Universitat Politècnica de Catalunya

Programa de Doctorat:

Automàtica, Robòtica i Visió

Tesi Doctoral

Automatic Control Advances in CPS Security

Carlos Trapiello Fernández

Directors: Prof. Vicenç Puig Cayuela i Dra. Gabriela Cembrano Gennari

Juny 2021

Acknowledgements

Now that I am finishing my Ph.D. studies, it is time to look back and thank the people who have helped me on this journey. Certainly, completing a doctoral thesis is a laborious undertaking that cannot be done alone, but requires the support and advice of many people. To all of them I express my gratitude below.

First of all, I would like to thank Prof. Vicenç Puig for his effort and dedication. During these years he has always given me the freedom and confidence to investigate, and the right advice to guide me back onto the correct path. Additionally, I am sincerely grateful to Dr. Gabriela Cembrano for the academic discussions and her invaluable guidance.

Second, I want to thank the people with whom I have worked with over the years and who in one way or another have left their mark on me. In particular, I am really grateful to my office colleague Luis Romero for patiently listening to all my complaints (which are not few) and turning them into laughter. Besides, I also want to thank the people of the Sistemes Avançats de Control (SAC) lab, those who are here, and those who like Dr. Daminao Rontondo, Dr. Bartomeu Rubí and Dr. Julen Cayero have left.

Finally, I would like to thank my family for the teachings of life and the support received. They have always been as good to me as only family could be. Furthermore, I would also like to thank my friends here and there, and those who, flowing with life, have spread throughout the world. Last but not least, my deepest thanks to Olaya Cossío: for so many things, for everything.

Continuing with the tradition, I would like to conclude this acknowledgements section with a quote from the former professional boxer Mike Tyson, whose raw words are able to sum up my Ph.D. journey

Everyone has a plan until they get punched in the face

Carlos Trapiello
Barcelona, Spain, 2021

Abstract

Cyber-physical systems (CPSs) constitute a new generation of control systems that seeks a better integration between computation and physical processes. However, the notable advances of these systems in terms of efficiency and modularity come at the price of introducing new security vulnerabilities. Consequently, the complementation of existing cybersecurity methods with model-based techniques that take into account the effect of attacks on the system dynamics, has become an urgent need and a major theoretical challenge. Therefore, this thesis presents several contributions in the security of CPSs from an automatic control perspective. Notably, motivated by the concerns they generate in the industrial sector, attention is focused in two particular problems: the detection of replay attacks launched against control systems, and the reconfiguration of hardware-redundant systems after an anomaly causes the nominal component configuration to be no longer admissible. During this thesis, the non-negligible effect of the uncertainty that remains when modelling a dynamical system is addressed assuming an unknown-but-bounded description.

Regarding the detection of replay attacks, the bounded uncertainty premise has motivated a set-based paradigm that allows to infer deterministically whether the system is under attack or not. In particular, the contributions presented in this field make use of zonotopic-sets in order to develop a set-invariance analysis on the detectability of the replay attack launched against the supervisory layer of a control system, as well as to derive explicit expressions regarding the attack detectability with respect the set-point output reference set from the supervision center. The advantages offered by zonotopes are exploited in the design of different physical watermarking schemes, which guarantee the detection of the attack either by injecting finite watermark sequences, or by imposing a continuous degradation of the system's operation. In addition, the relationship between the proposed zonotope-based deterministic techniques and the stochastic detection techniques, which are widely used in optimal control schemes, has been studied, obtaining analogous expressions that allow connecting both approaches.

On the other hand, techniques for exploiting the physical redundancy common in large-scale systems, which often include back-up elements other than those used in nominal operation, are studied. On this subject, the (in some sense) optimal selection of the new system configuration after an anomaly detection is particularized for flow-based networks. In the first instance, the problem is addressed at a planning level, proposing a generic methodology where the selection of the optimal configuration is posed as a multi-objective mixed-integer program (MIP), pruned with the offline assessment of necessary properties. The methodology is further extended with the inclusion of closed-loop stability guarantees and the consideration of unknown-but-bounded uncertainty. In this regard, the admissibility of each configuration is evaluated according to the average stage-cost of the optimal periodic trajectory towards which the flow-based network can be robustly steered to. Besides, the configuration selection is solved using a single-layer robust model predictive control (MPC) scheme, where the local controller is designed using the minimal configurations that guarantee an admissible performance. To that end, several algorithms for searching minimum cardinality configurations in the associated lattice of configurations are proposed.

Keywords: Secure control, replay attack, zonotopes, physical watermarking, set-invariance, uncertain systems, optimal control, fault tolerant control, back-up components, flow-based systems, large-scale systems.

Resumen

Los sistemas ciberfísicos (SCF) constituyen una nueva generación de sistemas de control que busca una mejor integración entre la parte computacional y los procesos físicos. Sin embargo, los notables avances que ofrecen estos sistemas en términos de eficiencia y modularidad, tienen como contrapartida la introducción de nuevas vulnerabilidades de seguridad. Como consecuencia, la complementación de los ya existentes métodos de ciberseguridad con técnicas basadas en modelo que analicen los posibles efectos que producen los ataques en las dinámicas del sistema, aparece como una necesidad imperiosa que, a su vez, plantea un gran reto académico. Por consiguiente, esta tesis presenta varias contribuciones en la seguridad de los SCF desde la perspectiva del control automático. En particular, motivado por su interés industrial, la atención se ha centrado en dos problemas concretos: la detección de ataques de repetición lanzados sobre un sistema de control, y la reconfiguración de un sistema que presenta redundancia de componentes físicos después de que una anomalía cause la inadmisibilidad de la configuración de elementos nominal. Durante esta tesis, el efecto no negligible de la incertidumbre que aparece al modelar un sistema físico es abordado mediante una descripción desconocida pero acotada.

En lo referente a la detección de los ataques de repetición, la premisa de considerar perturbaciones acotadas motiva el uso de técnicas basadas en conjuntos, las cuales permiten inferir de forma determinista si el sistema se encuentra, o no, bajo ataque. En particular, los conjuntos zonotópicos son utilizados para desarrollar un análisis de invariancia sobre la detectabilidad de los ataques de repetición que afectan la capa de supervisión de un sistema de control, así como para la obtención de expresiones analíticas que relacionan la consigna que es fijada desde el centro de supervisión con la detección del ataque. Las ventajas que ofrecen los zonotopos son aprovechadas en el diseño de distintos esquemas de marca de agua física mediante los cuales se puede garantizar la detección del ataque, bien a través de la inyección discreta de señales secuenciales, o bien mediante una degradación continua de la operación del sistema. Además, se ha estudiado la relación entre las técnicas deterministas propuestas y las técnicas de detección estocásticas, que son ampliamente usadas en los esquemas de control óptimos, obteniéndose expresiones análogas que permiten conectar ambas aproximaciones.

Por otro lado, también se han investigado distintas técnicas para el aprovechamiento inteligente de la redundancia física que es común en sistemas de gran escala. En concreto, se analiza como seleccionar los componentes de respaldo que dan lugar a una nueva configuración óptima tras la detección de una anomalía en una red de flujo. En una primera instancia, se aborda el problema desde la capa de planificación, proponiéndose una metodología general donde la selección de la configuración es formulada como una optimización multiobjetivo mixta en enteros, que es previamente filtrada con la evaluación fuera de línea de propiedades necesarias. Esta metodología es extendida con la inclusión de garantías de estabilidad en lazo cerrado y de incertidumbres. A este respecto, la admisibilidad de cada configuración es evaluada en función de la trayectoria periódica óptima hacia la que se puede hacer converger de forma robusta el sistema, mientras que la selección de la configuración es planteada usando esquemas de control de modelo predictivo de una capa, cuyo controlador local es diseñado usando la configuración de actuadores mínima que garantiza una operación admisible. Con ese fin, también se proponen distintos algoritmos para buscar tales configuraciones mínimas.

Palabras clave: Control seguro, ataques de repetición, marca de agua física, conjuntos invariantes, sistemas inciertos, control óptimo, control tolerante a fallos, elementos de respaldo, sistemas de flujo, sistemas de gran escala.

Common Notation

$\mathbb{R}$	Set of real numbers
$\mathbb{N}$ ($\mathbb{N}_+$)	Set of (positive) natural numbers
$\mathbb{R}^n$	Set of n-dimensional real vectors
$\mathbb{R}^{n\times m}$	Set of $n \times m$ real matrices
$\|x\|_p$	p-norm of a vector
$\|A\|_p, \|A\|_F$	p-norm and Frobenius norm of a matrix
a_{ij}	(i, j) entry of a matrix
x^T, A^T	Transpose of a vector/matrix
$\|A\|$	Absolute value of a matrix
$\mathcal{N}(A)$	Nullspace of a matrix
$rank(A)$	Rank of a matrix
$det(A)$	Determinant of a matrix
A^{-1}	Inverse of a matrix
$Tr[A]$	Trace of a matrix
$\rho(A)$	Spectral radius
I_n	$n \times n$ identity matrix
$O_{n\times m}$	$n \times m$ matrix with zero entries
$diag(\cdot)$	Diagonal matrix with appropriate dimensions
$\otimes$	Kronecker product
$\oplus$	Minkowski sum
$\ominus$	Pontryagin difference
$\in$	It belongs to
$\cap$	Intersection
$\subseteq$, $\subset$	Subset, strict subset
$\supseteq$, $\supset$	Superset, strict superset
$X \succ 0 (\prec 0)$	X is a positive (negative) definite matrix
$X \succeq 0 (\preceq 0)$	X is a positive (negative) semidefinite matrix
$\mathbf{B}^m$	Unitary box composed of m unitary intervals
x_k	The subindex k indicates the discrete time
$\boldsymbol{x}_{k:k+N}$	Sequence of discrete time variables in $[k,\ k+N-1]$ and $N \in \mathbb{N}_+$
$\langle c, H\rangle$	Zonotope with center c and generator matrix H
$cov(\mathcal{Z})$	Covariation of the zonotope $\mathcal{Z}$
$\downarrow_q$	Reduction operator
$\mathbb{I}_a$	Set of integer numbers contained in the interval $[0,\ a]$

Acronyms

BUM	Bottom-up Monotonous
CPS	Cyber-physical System
DAE	Difference-algebraic Equation
DARE	Discrete Algebraic Ricatti Equation
DoS	Denial of Service
DWTN	Drinking Water Transport Network
FD	Fault Diagnosis
FTC	Fault-tolerant Control
IID	Independent Identically Distributed
LP	Linear Program
LQG	Linear Quadratic Gaussian
LQR	Linear Quadratic Regulator
LQZ	Linear Quadratic Zonotopic
LTI	Linear time-invariant
MAC	Minimal Actuator Configuration
MIP	Mixed-integer Program
MLD	Mixed Logical Dynamical
MNAC	Minimal Necessary Actuator Configuration
MPC	Model Predictive Control
mRPI	Minimal Robust Positively Invariant
MTD	Moving Target Defense
RPI	Robust Positively Invariant
SIA	Set-invariance Approach
ZKF	Zonotopic Kalman Filter

Contents

Abstract ii

Resumen iii

Common Notation iv

Acronyms v

Contents v

List of Figures x

List of Tables xii

1 Introduction 1

1.1 Motivation 1
1.2 Thesis objectives 2
1.3 Thesis structure 3

I Replay attack detection 7

2 The replay attack 9

2.1 Introduction to replay attack detection 9
2.2 Literature review 10
2.2.1 Attack detection techniques 11
2.2.2 Set-theoretic techniques in attack characterization, detection and tolerant control 13
2.3 Replay attack against control systems 14
2.3.1 Specific notation 15

3 Zonotopic SIA for replay attacks affecting the supervisory layer 17

3.1 Introduction 17
3.2 Problem statement 18
3.2.1 Regulatory layer 20
3.2.2 Supervisory layer 20
3.2.2.1 Anomaly detector 21
3.3 Record phase 22
3.4 Replay attack detectability 22

3.4.1 Regulatory layer 23
3.4.2 Scenario I - Supervisory layer 23
3.4.2.1 Steady-state attack detectability 24
3.4.3 Scenario II - Supervisory layer 26
3.4.3.1 Steady-state attack detectability 26
3.5 Watermark sequences 29
3.5.1 Sequence design 30
3.5.2 Programming problem 31
3.5.3 Complexity reduction 34
3.5.3.1 Pair symmetry 34
3.5.3.2 mRPI outer-approximations 35
3.6 Case study 35
3.6.1 Steady-state analysis - Scenario I 36
3.6.2 Steady-state analysis - Scenario II 38
3.6.3 Watermark sequence design 41
3.7 Summary 45

4 A zonotopic-based watermarking design to detect replay attacks 47

4.1 Introduction 47
4.2 Problem statement 48
4.2.1 System description 49
4.2.2 Zonotopic KF 50
4.3 LQZ control 50
4.3.1 Finite horizon control 51
4.3.2 Infinite horizon control 53
4.3.3 Anomaly detector 55
4.4 Replay attack detectability 56
4.4.1 Performance loss induced by a zonotopically-bounded signal 57
4.5 Zonotopically bounded watermark signal design 58
4.5.1 Design of an extended state estimator 59
4.5.2 Watermarked system stability 61
4.5.3 Attack detection 62
4.6 Case study 64
4.6.1 Performance assessment 64
4.6.2 Watermark signal design 65
4.6.3 Attack simulation 66
4.7 Summary 70

II Reconfiguration with back-up components 71

5 System reconfiguration with back-up components 73

5.1 Introduction 73
5.2 System reconfiguration with back-up components in FTC 74
5.2.1 The system reconfiguration problem 74
5.2.2 Literature review 76
5.3 Problem statement 77
5.3.1 Actuator configurations 77
5.3.2 Selection of an admissible configuration 78

5.3.3 Specific notation . . . 79

6 Reconfiguration with back-up components - Planning 81

6.1 Introduction . . . 81
6.2 Methodology overview . . . 82
6.3 System admissibility . . . 83
6.3.1 System description . . . 84
6.3.2 Admissibility criterion . . . 85
6.3.3 Admissibility test for the faulty configuration . . . 86
6.4 Configuration selection . . . 86
6.4.1 MLD formulation . . . 87
6.4.2 Lexicographic optimality criterion . . . 88
6.4.3 MIP optimization . . . 88
6.5 Analysis of necessary properties . . . 89
6.5.1 Flow-based networks . . . 89
6.6 Pruning . . . 91
6.6.1 Minimal actuator configurations . . . 92
6.6.1.1 MACs as constraints . . . 92
6.6.2 Minimal necessary actuator configurations . . . 93
6.6.2.1 MNACs as constraints . . . 94
6.7 Case study . . . 94
6.7.1 Network model . . . 94
6.7.2 Reference model . . . 96
6.7.3 Offline tests . . . 96
6.7.4 Admissibility criterion . . . 97
6.7.5 Results . . . 97
6.8 Summary . . . 99

7 Reconfiguration with back-up components - Control 101

7.1 Introduction . . . 101
7.2 Problem statement . . . 103
7.2.1 Performance assessment . . . 105
7.2.2 Admissibility . . . 105
7.3 Stability property . . . 106
7.3.1 Robust planner . . . 106
7.3.2 Single-layer robust MPC . . . 107
7.3.3 Parametrized solution of the robust planner . . . 109
7.4 Performance property . . . 111
7.4.1 Search for minimal configurations . . . 111
7.5 Solution of the reconfiguration problem . . . 112
7.5.1 Solution based on minimal configurations . . . 113
7.5.2 MIP optimization . . . 114
7.5.3 Decision between sets . . . 115
7.6 Case study . . . 116
7.6.1 Management criteria . . . 116
7.6.2 Robust controller design for the nominal configuration . . . 118
7.6.3 Minimal configurations that satisfy the performance condition . . . 118
7.6.4 Criteria for configuration selection . . . 119
7.6.5 Fault scenario - Fault in 24 . . . 119

7.6.6 Fault scenario - Tank isolation. 122
7.7 Summary 124

III Conclusions and future research 125

8 Concluding remarks and further extensions 127

8.1 Contributions 127
8.2 Directions for future research 128

IV Appendices 131

Appendices 133

A Mathematical background 133

A.1 Matrix definitions and properties 133
A.2 Set definitions and properties 134
A.2.1 Polyhedral sets 135
A.2.2 Zonotopic sets 136
A.3 Zonotopic ϵ-approximation of the mRPI set 137
A.3.1 Initial RPI set 137
A.3.2 Forward propagation and stopping criterion 138
A.4 Lexicographic multi-objective optimization 140

B Case study description 141

B.1 Four-tank system 141
B.2 The Barcelona Drinking Water Transport Network 144
B.2.1 Control-oriented model of the DWTN 144
B.2.1.1 Tanks 144
B.2.1.2 Actuators 145
B.2.1.3 Nodes 145
B.2.1.4 Sector of consume 145
B.2.1.5 Network model 146
B.2.2 The aggregated case study 146

Bibliography 159

List of Figures

1.1 Roadmap of the thesis . . . 4

3.1 Overall scheme. Solid (dotted) lines represent local (remote) connections. . . . 19
3.2 Scenario I. . . . 36
3.3 Residual signal - Scenario I. . . . 37
3.4 Set-point imposed from the supervisory layer - Scenario I. . . . 38
3.5 Residuals at the supervisory layer - Scenario I. . . . 38
3.6 Scenario II. . . . 39
3.7 Temporal evolution of the system outputs - Scenario II. . . . 40
3.8 Temporal evolution of the residual signal - Scenario II. . . . 40
3.9 Residual signal - Scenario II. . . . 41
3.10 Effect of the data replay - Scenario II. . . . 41
3.11 Input space partitions. . . . 42
3.12 Watermark sequences. . . . 43
3.13 Temporal evolution of the watermark sequences. . . . 44
3.14 Temporal evolution of the system outputs - Watermark sequences. . . . 44
3.15 Residual signal - Watermark sequences. . . . 44

4.1 Overall scheme and replay attack representation. . . . 49
4.2 Watermark generator placement in the control loop. . . . 57
4.3 Temporal evolution of the watermark signal. . . . 67
4.4 Temporal evolution of the system outputs with watermark. . . . 67
4.5 Temporal evolution of the residual signal with watermark. . . . 68
4.6 Temporal evolution of the system outputs without watermark. . . . 69
4.7 Temporal evolution of the residual signal without watermark. . . . 69
4.8 Comparison of performance loss vs mean detection time. . . . 69

5.1 Lattice of actuators. . . . 78

6.1 Reconfiguration block placement. . . . 82
6.2 Reconfiguration block operation diagram. . . . 84
6.3 Schematic representation of actuator configurations. . . . 92
6.4 DWTN Graph. . . . 95
6.5 Non-recoverable elements. . . . 97
6.6 Solver time. . . . 98
6.7 Temporal evolution - Tank 12. . . . 98
6.8 Example of system reconfiguration. . . . 99

7.1 Lattice of $\mathbb{P}_{\tilde{C}_f}(C_F)$. . . . 114
7.2 Nominal / back-up components. . . . 117
7.3 Tank volumes evolution - Fault in 24. . . . 121
7.4 Actuators evolution - Fault in 24. . . . 121
7.5 Average stage cost - Fault 24. . . . 122

7.6 Tank volumes evolution - Tank isolation. 123
7.7 Actuators evolution - Tank isolation. 123
7.8 Average stage cost - Tank isolation. 124

B.1 Schematic representation of the four-tank process. 141
B.2 Aggregated Barcelona DWN. 147

List of Tables

3.1 Cost function for different combinations of (N, s). 42

4.1 Zonotopic RPI outer approximations of the mRPI set Φ. 65
4.2 Watermark signal design. 66

6.1 Elements partition. 95

7.1 Elements partition. 116
7.2 Minimal configurations. 119
7.3 Set of minimal configurations - Fault 24, $\beta = 1.2$. 120
7.4 Set of minimal configurations - Tank isolation, $\beta = 1.4$. 122

B.1 Four-tank system parameters. 142
B.2 Four-tank operating point. 143

Chapter 1

Introduction

This chapter presents an introduction to the Ph.D. thesis manuscript detailing all the work carried out to achieve the planned objectives. Thereby, in the first instance, the main motivations of the Ph.D. thesis are described in Section 1.1. Then, the thesis objectives are presented in Section 1.2. Finally, in Section 1.3, a brief outline of the structure of this dissertation is presented, which provides a summary of every chapter.

1.1 Motivation

The migration from traditional point-to-point control systems to the more complex cyber-physical systems (CPSs), which combine networked computing and sensing resources with physical control systems, appears as a natural step in order to increase efficiency, manage complexity, or provide convenience. Nevertheless, despite the fact that computer security has developed mature technologies to protect system information, the integration between cyber and physical components present in CPSs introduces new vulnerabilities for which cyber security tools may be insufficient [Byres and Lowe, 2004, Cárdenas et al., 2008, Sandberg et al., 2015]. In addition, since control systems are at the core of many critical infrastructures, a successful attack launched against these systems can have a major socio-economic impact as reflected in an ever-growing list of registered attacks [Abrams and Weiss, 2008, Lee et al., 2014, Case, 2016]. As a consequence, the interest aroused by cybersecurity aspects of control systems has experienced an exponential growth during the last decade (see Sánchez et al. [2019a] and the references therein).

At this point, it must be highlighted that security in control systems is not a new topic of study, in such a way that the fault diagnosis (FD) and fault tolerant control (FTC) problems have been thoroughly studied, and at present-day they constitute mature fields of study with well-grounded techniques. Nonetheless, unlike in the case of faults, cybersecurity related works are based on the implicit assumption that the adversary is a rational entity, and thus endowed with intelligence, that seeks to accomplish a specific objective in the most efficient manner. This substantial difference entails a paradigm shift with respect the fault-related literature, as evidenced by attacks like: replay attacks Mo and Sinopoli [2009], false data injection attacks Liu et al. [2011], etc. which are capable to disrupt the system operation while remaining undetected from standard anomaly detectors.

Consequently, it is of paramount importance the implementation of secure control techniques from both: the system design phase, conceiving vulnerability-free systems and installing hardware redundant components; and the operational phase, with the development of new techniques capable of first detecting, and then counteracting, complex attacks. This Ph.D. dissertation focuses on the operational phase, aiming attention at developing new techniques for detecting replay attacks (pre-detection operation), and devising efficient physical reconfiguration strategies in hardware-redundant systems (post-detection operation). The main motivations for focusing on these particular problems are reported below.

- **Detection of replay attacks:** Replay attacks constitute probably the most natural way to attack a control system without being detected. In this type of attack, an anomaly detector that monitors the system operation can be deceived by simply feeding it with previously recorded data. This, in turn, allows a malicious attacker to mask possible disruptive actions conducted over the system. In this regard, the damaging effects of replay attacks have been made evident in the famous Stuxnet case [Langner, 2011, 2013]. Therefore, the development of techniques that grant a fast and unambiguous detection of replay attacks in uncertain systems is presented as a major problem with immediate industrial applications.

  In addition, conversely to the most common stochastic uncertainty characterizations, unknown-but-bounded descriptions have proven to be a more adequate approach for representing certain types of uncertainties like modelling errors caused by the lack of knowledge about deterministic behaviours. On this subject, although rarely used for detecting attacks, the good results obtained by set-based techniques in the fault diagnosis problem present them as a solid approach for tackling the replay attack detection problem.

- **Reconfiguration strategies for systems with hardware redundancy:** Large-scale network systems typically present a large number of alternative connections and back-up elements, which differ from those used in nominal operation, and that could be brought into play after detecting an anomaly. In this scenario, a smart management of the hardware redundancy may extend the capabilities of the system to cope with unforeseen events. However, the majority of works that propose solution to enhance the system tolerance to disruptive event focus on exploiting the *analytical redundancy* of the system variables for a fixed structure of components, whereas there is a lack of an appropriate methodology that makes use of the *physical redundancy* of the system in an efficient manner. Solutions to this problem are aimed at answering the question of which is the optimal subset of interventions to be carried out after isolating a part of a large-scale system for security reasons.

1.2 Thesis objectives

In accordance with the stated in the previous section, two main objectives have been set in this Ph.D. thesis, namely:

Objective I Contributing to replay attack detection techniques using set-based approaches.

Objective II Tackling the problem of reconfiguring a system that has back-up elements.

Seeking to achieve such ambitious objectives, the main objectives of the thesis have been divided into the following specific objectives:

Objective I.1 Characterize the detectability of the replay attack using set-based techniques.

Objective I.2 Derive explicit expressions regarding the replay attack detectability under norm-bounded uncertainties.

Objective I.3 Design finite input sequences that force the detection of the replay attack.

Objective I.4 Study the analogy between stochastic and set-based approaches.

Objective I.5 Propose a guaranteed watermarking scheme.

Objective II.1 Introduce a methodology for addressing the system reconfiguration with back-up components.

Objective II.2 Analyse the necessary properties that must be satisfied by each possible system configuration.

Objective II.3 Include closed-loop guarantees in the configuration switch.

Objective II.4 Extend the methodology by considering system uncertainty.

1.3 Thesis structure

The remainder of this dissertation is divided into three parts: Part I is dedicated to the study of the replay attack detection problem, Part II is devoted to the analysis of how to reconfigure a system with back-up components and Part III presents the conclusions and possible future research directions that have been identified. In this regard, the roadmap of this document is presented in Figure 1.1.

Each of the parts that have been described previously has the following structure.

- *Part I* entitled *Replay attack detection* contains the following chapters:
  - **Chapter 2.** The replay attack.
  - **Chapter 3.** Zonotopic set-invariance approach for replay attacks affecting the supervisory layer.
  - **Chapter 4.** A zonotopic-based watermarking design to detect replay attacks.
- *Part II* entitled *Reconfiguration with back-up components* contains the following chapters:
  - **Chapter 5.** System reconfiguration with back-up components.
  - **Chapter 6.** Reconfiguration with back-up components - Planning.
  - **Chapter 7.** Reconfiguration with back-up components - Control.
- *Part III* entitled *Conclusions and future research* contains the following chapter:
  - **Chapter 8.** Concluding remarks and further extensions.

In addition, a brief description of the content and related papers for each chapter is presented below.

