–3), 231–249.

Heal, G. and Kunreuther, H. [2004], Interdependent security: A general model, Nber working papers, National Bureau of Economic Research, Inc.

Hespanha, J. P., Naghshtabrizi, P. and Xu, Y. [2007], ‘A survey of recent results in networked control systems’, *Proceedings of the IEEE* **95**(1), 138–162.

Hofmann, A. [2007], ‘Internalizing externalities of loss prevention through insurance monopoly: an analysis of interdependent risks’, *The GENEVA Risk and Insurance Review* **32**(1), 91–111.

Hokayem, P., Chatterjee, D. and Lygeros, J. [2009], On stochastic model predictive control with bounded control inputs, *in* ‘Proceedings of the IEEE Conference on Decision and Control’, Shanghai, China, pp. 6359–6364. Extended version available at http://arxiv.org/abs/0902.3944.

Hokayem, P. F. and Spong, M. W. [2006], 'Bilateral teleoperation: An historical survey', *Automatica* **42**(12), 2035–2057.

Horn, R. A. and Johnson, C. R. [1990], *Matrix Analysis*, Cambridge University Press, Cambridge, UK.

Hurd, S., Smith, R. and Leischner, G. [2008], Tutorial: Security in electric utility control systems, *in* '61st Annual Conference for Protective Relay Engineers', pp. 304–309.

Iftime, O. V. and Demetriou, M. A. [2009], 'Optimal control of switched distributed parameter systems with spatially scheduled actuators', *Automatica J. IFAC* **45**(2), 312–323.

Igure, V., Laughter, S. and Williams, R. [2006], 'Security issues in SCADA networks', *Computers & Security* **25**(7), 498–506.

Imer, O. C., Yüksel, S. and Başar, T. [2006], 'Optimal control of LTI systems over unreliable communication links', *Automatica* **42**(9), 1429–1439.

INL [2010], 'National SCADA Testbed Program', http://www.inl.gov/scada.

ISA [2007], 'Wireless systems for automation', *International Society of Automation* .

Jung, J., Paxson, V., Berger, A. and Balakrishan, H. [2004], Fast portscan detection using sequential hypothesis testing, *in* 'Proceedings of the 2004 IEEE Symposium on Security and Privacy', pp. 211–225.

Kailath, T. and Poor, H. V. [1998], 'Detection of stochastic processes', *IEEE Transactions on Information Theory* **44**(6), 2230–2258.

Koenig, D., Bedjaoui, N. and Litrico, X. [2005], Unknown input observers design for time-delay systems application to an open-channel, *in* 'Proc. 44th IEEE Conference on Decision and Control (CDC '05)', pp. 5794–5799.

Krause, A. and Guestrin, C. [2009], 'Optimizing sensing: From water to the web', *Computer* **42**, 38–45.

Kravets, D. [2009], 'Feds: Hacker disabled offshore oil platform leak-detection system', http://www.wired.com.

Krebs, B. [2008], *Cyber Incident Blamed for Nuclear Power Plant Shutdown*, Washington Post, http://www.washingtonpost.com.

Kreiss, O. [1970], 'Initial boundary value problems for hyperbolic partial differential equations', *Comm. on Pure and Appl. Math.* **23**(3), 277–298.

Kunreuther, H. and Heal, G. [2002], Interdependent security: The case of identical agents, Working Paper 8871, National Bureau of Economic Research.

Langner, R. [2010], 'Langner communications', http://www.langner.com/en/.

Lelarge, M. [2009], Economics of malware: epidemic risks model, network externalities and incentives, *in* 'Allerton'09: Proceedings of the 47th annual Allerton conference on Communication, control, and computing', IEEE Press, Piscataway, NJ, USA, pp. 1353–1360.

Lelarge, M. and Bolot, J. [2008], 'Network externalities and the deployment of security features and protocols in the internet', *SIGMETRICS Perform. Eval. Rev.* **36**(1), 37–48.

Leugering, G. and Schmidt, J.-P. G. [2002], 'On the modeling and stabilisation of flows in networks of open canals', *SIAM Journal of Control and Optimization* **37**(6), 1874–1896.

Leyden, J. [2008], 'Polish teen derails tram after hacking train network', *The Register* .
**URL:** *http://www.theregister.co.uk/2008/01/11/tram_hack/*

Li, T. T. [1994], *Global classical solutions for quasilinear hyperbolic systems*, Research in Applied Mathematics, Masson and Wiley, Paris, Milan, Barcelona.