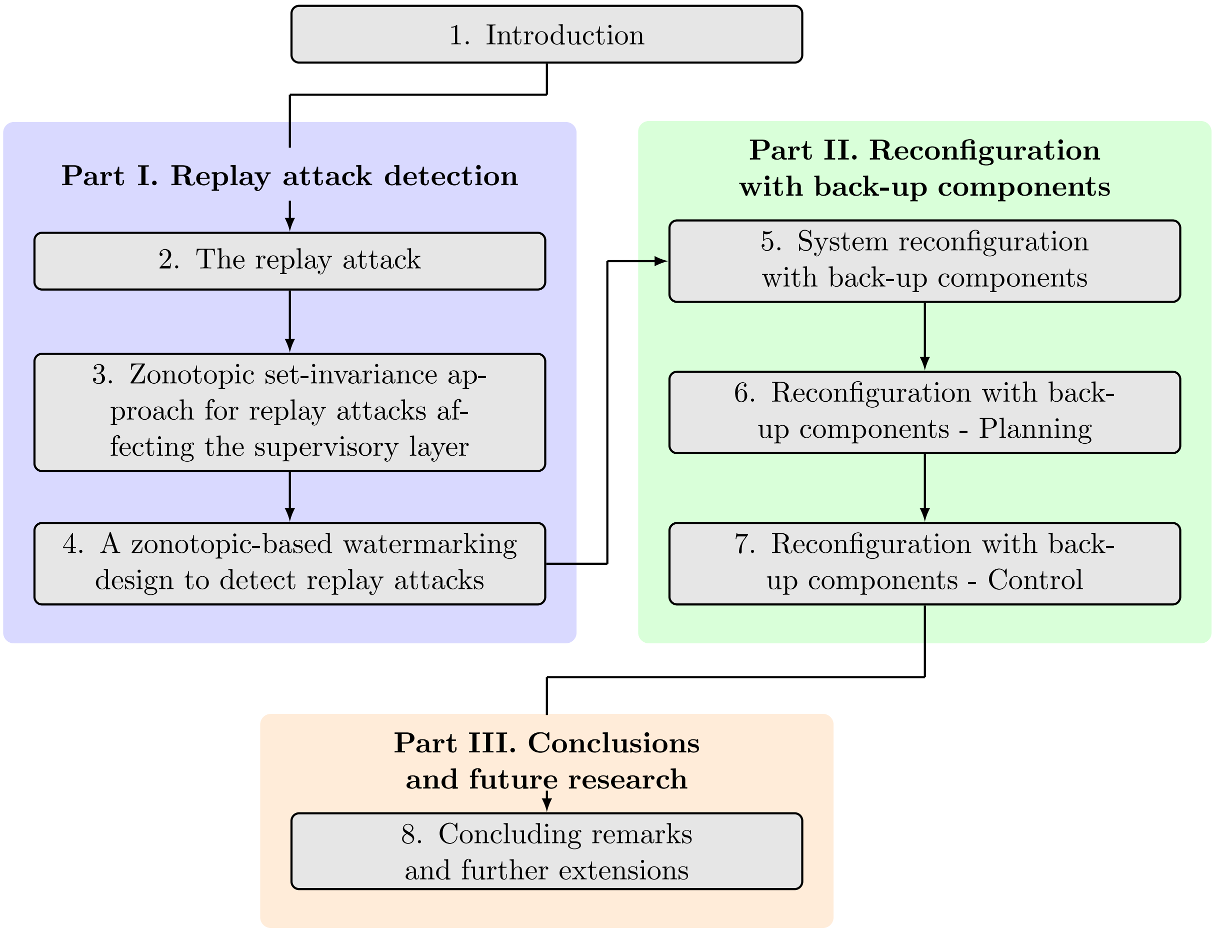

1. Introduction
Part I. Replay attack detection
2. The replay attack
3. Zonotopic set-invariance approach for replay attacks affecting the supervisory layer
4. A zonotopic-based watermarking design to detect replay attacks
Part II. Reconfiguration with back-up components
5. System reconfiguration with back-up components
6. Reconfiguration with back-up components - Planning
7. Reconfiguration with back-up components - Control
Part III. Conclusions and future research
8. Concluding remarks and further extensions

Figure 1.1: Roadmap of the thesis

Chapter 2. The replay attack

This chapter is an introduction to Part I of the thesis. Therefore, firstly, this chapter introduces the arguments that support a set-based approach to the replay attack detection problem. Then, a literature review on the central attack detection techniques and on the main set-theoretic works that have been developed within the cyber-security field, is presented. Finally, the main aspects of the replay attack are detailed.

The material presented in this chapter supports the developments presented in Chapter 3 and Chapter 4, which collect the main contributions of this first part of the thesis.

Chapter 3. Zonotopic set-invariance approach for replay attacks affecting the supervisory layer

This chapter presents a set-invariance approach (SIA) to the detectability of replay attacks affecting the communication network that serves the supervisory layer of complex control systems. In particular, by representing invariant sets as zonotopes, analytical expressions are derived for the steady-state regarding the attack detectability under the presence of bounded uncertainties. In addition, the proposed zonotopic characterization of the residual set is used for the offline design of a set of watermark sequences. A case study based on a four-tank system is used to validate the developments proposed in this chapter.

The material presented in this chapter has resulted in the following publications:

– C. Trapiello, V. Puig, and D. Rotondo. A zonotopic set-invariance analysis of replay attacks affecting the supervisory layer. Systems & Control Letters. *Under review.*

– C. Trapiello, and V. Puig. Optimal finite-time watermark signal design for replay attack detection using zonotopes. 5th Conference on Control and Fault Tolerant Systems (SysTol). *Under review.*

– C. Trapiello, and V. Puig. Input design for active detection of integrity attacks using set-based approach. 21th IFAC World Congress. IFAC-PapersOnLine, 53(2): 11094-11099, 2020.

– C. Trapiello, and V. Puig. Set-based replay attack detection in closed-loop systems using a plug & play watermarking approach. 4th Conference on Control and Fault Tolerant Systems (SysTol), pages 330-335. IEEE, 2019.

Chapter 4. A zonotopic-based watermarking design to detect replay attacks

This chapter exploits the recent analogy found between stochastic and zonotopic-based estimators to propose a deterministic counterpart of current approaches that study the replay attack in the context of stationary Gaussian processes. In this line, analogous expressions concerning the impact that a zonotopic/Gaussian watermark signal has on the system operation are derived. Additionally, a novel zonotopically bounded watermark signal that guarantees the attack detection by causing that the residual vector abandons the healthy residual set during the replay phase of the attack is presented. Finally, a four-tank system is used to both illustrate and discuss the effectiveness of the proposed approach.

The material presented in this chapter has resulted in the following publications:

– C. Trapiello, and V. Puig. A zonotopic-based watermarking design to detect replay attacks. IEEE/CAA Journal of Automatica Sinica. *Under review.*

– C. Trapiello, and V. Puig. Replay attack detection using a zonotopic KF and LQ approach. IEEE International Conference on Systems, Man, and Cybernetics (SMC), pages 3117-3122. IEEE, 2020.

– C. Trapiello, D. Rotondo, H. Sanchez, and V. Puig. Detection of replay attacks in CPSs using observer-based signature compensation. 6th International Conference on Control, Decision and Information Technologies (CoDIT), pages 1-6. IEEE, 2019.

Chapter 5. System reconfiguration with back-up components

This chapter serves as an introduction to Part II where the problem of, given a system with back-up components, select the (in some sense) optimal actuators configuration after an anomaly occurrence is studied. Consequently, the objective of this chapter is to motivate, introduce and formulate the problem, establishing thus the basis on which Chapter 6 and Chapter 7, which collect the contribution of this part of the thesis, will be developed.

Chapter 6. Reconfiguration with back-up components - Planning

This chapter addresses the reconfiguration problem with back-up components from a planner perspective. The admissibility of an actuator configuration is assessed according to its ability, at the moment that an anomaly has been detected, to generate a system trajectory that is feasible, and with an acceptable performance, for a horizon that is related with the time interval by which the system is expected to have restored its nominal service.

The configuration selection is posed as a multi-objective mixed-integer program (MIP) solved using a lexicographic approach. Aiming at reducing the worst-case execution time, the analysis of necessary properties for the existence of an admissible solution, as well as how to manage the information obtained by evaluating such properties to be included in the MIP, are investigated. Finally, a portion of a water transport network is used in order to validate the proposed solution.

The material presented in this chapter has resulted in the following publications:

- C. Trapiello, V. Puig, and G. Cembrano. Reconfiguration of large-scale systems using back-up components. Computers & Chemical Engineering, page 107288, 2021.
- C. Trapiello, V. Puig, and G. Cembrano. System reconfiguration of large-scale control systems using back-up actuators. 7th International Conference on Control, Decision and Information Technologies (CoDIT), volume 1, pages 335-340. IEEE, 2020.

Chapter 7. Reconfiguration with back-up components - Control

In this chapter, the reconfiguration of flow-based networks with back-up components that stabilize over a periodic reference trajectory is posed under the consideration of bounded uncertainties and the requirement of providing closed-loop stability guarantees in the configuration selection. The stability of the candidate configurations is imposed in the reconfiguration problem by means of a single-layer model predictive control (MPC) scheme. Besides, the optimal periodic trajectory that the system can reach with each actuator configuration is used in order to assess if the configuration performance is admissible. Furthermore, the local controller, in charge of attenuating the effect of unknown disturbances, is designed using the minimal actuators configuration that ensure an admissible performance. A case study based on a portion of a water transport network is used to validate the developments proposed in this chapter.

The material presented in this chapter has resulted in the following publication:

- C. Trapiello, V. Puig, and G. Cembrano. Reconfiguration of flow-based networks with back-up components using robust economic MPC. Journal of Process Control. *Under review.*

Part I

Replay attack detection

Chapter 2

The replay attack

This first part of the thesis addresses the replay attack detection problem following a set-theoretic framework. To that end, the consideration of norm-bounded uncertainties has been set as an initial premise that holds throughout this part of the dissertation. Accordingly, this chapter is intended to present the arguments that support a set-based approach to the detection problem, as well as to introduce the main aspects of the replay attack.

The remainder of this chapter is structured as follows: Section 2.1 introduces the secure control topic. Section 2.2 presents a literature review on the main techniques used for detecting attacks on control systems as well as on the main set-based proposals in the security field. Finally, Section 2.3 introduces the attack model that will be used throughout this part of the dissertation.

2.1 Introduction to replay attack detection

The security vulnerabilities that appear as a result of the integration between cyber and physical components in the CPSs, can be exploited by a malicious attacker whose objective is to alter the operation of the system. Those attack models that cause an effect on the system dynamics, and that are therefore studied within the field of automatic control, can be divided into three categories [Cardenas et al., 2008]: *denial of service (DoS) attacks*, where there is a lack of availability on the control and/or measurement signals; *direct attacks*, where the components of the system (plant, actuators, sensors) are physically attacked; *deception attacks*, where the control or measurement signals are altered. Focusing on the detection of the attacks presented above, although harmful, DoS attacks can be easily detected, especially when the attacker jams all the communication channels at the same time. Furthermore, the effect caused by a direct attack is similar to a fault in one of the components, and thus these attacks may be detected using the standard anomaly detectors. On the other hand, in the scenario where the integrity of the transmitted data is violated, the attack detection becomes a challenging problem that deserves an in-depth study.

Accordingly, in the sequel, *stealthy attacks* will be understood as those deception attack in which the attacker aims at remaining undetected by anomaly detectors. In order to achieve undetectability, the attacker must be able to feed the monitoring algorithms with data that is consistent with the nominal operation of the system. This can be done by deploying some of the most common stealthy attacks reported in the literature, namely: false data injection attacks Liu et al. [2011], zero dynamics attacks [Teixeira et al., 2012], covert attacks [Smith, 2011] or replay attacks [Mo and Sinopoli, 2009]. In this regard, each of the attacks listed above is based on different premises concerning the attacker's knowledge of the system, as well as its access to disclosure and disruption resources [Teixeira et al., 2015a].

In particular, this part of the thesis concentrates on the detection of replay attacks. The interest in this type of attack lies mainly in its simplicity of implementation, since the attacker does not require a detailed model of the system, but reading/writing access to the data received by the supervision algorithms. In this scenario, and under certain conditions, a malicious attacker is able to mask the effect of an unbounded attack deployed against the system while remaining undetected. As a consequence, the replay attack detection appears as a challenging problem that requires of active detection methods. In addition, the fact that replay attacks have been registered in the real world [Langner, 2011], further supports the interest in investigating them.

Moreover, a high percentage of security-related works, not only concerning attacks but also faults, follow a stochastic approach, where assumptions about the statistical properties of the uncertainties, mainly characterized through Gaussian probability distributions, are made. As remarked by Combastel [2016], this approach is often well suited to deal with measurement noises. However, on the other hand, modelling of disturbances mostly related to some lack of knowledge about deterministic behaviours (e.g. load torque of a motor under incompletely specified operating conditions) is often more representative using bounded errors than Gaussian distributions. As a matter of fact, such disturbances can successively vary arbitrarily, then temporarily remain constant but equal to unknown values, then vary again but differently, etc., and do not have any other stationary behaviour than that of remaining within specified bounds. Therefore, throughout this part of the thesis, a set-based (or norm-bounded) paradigm will be followed. It must be pointed out, that set-based techniques have proven useful in fault-related secure control, as they allow to construct sets for the system in healthy and faulty operations, so that it becomes possible to infer deterministically whether a system is working under nominal or faulty conditions [Stoican, 2011, Pourasghar-Lafmejani, 2019].

Finally, among the different set-representations used in the automatic control field, zonotopes are mainly used in the developments presented below. Zonotopic sets constitute a special type of symmetric polytopes whose shape is implicitly represented by a rectangular matrix (cf. Section A.2.2 of Appendix A). This set selection is supported by the strengths of the zonotopes, namely: a good trade-off between flexibility and reduced complexity, as well as the efficient computation of linear transformations and Minkowski sums [Le et al., 2013]. These advantages will be further exploited in the design of active detection algorithms aimed at exposing replay attacks.

2.2 Literature review

The objective of this section is twofold: Firstly, providing a literature review on the main techniques used for detecting stealthy attacks; secondly, due to its importance in the development of this dissertation, analysing the main works that use set-theoretic approaches within the context of cyber attacks. At this point, it should be noted that the literature review reflects the growing interest in aspects related to cyber-security in control systems, being the case that a large number of the referenced publications have emerged during the course of this thesis.

2.2.1 Attack detection techniques

Virtually all anomaly detection algorithms are based on the same principle: a residual signal is computed as the difference between the system measurements and its estimates. Then, the system operation is assessed according to the values adopted by this residual signal, in such a way that if it trespasses a user-defined threshold an alarm is raised. This threshold is related to the nominal, or healthy, operation of the system that the supervisor expects, and its value depends on the system modelling as well as on the characterization of the uncertainty sources that are considered coherent with a healthy behaviour.

Regarding the attack detection techniques, these follow the same classification as fault diagnosis algorithms, and can be divided into: *passive* algorithms, which observe the evolution of the system without causing any effect on it; *active* algorithms, which alter the state of the system while observing, accepting a small performance degradation in exchange for the ability to detect complex attacks [Weerakkody et al., 2016]. The main techniques that can be found in the literature concerning both detection strategies are detailed below.

Passive Methods

In the attack-related literature, passive detectors have been studied primarily to assess their limitations and investigate possible attacks that remain undetected [Pasqualetti et al., 2013, Fawzi et al., 2014, Teixeira et al., 2015a]. These vulnerabilities, have led to the proposal of different security metrics that evaluate the impact that a malicious attacker could cause on a control system without being detected [Bai et al., 2015, Milošević et al., 2019, Murguia et al., 2020], or that assess the smallest number of sensors and actuators that have to be compromised in order to successfully launch a stealthy attack [Teixeira et al., 2015b, Tang et al., 2019].

Passive detection algorithms can be mainly divided into: *stateless*, which only use the current value of the residual signal; *stateful*, which make use of previously measured residuals as well. Within the stochastic approaches, the most common stateless method is the chi-squared detector. This detector uses the normalized residual signal, which follows a chi-square distribution, to perform a statistical test. The vulnerabilities of the chi-square detector under different types of deception attacks have been extensively analysed [Mo and Sinopoli, 2009, Mo et al., 2010, Kwon et al., 2013, Mo and Sinopoli, 2015]. In addition, in Tunga et al. [2018] an intermediate stateless-stateful windowed chi-squared detector is proposed for detecting sensor attacks, whereas a chi-squared extension under non-Gaussian uncertainties has been presented in Hashemi and Ruths [2019].

Alternatively, the most widely studied stateful methods are the cumulative sum (CUSUM) detector and the multivariate exponentially weighted moving average (MEWMA) detector. Comparisons of the previous detectors, that have internal dynamics, with respect to the static chi-squared detector can be found in Murguia and Ruths [2016], Umsonst and Sandberg [2018]. In these works, it was shown that the stateful detectors are better at mitigating the impact of possible stealthy attacks on the system, albeit at the cost of a more complex implementation and an attack detection that requires more time. Additionally, in Porter et al. [2019], the aforementioned passive algorithms are compared against active techniques on a real platform.

Here, it is recalled that stealthy attacks are defined according to their capability to mislead a passive anomaly detector and achieve their objectives without being noticed. Hence, working under such a premise, the use of active detection methods becomes a necessity.

Active Methods

Most active methods fall into one of the following two categories: *moving target defense* (MTD), which changes system parameters to keep attackers from obtaining the current configuration; *watermarking-based methods*, which encrypt measurement signals with the addition of a watermark signal. Both techniques are analysed below.

- **Moving target defense:** The concept of MTD is a well-grounded topic within the field of cyber-security, and it includes actions like randomizing the order of code execution and physical memory storage locations [Jajodia et al., 2011]. Specifically within the automatic control field, that is, taking into account the system dynamics, the following works have been developed: in Teixeira et al. [2012], the authors studied how to modify the system structure (including actuators and sensors) in order to detect zero-dynamics attacks; in Giraldo et al. [2019], redundant measurements are exploited by an MTD mechanism that randomizes which sensor values an observer-based controller uses at a given time; in Kanellopoulos and Vamvoudakis [2019], the control law is randomly modified in a system with redundant actuators and sensors. Indeed, altering the control strategy is arguably similar to watermarking-based methods, whereas the methods that switch the measurement signals only work well if the attacker hijacks a subset of the sensors. Otherwise, its efficiency is similar to passive detectors.

  Alternative MTD methods suggest to extended the system dynamics through an auxiliary system whose dynamics are unknown to the adversary. On this subject, in Weerakkody and Sinopoli [2015], the auxiliary system is make up of Gaussian time-varying external states, whereas in Schellenberger and Zhang [2017] the auxiliary system is a discrete-time switched system. Besides, in Griffioen et al. [2020], three moving target designs are analysed, namely: a hybrid moving target; an auxiliary system with time-varying dynamics; the introduction of random non-linearities in the sensor measurements. On the other hand, in Ghaderi et al. [2019], a combination of a watermark signal and a non-linear auxiliary system has been proposed.

  Moving target defense allows the detection of almost all attack models, however the method makes certain assumptions about the system that is being defended. In this regard, the secret auxiliary system must either take the form of an additional physical system that is coupled with the plant, or of a simulated system that requires secure knowledge of the underlying system state. On the other hand, in the techniques that do not use an auxiliary system, it is assumed that the plant dynamics are changeable.

- **Watermarking-based methods:** The concept of physical watermarking was first proposed in Mo and Sinopoli [2009], where an independent identically distributed (IID) Gaussian signal was added to the control inputs with the intention of making the chi-squared detector robust to replay attacks. Later, a similar watermarking scheme was used in Weerakkody et al. [2014] for detecting more complex attacks (covert attacks), under the assumption of a secure subset of actuation channels. Different to previous works, the watermark signal proposed in Mo et al. [2015] is generated by a hidden Markov model, and thus auto-correlated. More recently, in Liu et al. [2020a], an IID Gaussian signal is used in a combined algorithm that deals with the online watermarking design and the unknown system parameters identification. Besides, in Fang et al. [2020], a periodical injection of a Gaussian signal is proposed for detecting discontinuous replay attacks. In addition, a stochastic game approach has been proposed in Miao et al. [2013] in order to derive an optimal control policy that switches between a control-cost optimal (but nonsecure) and watermarked (but cost-suboptimal) controllers.

The so-called dynamic watermarking techniques [Satchidanandan and Kumar, 2016, Hespanhol et al., 2017], also add an IID Gaussian signal in order to secure the systems against persistent attacks. To that end, these techniques evaluate the residuals contained in a temporally sliding window using a metric that relies on both the covariance of the residuals and the correlation between the residuals and the injected watermark. This dynamic watermarking scheme has been extended to a limited set of non-linear systems in Ko et al. [2016], and to linear time varying systems in Porter et al. [2020]. Furthermore, in Satchidanandan and Kumar [2019], the authors provide necessary and sufficient conditions that the statistics of the watermark should satisfy in order for the fundamental security guarantees to hold in a linear time-invariant system affected by noise having an arbitrary distribution. Additionally, dynamic watermarking has been used to detect attacks against the automatic generation control unit of power systems in Huang et al. [2018].

All the watermarking techniques presented so far, are based on the additive introduction of an exogenous signal on the control action. Conversely, other watermarking approaches address the detection of stealthy attacks by multiplicatively affecting the system outputs. In this regard, in Miao et al. [2014], the sensor outputs are coded by means of a multiplicative coding matrix that mixes the outputs. On the other hand, a modular multiplicative scheme where the outputs are independently secured by a watermarking generator allocated at the plant side of the communication network, and later the watermark is removed at the controller side of the network, has been proposed to detect false-data injection attacks [Teixeira and Ferrari, 2018], replay attacks [Ferrari and Teixeira, 2017a], and rerouting attacks [Ferrari and Teixeira, 2017b]. Recently this scheme was extended in Ferrari and Teixeira [2020], by designing a switching protocol with no communication overhead to allow the watermarking filters to synchronously update their parameters.

Other physical watermarking schemes like Ozel et al. [2017] propose the injection of Bernoulli packet drops into the control inputs, while in Weerakkody et al. [2017] the authors suggest a combination of Bernoulli packet drops altogether with an additive Gaussian signal. Moreover, in Khazraei et al. [2017], the attack is detected by destabilizing the residuals in a stationary Gaussian process. In Sánchez et al. [2019b], a sinusoidal signal with a time-varying frequency is devised as a possible watermark signal. Furthermore, in Romagnoli et al. [2019], the watermark proposal is based on B-splines added to feed-forward inputs.

2.2.2 Set-theoretic techniques in attack characterization, detection and tolerant control

Set-based anomaly detection algorithms follow exactly the same principle as stochastic ones. However, in this case, the threshold of the residual signal is derived by computing (or approximating) the reachable set of the healthy system. In particular, for steady-state operation, the guarantees of positive invariant sets can be exploited in order to assess the system operation in an efficient manner. This was done in the computation of security metrics presented in Murguia et al. [2020], where an ellipsoidal approximation of the robust positive invariant set of the residual signal was employed.

Set-theoretic techniques has been particularly used for evaluating attacks on the load frequency control loop of power networks. On this subject, in Esfahani et al. [2010] reachability methods were used to quantify the attack impact on a two-area power plant. Besides, the maximal invariant set for a closed-loop constrained system has been firstly used in Kontouras et al. [2017], and later in Kontouras et al. [2018], to detect bias injection attacks in a single-area/networked power plant, respectively. In the networked case, if the bias is injected in all the sensors of the grid, then the attack effect on the system states guarantees the detection for high values of the bias. On the other hand, if the attack only affects a subset of the areas, some states of the plant become unstable, guaranteeing thus the attack detection. At this point it must be highlighted that previous works assess the system operation using sets related with the control design stage, e.g. maximal invariant set for a constrained system, instead of generating the residual signal though a closed-loop estimation of the system states. This way of proceeding presents a double problem: I) it implicitly assumes knowledge of the system state at initial time; II) it does not base the attack detection on the tightest estimation of the healthy system evolution, being thus less sensitive to anomalies. Additionally, in Zhang and Zhu [2020], the interval hull of the residual zonotope is used to detect bias injection attacks. Observe that the stealthiness of the bias injection attacks is related with the magnitude of the injected bias and not with the violation of the data integrity, in other words, they can be analysed similarly to a components fault.

In addition, set-theoretic approaches have been used for designing attack resilient control strategies. In this regard, in Lucia et al. [2016], a secure control framework that combines a delayed set-theoretical receding horizon controller with an anomaly detector though for false data injection attacks is proposed. Furthermore, in Franzè et al. [2019], the authors present a resilient control strategy against replay attacks that couples a set-theoretic controller, that is resilient to certain delays, with an auxiliary MPC controller that comes into play when the attack is detected. Observe that the detection algorithms used in these two works, are also subject to the same arguments exposed in the paragraph before regarding the detection technique. Recently, in Abdelwahab et al. [2020], the authors combined a set-theoretic controller with a watermarking signal that imposes packet-drops providing stability guarantees in the attack-free case.

Another growing-body of set-theoretic works within the cyber-security field is related with the state estimation of systems under attack. Therefore, in Song et al. [2019], Zhang et al. [2020], Qu and Pang [2020] set-membership techniques have been used for secure state estimation in systems under bounded attacks. Moreover, in Shinohara and Namerikawa [2018], set-theoretic techniques are used to propose a secure estimator which minimizes the estimation error under a worst-case manipulation of a subset of sensors. Finally, a distributed set-membership state estimation scheme for replay attacks affecting a sensor network is presented in Liu et al. [2020b]. Here, the replay attack detection is assumed to be done on the cryptographic layer.

2.3 Replay attack against control systems

Replay attacks are a type of deception attacks in which the adversary replaces the real-time measurements coming from the sensors, or the control actions sent to the actuators, with previously recorded data. This attack is often depicted in movies, in which security videos are recorded and later replayed to hide thefts or sabotages.

Regarding the attack implementation, as remarked by Mo and Sinopoli [2009], the replay attacks constitutes a really plausible attack scenario since the adversary does not need any knowledge about the attacked system, except for being aware of the fact that the system itself will be in steady-state during the attack. In addition, any further knowledge of the system dynamics can be exploited in order to conduct an efficient physical attack on the system components, or in the smart design of an external signal injected in the plant inputs with malicious objectives.

Replay attacks constitute a two-phase attack, so that, in a first phase, a malicious attacker secretly records the data transmitted through the communication network. This first phase of the attack is also known as *eavesdropping attack*. Later, the recorded data are replayed with the intention of masking an attack conducted over the plant. Throughout this part of the thesis, it will be consider the case where an attacker is able to directly affect the system states through a physical attack, e.g., liquid theft from different tanks of a water network. Consequently, by considering that $k \in \mathbb{N}$ denotes the sampling time on a discrete-time system model, the following time windows are defined:

1. **Record window:** transmitted data are assumed to be recorded for $\mathcal{K}_{REC} = \{k \in \mathbb{N} : k \in [k_0, \;\; k_0 + l - 1]\}$, where $l \in \mathbb{N}$ denotes the size of the record window.
2. **Replay window:** real data are replaced for $\mathcal{K}_{REP} = \{k \in \mathbb{N} : k \in [k_1, \; k_1 + l - 1]\}$, with $k_1 \geq k_0 + l$.
3. **Physical attack window:** a physical attack against the plant is launched for $\mathcal{K}_{PHY} = \{k \in \mathbb{N} : k \in [k_2, \; k_3]\} \subseteq \mathcal{K}_{REP}$, i.e., $k_2 \geq k_1$ and $k_3 \leq k_1 + l - 1$.

Note that in the time windows defined above, it is implicitly assumed that the recorded sequence is only replayed once. However, registered attacks have demonstrated that a more realistic scenario is the case where the attacker records a short sequence of data which is then replayed in a loop [Langner, 2011]. Nevertheless, due to the linearity of the systems under study, the analysis developed in subsequent chapters holds for this latter scenario just by shifting the initial time instant k_1 to match the start of each replayed sequence.

2.3.1 Specific notation

Throughout this part of the thesis, whenever variables from different time windows are being compared, superscripts ^{r} and ^{a} will be used in order to differentiate that a system variable is in the record and in the replay phase, respectively. That is, for example, given a variable z_k, for all $k \in \mathcal{K}_{REP}$ then $z_k^a = z_k$, whereas $z_k^r = z_{k+(k_0-k_1)}$.

Chapter 3

Zonotopic SIA for replay attacks affecting the supervisory layer

This chapter presents a set-invariance approach (SIA) to the detectability of replay attacks affecting the communication network that serves the supervisory layer of complex control systems. Depending on the attacker access to the system resources, two scenarios are considered: I) Sensors and controller data can be counterfeited; II) Only sensor measurements can be counterfeited. The effect of a physical attack against the plant during the data replay is also taken into consideration. The representation of invariant sets as zonotopes, allows deriving analytical expressions for the steady-state regarding the attack detectability under the presence of bounded uncertainties. Moreover, the zonotopic characterization of the residual set is further exploited in the offline design of a set of watermark sequences. These sequences are intended to be injected asynchronously into the control loop, and their detection guarantees are related with the number thereof. The performance of the proposals is validated through simulations using a four-tank process.

3.1 Introduction

The vast majority of works that study the effects and the detectability of replay attacks on control systems, consider that the control loop is closed remotely by means of a communications network which is prone to be corrupted by a malicious attacker. On this subject, the standard replay attack formulation considers that the control loop has been affected either by deceiving the system controller with previously recorded output measurements [Mo and Sinopoli, 2009, Miao et al., 2013, Ferrari and Teixeira, 2017a, Fang et al., 2020], or by directly replaying back previous control actions to the system actuators [Zhu and Martinez, 2013a,b].

Nevertheless, in many industrial control systems the low-level controller, allocated in the *regulatory layer*, makes use of dedicated networks which are hard to access. Whereas, on the other hand, it is common that the *supervisory layer*, in charge of the system monitoring and set-point reference generation, operates remotely by means of a vulnerable communications network. This difference between regulatory and supervisory layers, which is normally done when analysing the effect of faults, has not been widely considered in the analysis of replay attacks. Accordingly, this chapter aims to analyse the different replay attack scenarios that may appear on systems with the regulation/supervision layer scheme described above.

As exposed in Chapter 2, while the majority of security-related works follow a stochastic paradigm, here, the replay attack detection problem is addressed following a set-based approach. In particular, note that the inherent steady-state conditions of replay attacks fit perfectly with the use of set-invariance concepts [Blanchini and Miani, 2008]. In this line, by means of the computation of the healthy/attacked residual invariant sets, the system evolution can be assessed, in such a way that guaranteed detection is achieved whenever the residual signal lies outside the healthy residual set.

Moreover, the invariant set computation is performed using zonotopic sets (cf. Appendix A). In this regard, the strengths of this particular set-representation are used for the efficient design of finite watermark signals that allow to impose the attack detection. Additionally, the zonotopic characterization carried out below is further exploited in the developments presented in Chapter 4.

A positive SIA has been followed in order to face the detectability of replays attack compromising the communication between the regulatory and supervisory layers of a control system, while the low-level controller at the regulatory layer remains unaffected. The main contributions of this chapter are the following:

- Analytical expressions concerning the attack detectability are obtained for the steady-state under the assumption of bounded disturbances. These expressions are derived as a function of the set-point reference signal imposed from the supervisory layer and an external attack conducted over the plant. The following scenarios have been considered: I) the attacker is able to record and replay sensors and controller data, causing the supervisory layer to completely operate based on false data; II) the attacker is able to replay sensors data to the supervisory layer while the low-level control actions are received unaffected. This scenario models the case where the supervisory layer operates based on a set of sensors installed for system monitoring and that may differ from the sensors used for control.

- The zonotopic set-invariance representation is used to compute a set of finite watermark sequences. Unlike other approaches that continuously degrade the system operation, these sequences are thought to be injected asynchronously into the control loop, and have been designed in such a way that guarantee the attack detection as long as the recording data does not contain the exact sequence that it is being injected at the current moment.

The remainder of this chapter is organized as follows: First, Section 3.2 describes the system in healthy operation. Then, Section 3.3 details the assumptions on the system operation during the record phase. In Section 3.4, the steady-state analytical expressions regarding attack detection are derived. Section 3.5 introduces the guaranteed watermark sequences design. An illustrative example is presented in Section 3.6. Finally, in Section 3.7 the main conclusions of the chapter are drawn.

3.2 Problem statement

This section focuses on discrete-time linear time-invariant (LTI) systems of the form

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Ew_k, \\ y_k &= Cx_k + Fv_k, \end{aligned} \tag{3.1}$$

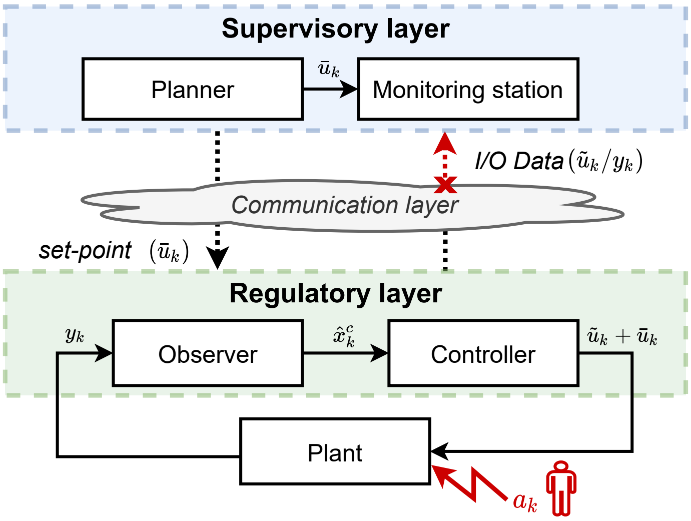


Figure 3.1: Overall scheme. Solid (dotted) lines represent local (remote) connections.

where A, B, C, E and F are the state-space matrices with adequate dimensions, $x_k \in \mathbb{R}^{n_x}$ is the state vector, $u_k \in \mathbb{R}^{n_u}$ is the applied control action and $y_k \in \mathbb{R}^{n_y}$ corresponds to the sensor measurements at time instant k. Furthermore, $w_k \in \mathbb{R}^{n_w}$ and $v_k \in \mathbb{R}^{n_v}$ represent the process disturbances and measurement noise, respectively.

Assumption 3.1. The pair (A, B) is asymptotically stabilizable and the pair (A, C) is asymptotically detectable.

Assumption 3.2. For all $k \in \mathbb{N}$, process disturbances and measurement noise satisfy $w_k \in \mathcal{W}$ and $v_k \in \mathcal{V}$, where

$$\mathcal{W} = \langle c_w, H_w \rangle, \qquad \mathcal{V} = \langle c_v, H_v \rangle, \tag{3.2}$$

and $H_w \in \mathbb{R}^{n_w \times m_w}$, $H_v \in \mathbb{R}^{n_v \times m_v}$ are known generators matrices while $c_w \in \mathbb{R}^{n_w}$, $c_v \in \mathbb{R}^{n_v}$ are known center vectors.

Regarding the validity of Assumption 3.2, note that element-wise constraints on the disturbances, i.e., compact interval sets, are also zonotopes which can be easily formulated as in (3.2). On the other hand, in Stoican et al. [2011], the authors discuss different methods in order to zonotopically approximate generic disturbance sets expressed as polytopic or ellipsoidal sets.

The control objective of the system is to regulate the plant tracking error, defined at each sample as $z_k = x_k - \bar{x}_k$, where $\bar{x}_k \in \mathbb{R}^{n_x}$ is the reference signal governed by

$$\bar{x}_{k+1} = A\bar{x}_k + B\bar{u}_k, \tag{3.3}$$

and $\bar{u}_k$ denotes the feedforward action that must be provided at each time instant in order to track the reference state trajectory $\bar{x}_k$. For a desired output set-point $\bar{y}_k = C\bar{x}_k$, the corresponding signal $\bar{u}_k$ can be computed by means of classical model inversion-based feedforward schemes [Franklin et al., 2002].

The system under study is displayed in Figure 3.1. In the regulatory layer, a local low-level controller is in charge of regulating the plant tracking error. On the other hand, in the supervisory layer, two tasks are carried out remotely: I) the generation of the set-point reference $\bar{u}_k$; II) the monitoring of the system operation by means of an anomaly detector.

3.2.1 Regulatory layer

In order to satisfy the control objective, the low-level controller is assumed to perform an estimate-feedback control action based on the estimates provided by its own state observer. Denoting the state estimates generated by the low-level observer as $\hat{x}_k^c \in \mathbb{R}^{n_x}$, then, under bounded uncertainties like (3.2), the estimation error $\eta_k = x_k - \hat{x}_k^c$ can be bounded using a set-based observer [Alamo et al., 2005, Combastel, 2015, Wang et al., 2018b]. In this regard, the following assumption concerning the bounds of η_k is introduced.

Assumption 3.3. For all $k \in \mathbb{N}$, the controller estimation error η_k lies within the zonotopic set

$$\eta_k \in \mathcal{H} = \langle c_\eta,\ H_\eta \rangle, \tag{3.4}$$

with generators matrix $H_\eta \in \mathbb{R}^{n_x \times m_\eta}$ and center $c_\eta \in \mathbb{R}^{n_x}$.

The fixed bound on n_k imposed in Assumption 3.3, has been considered in order to simplify the notation as well as the analysis that will be developed. Nevertheless, this assumption aligns with the steady-state conditions that are imposed in Assumption 3.4, which will be introduced later.

Hence, by taking into consideration the feedforward reference input $\bar{u}_k$ sent from the supervisory layer, the control signal is composed by

$$u_k = \bar{u}_k + \tilde{u}_k, \tag{3.5}$$

where $\tilde{u}_k$ denotes the estimate-feedback action computed by the low-level controller

$$\tilde{u}_k = -K(\hat{x}_k^c - \bar{x}_k) = -K(\hat{x}_k^c - x_k + x_k - \bar{x}_k) = -Kz_k + K\eta_k. \tag{3.6}$$

Under control law (3.5), the tracking error dynamics are governed by

$$z_{k+1} = x_{k+1} - \bar{x}_{k+1} = (A - BK)z_k + BK\eta_k + Ew_k, \tag{3.7}$$

with the controller gain K designed such that the matrix $A - BK$ is asymptotically stable, i.e., all its eigenvalues are strictly inside the unit disk. Accordingly, (3.7) represents an asymptotically stable dynamical system subject to bounded disturbances.

3.2.2 Supervisory layer

On the other hand, in the supervisory layer, an anomaly detector monitors the plant operation. For this purpose, a generic Luenberger observer is considered as follows

$$\hat{x}_{k+1} = (A - LC)\hat{x}_k + Bu_k + Ly_k, \tag{3.8}$$

where $\hat{x}_k \in \mathbb{R}^{n_x}$ represents the state estimation vector.

By denoting the estimation error as $e_k = x_k - \hat{x}_k$, then, comparing (3.1) and (3.8), it follows that the estimation error dynamics are governed by

$$e_{k+1} = x_{k+1} - \hat{x}_{k+1} = (A - LC)e_k + Ew_k - LFv_k, \tag{3.9}$$

with the observer gain L designed such that the matrix $A - LC$ is asymptotically stable. Accordingly, (3.9) represents an asymptotically stable dynamical system subject to bounded disturbances.

Assumption 3.4. The system is in stationary operation such that $k \geq k^*$, with $k^* \in \mathbb{N}$ a finite sample by which the trajectories of (3.7) and (3.9) have converged into their respective mRPI sets.

Note that, for linear asymptotically stable systems subject to bounded disturbances like (3.7) and (3.9), its system trajectories are guaranteed to converge into their respective mRPI sets [Kolmanovsky and Gilbert, 1998]. Hence, Assumption 3.4 imposes that the system has been operative for a sufficiently large period of time without being compromised. Observe that this assumption matches the inherent stationary assumption that is present in the analysis of replay attacks in stationary Gaussian processes [Mo and Sinopoli, 2009].

In the sequel, the mRPI set for system (3.9) is denoted as $\mathcal{E}_m$. In this regard, it is well-known that the exact computation the mRPI set can only be achieved under the restrictive assumption that the system dynamics are nilpotent [Mayne and Schroeder, 1997]. Hence, throughout the chapter, an RPI zonotopic outer-approximation $\mathcal{E} = \langle c_e, H_e \rangle \supseteq \mathcal{E}_m$ will be employed. In order to compute the set $\mathcal{E}$, Algorithm A.1 in Appendix A is used. The zonotope center and the generators matrix recursion used by the algorithm are

$$c_e = \Lambda^{-1}(Ec_w - LFc_v), \tag{3.10a}$$

$$H_{e,j+1} = [(A - LC)H_{e,j},\ EH_w,\ -LFH_v], \tag{3.10b}$$

where $\Lambda = I_{n_x} - (A - LC)$.

3.2.2.1 Anomaly detector

The presence of anomalies affecting the system is monitored based on the values adopted by the following residual signal

$$r_k = y_k - C\hat{x}_k = C(x_k - \hat{x}_k) + Fv_k = Ce_k + Fv_k. \tag{3.11}$$

Therefore, taking into account Assumption 3.4 and that $\mathcal{E}$ is a zonotope, in healthy operation the residual signal (3.11) is guaranteed to lie within the residual zonotopic set written as

$$\mathcal{R}^H = C\mathcal{E} \oplus F\mathcal{V} = \langle c_r^h, H_r^h \rangle, \tag{3.12}$$

where, by means of the set operations in Appendix A, it follows that

$$c_r^h = C\Lambda^{-1}(Ec_w - LFc_v) + Fc_v, \tag{3.13a}$$

$$H_r^h = [CH_e,\ FH_v]. \tag{3.13b}$$

Hence, from the online evaluation of the residual signal, the following can be established

$$\begin{cases} r_k \in \mathcal{R}^H & \implies \text{Healthy system,} \\ \text{otherwise} & \implies \text{Something is wrong.} \end{cases}$$

Remark 3.1. In order to test whether a point belongs to a given zonotope or not, the constraint satisfaction problem in Property A.10 can be solved. Also, other algorithms like the Gilbert-Johnson-Keerthi (GJK) algorithm as proposed in Lalami and Combastel [2006] can be used in order to test it.

From this point forward, use is made of the description of the attack and the nomenclature presented in Section 2.3.

3.3 Record phase

This section introduces several aspects concerning the system dynamics during the record phase that play a fundamental role in the detectability of the attack during the replay phase. Furthermore, depending on the capabilities of the attacker to access the different system resources the following attack scenarios are considered:

- **Scenario I:** the attacker has gained access to the input/output data sent by the low-level controller to the monitoring center. Thus, the recorded data sets are $\mathcal{Y} = \{y_k : k \in \mathcal{K}_{REC}\}$ and $\mathcal{U} = \{\tilde{u}_k : k \in \mathcal{K}_{REC}\}$.
- **Scenario II:** the attacker is able to access only the output sensors data, and thus the recorded data set is $\mathcal{Y} = \{y_k : k \in \mathcal{K}_{REC}\}$. Note that this scenario models also the case where the monitoring center operates based on a set of input/output sensors installed for monitoring the plant operation and the attacker has gained access to the output sensors only. These sensors may differ from the ones used by the controller to close the low-level control loop.

For both scenarios, it is considered that the attacker records the system in healthy operation and that the reference signal sent from the supervisory layer satisfies the following assumption.

Assumption 3.5. The set-point signal $\bar{u}^r_k$ is fixed.

Note that this is a reasonable assumption considering that a rational attacker aims at obtaining coherent sets of data. Consequently, this attacker would be interested in recording data of the system in steady-state, since the recording of system outputs during a transient would indeed yield the attack detection during the data replay.

Furthermore, from Assumption 3.4, it follows that the residual signal has converged into the healthy residual set (3.12), and thus

$$r_k \in \mathcal{R}^H, \quad \forall k \in \mathcal{K}_{REP}. \tag{3.14}$$

3.4 Replay attack detectability

This section analyses the detectability of the replay attack scenarios presented in Section 3.3. In this regard, it is considered that during the attack phase, the attacker replays previous measurement with the intention of masking a physical attack conducted over the plant.

Henceforth, the effect of the attacker in the plant dynamics is modelled by means of the attack signal $a_k \in \mathbb{R}^{n_x}$. According to the time windows defined in Section 2.3, the attack vector satisfies

$$a_k \begin{cases} \neq 0 & \text{if } \ k \in \mathcal{K}_{PHY}, \\ = 0 & \text{otherwise}, \end{cases} \tag{3.15}$$

and thus, for all $k \in \mathcal{K}_{REP}$, the system dynamics are governed by

$$x^a_{k+1} = Ax^a_k + Bu^a_k + Ew^a_k + a_k, \tag{3.16}$$

starting at the initial state $x^a_{k_1} = x_{k_1}$.

Remark 3.2. The analysis performed below is also applicable to the case in which a_k is injected through the input matrix, i.e., substituting a_k in (3.16) for Ba'_k with $a'_k \in \mathbb{R}^{n_u}$. This case would describe cyber-attacks that are able to modify the input signal sent from the supervisory layer.

3.4.1 Regulatory layer

Accordingly with the attack model studied in this chapter, it is considered that the attacker is unable to access the dedicated network of the low-level controller, and thus the regulatory control loop remains healthy. However, the effect of the attack signal a_k on the system dynamics will affect the controller capability to regulate the tracking error.

In this regard, the control signal injected during the replay phase is

$$u_k^a = \bar{u}_k^a + \tilde{u}_k^a,$$

where $\bar{u}_k^a$ denotes the reference signal that is being sent from the supervisory layer, and

$$\tilde{u}_k^a = -Kz_k^a + K\eta_k^a, \tag{3.17}$$

with $\eta^a \in \mathcal{H}$.

Accordingly, the tracking error dynamics are governed by

$$z_{k+1}^a = x_{k+1}^a - \bar{x}_{k+1}^a = (A - BK)z_k^a + Ew_k^a + BK\eta_k^a + a_k, \tag{3.18}$$

that is, since the controller remains healthy it will react to the attack signal a_k.

3.4.2 Scenario I - Supervisory layer

The first attack scenario considers that the malicious attacker replays back the data sets $\mathcal{Y}$ and $\mathcal{U}$. Therefore, from the supervisory layer point of view, the received signals are

$$\begin{cases} y_k^a &= y_k^r, \\ \tilde{u}_k^a &= \tilde{u}_k^r, \end{cases} \qquad k \in \mathcal{K}_{REP}.$$

Consequently, the residual signal during the replay phase can be written as

$$r_k^a = y_k^r - C\hat{x}_k^a = (y_k^r - C\hat{x}_k^r) + (C\hat{x}_k^r - C\hat{x}_k^a) = r_k^r + C(\hat{x}_k^r - \hat{x}_k^a), \tag{3.19}$$

with the dynamics of the observer during the record and replay phases evolving as

$$\hat{x}_{k+1}^r = (A - LC)\hat{x}_k^r + B\tilde{u}_k^r + B\bar{u}_k^r + Ly_k^r, \tag{3.20a}$$
$$\hat{x}_{k+1}^a = (A - LC)\hat{x}_k^a + B\tilde{u}_k^r + B\bar{u}_k^a + Ly_k^r, \tag{3.20b}$$

and starting at $\hat{x}_{k_1}^a = \hat{x}_{k_1}$ and $\hat{x}_{k_1}^r = \hat{x}_{k_0}$.

Hereafter, the difference in the estimation between record and replay phases is denoted as $p_k = \hat{x}_k^r - \hat{x}_k^a$. Thus, from the comparison of (3.20a) and (3.20b), it follows

$$p_{k+1} = \hat{x}_{k+1}^r - \hat{x}_{k+1}^a = (A - LC)p_k + B\Delta\bar{u}_k, \tag{3.21}$$

where $\Delta \bar{u}_k = \bar{u}_k^r - \bar{u}_k^a$ represents the difference in the reference signal between phases.

Since system (3.21) is not affected by uncertainties, its evolution can be rewritten in zonotopic form as $\mathcal{P}_k = \langle p_k, 0 \rangle$. Hence, by taking into consideration (3.14), from (3.19) it follows that the residual set under attack is

$$\mathcal{R}_k^A = \mathcal{R}^H \oplus C\mathcal{P}_k = \langle c_r^h, H_r^h \rangle \oplus \langle Cp_k, 0 \rangle = \langle c_r^h + \delta c_k, H_r^h \rangle, \tag{3.22}$$

where $\delta c_k = c_{r,k}^a - c_r^h = Cp_k$ denotes the center difference, which, for all $k \in \mathcal{K}_{REP}$, evolves according to

$$\delta c_k = C(A - LC)^{k-k_1}(\hat{x}_{k_1}^r - \hat{x}_{k_1}^a) + C \sum_{i=1}^{k-k_1} (A - LC)^{i-1} B \Delta \bar{u}_{k-i}. \tag{3.23}$$

At this point, the following definition regarding attack detectability is introduced.

Definition 3.1. Guaranteed attack detection is achieved if and only if

$$\mathcal{R}^H \cap \mathcal{R}_k^A = \emptyset,$$

at some $k \in \mathcal{K}_{REP}$.

Note that, if the reference signal between record and replay phases is the same, i.e., $\Delta \bar{u}_k = 0$, then from the asymptotic stability of matrix $A - LC$ it follows that the center displacement vector in (3.23) will settle at $\delta c_k = 0$, and thus $\mathcal{R}_k^A$ tends to $\mathcal{R}^H$. In other words, unless the attack is detected in the transient induced by the data substitution, healthy and attacked residual sets would match perfectly generating a completely undetectable attack.

3.4.2.1 Steady-state attack detectability

In this section, the set invariance properties are exploited with the intention of deriving analytical expressions regarding the separability of the residual sets. To that end, a fixed set-point modification between phases is considered. That is, $\bar{u}_k^a$ is a constant reference signal imposed in the attack phase, and thus, by taking into account Assumption 3.5, it follows that $\Delta \bar{u}_k = \Delta \bar{u} =$ const.

Furthermore, since this chapter uses zonotopic sets, the set separation condition introduced in Definition 3.1 will be assessed by means of the zonotopic interpretation of Lemma 2.1 in Dobkin et al. [1993]. That lemma is rewritten in zonotopic form in Scott et al. [2014], and it is recalled below for completeness.

Lemma 3.1. Let $\mathcal{Z} = \langle a_z + b_z, H_z \rangle$ and $\mathcal{Y} = \langle a_y + b_y, H_y \rangle$. Then, $\mathcal{Z} \cap \mathcal{Y} = \emptyset$ if and only if $a_y - a_z \notin \langle b_z, H_z \rangle \oplus \langle -b_y, H_y \rangle$.

Below, the attack separability will be analysed with respect the output set-point imposed from the supervisory layer. This responds to the fact that, by referring the separability conditions to the output reference imposed during the record and replay phases, then the attack detectability can be easily related with the induced performance loss with respect a target set-point. In this regard, the following proposition is obtained.

Proposition 3.1. Guaranteed attack detection is achieved in the steady-state if the output set-point difference between record and replay phases $\Delta\bar{y}_k = \bar{y}_k^r - \bar{y}_k^a$ fulfills

$$\Delta\bar{y}_k \notin \langle 0, 2M^{-1}H_r^h \rangle, \tag{3.24}$$

with $M = (I_{n_y} - C\Lambda^{-1}L)$.

Proof. From $\Delta\bar{u}_k = \Delta\bar{u} = \text{const.}$ and by taking into consideration the asymptotic stability of $A - LC$, then from (3.23) it follows that the displacement of the residual set center settles at

$$\delta c = c_r^a - c_r^h = C\Lambda^{-1}B\Delta\bar{u}. \tag{3.25}$$

Besides, by denoting as $\bar{y}_k^a$ the fixed set-point generated by $\bar{y}_k^a = C(I_{n_x} - A)^{-1}B\bar{u}_k^a$, from the linearity of the reference model (3.3) it follows that

$$\Delta\bar{y} = C(I_{n_x} - A)^{-1}B\Delta\bar{u}. \tag{3.26}$$

Hence, by taking into consideration the equality[2]

$$\Lambda^{-1} = (I_{n_x} - A)^{-1} - \Lambda^{-1}LC(I_{n_x} - A)^{-1}, \tag{3.27}$$

by means of (3.26) and (3.27), (3.25) can be rewritten as

$$\begin{aligned} \delta c &= C((I_{n_x} - A)^{-1} - \Lambda^{-1}LC(I_{n_x} - A)^{-1})B\Delta\bar{u} = \\ &= (I_{n_y} - C\Lambda^{-1}L)\Delta\bar{y} = \\ &= M\Delta\bar{y}. \end{aligned} \tag{3.28}$$

Therefore, by recalling that the steady-state attacked residual set is $\mathcal{R}^A = \langle c_r^h + \delta c, H_r^h \rangle$ and by realizing that $\langle 0, H_r^h \rangle \oplus \langle 0, H_r^h \rangle = \langle 0, 2H_r^h \rangle$, then it follows that the satisfaction of (3.24) guarantees $\mathcal{R}^H \cap \mathcal{R}^A = \emptyset$ from Lemma 3.1. □

Proposition 3.1 provides and analytic bound for switching the system set-points in such a way that the incoherence in the data induced by the temporal mismatch between set-points guarantees the attack detectability. It must be highlighted that in attacks affecting the supervisory layer, the stability of the system is preserved despite the modifications in the operation point, since the regulatory layer remains unaffected. Note that this may be not true if the attack affects the regulatory layer, where the data replay breaks the control loop, and thus no guarantees regarding the system stability can be given. Besides, it must be remarked that the vector δc_k is independent of the physical attack a_k, and thus its presence remains masked to the supervisory layer.