Lin, C., Wang, Q.-G. and Lee, T. [2006], 'A less conservative robust stability test for linear uncertain time-delay systems', *IEEE Transactions on Automatic Control* **51**(1), 87–91.

Lin, H. and Antsaklis, P. J. [2009], 'Stability and stabilizability of switched linear systems: A survey of recent results', *IEEE Transactions on Automatic Control* **54**(2), 308–322.

Litrico, X. and Fromion, V. [2004*a*], 'Analytical approximation of open-channel flow for controller design', *Applied Mathematical Modelling* **28**(7), 677–695.

Litrico, X. and Fromion, V. [2004*b*], 'Frequency modeling of open-channel flow', *Journal of Hydraulic Engineering* **130**(8), 806–815.

Litrico, X. and Fromion, V. [2005], Design of structured multivariable controllers for irrigation canals, *in* 'Proc. of the 44th IEEE Conference on Decision and Control (CDC '05)', pp. 1881–1886.

Litrico, X. and Fromion, V. [2009*a*], 'Boundary control of hyperbolic conservation laws using a frequency domain approach', *Automatica* **45**(3), 647–656.

Litrico, X. and Fromion, V. [2009*b*], *Modeling and Control of Hydrosystems*, Springer Verlag.

Litrico, X., Malaterre, P.-O., Baume, J.-P. and Ribot-Bruno, J. [2008], 'Conversion from discharge to gate opening for the control of irrigation canals', *Journal of Irrigation and Drainage Engineering* **134**(3), 305–314.

Litrico, X., Malaterre, P.-O., Baume, J.-P., Vion, P.-Y. and Ribot-Bruno, J. [2007], 'Automatic tuning of PI controllers for an irrigation canal pool', *Journal of irrigation and drainage engineering* **133**(1), 27–37.

Liu, Y., Reiter, M. K. and Ning, P. [2009], False data injection attacks against state estimation in electric power grids, *in* 'CCS '09: Proceedings of the 16th ACM conference on Computer and communications security', ACM, New York, NY, USA, pp. 21–32.

Malaterre, P. and Chateau, C. [2007], SCADA interface of the SIC software for easy real time application of advanced regulation algorithms, *in* 'Second Conference on SCADA and Related Technologies for Irrigation System Modernization', Denver, CO.

Michel, A. N., Sun, Y. and Molchanov, A. P. [2005], 'Stability analysis of discontinuous dynamical systems determined by semigroups', *IEEE Transactions on Automatic Control* **50**(9), 1277–1290.

Mo, Y. and Sinopoli, B. [2009], Secure control against replay attacks, *in* 'Proc. 47th Annual Allerton Conf.', Monticello, IL, pp. 911–918.

Morse, A. S. [1996], 'Supervisory control of families of linear set-point controlles. part 1: Exact matching', *IEEE Transactions on Automatic Control* **41**(10), 1413–1431.

Mounzer, J., Alpcan, T. and Bambos, N. [2010], Dynamic control and mitigation of interdependent IT security risks, *in* 'Proceedings of the IEEE Conference on Communication (ICC)', IEEE Communications Society.

Muirhead, R. J. [1982], *Aspects of Multivariate Statistical Theory*, John Wiley & Sons.

Nair, G. N. and Evans, R. J. [2004], 'Stabilizability of stochastic linear systems with finite feedback data rates', *SIAM Journal on Control and Optimization* **43**(2), 413–436.

Nair, G. N., Fagnani, F., Zampieri, S. and Evans, R. J. [2007], 'Feedback control under data rate constraints: An overview', *Proceedings of the IEEE* **95**(1), 108–137.

Negenborn, R., Sahin, A., Lukszo, Z., De Schutter, B. and Morari, M. [2009], A non-iterative cascaded predictive control approach for control of irrigation canals, *in* 'Proc. of the IEEE International Conference on Systems, Man and Cybernetics (SMC '09)', pp. 3552–3557.

NERC-CIP [2008], 'Critical infrastructure protection', http://www.nerc.com/cip.html.

Nguyen, K., Alpcan, T. and Basar, T. [2008], A decentralized bayesian attack detection algorithm for network security, *in* S. Jajodia, P. Samarati and S. Cimato, eds, '23rd Intl. Information Security Conf', Vol. 278 of *IFIP International Federation for Information Processing*, Springer Boston, pp. 413–428.