[2] Equality derivation:

$$\begin{aligned} I_{n_x} - A &= (I_{n_x} - (A - LC)) - LC \\ (I_{n_x} - (A - LC))^{-1}(I_{n_x} - A) &= I_{n_x} - (I_{n_x} - (A - LC))^{-1}LC \\ (I_{n_x} - (A - LC))^{-1} &= (I_{n_x} - A)^{-1} - (I_{n_x} - (A - LC))^{-1}LC(I_{n_x} - A)^{-1} \end{aligned}$$

3.4.3 Scenario II - Supervisory layer

According to Section 3.3, in scenario II, the attacker only has access to the output sensors data, and thus replays back the data set $\mathcal{Y}$. Therefore, from the supervisory layer point of view, the received signals are

$$\begin{cases} y_k^a &= y_k^r, \\ \tilde{u}_k^a &= \tilde{u}_k^a, \end{cases} \qquad k \in \mathcal{K}_{REP}.$$

Accordingly, the residual signal during the replay phase is given by

$$r_k^a = y_k^r - C\hat{x}_k^a = C(x_k^r - \hat{x}_k^a) + Fv_k^r. \tag{3.29}$$

Below, the difference between the system states during the record phase and the estimations generated by the observer during the replay phase is denoted as $\breve{x}_k = x_k^r - \hat{x}_k^a$. The dynamics of this signal evolve according to

$$\breve{x}_{k+1} = x_{k+1}^r - \hat{x}_{k+1}^a = (A - LC)\breve{x}_k + B\Delta\bar{u}_k + B(\tilde{u}_k^r - \tilde{u}_k^a) + Ew_k^r - LFv_k^r. \tag{3.30}$$

Note that, in (3.30) the healthy (resp. attacked) low-level control signal $\tilde{u}_k^r$ in (3.6) (resp. $\tilde{u}_k^a$ in (3.17)) depends on the tracking error dynamics given by

$$z_{k+1}^r = (A - BK)z_k^r + Ew_k^r + BK\eta_k^r, \tag{3.31a}$$
$$z_{k+1}^a = (A - BK)z_k^a + Ew_k^a + BK\eta_k^a + a_k, \tag{3.31b}$$

and thus the presence of several interconnected dynamic systems makes it difficult to obtain expressions regarding the detectability for the transient.

In this regard, a set invariance analysis like the one performed in Section 3.4.2.1 is carried out. To that end, with the purpose of analysing the evolution of the interconnected systems (3.30), (3.31a) and (3.31b), its dynamics are gathered in the extended vector $q_k = \begin{bmatrix}\breve{x}_k^T & z_k^{rT} & z_k^{aT}\end{bmatrix}^T$, such that

$$q_{k+1} = \Theta q_k + \Pi d_k + \Sigma\Delta\bar{u}_k + \Phi a_k, \tag{3.32}$$

where vector $d = \begin{bmatrix}w_k^{rT} & w_k^{aT} & v_k^{rT} & \eta_k^{rT} & \eta_k^{aT}\end{bmatrix}^T$ encompasses the different disturbances and the augmented system matrices are

$$\begin{aligned} \Theta &= \begin{bmatrix} A-LC & -BK & BK \\ 0 & A-BK & 0 \\ 0 & 0 & A-BK \end{bmatrix}, & \Sigma &= \begin{bmatrix} B \\ 0 \\ 0 \end{bmatrix}, \\ \Pi &= \begin{bmatrix} E & 0 & -LF & -BK & BK \\ E & 0 & 0 & -BK & 0 \\ 0 & E & 0 & 0 & -BK \end{bmatrix}, & \Phi &= \begin{bmatrix} 0 \\ 0 \\ I_{n_x} \end{bmatrix}. \end{aligned} \tag{3.33}$$

3.4.3.1 Steady-state attack detectability

Similar to Section 3.4.2.1, this section considers that $\bar{u}_k^a$ is a constant reference, and thus, from Assumption 3.5, that $\Delta\bar{u}_k = \Delta\bar{u} = \text{const}$. Furthermore, the following assumption is introduced regarding the attack signal.

Assumption 3.6. The physical attack a_k is performed abruptly and is kept constant over the attack set $\mathcal{K}_{PHY}$, i.e., $a_k = \bar{a} = \text{const.}$ for all $k \in \mathcal{K}_{PHY}$.

Consequently, since matrix Θ in (3.32) is an upper triangular block matrix with asymptotically stable matrices in the main diagonal, a zonotopic over-approximation $\mathcal{Q} = \langle c_q, H_q \rangle \supseteq \mathcal{Q}_m$ of the mRPI set $\mathcal{Q}_m$ for system (3.32) can be computed through Algorithm A.1, obtaining

$$\begin{aligned} c_q &= \left[I_{3n_x} - \Theta\right]^{-1} (\Pi c_d + \Sigma \Delta \bar{u}_k + \Phi a_k), \\ H_{q,j+1} &= \left[\Theta H_{q,j}, \quad \Pi \, diag(H_w, H_w, H_v, H_\eta, H_\eta)\right], \end{aligned} \tag{3.34}$$

where $c_d = \left[c_w^T \; c_w^T \; c_v^T \; c_\eta^T \; c_\eta^T\right]^T$ and

$$\left[I_{3n_x} - \Theta\right]^{-1} = \begin{bmatrix} \Lambda^{-1} & -\Lambda^{-1}BK\Gamma^{-1} & \Lambda^{-1}BK\Gamma^{-1} \\ 0 & \Gamma^{-1} & 0 \\ 0 & 0 & \Gamma^{-1} \end{bmatrix},$$

with $\Gamma = I_{n_x} - (A - BK)$.

Therefore, the set $\mathcal{Q}$ can be used in order to derive analytical expressions regarding the attack detectability in terms of the output set-point difference between record and replay phases $\Delta \bar{y}_k$, and the magnitude of the attack signal $\bar{a}$.

Proposition 3.2. Guaranteed attack detection is achieved in the steady-state if the output set-point difference $\Delta \bar{y}_k$ and attack vectors $\bar{a}$ satisfy

$$M \Delta \bar{y}_k + C\Lambda^{-1}BK\Gamma^{-1}\bar{a} \notin \langle 0, \left[H_r^h, \; H_r^a\right] \rangle, \tag{3.35}$$

with $H_r^a = [CPH_q, \; FH_v]$ and $P = \left[I_{n_x}, \; 0, \; 0\right]$.

Proof. Given an attack vector that satisfies Assumption 3.6 and $\Delta \bar{u}_k = \text{const.}$, then, by defining the projection matrix $P = \left[I_{n_x}, \; 0, \; 0\right]$, the trajectories of (3.30) will converge into the zonotopic set $\breve{\mathcal{X}} = P\mathcal{Q} = \langle c_{\breve{x}}, H_{\breve{x}} \rangle = \langle Pc_q, PH_q \rangle$ with

$$c_{\breve{x}} = \Lambda^{-1}\big(Ec_w - LFc_v + B\Delta \bar{u} + BK\Gamma^{-1}\bar{a}\big). \tag{3.36}$$

Hence, from (3.29), the residuals under attack will settle in the set

$$\mathcal{R}^A = C\breve{\mathcal{X}} \oplus E_v \mathcal{V} = \langle c_r^a, H_r^a \rangle = \langle Cc_{\breve{x}} + Fc_v, [CH_{\breve{x}}, \; FH_v] \rangle. \tag{3.37}$$

Accordingly, by recalling that $c_r^h = C\Lambda^{-1}(Ec_w - LFc_v) + Fc_v$, the center of the residual set under attack can be rewritten as a function of the healthy center as $c_r^a = c_r^h + \delta c$, where

$$\delta c = C\Lambda^{-1}(B\Delta \bar{u} + BK\Gamma^{-1}\bar{a}). \tag{3.38}$$

Therefore, adapting the steps given in proof of Proposition 1, it follows that the satisfaction of (3.35) guarantees that $\mathcal{R}^H \cap \mathcal{R}^A = \emptyset$. □

According to Proposition 3.2, the following discussion may be given regarding the attack detectability.

- **Residual set size:** Note that for the attacked case, the generators matrix H_r^a contains additional terms with respect those included in the healthy case. This is a direct consequence of the fact that the cause-effect relationship between the injected control signal and the obtained measurements during the attack is lost, i.e., the healthy control signal $\tilde{u}_k^a$ and the replayed output y_k^r take independent values during the attack. The bigger size of $\mathcal{R}^A$ with respect to $\mathcal{R}^H$ has two consequences: I) it is possible to detect the attack even without forcing the center displacement; II) the bigger size of the attacked set requires a bigger center displacement in order to fulfill condition (3.35).

- **Center displacement**: Note that the attack vector $\bar{a}$ appears explicitly in the detectability condition (3.35). If the output set-point is maintained constant between phases, i.e., $\Delta\bar{y}_k = 0$, the effect of the injected vector $\bar{a}$ is particularly critical along the directions that belong to the null space of the matrix $C\Lambda^{-1}BK\Gamma^{-1}$, that is, $\bar{a} \in \mathcal{N}(C\Lambda^{-1}BK\Gamma^{-1})$, since these attacks would not cause a displacement of δc_k. In other words, a malicious attacker could carry out an unbounded attack $\bar{a}$ for which there are no detectability guarantees from the defender's point of view. Besides, attacks performed along directions associated to small singular values would require a larger norm value in order to achieve the separability condition expressed by (3.35), and thus they are also harmful for the system operation.

Regarding the existence of $\mathcal{N}(C\Lambda^{-1}BK\Gamma^{-1})$, the following proposition can be derived.

Proposition 3.3. The dimension d of $\mathcal{N}(C\Lambda^{-1}BK\Gamma^{-1})$ is lower bounded by

$$d = n_x - rank(C\Lambda^{-1}BK\Gamma^{-1}) \geq n_x - min\{rank(C), rank(BK)\}.$$

Proof. The proof is based on well-known matrix rank properties (cf. Section A.1 of Appendix A). By denoting $X = C\Lambda^{-1}$, $Y = BK\Gamma^{-1}$ such that $\mathcal{N}(C\Lambda^{-1}BK\Gamma^{-1}) = \mathcal{N}(XY)$, the following holds

$$\begin{aligned} rank(XY) &\leq min\{rank(X), rank(Y)\}, \\ rank(X) &= rank(C), \\ rank(Y) &= rank(BK). \end{aligned}$$

Therefore, it can be concluded that

$$rank(XY) \leq min\{rank(C), rank(BK)\}.$$

Finally, using the fact that dimension of $\mathcal{N}(XY)$ is $d = n_x - rank(XY)$, the proof is completed. □

Note that, given a null space $\mathcal{N}(C\Lambda^{-1}BK\Gamma^{-1})$ of dimension d, a malicious attacker could introduce an attack $\bar{a} \in \mathcal{N}(C\Lambda^{-1}BK\Gamma^{-1})$ with $n_x - d + 1$ components different that zero, i.e., the attacker needs to have access only to $n_x - d + 1$ states to carry out this attack. Moreover, the lower bound on d does not depend on the supervisory observer gain L. Consequently, this motivates the need of modifying $\Delta\bar{y}_k$ in order to detect these possible unbounded attacks.

Remark 3.3. Observe that, since the analysis developed in Section 3.4 considers the steady-state operation of the system, the analytical expressions obtained in Proposition 3.1 and Proposition 3.2 are independent of the record/replay starting times.

3.5 Watermark sequences

So far, this chapter has focused on the characterization of the replay attack and the derivation of analytical expression that guarantee the attack detection by means of a set-invariance analysis. Nevertheless, while this type of analysis offers a good insight into the sensitivity of the anomaly detector to replay attacks, the inherent steady-state conditions required for the derivation of analytical expressions motivate the study of more efficient defensive techniques. Consequently, this section focuses on the design of efficient physical watermark signals which, injected from the supervisory layer, are able to guarantee the attack detection. To that end, the watermark proposal takes into account the transient behaviour of the residual signal during the replay phase of the attack. Because Scenario I presents a greater challenge for detection, since it can mask an attack introduced in any direction, the developments presented below are focused on this first attack scenario.

In the sequel, bold letters are used to designate sequence related with the time instants expressed in the subscript, e.g., for the system states $\boldsymbol{x}_{0:k} = \{x_0, x_1, ..., x_{k-1}\} \in \mathbb{R}^{kn_x}$. Accordingly, denoting the watermark signal as $\xi_k \in \mathbb{R}^{n_u}$, the injection of an N-sample sequence at a generic time instant k' has the form

$$\boldsymbol{\xi}_{k':k'+N} = \{\xi_{k'}, \xi_{k'+1}, ..., \xi_{k'+N-1}\} \in \mathbb{R}^{Nn_u}. \tag{3.39}$$

Following the same developments presented in Section 3.4.2 (cf. (3.22) and (3.23)), it is easy to see that the injection of an exogenous signal ξ_k causes that, for $k \in \mathcal{K}_{REP}$, the residual signal evolves as

$$r_k^a = r_k^r + C(A - LC)^{k-k_1}(\hat{x}_{k_1}^r - \hat{x}_{k_1}^a) + C\sum_{i=1}^{k-k_1}(A - LC)^{i-1}B(\xi_{k-i}^r - \xi_{k-i}^a), \tag{3.40}$$

starting at $\hat{x}_{k_1}^a = \hat{x}_{k_1}$ and $\hat{x}_{k_1}^r = \hat{x}_{k_0}$.

The remainder of this section is built upon the premise that the replay attack is undetectable by the anomaly detector. This worst-case detectability scenario is introduced through the following assumption.

Assumption 3.7. For all $k \in \mathcal{K}_{REP}$, the residual signal satisfies

$$r_k^r + C(A - LC)^{k-k_1}(\hat{x}_{k_1}^r - \hat{x}_{k_1}^a) \in \mathcal{R}^H.$$

Recall that during the record phase, the system is in healthy operation, hence from Assumption 3.4 it follows that $r_k^r \in \mathcal{R}^H$. On the other hand, since $A - LC$ is an asymptotically stable matrix, then $C(A - LC)^{k-k_1}(\hat{x}_{k_1}^r - \hat{x}_{k_1}^a) \to 0$ as $t \to \infty$. In other words, Assumption 3.7 states that the attack is not detected in the transient induced by the data substitution, and thus the attack detection must be forced by imposing a temporal mismatch in the exogenous signal ξ_k. For the sake of simplified notation, in the following it is considered that the replay phase starts at $k_1 = 0$ and that the watermark sequence is injected at the same instant, i.e., $k' = k_1 = 0$. In this regard, it must be remarked that the analysis developed below holds for any $k' \geq k_1$, as long as the attack is not detected in the time interval $[k_1,\ k']$.

Working under Assumption 3.7, from (3.40) it follows that (for $k \in K_{REP}$) the residual signal satisfies $r_k^a \in \mathcal{R}_k^A$, with

$$\mathcal{R}_k^A = \mathcal{R}^H \oplus M_k(\boldsymbol{\xi}_{0:k}^r - \boldsymbol{\xi}_{0:k}^a), \tag{3.41}$$

which, making use of the zonotopic representation of the healthy residual set $\mathcal{R}^H = \langle c_r, H_r \rangle$, can be written as

$$\mathcal{R}_k^A = \langle c_r + M_k(\boldsymbol{\xi}_{0:k}^r - \boldsymbol{\xi}_{0:k}^a), H_r \rangle. \tag{3.42}$$

where (for $k \geq 1$) M_k is given by the recursions

$$\tilde{M}_{k+1} = [(A - LC)\tilde{M}_k,\ B], \qquad M_k = C\tilde{M}_k \tag{3.43}$$

and $\tilde{M}_1 = B$.

Thus, given a pre-specified horizon N, the idea is to design a finite sequence $\boldsymbol{\xi}_{0:N}$ such that, if injected during $[0,\ N-1] \in \mathcal{K}_{REP}$, enforces the guaranteed detectability condition formulated in Definition 3.1 at $k = N$, i.e., such that guarantees the satisfaction of

$$\mathcal{R}^H \cap \mathcal{R}_N^A = \emptyset. \tag{3.44}$$

Nevertheless, the expression (3.42) illustrates the dependence of the effect of the injected watermark signal during the replay phase $\boldsymbol{\xi}_{0:k}^a$, on the values adopted by the signal during the record phase $\boldsymbol{\xi}_{0:k}^r$. This problem will be addressed below.

3.5.1 Sequence design

In order to generate a temporal mismatch between record and replay phases, $s \in \mathbb{N}_+$ different sequences of N steps are designed. These sequences are denoted as

$$\boldsymbol{\xi}_{0:N}[i], \quad \forall i \in S = \{1, ..., s\}.$$

Additionally, each sequence $\boldsymbol{\xi}_{0:N}[i]$ must satisfy that, if injected during the replay phase, enforces $\mathcal{R} \cap \mathcal{R}_N^A = \emptyset$ for any $\xi_k[j] \in \boldsymbol{\xi}_{0:N}^r$ such that $j \in S \setminus i$. Accordingly, the number of sequences is related with the security of the watermarking scheme, so the bigger the s, the lower the probability that there is a $\xi_k[j] \in \boldsymbol{\xi}_{0:N}^r$ with $j = i$.

At this point, it must be highlighted the large number of watermark combinations that may appear during the recordings. In order to illustrate this point, consider the design of $s = 3$ sequences with length $N = 3$ denoted as $\boldsymbol{\xi}_{0:3}[1]$, $\boldsymbol{\xi}_{0:3}[2]$, $\boldsymbol{\xi}_{0:3}[3]$. Hence, $\boldsymbol{\xi}_{0:3}[1]$ should be designed to guarantee the attack detectability for the following possible combinations that may appear in the recordings during the interval [0, 2]:

$$\begin{aligned}
&\boldsymbol{\xi}_{0:3}^r = \{\xi_0[2], \xi_1[2], \xi_2[2]\}, \quad \boldsymbol{\xi}_{0:3}^r = \{\xi_0[3], \xi_1[2], \xi_2[2]\}, \quad \boldsymbol{\xi}_{0:3}^r = \{\xi_0[3], \xi_1[3], \xi_2[2]\},\\
&\boldsymbol{\xi}_{0:3}^r = \{\xi_0[3], \xi_1[3], \xi_2[3]\}, \quad \boldsymbol{\xi}_{0:3}^r = \{\xi_0[2], \xi_1[3], \xi_2[3]\}, \quad \boldsymbol{\xi}_{0:3}^r = \{\xi_0[2], \xi_1[2], \xi_2[3]\},
\end{aligned}$$

in addition, $\boldsymbol{\xi}_{0:3}[2]$ and $\boldsymbol{\xi}_{0:3}[3]$ should be designed similarly.

Motivated by the fact that in many real-world applications it is not desired to continuously degrade the system performance in order to protect the plant against an exceptional event like an attack, the design of the watermark sequences is performed under the following consideration.

Design criterion: Between the injection of two N steps watermark sequences, no watermark is injected during at least N steps, that is, the watermark sequences are injected following the scheme

$$\{..., \boldsymbol{\xi}_{0:N}[i],\ \mathbf{0}_{N'},\ \boldsymbol{\xi}_{0:N}[j],\ \mathbf{0}_{N'},\ ...\},$$

where $\mathbf{0}_{N'}$ is a concatenation of $N' \in \mathbb{N}_+$ null vectors with $N' \geq N$ and the injection time k' of each sequence is reset to $k' = 0$ for simplicity.

Observe that by means of the previous design criterion it is imposed that the record sequence $\boldsymbol{\xi}^r_{0:N}$ can only be composed of combinations of a unique sequence $\boldsymbol{\xi}_{0:N}[i]$ and/or the null sequence $\mathbf{0}_{N'}$. Hence, in order to include all the combinations that arise, the input space is partitioned into s sets $\mathcal{Z}_i$, such that each one of them satisfies

$$\boldsymbol{\xi}_{0:N}[i] \in \mathcal{Z}_i, \ 0 \in \mathcal{Z}_i, \quad \forall i \in S. \tag{3.45}$$

Consequently, from the design criterion and (3.45) it follows that

$$\boldsymbol{\xi}^r_{0:N} \in \tilde{\mathcal{Z}}_i, \text{ for any } i \in S,$$

with $\tilde{\mathcal{Z}}_i = \mathcal{Z}_i \times ... \times \mathcal{Z}_i$ an N Cartesian product.

Hereafter, the term $\mathcal{R}^A_N(\boldsymbol{\xi}^a_{0:N}[i], j)$ is employed to denote the attacked residual set at time N for the injected sequence $\boldsymbol{\xi}^a_{0:N}[i]$ and for any possible sequence satisfying $\boldsymbol{\xi}^r_{0:N} \in \tilde{\mathcal{Z}}_j$. From (3.41), the set $\mathcal{R}^A_N(\tilde{\xi}^a_{0:N}[i], j)$ can be written as

$$\mathcal{R}^A_N(\boldsymbol{\xi}^a_{0:N}[i], j) = \mathcal{R}^H \oplus M_N \tilde{\mathcal{Z}}_j \oplus M_N(-\boldsymbol{\xi}^a_{0:N}[i]), \tag{3.46}$$

with $M_N = [C(A - LC)^{N-1}B, \ \ ..., \ \ CB]$.

At this point, a new definition concerning the sets $\mathcal{Z}_i$ is introduced.

Definition 3.2 (N-admissible ordered pair)**.** Given the pair of sets $\mathcal{Z}_i$ and $\mathcal{Z}_j$, with $(i, j) \in S$ and $i \neq j$, the ordered pair $(\mathcal{Z}_i, \mathcal{Z}_j)$ is said to be N-admissible if

$$\exists \boldsymbol{\xi}^a_{0:N}[i] \in \tilde{\mathcal{Z}}_i : \ \mathcal{R}^H \cap \mathcal{R}^A_N(\boldsymbol{\xi}^a_{0:N}[i], j) = \emptyset. \tag{3.47}$$

Note that in the preceding definition the order of the pairs must be taken into account, that is, given an N-admissible pair $(\mathcal{Z}_i, \mathcal{Z}_j)$ this does not imply that $(\mathcal{Z}_j, \mathcal{Z}_i)$ is also N-admissible. Hence, the set $S = \{1, ..., s\}$ gives rise to $n_l = 2\binom{s}{2}$ different pair combinations. Below, the set $\mathcal{L}$ is used to characterize the possible n_l pair combinations, in such a way that $\forall (i, j) \in S$, with $i \neq j$, the pair $(\mathcal{Z}_j, \mathcal{Z}_i)$ maps to a different element of $\mathcal{L}$.

3.5.2 Programming problem

The design of s sets such that all the possible pair $(\mathcal{Z}_i, \mathcal{Z}_j) \in \mathcal{L}$ is N-admissible, is detailed in this section. At this point, the following condition in the watermark sequences will be imposed: each watermark sequence $\boldsymbol{\xi}_{0:N}[i]$ is associated with one input direction $h_i \in \mathbb{R}^{n_u}$, i.e.,

$$\boldsymbol{\xi}_{0:N}[i] = \{\beta^i_0 h_i, ..., \beta^i_{N-1} h_i\}.$$

Accordingly, the sets $\mathcal{Z}_i$ that satisfy the requirements in (3.45), can be parametrized as the zonotope

$$\mathcal{Z}_i = \langle c_i, \beta_i h_i \rangle \subset \mathbb{R}^{n_u}, \quad i \in S, \tag{3.48}$$

where $c_i \in \mathbb{R}^{n_u}$ and $\beta_i \in \mathbb{R}$ are free design parameters.

Making use of the zonotopic representation (3.48), the set $\tilde{\mathcal{Z}}_j$ can be written as the extended zonotope

$$\tilde{\mathcal{Z}}_i = \langle \tilde{c}_i, \tilde{H}_i \tilde{\beta}_i \rangle, \tag{3.49}$$

with

$$\begin{aligned} \tilde{c}_i &= [c_i^T,\ \ldots,\ c_i^T]^T, \\ \tilde{H}_i &= diag(h_i,\ \ldots,\ h_i), \\ \tilde{\beta}_i &= [\beta_i,\ \ldots,\ \beta_i]^T, \end{aligned} \tag{3.50}$$

and thus, the residual set under attack at sampling time N, is given by

$$\mathcal{R}_N^A(\boldsymbol{\xi}^a_{0:N}[i], j) = \big\langle c_r + M_N(\tilde{c}_j - \boldsymbol{\xi}^a_{0:N}[i]), [H_r,\ M_N \tilde{H}_j \tilde{\beta}_j] \big\rangle. \tag{3.51}$$

Recalling Lemma 3.1 and the developments presented in the proof of Proposition 3.1, it follows that the pair $(\mathcal{Z}_i, \mathcal{Z}_j) \in \mathcal{L}$ is N-admissible if and only if the following condition is satisfied

$$\exists \boldsymbol{\xi}^a_{0:N}[i] \in \tilde{\mathcal{Z}}_i \quad : \quad M_N \boldsymbol{\xi}^a_{0:N}[i] \notin \langle M_N \tilde{c}_j, [2H_r,\ M_N \tilde{H}_j \tilde{\beta}_j] \rangle. \tag{3.52}$$

On the other hand, note that the zonotope in the right-hand side of (3.52) can be rewritten as the set

$$\{ M_N \tilde{c}_j + 2H_r z_1 + M_N \tilde{H}_j z_2 : z_1 \in \mathbf{B}^{n_r},\ |z_2| \le \tilde{\beta}_j \}. \tag{3.53}$$

In the sequel, it is considered that the design of the N-admissible ordered pairs is done under some optimality criterion that relates with the size of the sets and that is assessed by means of the cost function $J(c_i, \beta_i)$ (cf. Section 3.6). Therefore, an optimization program P_1 (master problem), with decision variables (c_i, β_i), is formulated in such a way that minimizes $J(c_i, \beta_i)$ subject to the constraint that each pair $(\mathcal{Z}_i, \mathcal{Z}_j) \in \mathcal{L}$ is N-admissible.

In order to impose the N-admissibility constraint, from [Scott et al., 2014, Lemma 4] it is known that, when dealing with zonotopic sets, the design of an input sequence that satisfies a set separation condition like (3.47) can be formulated as an inequality constraint on the optimal objective value of a linear program (LP). Below, the different LPs that arise are denoted as $P_2^{[l]}$, with $l \in \mathcal{L}$, such that each $P_2^{[l]}$ constitutes a subproblem with its own decision variables $(z^{[l]}, \delta^{[l]})$, which are set after the master variables (c_i, β_i) are announced [Fortuny-Amat and McCarl, 1981].

Hence, the set design is posed as the two-level programming problem

$$P_1(\epsilon) \left\{ \begin{array}{lll} \underset{c_i, \beta_i, \tilde{\xi}[i], \eta_i, \rho_i, \delta^{[l]}, z^{[l]}}{\min.} \quad J(c_i, \beta_i), & & \text{(3.54a)} \\ \text{s.t.} \quad c_i + h_i \eta_i = 0, & i \in S, & \text{(3.54b)} \\ \qquad \boldsymbol{\xi}_{0:N}[i] = \tilde{c}_i + \tilde{H}_i \rho_i, & i \in S, & \text{(3.54c)} \\ \qquad |\eta_i| \le \beta_i, \quad |\rho_i| \le \tilde{\beta}_i, \quad \beta_i \ge 0, & i \in S, & \text{(3.54d)} \\ \qquad \hat{\delta}^{[l]} \ge \epsilon, & l \in \mathcal{L}, & \text{(3.54e)} \\ \qquad P_2^{[l]} \left\{ \begin{array}{l} \hat{\delta}^{[l]} \equiv \min._{\delta^{[l]}, z^{[l]}} \ \delta^{[l]}, \\ s.t. \quad f(z^{[l]}, \delta^{[l]}) = 0, \\ \qquad\ g(z^{[l]}, \delta^{[l]}) \le 0, \end{array} \right. & l \in \mathcal{L}, & \text{(3.54f)} \end{array} \right.$$

where (3.54b)-(3.54d) impose the satisfaction of the requirements expressed in (3.45), whereas (3.54e)-(3.54f) represents the constraints on the LP subproblems objective function that enforces the satisfaction of (3.52).

In this regard, using the zonotope characterization (3.53), from [Scott et al., 2014, Lemma 4] it follows that the set separation condition expressed in (3.52) is satisfied if and only if $\hat{\delta}^{[l]} > 0$, where $\hat{\delta}^{[l]} \in \mathbb{R}$ is the objective value of the following LP

$$P_2^{[l]} \left\{ \begin{array}{l} \hat{\delta}^{[l]} \equiv \min._{\delta^{[l]}, z^{[l]}} \;\; \delta^{[l]}, \\ \text{s.t.} \;\; M_N \boldsymbol{\xi}_{0:N}[i] = M_N \tilde{c}_j + 2 H_r z_1^{[l]} + M_N \tilde{H}_j z_2^{[l]} \\ \qquad\; |z_1^{[l]}| \leq \mathbf{1}_{n_r} + \mathbf{1}_{n_r} \delta^{[l]} \quad |z_2^{[l]}| \leq \tilde{\beta}_j^{[l]} + \mathbf{1}_N \delta^{[l]} \end{array} \right. \tag{3.55}$$

with $l \in \mathcal{L}$ and $\mathbf{1}_n$ denotes an n component column vector of ones and n_r is the number of columns of H_r.

Following the proposals of Fortuny-Amat and McCarl [1981], Scott et al. [2014], the replacement of each of the n_l different LPs by their necessary and sufficient conditions of optimality [Bertsekas, 1997], allows to reformulate the two-level program as single level program. Accordingly, the satisfaction of (3.52) can be posed as the existence of $(\delta^{[l]}, \tilde{\xi}_{0:N}[i], \lambda^{[l]}, \mu_1^{[l]}, \mu_2^{[l]}, \mu_3^{[l]}, \mu_4^{[l]})$ such that the equations expressed in (3.56) hold

$$\hat{\delta}_m^{[l]} \geq \delta^{[l]} \geq \epsilon, \tag{3.56a}$$

$$M_N \boldsymbol{\xi}_{0:N}[i] = M_N \tilde{c}_j + 2 H_r z_1^{[l]} + M_N \tilde{H}_j z_2^{[l]}, \tag{3.56b}$$

$$|z_1^{[l]}| \leq \mathbf{1}_{n_r} + \mathbf{1}_{n_r} \delta^{[l]}, \quad |z_2^{[l]}| \leq \tilde{\beta}_j + \mathbf{1}_N \delta^{[l]}, \tag{3.56c}$$

$$\begin{bmatrix} 2H_r^T \\ \tilde{H}_j^T \tilde{M}_N^T \end{bmatrix} \lambda^{[l]} = \begin{bmatrix} \mu_1^{[l]} \\ \mu_3^{[l]} \end{bmatrix} - \begin{bmatrix} \mu_2^{[l]} \\ \mu_4^{[l]} \end{bmatrix}, \tag{3.56d}$$

$$1 = (\mu_1^{[l]} + \mu_2^{[l]})^T \mathbf{1}_{n_r} + (\mu_3^{[l]} + \mu_4^{[l]})^T \mathbf{1}_N, \tag{3.56e}$$

$$\mu_1^{[l]}, \mu_2^{[l]}, \mu_3^{[l]}, \mu_4^{[l]} \geq 0, \tag{3.56f}$$

$$0 = \mu_{1,k}^{[l]} (z_{1,k}^{[l]} - 1 - \delta^{[l]}), \qquad \forall k = \{1, ..., n_r\}, \tag{3.56g}$$

$$0 = \mu_{2,k}^{[l]} (z_{1,k}^{[l]} + 1 + \delta^{[l]}), \qquad \forall k = \{1, ..., n_r\}, \tag{3.56h}$$

$$0 = \mu_{3,m}^{[l]} (z_{2,m}^{[l]} - \tilde{\beta}_j - \delta^{[l]}), \qquad \forall m = \{1, ..., N\}, \tag{3.56i}$$

$$0 = \mu_{4,m}^{[l]} (z_{2,m}^{[l]} + \tilde{\beta}_j + \delta^{[l]}), \qquad \forall m = \{1, ..., N\}, \tag{3.56j}$$

where $\hat{\delta}_m^{[l]}$ represents an upper bound on $\delta^{[l]}$ and $\epsilon > 0$.

Notice that the nonconvex complementary constraints (3.56g)-(3.56j) can be reformulated as linear constraints with the inclusion of the $(p_1^{[l]}, p_2^{[l]}) \in \{0,1\}^{n_r}$, $(p_3^{[l]}, p_4^{[l]}) \in \{0,1\}^N$ binary variables [Fortuny-Amat and McCarl, 1981]. Hence, (3.56g)-(3.56j) are rewritten as

$$\mu_{1,k}^{[l]} \leq \sigma p_{1,k}^{[l]}, \quad \mu_{2,k}^{[l]} \leq \sigma p_{2,k}^{[l]}, \tag{3.57a}$$

$$\mu_{3,l}^{[l]} \leq \sigma p_{3,l}^{[l]}, \quad \mu_{4,l}^{[l]} \leq \sigma p_{4,l}^{[l]}, \tag{3.57b}$$

$$-(1 - p_{1,k}^{[l]})\sigma \leq z_{1,k}^{[l]} - 1 - \delta^{[l]} \leq 0, \tag{3.57c}$$

$$0 \leq z_{1,k}^{[l]} + 1 + \delta^{[l]} \leq (1 - p_{2,k}^{[l]})\sigma, \tag{3.57d}$$

$$-(1 - p_{3,m}^{[l]})\sigma \leq z_{2,m}^{[l]} - \tilde{\beta}_j - \delta^{[l]} \leq 0, \tag{3.57e}$$

$$0 \leq z_{2,m}^{[l]} + \tilde{\beta}_j + \delta^{[l]} \leq (1 - p_{4,m}^{[l]})\sigma, \tag{3.57f}$$

where $\sigma = 2(1+\hat{\delta}_m)$. On this subject, the bounds imposed in (3.57a)-(3.57b) are a consequence of (3.56e)-(3.56f), and thus are not restrictive. On the other hand, the bounds on (3.57c)-(3.57f) are a consequence of (3.56a) and (3.56c). Note that each LP program requires the introduction of $2(n_r + N)$ binary variables.

Hence, the two-level program (3.54) in charge of the set design is equivalent to the single level mixed integer program

$$P_1(\epsilon)\left\{\begin{array}{lll} \underset{c_i,\beta_i,\tilde{\xi}[i],\eta_i,\rho_i,\delta^{[l]},\lambda^{[l]},\mu_j^{[l]},p_j^{[l]}}{\min.} \quad J(c_i,\beta_i) & & (3.58a) \\ \text{s.t.} \quad c_i + h_i\eta_i = 0, & i \in S, & (3.58b) \\ \qquad \boldsymbol{\xi}_{0:N}[i] = \tilde{c}_i + \tilde{H}_i\rho_i, & i \in S, & (3.58c) \\ \qquad |\eta_i| \leq \beta_i, \quad |\rho_i| \leq \tilde{\beta}_i, \quad \beta_i \geq 0, & i \in S, & (3.58d) \\ \qquad \left\{\begin{array}{l} (3.56a) - (3.56f), \\ (3.57a) - (3.57f), \\ p_1^{[l]}, p_2^{[l]} \in \ \{0,1\}^{n_r}, \\ p_3^{[l]}, p_4^{[l]} \in \ \{0,1\}^{N}, \end{array}\right. & l \in \mathcal{L}. & (3.58e) \end{array}\right.$$

Remark 3.4. Notice that the optimizer of the previous program allows to retrieve the information on the optimal set partitions (c_i^*, β_i^*), $\forall i \in S$, as well as the (probably not unique) separation sequences $\boldsymbol{\xi}_{0:N}^*[i] \in \tilde{\mathcal{Z}}_i$, $\forall i \in S$.

3.5.3 Complexity reduction

Although the optimization problem (3.58) is intended to be run offline, the total amount of binary variables introduced in the MIP reformulation is $2(n_r + N)n_l$, with $n_l = 2\binom{s}{2}$. Below, different strategies are discussed for, at the cost of introducing conservatism, reducing the number of binary variables while still being able to compute N-admissible pairs of sets.

3.5.3.1 Pair symmetry

The first approach is based on exploiting the symmetry of the pairs of sets. On this subject, selecting an even security index s, then the different subsets can be grouped in $s/2$ pairs in such a way that each pair $(\mathcal{Z}_i, \mathcal{Z}_i')$ satisfies $\mathcal{Z}_i = -\mathcal{Z}_i'$, $\forall i \in \{1, ..., s/2\}$.

Proposition 3.4. Under the pair symmetry constraint, the number of pair combinations n_l that must be examined can be halved.

Proof. The proof follows from the fact that if there exists $\boldsymbol{\xi}_{0:N}[i] \in \tilde{\mathcal{Z}}_i$ such that $(\mathcal{Z}_i, \mathcal{Z}_j)$ are N-admissible, then, by the center symmetry of the zonotopes, it follows that there exists $-\boldsymbol{\xi}_{0:N}[i] \in -\tilde{\mathcal{Z}}_i$ such that $(-\mathcal{Z}_i, -\mathcal{Z}_j)$ are N-admissible. Hence, by designing $\mathcal{Z}_i$ in such a way that $(\mathcal{Z}_i, -\mathcal{Z}_i)$, $(\mathcal{Z}_i, \mathcal{Z}_j)$, $(\mathcal{Z}_i, -\mathcal{Z}_j)$ are N-admissible pairs, it can be deduced that by imposing $\mathcal{Z}_j = -\mathcal{Z}_j$ the pairs $(-\mathcal{Z}_i, \mathcal{Z}_i)$, $(-\mathcal{Z}_i, -\mathcal{Z}_j)$, $(-\mathcal{Z}_i, \mathcal{Z}_j)$ are also N-admissible.

Consequently, for each pair $(\mathcal{Z}_i, \mathcal{Z}_i')$, only the N-admissibility of $\mathcal{Z}_i$ with respect the remaining sets must be imposed, halving thus the number of combinations n_l that must be taken into consideration in the set design process. □

3.5.3.2 mRPI outer-approximations

In the second approach, rather than reducing the number of combinations to be examined, the complexity of each set is reduced. To that end, observe that the number of binary variables required to transform the optimality conditions of each LP are related with the number of generators of the zonotope (see (3.52))

$$\langle M_N \tilde{c}_j, [2H_r,\ M_N \tilde{H}_j \tilde{\beta}_j] \rangle,$$

which can be rewritten the Minkowski sum of two zonotopes $\mathcal{Z}_N^{1,j} \oplus \mathcal{Z}_N^{2,j}$ such that

$$\mathcal{Z}_N^{1,j} = \langle M_N \tilde{c}_j, 2H_r \rangle, \qquad \mathcal{Z}_N^{2,j} = \langle 0, M_N \tilde{H}_j \tilde{\beta}_j \rangle.$$

Additionally, making use of zonotope properties, it follows that

$$\mathcal{Z}_N^{2,j} = \langle 0, M_N \tilde{H}_j \tilde{\beta}_j \rangle = \bigoplus_{i=1}^{N} C(A - LC)^{i-1} B h_j \beta_j. \tag{3.59}$$

At this point, from the definition of the mRPI set (cf. Appendix A) and the asymptotic stability of $A - LC$, it follows that each $\mathcal{Z}_N^{2,j}$ is contained within the corresponding mRPI set which is denoted as $\mathcal{Z}_m^{2,j}$ (i.e., $\mathcal{Z}_N^{2,j} \subseteq \mathcal{Z}_m^{2,j}, \forall N \in \mathbb{N}$), and that $\mathcal{Z}_N^{2,j} \to \mathcal{Z}_m^{2,j}$ as $N \to \infty$. Consequently, for optimizations with a large N, and thus with a large number of binary variables, $\mathcal{Z}_N^{2,j}$ can be replaced by a fixed order set $\bar{Z}^{2,j}$ that over-approximates the set $\mathcal{Z}_m^{2,j}$, that is, such that $\bar{\mathcal{Z}}^{2,j} \supseteq \mathcal{Z}_m^{2,j} \supseteq \mathcal{Z}_N^{2,j}$, while preserving the N-admissibility of the designed sets.

The proposed over-approximation becomes particularly useful for the case where $\bar{\mathcal{Z}}^{2,j}$ can be computed as a first-order zonotope (cf. Section A.3.1 of Appendix A) that can be written as an affine parametrization of the disturbance set Δ_j. That is, $\bar{\mathcal{Z}}^{2,j} = \langle 0, \bar{H} \tilde{\beta}_j \rangle$, with $\bar{H}$ a square matrix. In this case, the substitution of $\mathcal{Z}_N^{2,j}$ by the fixed order zonotope $\bar{\mathcal{Z}}^{2,j}$ in the optimization problem (3.58), reduces the number of binary constraint from $2(n_r + N)n_l$ to the fixed value $2(n_r + n_y)n_l$ for any $N \in \mathbb{N}$.

Remark 3.5. Note the influence of the observer gain L, which appears in the displacement matrix M_k, on the performance of the detection method. Accordingly, in order to achieve an effective detection, matrix L should be designed in such a way that optimizes the trade-off between the size of the residual set $\mathcal{R}^H$ and the gain of the displacement matrix M_k. In this regard, since in the case that the system is under attack the injection of an exogenous sequence $\boldsymbol{\xi}_{0:N}$ causes an effect analogous to a sensors fault, a convenient approach in the designing of the observer gain is to enhance its sensitivity to faults with respect to disturbances/noises. On this subject, in Pourasghar et al. [2019] the authors followed the zonotope bounded-disturbances paradigm in the design of a fault sensitive observer gain.

Remark 3.6. Observe that the injection directions h_i of each sequence $\boldsymbol{\xi}_{0:N}[i]$ has been left as a free parameter that is specified by the user. The selection of directions that minimizes the impact on the operation of the system should be studied in detail.

3.6 Case study

The case study proposed in order to validate the results obtained in previous sections is a four-tank process regulated using a low-level state estimated-feedback controller and supervised by means of a state estimator.

The four-tank system is a multi-input/multi-output process proposed by Johansson [2000], which constitutes a well-known benchmark used to evaluate control and supervision strategies. The system description and the computation of a discrete-time LTI model is presented in Section B.1 of Appendix B. Throughout all the simulations carried out in this section, it has been considered that the process disturbances w_k and measurement noise v_k are random white noise signals bounded within the zonotopes $w_k \in \langle 0, I_4 \rangle$ and $v_k \in \langle 0, I_2 \rangle$, and affecting the system states and output signal through the following distribution matrices

$$E = \begin{bmatrix} 0.01 & 0 & 0 & 0 \\ 0.01 & 0.01 & 0 & 0 \\ 0.01 & 0 & 0.01 & 0 \\ 0.01 & 0 & 0 & 0.01 \end{bmatrix}, \qquad F = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}.$$

On the one hand, the considered low-level controller is a linear-quadratic regulator (LQR) designed for the state and input weight matrices $Q_K = I_4$ and $R_K = I_2$. On the other hand, at the supervisory layer, the observer in charge of the plant monitoring is computed also following the dual LQR design with weights $Q_L = I_4$ $R = 10I_2$. The obtained controller and observer gains are

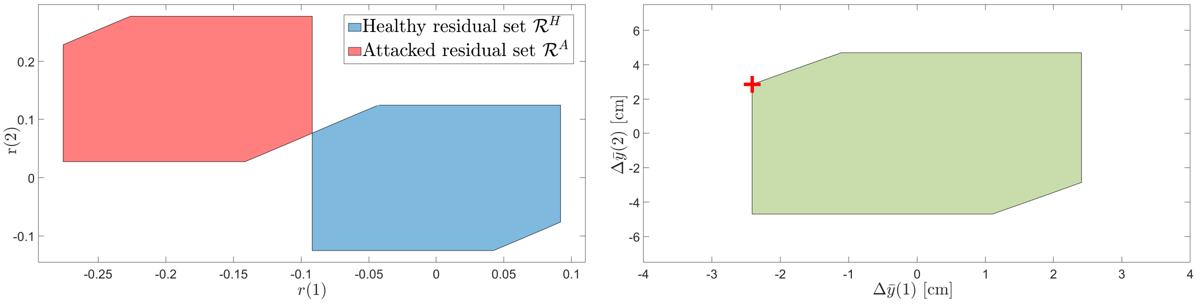
$$K = \begin{bmatrix} 0.8332 & -0.0060 & 0.1481 & 0.0228 \\ 0.0129 & 0.7710 & 0.2639 & 0.2831 \end{bmatrix}, \qquad L = \begin{bmatrix} 0.2909 & 0 & 0.0919 & 0 \\ 0 & 0.2948 & 0 & 0.0981 \end{bmatrix}^T.$$

Additionally, the error for the state estimate used by the low-level controller is considered to be constrained within the zonotopic set $\eta_k \in \langle 0, 0.01 I_4 \rangle$.

At this point, an ϵ-approximation $\mathcal{E} \supseteq \mathcal{E}_m$ of the mRPI set $\mathcal{E}_m$ for the estimation error dynamics is computed by means of Algorithm A.1 in Appendix A. The approximation parameter of the algorithm is set to $\epsilon = 10^{-4}$. After computing the set $\mathcal{E}$, the healthy residual zonotopic set $\mathcal{R}^H = \langle c_r^h, H_r^h \rangle$ has been obtained using the expression (3.12).

3.6.1 Steady-state analysis - Scenario I

Scenario I considers that a malicious attacker is able to record and replay the input/output data transmitted from the regulatory to the supervisory layer. Below, in order to be able to graphically represent the sets, the order of the healthy zonotope is truncated by means of the reduction operator described in Property A.9. Accordingly, the set $\mathcal{R}^H$ is over-approximated by the zonotope $\langle c_r^h, \tilde{H}_r^h \rangle$, with $\tilde{H}_r^h = \downarrow_{15} (H_r^h)$, yielding the blue set depicted in Figure 3.2a.



(a) Residuals space.

(b) Output set-point space $\Delta\bar{y}$.

Figure 3.2: Scenario I.

On the other hand, Figure 3.2b represents the zonotope $\langle 0, 2M^{-1}\tilde{H}_r^h \rangle$, such that, according to Proposition 3.1, all the set-point differences $\Delta\bar{y}$ that do not belong to this zonotope guarantee the replay attack detection in the steady-state. In order to illustrate this point, the signal $\Delta\bar{y} = [-2.411,\ 2.860]^T$, highlighted by means of a red cross in Figure 3.2b, is selected. Note that this particular selection of $\Delta\bar{y}$ shifts the attacked residual set $\mathcal{R}^A$ to the position displayed by the red zonotope in Figure 3.2a, in such a way that the set separability is enforced and thus the attack detection can be guaranteed.

Here, a replay attack is reproduced in order to illustrate the performance of the approach. The simulated attack consists of the following time windows

$$\mathcal{K}_{REC} = [100,\ 300],$$
$$\mathcal{K}_{REP} = \mathcal{K}_{PHY} = [400,\ 1000].$$

Observe that the recording time window encompasses the repetition of 3 times the recorded data set. For all $k \in \mathcal{K}_{PHY}$, the system is perturbed by means of the attack signal $a_k = [1, 1, 1, 1]^T$. Furthermore, at $k = 500$ the aforementioned output set-point modification $\Delta\bar{y} = [-2.411,\ 2.860]^T$ is imposed from the supervisory layer.

For the attack scenario described above, Figure 3.3 depicts with black crosses the evolution of the residual signal. This figure shows the residuals initially confined within $\mathcal{R}^H$ (blue zonotope), and how, after a transient induced by the output set-point modification, the residual signal converges into the set $\mathcal{R}^A$ (red zonotope). Hence, the separability condition between the sets allows to unambiguously assess that the system is under attack.

Moreover, Figure 3.4 illustrates the temporal evolution of the output set-point signal $\bar{y} = [\bar{y}_1, \bar{y}_2]^T$ sent from the supervisory layer. Besides, Figure 3.5 represents the temporal evolution of the components of the residual signal. In this regard, in order to generate the healthy residual bounds that are represented by means of the red dashed lines in Figure 3.5, the interval hull (cf. Definition A.20) of $\mathcal{R}^H$ has been computed. Note that, since the attacked set $\mathcal{R}^A$ is shifted strictly to the left of the healthy set (cf. Figure 3.3), then this interval interpretation is still valid for the first component of the residual. Additionally, observe that in the time interval $k \in [400,\ 500]$ the presence of the physical attack is completely masked to the supervisory layer, and how the set-point modification that starts at $k = 500$ forces the residual signal to stabilize strictly outside the healthy residual set.

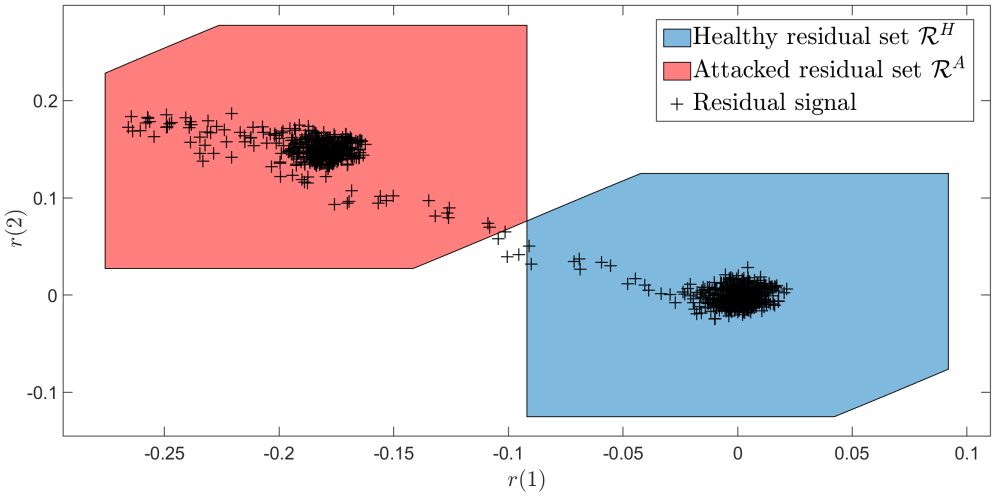


Figure 3.3: Residual signal - Scenario I.

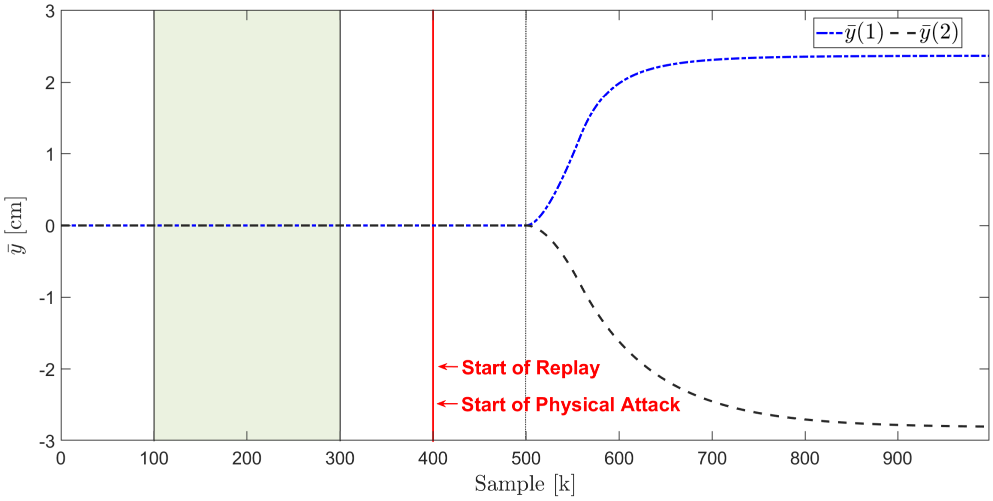


Figure 3.4: Set-point imposed from the supervisory layer - Scenario I.