Oman, P., Schweitzer, E. and Frincke, D. [2000], Concerns about intrusions into remotely accessible substation controllers and SCADA systems, *in* 'Proceedings of the Twenty-Seventh Annual Western Protective Relay Conference', Vol. 160.

Pemantle, R. and Rosenthal, J. S. [1999], 'Moment conditions for a sequence with negative drift to be uniformly bounded in $L^r$', *Stochastic Processes and their Applications* **82**(1), 143–155.

Peterson, D. [2010], 'Digital bond: Weisscon and stuxnet', http://www.digitalbond.com/index.php/2010/09/22/weisscon-and-stuxnet/.

Pinar, A., Meza, J., Donde, V. and Lesieutre, B. [2010], 'Optimization strategies for the vulnerability analysis of the electric power grid', *SIAM Journal on Optimization* **20**, 1786–1810.

Plusquellec, H. [2009], 'Modernization of large-scale irrigation systems: is it an achievable objective or a lost cause?', *Irrigation and Drainage* **58**(1), 104–120.

Prieur, C., Winkin, J. and Bastin, G. [2008], 'Robust boundary control of systems of conservation laws', *Math. Control Signals Systems* **20**(2), 173–197.

Primbs, J. and Sung, C. H. [2009], 'Stochastic receding horizon control of constrained linear systems with state and control multiplicative noise', *Automatic Control, IEEE Transactions on* **54**(2), 221 –230.

Quin, S. J. and Badgwell, T. A. [2003], 'A survey of industrial model predictive control technology', *Control Engineering Practice* **11**(7), 733–764.

Quinn-Judge, P. [2002], 'Cracks in the system', *TIME Magazine* .

Ralston, P., Graham, J. and Hieb, J. [2007], 'Cyber security risk assessment for SCADA and DCS networks', *ISA transactions* **46**(4), 583–594.

Ramponi, F., Chatterjee, D., Milias-Argeitis, A., Hokayem, P. and Lygeros, J. [2010], 'Attaining mean square boundedness of a marginally stable stochastic linear system with a bounded control input', *IEEE Transactions on Automatic Control* **55**(10), 2414–2418.

Åström and Hägglund [1995], *PID controllers: Theory, design, and tuning*, Instrument society of America.

Rauch, J. and Taylor, M. [1974], 'Exponential decay of solutions to hyperbolic equations in bounded domain', *Indiana University Mathematics Journal* **24**(1), 79–86.

Rawlings, J. [2000], 'Tutorial overview of model predictive control', *Control Systems Magazine, IEEE* **20**(3), 38–52.

Reed, T. [2004], *At the Abyss: An Insider's History of the Cold War*, Presidio Press.

Ricker, N. [1993], 'Model predictive control of a continuous, nonlinear, two-phase reactor', *JOURNAL OF PROCESS CONTROL* **3**, 109–109.

Rijo, M. and Arranja, C. [2010], 'Supervision and water depth automatic control of an irrigation canal', *Journal of Irrigation and Drainage Engineering* **136**(1), 3–10.

Rrushi, J. [2009], Composite Intrusion Detection in Process Control Networks, PhD thesis, Universita Degli Studi Di Milano, Italy.

Saberi, A., Stoorvogel, A. A. and Sannuti, P. [1999], *Control of Linear Systems with Regulation and Input Constraints*, 2 edn, Springer-Verlag, New York.

Salmeron, J., Wood, K. and Baldick, R. [2004], 'Analysis of electric grid security under terrorist threat', *IEEE Transactions on Power Systems* **19**, 905–912.

Sandberg, H., Teixeira, A. and Johansson, K. H. [2010], On security indices for state estimators in power networks, *in* '1st Workshop on Secure Control Systems, CPS Week 2010', Stockholm, Sweden.

Sasane, A. [2005], 'Stability of switching infinite-dimensional systems', *Automatica* **41**(1), 75–78.

Schechter, S. and Berger, J. J. A. [2004], Fast detection of scanning worm infections, *in* 'Proc. of the Seventh International Symposium on Recent Advances in Intrusion Detection (RAID)'.

Schenato, L., Sinopoli, B., Franceschetti, M., Poolla, K. and Sastry, S. S. [2007], 'Foundations of control and estimation over lossy networks', *Proceedings of the IEEE* **95**, 163–187.