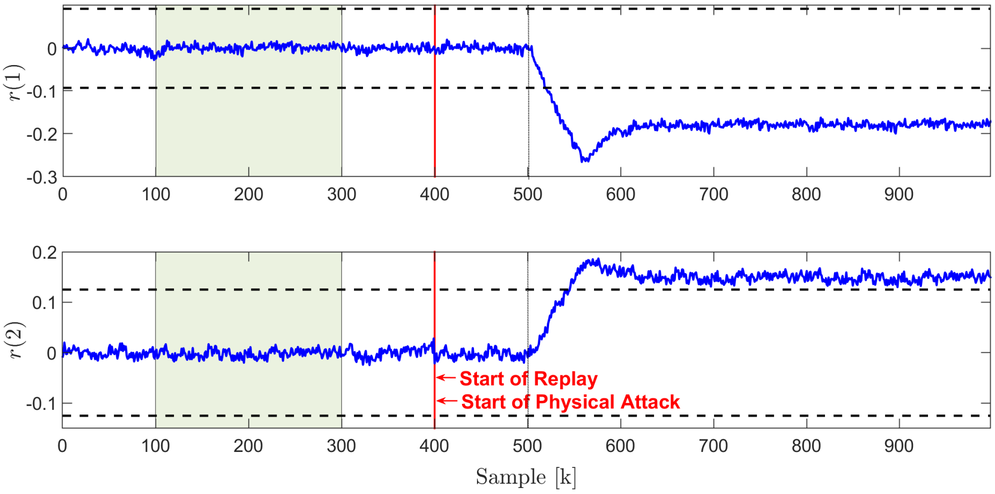


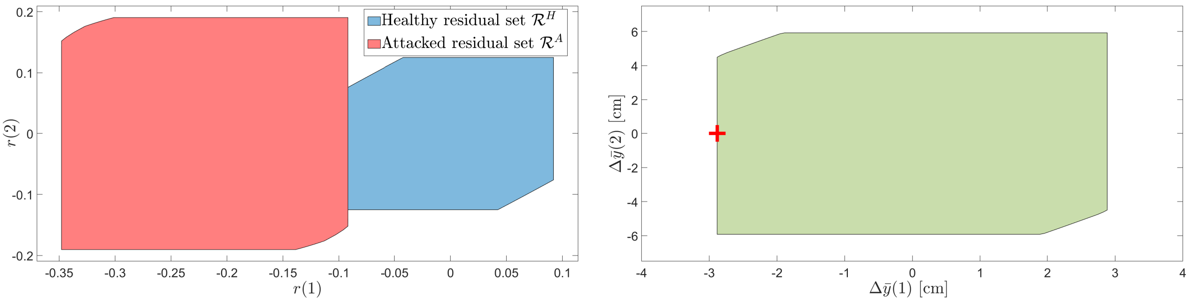
Figure 3.5: Residuals at the supervisory layer - Scenario I.

3.6.2 Steady-state analysis - Scenario II

Scenario II considers that a malicious attacker is only able to corrupt the output data sent from the regulatory to the supervisory layer. In this regard, an ϵ-approximation $\mathcal{Q} \supseteq \mathcal{Q}_m$ of the mRPI set $\mathcal{Q}_m$ described in Section 3.4.3.1 has been computed by means of Algorithm A.1, thus obtaining the shape of attacked residual set $\mathcal{R}^A$. As it has been done for the healthy residual set, the order of $\mathcal{R}^A$ is also reduced in order to graphically represent it.

Following Section 3.4.3.1, for this second scenario an output set-point difference must be established in order to protect the system against unbounded attacks satisfying $C\Lambda^{-1}BK\Gamma^{-1}\bar{a} = 0$. In this regard, Figure 3.6b represents the zonotope $\langle 0, M^{-1}[H_r^h,\ H_r^a]\rangle$, for which, according to Proposition 3.2, all the set-point differences $\Delta\bar{y}$ that do not belong to it guarantee the detectability of data replay and the presence of possible unbounded attacks.

The output set-point difference $\Delta\bar{y} = [-2.884,\ 0]^T$, highlighted by means of a red cross in Figure 3.6b, is selected. The imposition of this $\Delta\bar{y}$ displaces the attacked residual set to the position displayed by the red zonotope in Figure 3.6a, guaranteeing thus the set separation.



(a) Residual space.

(b) Output set-point space $\Delta\bar{y}$.

Figure 3.6: Scenario II.

Moreover, note that, consistently with the discussion of Section 3.4.3.1, the size of the attacked residual set is bigger than the healthy one.

Below, a replay attack for Scenario II is simulated. The considered time windows are

$$\begin{aligned}\mathcal{K}_{REC} &= [100,\ 300],\\ \mathcal{K}_{REP} &= [400,\ 2500],\\ \mathcal{K}_{PHY} &= [500,\ 2500].\end{aligned}$$

For the system under study, the null space of matrix $C\Lambda^{-1}BK\Gamma^{-1}$ has dimension $d = 2$, and thus an attacker would only need access to $n_x - d + 1 = 3$ states in order to carry out an attack belonging to such null space. Accordingly, in order to illustrate that no steady-state guarantees can be given for attack vectors belonging to $\mathcal{N}(C\Lambda^{-1}BK\Gamma^{-1})$, the system is attacked introducing an exogenous signal following the direction

$$\bar{a} = \begin{bmatrix}-0.5286 & -0.0445 & 0.4950 & 0\end{bmatrix}^T \in \mathcal{N}(C\Lambda^{-1}BK\Gamma^{-1}).$$

Note that $\bar{a}$ is selected in such a way that only the first three states are compromised, while the fourth component of the vector is set to zero. This attack signal is introduced incipiently in the time interval $[500,\ 1250]$ and later maintained constant at the magnitude $1.5\bar{a}$, for all $k \in [1251,\ 2500]$. In addition, at $k = 2000$, the aforementioned output set-point modification $\Delta\bar{y} = [-2.884,\ 0]^T$ is imposed from the supervisory layer.

The temporal evolution of the outputs is depicted in Figure 3.7. In this figure, the data received at the supervisory layer is plotted in red for all $k \in \mathcal{K}_{REP}$, while the real output data is plotted in blue. The effect caused by the injection of the, firstly incipient and later constant, attack signal $\bar{a}$ in the real system outputs can also be appreciated. Additionally, the figure also shows the set-point modification imposed from the supervisory layer at $k = 2000$.

On the other hand, Figure 3.8 plots the residual signals that are generated at the supervisory layer throughout the attack scenario. Similar to Scenario I, an interval hull approximation has been computed in order to obtain the healthy residual set limits, and being able to plot the temporal evolution of the residual signal. In this figure, it can be seen how the incipient injection of the attack signal causes a small transient in the residuals for the interval $[500,\ 1250)$. Note that no detectability conditions are given for this transient. However, since the vector $\bar{a}$ is in the null space of $C\Lambda^{-1}BK\Gamma^{-1}$, the constant injection of the attack signal during the interval $[1250,\ 2000]$ provokes that the attacked residual set is centered again at the same point than the healthy residual set. Finally, the effect of the set-point modification $\Delta\bar{y} = [-2.884,\ 0]^T$ in

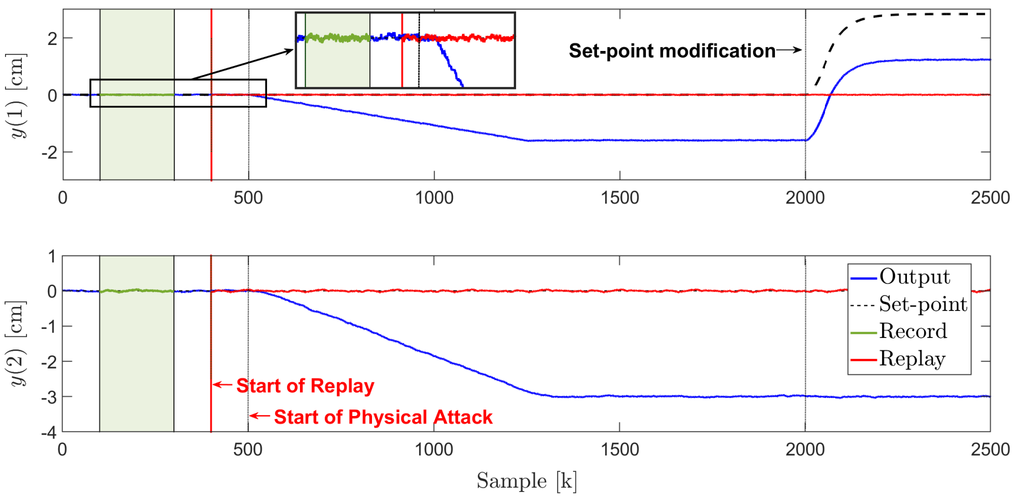


Figure 3.7: Temporal evolution of the system outputs - Scenario II.

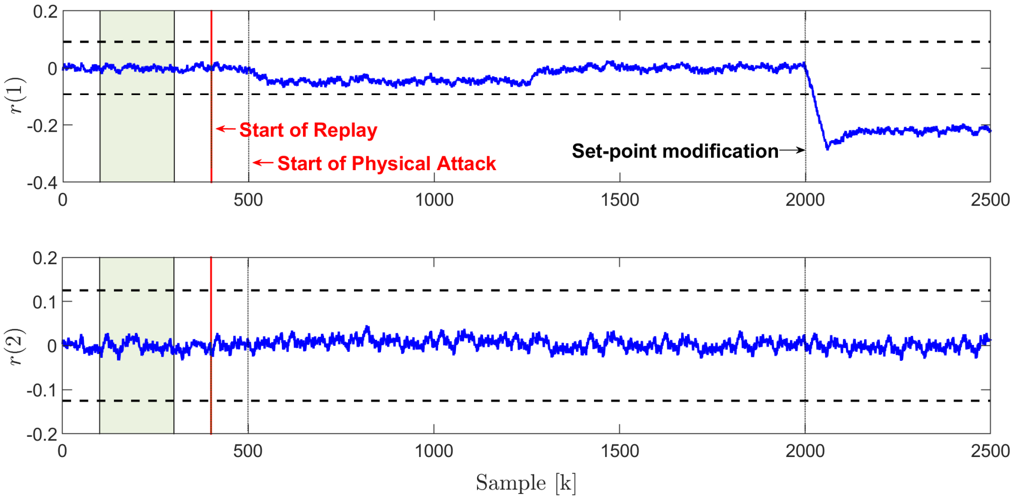


Figure 3.8: Temporal evolution of the residual signal - Scenario II.

the residuals, and the guaranteed attack detection induced by the residual signal exiting the healthy set, can be appreciated after $k = 2000$.

Moreover, Figure 3.9 shows the evolution of the residual signal during the previous scenario. This figure clearly shows how, by means of the imposed set-point modification, the residual signal converges towards the attacked residual set, and thus, due to the null intersection of the sets, the attack detection is guaranteed.

In addition, the capability of this type of replay attacks to mask the effect of an attack signal $\bar{a}$ is illustrated. In this regard, on the one hand, Figure 3.10a shows the evolution of the residual signal in the time interval $[0,\ 2000)$ for the case scenario described above, that is, with a concatenation of the replay of the output data and the injection of the attack signal $\bar{a}$. On the other hand, Figure 3.10b shows the evolution of the residual signal for the case where only the attack signal $\bar{a}$ is injected, i.e., the attacker does not replay the output data. It can be seen how, in the latter case, the residual signal exits the healthy set triggering thus the alarm, whereas, in the former case, the attack remains undetected from the supervisory layer point of view.

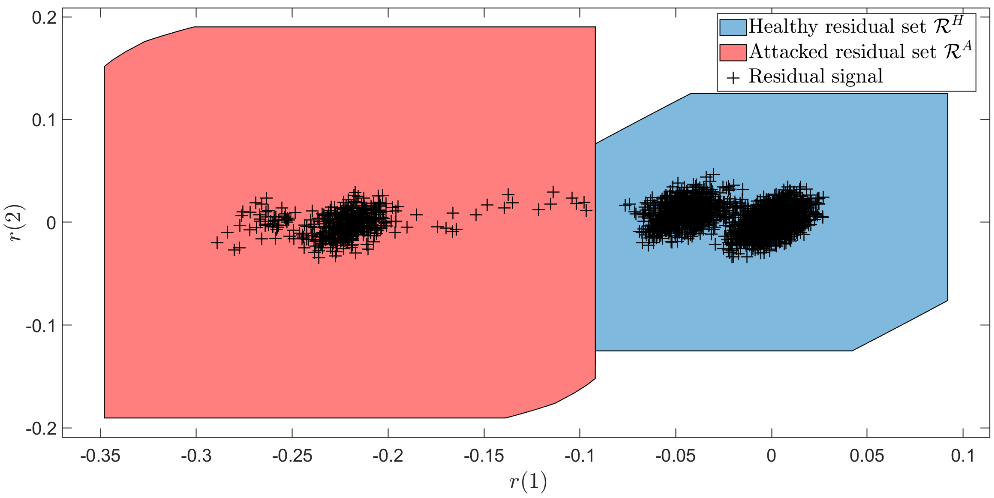
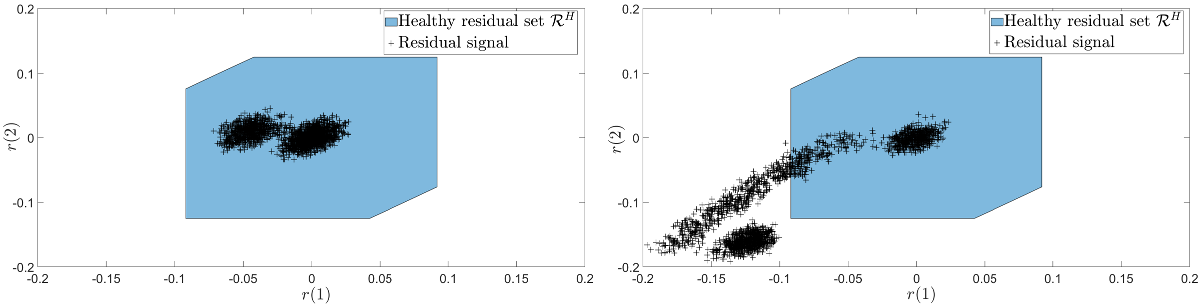


Figure 3.9: Residual signal - Scenario II.



(a) Replay of the output data.

(b) No replay of the output data.

Figure 3.10: Effect of the data replay - Scenario II.

3.6.3 Watermark sequence design

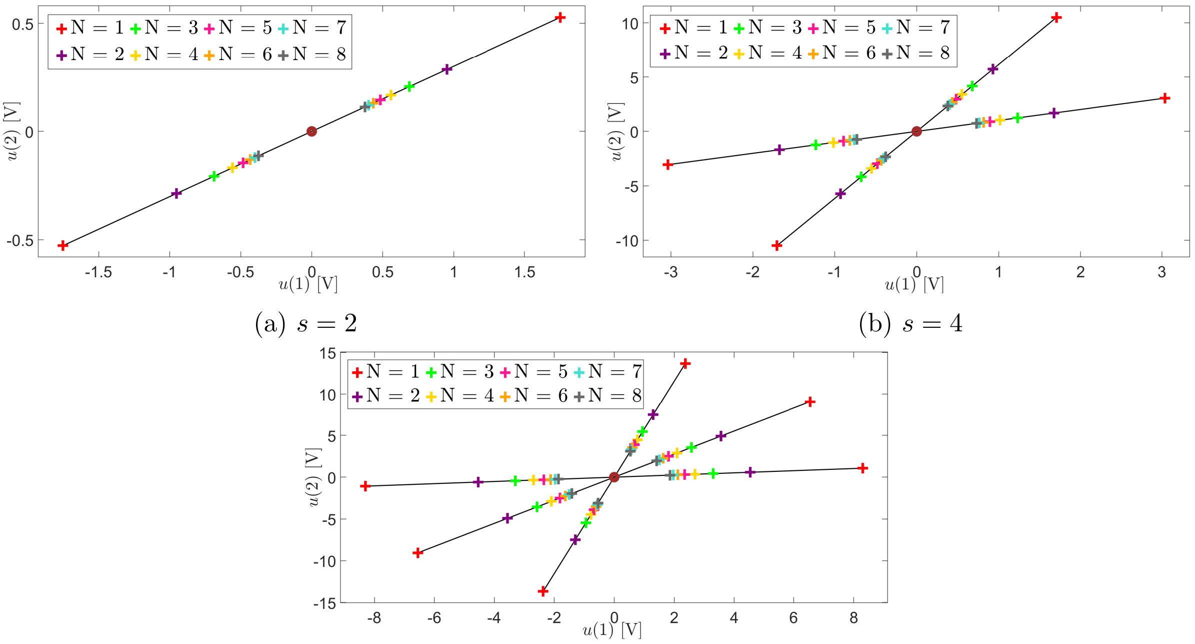
Here, the design of a set of N steps watermark sequences is addressed. As discussed in Section 3.5, these sequences have been particularized to guarantee the attack detection for the first attack scenario.

Henceforth, the single-level optimization problem in charge of the s sets design is posed to minimize the following cost function

$$J(c_i, \beta_i) = \sum_{i=1}^{s} \|\beta_i h_i\|_\infty,$$

that is, the peak value of the input space required by the watermark sequences is intended to be minimized.

The MIP in (3.58) has been run for different security indexes s and parameters $\hat{\delta}_m = 50$ and $\epsilon = 10^{-3}$. In this regard, Figure 3.11 shows the set partitions obtained for security indexes $s = \{2, 4, 6\}$ and time steps N ranging from 1 to 8. In those simulations, the zonotopes directions h_i have been randomly selected, and the pair symmetry as discussed in Section 3.5.3.1 imposed. All the obtained sets present one vertex in the origin (red dot) and the other vertex is represented by a cross with the corresponding colour. For all the cases, it can be seen the existing trade-off between the length of the watermark sequence N and the set size. Moreover, it can be



(a) $s = 2$

(b) $s = 4$

(c) $s = 6$

Figure 3.11: Input space partitions.

appreciated how, as the security index s increases, the size of the computed sets also increases. The discussion performed above is evidenced in Table 3.1, where the cost functions obtained for different combinations of s and N are reported.

Below a particular case with parameters $N = 8$ and $s = 4$ is illustrated. All the watermark sequences retrieved from solving the optimization problem (3.58) present a constant structure, i.e., $\boldsymbol{\xi}_{0:8} = (\boldsymbol{\xi}_{1\times 8})$. The obtained sequences are

$$\begin{aligned} \boldsymbol{\xi}_{0:8}[1] &= \left(\begin{bmatrix} 0.7296 \\ 0.7353 \end{bmatrix}_{1\times 8} \right), \quad \boldsymbol{\xi}_{0:8}[2] = \left(\begin{bmatrix} -0.7296 \\ -0.7353 \end{bmatrix}_{1\times 8} \right), \\ \boldsymbol{\xi}_{0:8}[3] &= \left(\begin{bmatrix} 0.3803 \\ 2.3424 \end{bmatrix}_{1\times 8} \right), \quad \boldsymbol{\xi}_{0:8}[4] = \left(\begin{bmatrix} -0.3803 \\ -2.3424 \end{bmatrix}_{1\times 8} \right). \end{aligned} \tag{3.60}$$

		$\boldsymbol{s = 2}$	$\boldsymbol{s = 4}$	$\boldsymbol{s = 6}$
N	**1**	3.5042	27.1360	62.0543
	2	1.9034	14.8495	33.8830
	3	1.3759	10.8168	24.6214
	4	1.1168	8.8492	20.0928
	5	0.9651	7.7090	17.4623
	6	0.8671	6.9833	15.7806
	7	0.7998	6.4948	14.6441
	8	0.7516	6.1554	13.8481

Table 3.1: Cost function for different combinations of (N, s).

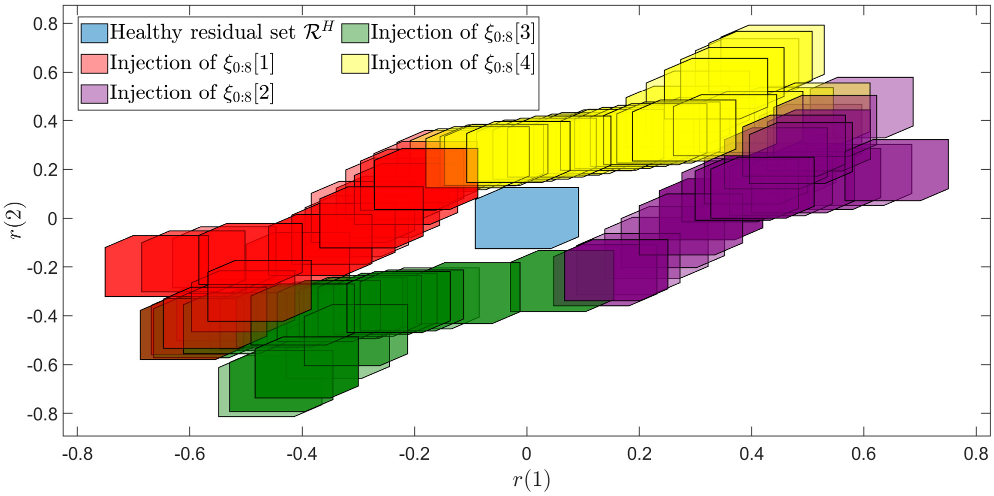


Figure 3.12: Watermark sequences - $(N, s) = (8, 4)$.

In this regard, Figure 3.12 shows the displacement of the attacked residual set $\mathcal{R}_8^A$ under the injection of the above sequences for the different combinations of $\boldsymbol{\xi}_{0:8}^r$ that may arise, as long as the recordings does not contain fragments of the same injected sequence (see Section 3.5.1). For all these cases, the attack detection is guaranteed as a consequence of the null intersection of the healthy and attacked residual sets.

At this point, a replay attack is simulated. The studied attack scenario considers that the input and output data sets sent from regulatory to supervisory layer are compromised (Scenario I). Additionally, in this case, the low-level controller is designed as an LQR controller with weighting matrices $Q_K = 10I_4$ and $R_K = I_2$, that yields the following controller gain

$$K = \begin{bmatrix} 2.2171 & 0.0004 & 0.1650 & 0.0744 \\ -0.0016 & 2.4915 & 1.0060 & 0.4850 \end{bmatrix}.$$

The considered time intervals are

$$\mathcal{K}_{REC} = [120,\ 150],$$
$$\mathcal{K}_{REP} = [400,\ 500],$$

and thus no attack vector a_k is contemplated in this case in order to better illustrate the effect of the watermark sequences on the system outputs.

The designed watermark sequences Eq. (3.60) are injected at random time instants, satisfying the design criterion that no watermark signal is added during at least $N = 8$ samples. Furthermore, these sequences are included in the control loop following a random pattern.

Figure 3.13 shows the temporal evolution of the watermark sequences that are injected in the control loop. Note that, in this case, the recorded sequence (green background) does not contain any watermark signal, i.e., $\boldsymbol{\xi}_{k':k'+8}^r = 0_{1\times 8}$. Accordingly, the attack is not detected until (at $k' = 467$) the sequence $\xi_{k':k'+8}[4]$ is injected, guaranteeing that in the worst-case the attack would be detected at $k' + N = 475$. Figure 3.14 shows the effect of the above watermark sequences on the system outputs.

Finally, Figure 3.15 depicts the temporal evolution of the residual signal. In this figure it can be seen that during the time interval $[k_1,\ k'] = [400,\ 467]$ the attack is no detected (red

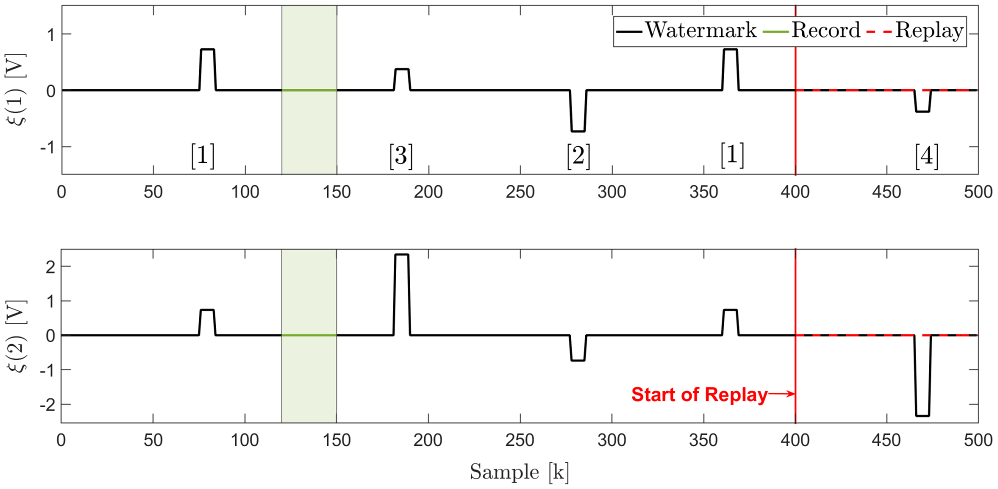


Figure 3.13: Temporal evolution of the watermark sequences.

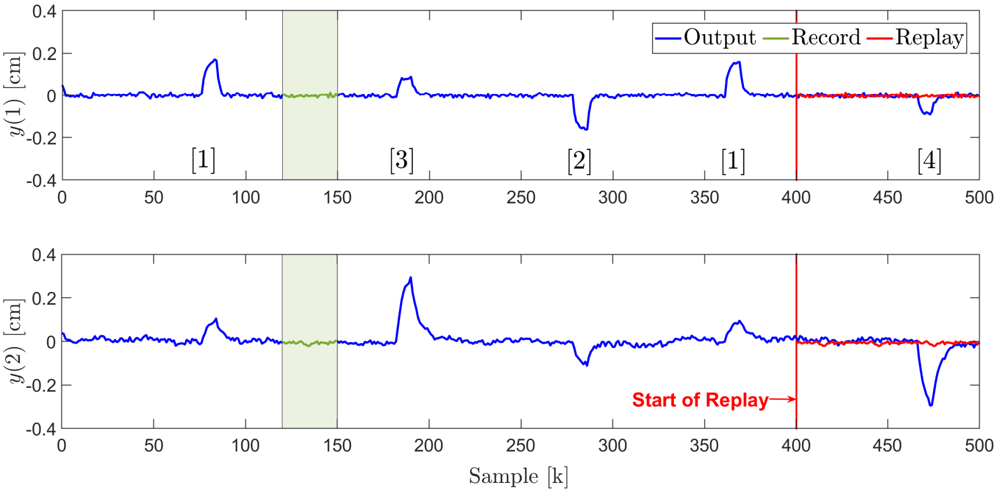


Figure 3.14: Temporal evolution of the system outputs - Watermark sequences.

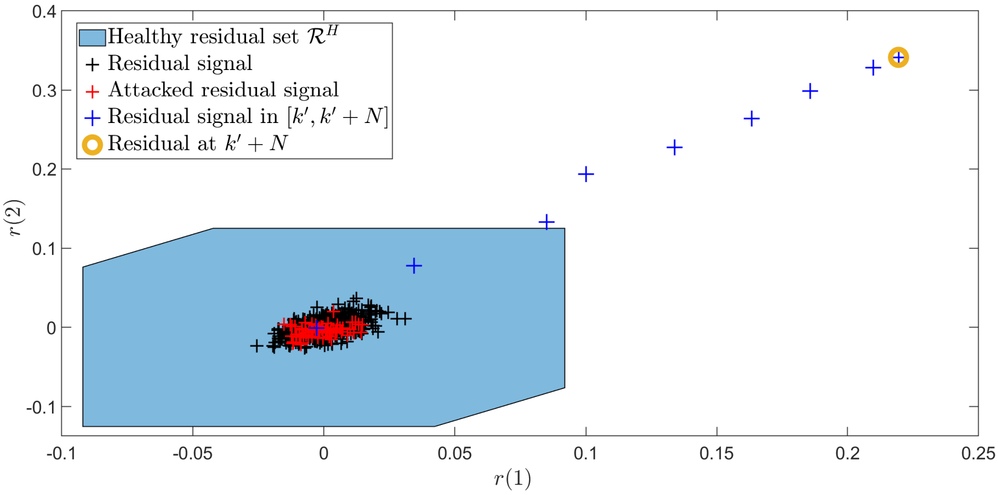


Figure 3.15: Residual signal - Watermark sequences.

crosses), and that the injection of $\boldsymbol{\xi}_{k':k'+8}[4]$, with $\boldsymbol{\xi}^r_{k':k'+8} = 0_{1\times 8}$, causes the residual signal to exit the healthy residual set. The residual signal at $k' + N$ is highlighted in yellow. Note that in this case, attack detection is achieved before $k' + N$ (blue crosses).

3.7 Summary

In this chapter, zonotopic sets have been used in order to develop a set-invariance analysis on the detectability of replay attacks against the supervisory layer. In spite of its inherent conservativeness, invariance analysis is presented as an interesting tool for assessing the attack detectability since it allows to derive analytical expressions that guarantee the detection. Two different attack scenarios that depend on the attacker's capabilities have been studied. It was shown how, even in the case where the attacker is able to replay only sensor measurements, no guarantees regarding attack detectability can be given unless a temporal mismatch between record and replay phases is forced by means of a signal sent from the supervisory layer. Moreover, the strengths of the zonotopic set representation have been exploited for designing efficient watermark sequences that take into account the transient behaviour of the residual signal. The next chapter delves into the physical watermarking problem and, in particular, analyses the relationship between zonotopic-based watermarking approaches and stochastic approaches within the context of optimal control.

Chapter 4

A zonotopic-based watermarking design to detect replay attacks

This chapter further exploits the use of zonotopes for the design of watermark signals. The proposed approach makes use of the recent analogy found between stochastic and zonotopic-based estimators to propose a deterministic counterpart of current approaches that study the replay attack in the context of stationary Gaussian processes. In this regard, the zonotopic analogous case where the control loop is closed based on the estimates of a Zonotopic Kalman Filter, is analysed. This formulation allows generating a new performance metric that is related to the Frobenius norm of the prediction zonotope. Hence, the steady-state operation of the system can be related with the size of the minimal Robust Positive Invariant set of the estimation error. Furthermore, analogous expressions concerning the impact that a zonotopic/Gaussian watermark signal has on the system operation, are derived. Finally, a novel zonotopically bounded watermark signal that guarantees the attack detection by causing that the residual vector abandons the healthy residual set during the replay phase of the attack, is presented. The proposed approach is illustrated in simulation using a four-tank process.

4.1 Introduction

This chapter delves into the design of additive watermark signals that guarantee the replay attack detection. On this subject, as discussed in Chapter 2, in the pioneering work of Mo and Sinopoli [2009], the authors not only model and analyse the effect of replay attacks in the framework of stationary Gaussian processes, but also propose the addition of an exogenous signal to the control loop in order to improve the detectability of the attack in those scenarios where the attack remains undetected. From this moment on, the design of those exogenous signals, which are denoted with the term *physical watermarking* [Mo et al., 2015], has been intensively studied until becoming a well-adopted technique for the detection of replay attacks.

At this point, it must be highlighted that the vast majority of works in the literature have addressed the physical watermarking design problem from a stochastic point of view and, in particular, within the context of stationary Gaussian processes. The reader is referred to Section 2.2.1, where a detailed analysis on the watermarking-based techniques presented in the literature has been carried out. Conversely, the main aim of this chapter is to present a deterministic set-based counterpart to such stochastic watermarking schemes.

To that end, zonotopes will be used motivated by the results presented in Combastel [2015]. In this paper, the author introduces the notion of covariation matrix of a zonotope in order to propose a robust state observer. Furthermore, by minimizing the weighted Frobenius norm of the prediction zonotope, expressions analogous to the standard stochastic Kalman filter are obtained. Hence, this analogy between zonotope-based set-membership approaches and stochastic approaches will be further exploited for addressing the watermarking design problem. In this regard, with the final intention of assessing the impact that the injection of a bounded signal has on an optimal control scheme, the system performance is analysed for the case where the control loop is closed by means of a zonotopic Kalman filter (ZKF), which gives rise to the zonotopic based analogue of the Linear Quadratic Gaussian (LQG) control.

Note that, the main strength of posing the watermarking design under a set-based paradigm is that it becomes possible to deterministically infer the system operation by testing online whether or not a residual signal belongs to a set associated with the healthy operation of the system. Hence, this binary condition in the detection, motivates the design of an additive watermark signal that forces the residual signal out of its healthy residual set, thus generating a guaranteed replay attack detection mechanism.

The contributions of this chapter are as follows

1. An optimal control (in the LQR sense) that operates based on the estimates of a ZKF is proposed. It is proven that this control scheme minimizes a new cost function that considers the weighted Frobenius norm of the system states, in what can be seen as the zonotopic counterpart of the expected value of a quadratic cost function in an LQG scheme.
2. The new zonotope-based cost function is used to evaluate the impact that a zonotopically bounded watermark signal has on the stationary closed-loop operation of the system. Analogous expressions to the performance loss induced by an exogenous Gaussian signal in a LQG system, are obtained. This analogy serves as a bridge to relate both ways of facing the problem.
3. A zonotopically bounded watermark signal that minimizes the new cost function is proposed. This detection scheme, injects a known signal in the system inputs, while filtering its effect through the estimations generated using the outputs data. Since the replay phase of the attack entails that the known signal is no longer observable, then the estimation error generates a new residual signal that is destabilized whenever the output data are being replayed back.

The remainder of this chapter is organized as follows: Section 4.2 introduces the system under study. Then, in Section 4.3 the optimal finite/infinite horizon control problem based on the estimates of a ZKF is analysed. Section 4.4 analyses the impact of a bounded watermark signal on the previous optimal control loop. Besides, in Section 4.5 a new guaranteed replay attack detection scheme is introduced whereas in Section 4.6 the proposed results are validated in simulation using a well-known control benchmark. Finally, in Section 4.7 the main conclusion of the chapter are drawn.

4.2 Problem statement

In order to propose a set-based counterpart to the works in the literature mainly dealing with stochastic approaches, this chapter focuses on replay attacks where the control loop is compro-

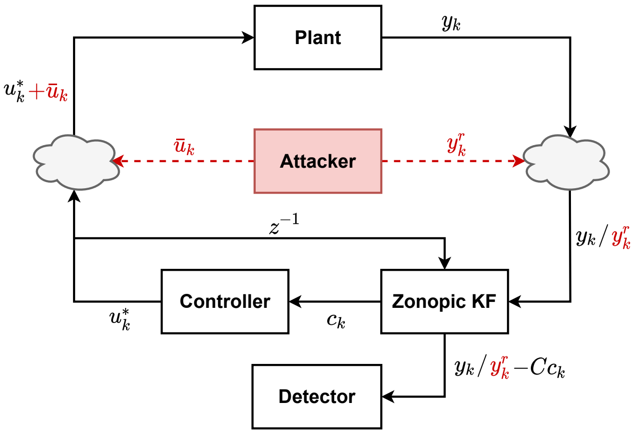


Figure 4.1: Overall scheme and replay attack representation.

mised. Accordingly, the control scheme under consideration is illustrated in Figure 4.1, in such a way that, conversely to Chapter 3, the control loop is closed remotely by means of a state estimate feedback control policy. In this scenario, a malicious attacker can access the communications layer and spoof the output data received on the controller side of the network, causing the controller to operate on fake data.

It should be noted that, considering bounded process disturbances and measurements noise, and that the state estimates are generated by a ZKF, this attack scenario becomes a set-based analogous of the standard replay attack formulated in the context of LQG systems. On this subject, the different elements of the scheme in Figure 4.1, including the subsequent addition of a watermarking block for active attack detection, will be analysed from a set-based perspective using zonotopic set-representations. In order to do so, the remainder of this section is devoted to introducing the system model as well as the zonotopic KF.

4.2.1 System description

This chapter considers that the plant illustrated in Figure 4.1 is modelled by means of a discrete-time LTI system with disturbance of the form

$$x_{k+1} = Ax_k + Bu_k + Ew_k, \tag{4.1a}$$
$$y_k = Cx_k + Fv_k, \tag{4.1b}$$

where A, B, C, E and F are the state-space matrices with adequate dimensions, $x_k \in \mathbb{R}^{n_x}$ is the state vector, $u_k \in \mathbb{R}^{n_u}$ is an exogenous input signal and $y_k \in \mathbb{R}^{n_y}$ is the output vector. The system state at $k = 0$ is assumed to satisfy $x_0 \in \langle c_0, H_0\rangle$, where $c_0 \in \mathbb{R}^{n_x}$ and $H_0 \in \mathbb{R}^{n_x \times m_0}$.

For the sake of simplified notation, it is considered that for all $k \in \mathbb{N}$ process disturbances $w_k \in \mathbb{R}^{n_w}$ and sensor noise $v_k \in \mathbb{R}^{n_v}$ satisfy

$$w_k \in \langle 0, I_{n_w}\rangle, \qquad v_k \in \langle 0, I_{n_v}\rangle. \tag{4.2}$$

In this regard, as long as the uncertainties are bounded within zonotopes, a zero-centered unitary box representation like (4.2) can be obtained by performing: I) a change of coordinates that shifts the uncertainty center to the zero; II) a coherent modification of the distribution matrices E and F, e.g., given $\tilde{\mathcal{W}} = \langle 0, H_w\rangle$, with $H_w \in \mathbb{R}^{n_w \times m_w}$, using Property A.8 it follows that $E\tilde{\mathcal{W}} = \langle 0, [EH_w]\rangle = \tilde{E}\mathcal{W}$, with $\tilde{E} = EH_w$ and $\mathcal{W} = \langle 0, I_{m_w}\rangle$.

Assumption 4.1. The pairs (A, B) and (A, C) are assumed to be stabilizable and detectable, respectively.

4.2.2 Zonotopic KF

The system model (4.1) and the bounded uncertainties (4.2), are used in order to generate the zonotope-based state estimator presented below. As shown in Figure 4.1, this observer is used for closing the control loop through an optimal state estimate control policy, as well as for generating the residual signal that is used for supervising the system operation. Both uses are discussed in more detail throughout the chapter.

According to Combastel [2015], given an initial state $x_0 \in \langle c_0,\, H_0\rangle$, then recursively defining the center estimate c_k and the generators matrix H_k as

$$c_{k+1} = (A - G_k C)c_k + Bu_k + Gy_k, \tag{4.3a}$$

$$H_{k+1} = [(A - G_k C)\bar{H}_k,\ E,\ -G_k F], \tag{4.3b}$$

$$\bar{H}_k = \downarrow_{q,W} (H_k), \tag{4.3c}$$

the state inclusion property $x_k \in \langle c_k,\, H_k\rangle$ holds for all $k \geq 0$.

Moreover, the optimal observer gain that minimizes the F_W-radius of the prediction zonotope $\langle c_{k+1}, H_{k+1}\rangle$, i.e., $G_k^* = \arg\min_{.G} \|H_{k+1}\|_{F,W}^2$, is computed as

$$G_k^* = AK_k^* = A\bar{P}_k C^T (C\bar{P}_k C^T + Q_v)^{-1}, \tag{4.4}$$

where P_k, $\bar{P}_k$, $Q_w \in \mathbb{R}^{n_x \times n_x}$ and $Q_v \in \mathbb{R}^{n_y \times n_y}$ are the covariation matrices

$$P_k = H_k H_k^T, \quad \bar{P}_k = \bar{H}_k \bar{H}_k^T, \quad Q_w = EE^T, \quad Q_v = FF^T,$$

and matrix P_k satisfies

$$P_{k+1} = A\bar{P}_k A^T + Q_w - A\bar{P}_k C^T (C\bar{P}_k C^T + Q_v)^{-1} C\bar{P}_k A^T. \tag{4.5}$$

Observe that the weighted criterion in the F_W-radius is selected in order to be consistent with the reduction operator $\downarrow_{q,W}$ presented in Property A.9. Furthermore, if the generators matrix is sorted as $H_k = [H_k^a,\ H_k^b]$, in such a way that $\bar{H}_k = \downarrow_{q,W} (H_k) = [H_k^a,\ b(H_k^b)]$, then the difference in the covariation matrices induced by the reduction operator can be rewritten as

$$\bar{P}_k = P_k - H_k^b H_k^{bT} + b(H_k^b)^2. \tag{4.6}$$

4.3 LQZ control

This section analyses the optimal control problem for the case in which the feedback loop is closed using the estimates generated by a ZKF. In this regard, the following performance criterion is introduced in order to assess the system operation

Definition 4.1 (Zonotopic quadratic performance)**.** Given a matrix $S \in \mathbb{R}^{n\times n}$ and the unknown but zonotopically bounded vector $x \in \langle c, H\rangle \subset \mathbb{R}^n$. The performance of x is assessed as

$$\mathcal{Q}[x^T Sx] = c^T Sc + \|H\|_{F,S}^2 = c^T Sc + Tr[SP], \tag{4.7}$$

with $P = cov(\langle c, H\rangle) = HH^T$.

It must be highlighted that Definition 4.1 matches the expected value for a Gaussian random vector centered at c with covariance P, i.e., $x \sim \mathcal{N}(c, P)$. Furthermore, from Definition 4.1, it is straightforward to see that the operator $\mathcal{Q}[\cdot]$ satisfies the following property.

Proposition 4.1 (Distributive property)**.** Given the variables $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ such that $x \in \langle c_x, H_x \rangle \subset \mathbb{R}^n$ and $y \in \langle c_y, H_y \rangle \subset \mathbb{R}^m$, then

$$\mathcal{Q}[x^T S_x x + y^T S_y y] = \mathcal{Q}[x^T S_x x] + \mathcal{Q}[y^T S_y y], \tag{4.8}$$

with S_x and S_y real matrices of appropriate dimensions.

Proof. The left-hand side of (4.8) can be rewritten as

$$\mathcal{Q}\left[\begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}\right], \quad \text{with} \quad \begin{bmatrix} x \\ y \end{bmatrix} \in \left\langle \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \begin{bmatrix} H_x & 0 \\ 0 & H_y \end{bmatrix} \right\rangle.$$

Therefore, from Definition 4.1, it follows

$$\begin{aligned} \mathcal{Q}\left[\begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}\right] &= \begin{bmatrix} c_x^T & c_y^T \end{bmatrix} \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \begin{bmatrix} c_x \\ c_y \end{bmatrix} + Tr\left[\begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \begin{bmatrix} H_x & 0 \\ 0 & H_y \end{bmatrix} \begin{bmatrix} H_x^T & 0 \\ 0 & H_y^T \end{bmatrix}\right] \\ &= c_x^T S_x c_x + Tr[S_x P_x] + c_y^T S_y c_y + Tr[S_y P_y] \\ &= \mathcal{Q}[x^T S_x x] + \mathcal{Q}[y^T S_y y]. \end{aligned}$$

□

Below, the finite and infinite horizon control problems are analysed.

4.3.1 Finite horizon control

Given a non-measurable state for which a zonotopic state estimation $\langle c_k, \bar{H}_k \rangle \supseteq \langle c_k, H_k \rangle$ is generated by means of the ZKF described in Section 4.2.2. The Finite Horizon Linear Quadratic control is posed as computing the inputs sequence $\{u_k : k = 0, ..., N-1\}$ which minimizes the performance criterion

$$J(N, x_0, u) = \mathcal{Q}\Big[\sum_{k=0}^{N-1} (x_k^T W x_k + u_k^T U u_k) + x_N^T W_f x_N\Big], \tag{4.9}$$

where $W = W^T \succeq 0$, $W_f = W_f^T \succeq 0$, $U = U^T \succ 0$ and $x_0 \in \langle c_0, H_0 \rangle$.

At a generic instant k_1 and state estimate $x_{k_1} \in \langle c_{k_1}, \bar{H}_{k_1} \rangle$, the optimal *cost-to-go* of (4.9) is given by the expression

$$V_{k_1}(x_{k_1}) = \min_{u_{k_1}^*, \dots, u_{N-1}^*} \left\{ \mathcal{Q}\Big[\sum_{k=k_1}^{N-1} (x_k^T W x_k + u_k^T U u_k) + x_N^T W_f x_N\Big]\right\}.$$

Moreover, from dynamic programming, it is known that

$$V_k(x_k) = \min_{u_k^*} \Big\{ \mathcal{Q}[x_k^T W x_k + u_k^T U u_k] + V_{k+1}(x_{k+1}) \big| \hat{X}_k \Big\},$$

where $\hat{X}_k$ represents the set of estimations $\hat{X}_k = \{x_0, ..., x_k\}$, such that $x_k \in \langle c_k, \bar{H}_k \rangle$, and $V_{k+1}(x_{k+1})|\hat{X}_k$ is the *cost-to-go* obtained at $k+1$ given the sequence $\hat{X}_k$. In this regard, the following proposition can be derived.

Proposition 4.2. At each time instant, $V_k(x_k)$ is given by

$$V_k(x_k) = \mathcal{Q}[c_k^T S_k c_k] + s_k, \tag{4.10}$$

with S_k and s_k defined by the backwards recursions

$$S_k = A^T S_{k+1} A + W - A^T S_{k+1} B (B^T S_{k+1} B + U)^{-1} B^T S_{k+1} A, \tag{4.11a}$$
$$s_k = s_{k+1} + Tr[W \bar{P}_k] + Tr[S_{k+1}(G_k^*(C\bar{P}_k C^T + Q_v)G_k^{*T})], \tag{4.11b}$$

starting at $S_N = W_f$ and $s_N = Tr[W_f \bar{P}_N]$.

Proof. Backward induction will be used to prove (4.10). Given $x_N \in \langle c_N, \bar{H}_N \rangle$, since its center can be rewritten as $c_N = \langle c_N, 0 \rangle$, from Definition 4.1 it follows that $c_N^T W_f c_N = \mathcal{Q}[c_N^T W_f c_N]$. Accordingly, the expression

$$V_N(x_N) = \mathcal{Q}[x_N^T W_f x_N] = c_N^T W_f c_N + Tr[W_f \bar{P}_N],$$

satisfies (4.10). Thus, if $V_{k+1}(x_{k+1})$ satisfies (4.10), then, by means of Proposition 4.1, at a generic time instant

$$V_k(x_k) = \min_{u_k^*} \left\{ \mathcal{Q}[x_k^T W x_k] + \mathcal{Q}[u_k^T U u_k] + \mathcal{Q}[c_{k+1}^T S_{k+1} c_{k+1}] + s_{k+1} \right\},$$

with $x_k \in \langle c_k, \bar{H}_k \rangle$ and $u_k \in \langle u_k, 0 \rangle$.

Furthermore, from the outputs equation (4.1b), it follows that

$$y_k \in C\langle c_k, \bar{H}_k \rangle \oplus F\langle 0, I_{n_v} \rangle = \langle Cc_k, [C\bar{H}_k, \ \ F] \rangle,$$

recalling the equation of the zonotope center (4.3a) generated by the ZKF

$$c_{k+1} = (A - G_k^* C)c_k + Bu_k + G_k^* y_k.$$

Then, the prediction zonotope center satisfies

$$c_{k+1} \in \langle (Ac_k + Bu_k), [G_k^* C \bar{H}_k, \ \ G_k^* F] \rangle.$$

Therefore, from Definition 4.1, the following expressions are obtained

$$\begin{aligned} \mathcal{Q}[x_k^T W x_k] &= c_k^T W c_k + Tr[W \bar{P}_k], \\ \mathcal{Q}[u_k^T U u_k] &= u_k^T U u_k, \\ \mathcal{Q}[c_{k+1}^T S_{k+1} c_{k+1}] &= (Ac_k + Bu_k)^T S_{k+1} (Ac_k + Bu_k) + Tr[S_{k+1}(G_k^*(C\bar{P}_k C^T + Q_v)G_k^{*T})], \end{aligned}$$

with $\bar{P}_k = \bar{H}_k \bar{H}_k^T$, $\bar{H}_k = \downarrow_{q,W} (H_k)$ and $Q_v = FF^T$.

Using the expressions above, the *cost-to-go* can be rewritten as

$$\begin{aligned} V_k(x_k) = \min_{u_k^*} \Big\{ & u_k^{*T}(B^T S_{k+1} B + U)u_k^* + 2u_k^{*T} B^T S_{k+1} A c_k + c_k^T (W + A^T S_{k+1} A) c_k \\ & + Tr[W \bar{P}_k] + Tr[S_{k+1}(G_k^*(C\bar{P}_k C^T + Q_v)G_k^{*T})] + s_{k+1} \Big\}, \end{aligned}$$

and thus the optimal u_k^* that minimizes $V_k(x_k)$ is

$$u_k^* = -(B^T S_{k+1} B + U)^{-1} B^T S_{k+1} A c_k.$$

Consequently, $V_k(x_k)$ results in

$$\begin{aligned} V_k(x_k) =& c_k^T \big(W + A^T S_{k+1} A - A^T S_{k+1} B (B^T S_{k+1} B + U)^{-1} B^T S_{k+1} A \big) c_k \\ & + Tr[W \bar{P}_k] + Tr[S_{k+1} (G_k^* (C \bar{P}_k C^T + Q_v) G_k^{*T})] + s_{k+1} = c_k^T S_k c_k + s_k. \end{aligned}$$

Since at the given time instant $x_k \in \langle c_k,\ \bar{H}_k \rangle$, then the center satisfies $c_k = \langle c_k, 0 \rangle$ and thus $c_k^T S_k c_k = \mathcal{Q}[c_k^T S_k c_k]$, which concludes the proof. □

According to the development presented in the proof of Proposition 4.2, the control law that minimizes the cost function (4.9) is the LQR controller

$$u_k^* = -(B^T S_{k+1} B + U)^{-1} B^T S_{k+1} A c_k = -L_k^* c_k, \tag{4.12}$$

which operates based on the center of the ZKF estimator.

In order to set it apart from the stochastic LQG control, in the sequel the zonotopic version will be denoted as Linear Quadratic Zonotopic (LQZ) control. For this case, from (4.10)-(4.11) it follows that the optimal value of the cost function is

$$J_N = V_0(x_0) = \mathcal{Q}[c_0^T S_0 c_0] + Tr[\sum_{k=0}^{N-1} W \bar{P}_k] + Tr[\sum_{k=0}^{N-1} S_{k+1} (G_k^* (C \bar{P}_k C^T + Q_v) G_k^{*T})] + Tr[W_f \bar{P}_N], \tag{4.13}$$

with $\mathcal{Q}[c_0^T S_0 c_0] = c_0^T S_0 c_0$.

At this point, it is worth noting the fact that if no over-approximation is introduced during the horizon N, i.e., if $\bar{P}_k = P_k$, then (4.13) is equal to the cost function obtained in an LQG system but substituting the covariation matrices for the appropriate covariance matrix of the Gaussian variables. Therefore, it can be concluded that an optimal control scheme has the same impact on the expected value of the state variables when the uncertainties are modelled as Gaussian variables, that on the F_W-radius of the state bounding zonotope under bounded uncertainties.

Remark 4.1. According to Theorem 11 in Combastel [2015], if the zonotope $\langle 0, H_k \rangle = \langle 0, [H_k^a,\ H_k^b] \rangle$, whose columns have been sorted in decreasing weighted norm, is over-approximated by $\langle 0, \bar{H}_k \rangle = \langle 0, \downarrow_{q,W} (H_k) \rangle = \langle 0, [H_k^a\ b(H_k^b)] \rangle$, then there exist a value $q_0 \in \mathbb{N}_+$ such that by selecting $q \geq q_0$ it can be ensured that the resulting sequence of zonotopes has bounded F_W-radius. Nevertheless, under the effect of the reduction operator, the gain G_k^* and covariation $\bar{P}_k$ matrices may not converge to some fixed values. This is due to the coupling between: I) the convergence of the over-approximated zonotope $\langle 0, \bar{H}_k \rangle$ to a fixed structure, i.e., such that $b(H_k^b)^2 - H_k^b H_k^{bT}$ becomes constant; II) the evolution of the covariation recursion (4.5) which depends on the relationship $\bar{P}_k - P_k = b(H_k^b)^2 - H_k^b H_k^{bT}$.

4.3.2 Infinite horizon control

Since the analysis of replay attacks focuses on control systems that operate for long periods of time, the remainder of this chapter will focus on systems whose control law is obtained by solving the infinite problem that is formulated as

$$J_\infty = \lim_{N\to\infty} \frac{1}{N} \mathcal{Q}\Big[\sum_{k=0}^{N-1} x_k^T W x_k + u_k^T U u_k\Big], \tag{4.14}$$

where X and U are SPD matrices and the state x_k is not measurable. Because the infinite horizon problem does not depend on the time to go, i.e., it is shift invariant, the following well-known time-invariant controller and observer gains are obtained:

- From Assumption 4.1 and the consideration of $(A, W^{1/2})$ being detectable, the optimal infinite horizon controller matches the steady-state finite horizon control given by the constant control gain
$$L^* = (B^T S_\infty B + U)^{-1} B^T S_\infty A,$$
where S_∞ is the unique positive definite solution of the Discrete Algebraic Ricatti Equation (DARE)
$$S_\infty = A^T S_\infty A + W - A^T S_\infty B(B^T S_\infty B + U)^{-1} B^T S_\infty A, \tag{4.15}$$
such that the matrix $A - BL^*$ is asymptotically stable [Bitmead and Gevers, 1991].

- From Assumption 4.1 and under the consideration that $Q_v = Q_v^T \succ 0$ and that $(A, Q_w^{1/2})$ is stabilizable, the optimal infinite horizon observer gain is
$$G^* = AP_\infty C^T (CP_\infty C^T + Q_v)^{-1},$$
with P_∞ the unique positive definite solution of the DARE
$$P_\infty = AP_\infty A^T + Q_w - AP_\infty C^T (CP_\infty C^T + Q_v)^{-1} CP_\infty A^T, \tag{4.16}$$
such that the matrix $A - G^* C$ is asymptotically stable.

Henceforth, the estimation error is defined as $e_k = x_k - c_k$. Thus, from the comparison of (4.1a) and (4.3a) for the fixed gain G^*, it follows that the dynamics of the estimation error are governed by

$$e_{k+1} = (A - G^* C)e_k + Ew_k - G^* F v_k, \tag{4.17}$$

such that the inclusion property $e_{k+1} \in \langle 0, H_{k+1}\rangle$ is satisfied for

$$H_{k+1} = [(A - G^* C)H_k,\ E,\ -G^* F], \tag{4.18}$$

and any initial $e_0 \in \langle 0, H_0\rangle$. Regarding the evolution of the estimation error, the following proposition is introduced.