Seidman, T. I. [2009], Feedback modal control of partial differential equations, *in* K. Kunisch, G. Leugering, J. Sprekels and F. Tröltzsch, eds, 'Optimal Control of Coupled Systems of Partial Differential Equations', Birkhäuser, Basel, pp. 239–254.

Shampine, L. F. [2005], 'Solving Hyperbolic PDEs in Matlab', *Appl. Numer. Anal. & Comput. Math.* (2), 346–358.

Shorten, R., Wirth, F., Mason, O., Wulff, K. and King, C. [2007], 'Stability criteria for switched and hybrid systems', *SIAM Review* **49**(4), 545–592.

Skaf, J. and Boyd, S. [2010], 'Design of affine controllers via convex optimization', *IEEE Transactions on Automatic Control* **55**(11), 2476 –2487.

Slay, J. and Miller, M. [2007], Lessons learned from the Maroochy Water Breach, *in* 'Critical Infrastructure Protection', Vol. 253/2007, Springer, pp. 73–82.

Stouffer, K., Falco, J. and Kent, K. [2006], Guide to supervisory control and data acquisition (SCADA) and industrial control systems security, Sp800-82, NIST.

Toivonen, H. T. [1983], 'Suboptimal control of discrete stochastic amplitude constrained systems', *International Journal of Control* **37**(3), 493–502.

Tsang, P. P. and Smith, S. W. [2008], YASIR: A low-latency high-integrity security retrofit for lecacy SCADA systems, *in* '23rd International Information Security Conference (IFIC SEC)', pp. 445–459.

Turk, R. J. [2005], Cyber incidents involving control systems, Technical Report INL/EXT-05-00671, Idaho National Laboratory.

US-CERT [2008], *Control Systems Security Program*, US Department of Homeland Security, http://www.us-cert.gov/control_systems/index.html.

van Hessem, D. and Bosgra, O. [2003], A full solution to the constrained stochastic closed-loop mpc problem via state and innovations feedback and its receding horizon implementation, *in* 'Decision and Control, 2003. Proceedings. 42nd IEEE Conference on', Vol. 1, pp. 929 – 934.

W.A. Owens, K. D. and Lin, H. [2009], *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities*, Committee on Offensive Information Warfare, National Research Council, Philadelphia, PA.

Wald, A. [1947], *Sequential Analysis*, J. Wiley & Sons, New York, NY.

Wang, Y. and Boyd, S. [2009], 'Performance bounds for linear stochastic control', *Systems & Control Letters* **58**(3), 178–182.

Weiss, J. [2010], *Protecting Industrial Control Systems from Electronic Threats*, second edition edn, Momentum Press, Philadelphia.

Weyer, E. and Bastin, G. [2008], Leak detection in open water channel, *in* 'Proc. 17th IFAC World Congress', pp. 7913–7918.

Wonham, W. M. and Cashman, W. F. [1969], 'A computational approach to optimal control of stochastic saturating systems', *International Journal of Control* **10**(1), 77–98.

Wright, A. K., Kinast, J. A. and McCarty, J. [2004], Low-latency cryptographic protection for SCADA communications, *in* 'Applied Cryptography and Network Security (ACNS)', pp. 263–277.

WSCC-CSWG [2008], Roadmap to secure control systems in the water sector, Technical report, AWWA and DHS.

Wu, F. F. [1990], 'Power system state estimation: a survey', *Int. J. Elec. Power and Energy Systems* (2).

Wu, F. F. and Liu, W.-H. E. [1989], 'Detection of topology errors by state estimation', *IEEE Trans. Power Syst.* (1).

Xie, M., Yin, H. and Wang, H. [2006], An effective defense against email spam laundering, *in* 'Proceedings of the 13th ACM Conference on Computer and Communications Security', pp. 179–190.

Yang, Y. D., Sontag, E. D. and Sussmann, H. J. [1997], 'Global stabilization of linear discrete-time systems with bounded feedback', *Systems and Control Letters* **30**(5), 273–281.

Yu, Z. H., Li, W., Lee, J. H. and Morari, M. [1994], 'State estimation based model predictive control applied to shell control problem: a case study', *Chemical Engineering Science* **49**(3), 285 – 301.

Zhang, J., Johansson, K. H., Lygeros, J. and Sastry, S. [2001], 'Zeno hybrid systems', *International Journal of Robust and Nonlinear Control* **11**(5), 435–451.

Zuazua, E. [2011], 'Switching control', *J. Eur. Math. Soc. (JEMS)* **13**(1), 85–117.