Proposition 4.3. The F-radius of the mRPI set Φ for (4.17) is $\|\Phi\|_F^2 = Tr[P_\infty]$.

Proof. Define the following matrices and zonotope

$$\hat{A} = A - G^* C, \qquad \hat{B} = [E,\ -G^* F], \qquad \hat{W} = \left\langle \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_{n_w} & 0 \\ 0 & I_{n_v} \end{bmatrix} \right\rangle,$$

where $\hat{A}$ is asymptotically stable. Therefore, since $\hat{W}$ is a zonotope and making use of Property A.7 and Property A.8, it follows that

$$\bigoplus_{i=0}^{k} \hat{A}^i \hat{B} \tilde{W} = \langle 0, \hat{H}_{k+1}\rangle = \langle 0, [\hat{A}^k \hat{B},\ \hat{A}^{k-1} \hat{B},\ \dots,\ \hat{B}]\rangle,$$

with the covariation matrix $P_{k+1} = \hat{H}_{k+1}\hat{H}_{k+1}^T = \sum_{i=0}^{k} \hat{A}^i \hat{B}\hat{B}^T \hat{A}^{Ti}$.

Given an asymptotically stable system, from [Kolmanovsky and Gilbert, 1998, Section 4] it is known that the mRPI set Φ exist, its unique and defined as $\Phi = \bigoplus_{i=0}^{\infty} \hat{A}^i \hat{B}\hat{W}$. Accordingly, $P_\Phi = \lim_{k\to\infty} P_{k+1} = \sum_{i=0}^{\infty} \hat{A}^i \hat{B}\hat{B}^T \hat{A}^{Ti}$ is the covariation matrix of Φ, which, since $\hat{A}$ is asymptotically stable, is the unique solution of Lyapunov equation $P_\Phi = \hat{A}P_\Phi\hat{A}^T + \hat{B}\hat{B}^T$ [Hamilton, 2020].

On the other hand, selecting $G^* = AP_\infty C^T(CP_\infty C^T + Q_v)^{-1}$, matrix P_∞ is the unique positive definite solution of (4.16), which can be rewritten in the form $P_\infty = \hat{A}P_\infty\hat{A}^T + \hat{B}\hat{B}^T$. Hence, it follows that $P_\Phi = P_\infty$, and thus $\|\Phi\|_F^2 = Tr[P_\Phi] = Tr[P_\infty]$. □

As discussed in Chapter 3, the exact computation of the mRPI set can only be achieved under the restrictive assumption that the system dynamics are nilpotent [Mayne and Schroeder, 1997]. Hence, in the sequel the mRPI set Φ will be outer-approximated through an RPI set. Note that, by means of recursively refining an outer RPI set, Φ can be approximated with arbitrarily precision at the price of increasing the complexity of the over-approximating set [Rakovic et al., 2005, Olaru et al., 2010]. The computation of an initial zonotopic RPI set and its recursive refinement in a way that guarantees an ϵ-approximation to the mRPI set is discussed in Section A.3 of the Appendix A.

In accordance with the stated on the preceding paragraph, below it will be considered that the system monitoring in the steady-state is performed using a reduced order RPI zonotopic set $\langle 0, \tilde{H}\rangle$, such that $\langle 0, \tilde{H}\rangle \supseteq \Phi$. Furthermore, the asymptotic stability of the closed-loop system (4.17) guarantees the convergence of its trajectories towards its mRPI set. Accordingly, there exist a finite time instant k^* for which $e_{k^*} \in \langle 0, \tilde{H}\rangle$, and thus $e_k \in \langle 0, \tilde{H}\rangle$ for all $k \geq k^*$.

For the ideal scenario where the mRPI set Φ is used in order to bound the estimation error steady-state value, the optimal infinite horizon cost function is given by

$$\begin{aligned} J_\infty &= \lim_{N\to\infty} \frac{1}{N} \mathcal{Q}\Big[\sum_{k=0}^{N-1} x_k^T W x_k + u_k^T U u_k\Big] \\ &= Tr[WP_\infty] + Tr[S_\infty(G^*(CP_\infty C^T + Q_v)G^{*T})]. \end{aligned} \tag{4.19}$$

On the other hand, if the zonotopic over-approximation $\langle 0, \tilde{H}\rangle \supset \Phi$ is used instead, the cost function is

$$\tilde{J}_\infty = Tr[W\tilde{P}] + Tr[S_\infty(G^*(C\tilde{P}C^T + Q_v)G^{*T})], \tag{4.20}$$

with $\tilde{P} = \tilde{H}\tilde{H}^T \neq P_\infty$.

4.3.3 Anomaly detector

The system operation is assumed to be monitored by means of an anomaly detector similar to the one presented in Chapter 3. The structure of this detector is recalled here for a better exposure of later developments. In this regard, the system operation is assessed based on the values adopted by the residual signal

$$r_k = y_k - Cc_k = Ce_k + Fv_k, \tag{4.21}$$

with $e_k = x_k - c_k \in \langle 0, H_k\rangle$.

Furthermore, in order to assess the system performance in the steady-state the zonotopic RPI set $\langle 0, \tilde{H} \rangle$ is used. Therefore, in healthy operation, the residual signal satisfies

$$r_k \in \mathcal{R}^H, \quad \forall k \geq k^*,$$

where the healthy residual set $\mathcal{R}^H$ is computed as

$$\mathcal{R}^H = \langle c_r, H_r \rangle = C\langle 0, \tilde{H} \rangle \oplus F\langle 0, I_{n_v} \rangle = \langle 0, [C\tilde{H},\ F] \rangle.$$

Hence, the presence of anomalies is assessed according to

$$\begin{cases} r_k \in \mathcal{R}^H & \implies \text{Healthy operation,} \\ \text{otherwise} & \implies \text{Something is wrong.} \end{cases} \tag{4.22}$$

4.4 Replay attack detectability

Below, Assumption 3.4, which establishes that the system under study satisfies $k \geq k^*$, is considered. Furthermore, the attack time windows and notation introduced in Section 2.3 are employed. Accordingly, by carrying out a development similar to the one followed in Chapter 3, the residual signal during the replay phase of the attack is governed by

$$r_k^a = y_k^r - Cc_k^a = y_k^r - Cc_k^r + C(c_k^r - c_k^a) = r_k^r + C(c_k^r - c_k^a), \tag{4.23}$$

such that from Assumption 3.4 it follows that $r_k^r \in \mathcal{R}^H$.

Besides, the last term of (4.23) represents the difference between the centers of the estimates (cf. (4.3a)) in the record c_k^r and replay c_k^a phases. In such a way that, under the optimal control law $u_k^* = -L^* c_k$, the evolution of the centers is given by

$$c_{k+1}^a = (A - G^*C - BL^*)c_k^a + G^* y_k^r, \tag{4.24a}$$

$$c_{k+1}^r = (A - G^*C - BL^*)c_k^r + G^* y_k^r, \tag{4.24b}$$

starting at $c_{k_1}^a = c_{k_1}$ and $c_{k_1}^r = c_{k_0}$.

Consequently, for $k \in \mathcal{K}_{REP}$, the attacked residuals are governed by the equation

$$r_k^a = r_k^r + C\tilde{A}^{k-k_1}(c_{k_1}^r - c_{k_1}^a), \tag{4.25}$$

where $\tilde{A} = A - G^*C - BL^*$.

The remainder of this chapter focuses on the case in which the attack is undetectable for an unprotected system like the one represented in Figure 4.1. Then, the following assumption will be considered.

Assumption 4.2. The attack is undetectable by the passive detector (4.22). That is, for all $k \in \mathcal{K}_{REP}$, the residual signal satisfies

$$r_k^r + C\tilde{A}^{k-k_1}(c_{k_1}^r - c_{k_1}^a) \in \mathcal{R}^H.$$

In this regard, it is straightforward to see that a necessary condition for the satisfaction of Assumption 4.2 is that the unstable modes of $\tilde{A}$ are unobservable with respect to the matrix C [Ferrari and Teixeira, 2017a].

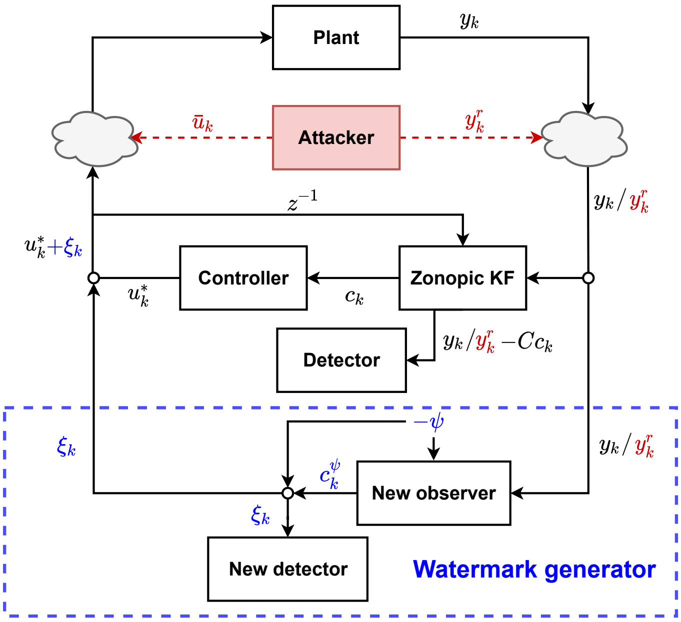


Figure 4.2: Watermark generator placement in the control loop.

Furthermore, note that, if $\tilde{A}$ is an asymptotically stable matrix, the detectability of the replay attack in case the control loop is compromised, is directly related with the Scenario I studied in Chapter 3, cf. Eq. (3.23), (with the exception that the analysis developed in Chapter 3 has been carried out under the assumption that the control objective is to regulate the tracking error with respect to a reference model). In other words, the attacker can generate pairs of input-output data consistent with the nominal operation either by directly counterfeiting the I/O data sent to the detector, or, in the case that $\tilde{A}$ is stable, by making that the controller operates based on false measurements generating thus inputs consistent with the false data.

Note that the consideration of Assumption 4.2 motivates the injection of an additive watermark signal ξ_k in the system inputs as presented in Figure 4.2. Therefore, the new input signal becomes $u_k = u_k^* + \xi_k$. In this regard, conversely to the stochastic approaches that inject an exogenous Gaussian signal [Mo and Sinopoli, 2009], in what follows, signal ξ_k will be assumed to be bounded within the known zonotope $\xi_k = \langle 0, H_\xi \rangle$. Accordingly, the first step is to analyse the impact on the system performance that entails the injection of the suboptimal signal $u_k = u_k^* + \xi_k$.

4.4.1 Performance loss induced by a zonotopically-bounded signal

For the case where the optimal control law derived in Section 4.3 is extended with the inclusion of the additive signal $\xi_k \in \langle 0, H_\xi \rangle$, the following proposition can be established

Proposition 4.4. The injection of the control law $u_k = -L_k^* c_k + \xi_k$, with $\xi_k \in \langle 0, H_\xi \rangle$, yields the same optimal cost-to-go than in Proposition 2 but for the recursion s_k which, for this case, satisfies

$$s_k = s_{k+1} + Tr[W\bar{P}_k] + Tr[S_{k+1}(G_k^*(C\bar{P}_kC^T + Q_v)G_k^{*T})] + Tr[(U + B^T S_{k+1} B)P_\xi],$$

starting at $s_N = Tr[W_f \bar{P}_N]$ and with $P_\xi = H_\xi H_\xi^T$.

Proof. The addition of the exogenous signal $\xi_k \in \langle 0, H_\xi \rangle$ implies that input signal satisfies $u_k = u_k^* + \xi_k \in \langle u_k^*, H_\xi \rangle$. Therefore, the proof is obtained by following the same steps as for Proposition 4.2 but taking into consideration that for this case

$$\mathcal{Q}[u_k^T U u_k] = \mathcal{Q}[(u_k^* + \xi_k)^T U (u_k^* + \xi_k)] = u_k^{*T} U u_k^* + Tr[UP_\xi],$$

with $P_\xi = H_\xi H_\xi^T$, and that the center of the state-estimator (4.3a) now satisfies

$$c_{k+1} \in \langle (Ac_k + Bu_k^*), [G_k^* C H_k,\ G_k^* F,\ BH_\xi] \rangle.$$

□

According to Proposition 4.4, since the solution of the infinite horizon control problem is given by the limit of the finite horizon solution, the new performance loss obtained with the inclusion of the bounded watermarking signal J_∞^{wm} can be assessed as

$$\begin{aligned} J_\infty^{wm} &= Tr[WP_\infty] + Tr[S_\infty(G^*(CP_\infty C^T + Q_v)G^{*T}] + Tr[(U + B^T S_\infty B)P_\xi] \\ &= J_\infty + Tr[(U + B^T S_\infty B)P_\xi] \\ &= J_\infty + \Delta J\,. \end{aligned} \tag{4.26}$$

It must be highlighted that the extra term in the cost function

$$\Delta J = Tr[(U + B^T S_\infty B)P_\xi] = \|H_\xi\|_{F,W_\xi}^2, \tag{4.27}$$

with $W_\xi = U + B^T S_\infty B$, matches the performance loss that induces the injection of a random Gaussian variable $\xi_k \sim \mathcal{N}(0, P_\xi)$ into an LQG control scheme [Mo and Sinopoli, 2009]. That is, there is clear a relationship in the effect that causes in the system performance to address the watermarking design by means of injecting an exogenous Gaussian signal or through its zonotopic counterpart.

Remark 4.2. It must be pointed out that the decision to model process disturbances and noises either as Gaussian random variables or following an unknown-but-bounded paradigm must be done according to the description that best suits the data retrieved from the real system. This claim is evidenced by the difficulty to find an unbiased metric to compare both approaches, since any quantitative comparison could only be attained by breaking the premises based on which is build each approach, i.e., either simulating a set-based approach using as disturbances random Gaussian signals (breaking the premise of bounded uncertainty); or simulating Gaussian detectors using bounded uncertainties with an arbitrary probability distribution function (breaking the premise of Gaussian uncertainty). On the other hand, note that both approaches sets some limits on implausible realizations of the disturbances. In this regard, set-based methods stablish explicitly this limit disregarding any specific probability distribution, whereas Gaussian stochastic detectors set a threshold on the implausibility of the data through the imposition of a specific alarm rate.

4.5 Zonotopically bounded watermark signal design

This section deals with the design of a watermark signal that can adopt any value within a given zonotope. Working under a set-based paradigm, this watermark signal must be designed in such a way that it guarantees the attack detection. To that end, the proposed scheme takes advantage of the fact that the observability of a signal that is known by the defender is lost during the replay phase of the attack. Hence, this fact is used in order to destabilize the estimation error of such known signal, for which, a zonotopic observer provides explicit bounds.

In order to implement the scheme described in the preceding paragraph, the closed-loop dynamics of the system under study (see Figure 4.1) are gathered in the model

$$\begin{aligned} z_{k+1} &= \mathcal{A} z_k + \mathcal{B}\xi_k + \mathcal{E}\alpha_k, \\ y_k &= \mathcal{C} z_k + F v_k, \end{aligned} \tag{4.28}$$

with $z_k = [x_k^T,\ e_k^T]^T$, $\alpha_k = [w_k^T,\ v_k^T]^T$, the new system matrices $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{E}$ conformed by

$$\mathcal{A} = \begin{bmatrix} A - BL^* & BL^* \\ 0 & A - G^*C \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}, \quad \mathcal{E} = \begin{bmatrix} E & 0 \\ E & -G^*F \end{bmatrix}, \quad \mathcal{C} \ = \begin{bmatrix} C & 0 \end{bmatrix},$$

and the initial conditions

$$z_0 = \begin{bmatrix} x_0 \\ e_0 \end{bmatrix} \in \langle c_0^z, H_0^z \rangle = \left\langle \begin{bmatrix} c_0 \\ 0 \end{bmatrix}, \begin{bmatrix} H_0 & 0 \\ 0 & H_0 \end{bmatrix} \right\rangle.$$

Hereafter, the watermark signal $\xi_k \in \mathbb{R}^{n_u}$ is defined as the difference

$$\xi_k = \psi - c_k^{\psi}, \tag{4.29}$$

where, for simplicity, $\psi \in \mathbb{R}^{n_u}$ is a nonzero constant vector set by the defender, i.e., $\psi \neq 0$, and $c_k^{\psi} \in \mathbb{R}^{n_u}$ is the center of a zonotopic observer that will be introduced later.

Let the constant vector ψ be rewritten as

$$\psi = M\psi - (M - I)\psi,$$

for any given matrix $M \in \mathbb{R}^{n_u \times n_u}$ which has been included as a tuning parameter. Thus, the injection of the exogenous signal (4.29) in the closed-loop system (4.28), can be written as

$$\begin{bmatrix} z_{k+1} \\ \psi \end{bmatrix} = \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ 0 & M \end{bmatrix} \begin{bmatrix} z_k \\ \psi \end{bmatrix} - \begin{bmatrix} \mathcal{B} c_k^{\psi} \\ (M - I)\psi \end{bmatrix} + \begin{bmatrix} \mathcal{E} \\ 0 \end{bmatrix} \alpha_k, \tag{4.30a}$$

$$y_k = \begin{bmatrix} \mathcal{C} & 0 \end{bmatrix} \begin{bmatrix} z_k \\ \psi \end{bmatrix} + F v_k, \tag{4.30b}$$

$$\begin{bmatrix} z_0 \\ \psi \end{bmatrix} \in \left\langle \begin{bmatrix} c_0^z \\ c_0^{\psi} \end{bmatrix}, \mathcal{H}_0 \right\rangle = \left\langle \begin{bmatrix} c_0^z \\ \psi \end{bmatrix}, \begin{bmatrix} H_0^z \\ 0 \end{bmatrix} \right\rangle \subset \mathbb{R}^{2n_x + n_u}. \tag{4.30c}$$

4.5.1 Design of an extended state estimator

In order to design a new observer for estimating the value of ψ, the time-varying matrix $\mathcal{G}_k \in \mathbb{R}^{(2n_x+n_u)\times n_y}$ is introduced in (4.30) yielding the equivalent system

$$\begin{bmatrix} z_{k+1} \\ \psi \end{bmatrix} = \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ 0 & M \end{bmatrix} \begin{bmatrix} z_k \\ \psi \end{bmatrix} - \begin{bmatrix} \mathcal{B} c_k^{\psi} \\ (M - I)\psi \end{bmatrix} + \begin{bmatrix} \mathcal{E} \\ 0 \end{bmatrix} \alpha_k + \mathcal{G}_k \left(y_k - \begin{bmatrix} \mathcal{C} & 0 \end{bmatrix} \begin{bmatrix} z_k \\ \psi \end{bmatrix} - F v_k \right). \tag{4.31}$$

Proposition 4.5. Given the system (4.30a)-(4.30b) with the initial conditions satisfying the inclusion (4.30c). By recursively defining the center and the generators matrix

$$\begin{bmatrix} c^z_{k+1} \\ c^\psi_{k+1} \end{bmatrix} = \left(\begin{bmatrix} \mathcal{A} & 0 \\ 0 & M \end{bmatrix} - \mathcal{G}_k \begin{bmatrix} \mathcal{C} & 0 \end{bmatrix} \right) \begin{bmatrix} c^z_k \\ c^\psi_k \end{bmatrix} - \begin{bmatrix} 0 \\ (M-I)\psi \end{bmatrix} + \mathcal{G}_k y_k, \tag{4.32a}$$

$$\mathcal{H}_{k+1} = \left[\left(\begin{bmatrix} \mathcal{A} & \mathcal{B} \\ 0 & M \end{bmatrix} - \mathcal{G}_k \begin{bmatrix} \mathcal{C} & 0 \end{bmatrix} \right) \mathcal{H}_k, \quad \begin{bmatrix} \mathcal{E} \\ 0 \end{bmatrix}, \quad -\mathcal{G}_k F \right], \tag{4.32b}$$

then the inclusion property

$$\begin{bmatrix} z_k \\ \psi \end{bmatrix} \in \left\langle \begin{bmatrix} c^z_k \\ c^\psi_k \end{bmatrix}, \mathcal{H}_k \right\rangle,$$

holds for all $k \geq 0$.

Proof. Assuming that the inclusion is satisfied at time k, which according to (4.30c) is true for $k = 0$, then using (4.31) it follows that

$$\begin{bmatrix} z_k \\ \psi \end{bmatrix} \in \left\langle \begin{bmatrix} c^z_k \\ c^\psi_k \end{bmatrix}, \mathcal{H}_k \right\rangle \implies \begin{bmatrix} z_{k+1} \\ \psi \end{bmatrix} \in \left\langle \begin{bmatrix} c^z_{k+1} \\ c^\psi_{k+1} \end{bmatrix}, \mathcal{H}_{k+1} \right\rangle,$$

where

$$\begin{aligned} \left\langle \begin{bmatrix} c^z_{k+1} \\ c^\psi_{k+1} \end{bmatrix}, \mathcal{H}_{k+1} \right\rangle = & \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ 0 & M \end{bmatrix} \left\langle \begin{bmatrix} c^z_k \\ c^\psi_k \end{bmatrix}, \mathcal{H}_k \right\rangle \oplus \begin{bmatrix} 0 & -\mathcal{B} \\ 0 & 0 \end{bmatrix} \left\langle \begin{bmatrix} c^z_k \\ c^\psi_k \end{bmatrix}, 0 \right\rangle \\ & \oplus \begin{bmatrix} 0 \\ (M-I)\psi \end{bmatrix} \oplus \begin{bmatrix} \mathcal{E} \\ 0 \end{bmatrix} \langle 0, I_{n_w+n_v} \rangle \oplus \mathcal{G}_k \langle y_k, 0 \rangle \\ & \oplus -\mathcal{G}_k \begin{bmatrix} \mathcal{C} & 0 \end{bmatrix} \left\langle \begin{bmatrix} c^z_k \\ c^\psi_k \end{bmatrix}, \mathcal{H}_k \right\rangle \oplus -\mathcal{G}_k F \langle 0, I_{n_v} \rangle. \end{aligned}$$

Therefore, using Property A.7 and Property A.8 in Appendix A, the expressions for the center (4.32a) and generators matrix (4.32b) are retrieved, and thus the inclusion property is satisfied at $k+1$. This gives the proof by induction. □

Accordingly, from (4.29) and Proposition 4.5 it follows that

$$\begin{bmatrix} z_k \\ \psi \end{bmatrix} \in \left\langle \begin{bmatrix} c^z_k \\ c^\psi_k \end{bmatrix}, \mathcal{H}_k \right\rangle \implies \begin{bmatrix} z_k - c^z_k \\ \xi_k \end{bmatrix} \in \langle 0, \mathcal{H}_k \rangle, \tag{4.33}$$

and as a result, if the projection matrix $N = [0, \ I_{n_u}] \in \mathbb{R}^{n_u \times (2n_x + n_u)}$ is defined, the watermark signal ξ_k satisfies

$$\xi_k \in N\langle 0, \mathcal{H}_k \rangle = \langle 0, N\mathcal{H}_k \rangle. \tag{4.34}$$

Additionally, with regard to the design of a matrix $\mathcal{G}_k$ that minimizes the performance loss induced by the signal (4.34) (cf. Section 4.4.1), the following proposition is introduced.

Proposition 4.6. Given $W_\xi = W_\xi^T \succ 0$ and the generators matrix $\mathcal{H}_{k+1}$ as in (4.32b). Then, the additional term $\Delta J = \|N\mathcal{H}_{k+1}\|^2_{F,W_\xi}$ induced by the injection of signal ξ_k is minimized for the observer matrix $\mathcal{G}^*_k = \mathcal{A}_2 \mathcal{P}_k \mathcal{C}_2^T (\mathcal{C}_2 \mathcal{P}_k \mathcal{C}_2^T + Q_v)^{-1}$, with N the projection matrix in (4.34) and

$$\mathcal{A}_2 = \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ 0 & M \end{bmatrix}, \quad \mathcal{C}_2 = \begin{bmatrix} \mathcal{C} & 0 \end{bmatrix}, \quad Q_v = FF^T, \quad \mathcal{P}_k = \mathcal{H}_k \mathcal{H}_k^T.$$

Proof. From Definition A.23 it follows that

$$\|\langle N\mathcal{H}_{k+1}\rangle\|_{F,W}^2 = Tr[\mathcal{H}_{k+1}^T N^T W_\xi N\mathcal{H}_{k+1}] = Tr[\mathcal{H}_{k+1}^T \mathcal{W}\mathcal{H}_{k+1}],$$

with matrix $\mathcal{W}$ defined as

$$\mathcal{W} = N^T W_\xi N = \begin{bmatrix} 0 & 0 \\ 0 & W_\xi \end{bmatrix}.$$

Hence, since $Tr[\mathcal{H}_{k+1}^T \mathcal{W}\mathcal{H}_{k+1}]$ is convex with respect to $\mathcal{G}_k$, following the same steps than in the proof of Theorem 5 in Combastel [2015], it follows that the matrices $\mathcal{G}_k^* = \arg\min_{\mathcal{G}} Tr[\mathcal{H}_{k+1}^T \mathcal{W}\mathcal{H}_{k+1}]$ must satisfy the equation

$$\mathcal{W}\mathcal{G}_k^* = \mathcal{W}\mathcal{A}_2\mathcal{P}_k\mathcal{C}_2^T(\mathcal{C}_2\mathcal{P}_k\mathcal{C}_2^T + Q_v)^{-1}.$$

Considering this condition, and taking into consideration that $\mathcal{W}$ symmetric positive semi-definite, it follows that $\mathcal{G}_k^* = \mathcal{A}_2\mathcal{P}_k\mathcal{C}_2^T(\mathcal{C}_2\mathcal{P}_k\mathcal{C}_2^T + Q_v)^{-1}$ is one solution that satisfies the previous equation. □

4.5.2 Watermarked system stability

Analogously to Section 4.3.2, in the sequel the fixed steady-state observer gain $\mathcal{G}^*$ is considered. Next, the stability of the closed-loop system (4.28) subject to the injection of the watermark signal ξ_k is analysed. To that end, through the comparison of (4.30a) with (4.32a), it can be seen that the estimation error produced by the new observer is governed by the dynamics

$$\tilde{e}_{k+1} = (\mathcal{A}_2 - \mathcal{G}^*\mathcal{C}_2)\tilde{e}_k + \mathcal{E}_2\alpha_k - \mathcal{G}^*Fv_k, \tag{4.35}$$

with

$$\tilde{e}_k = \begin{bmatrix} z_k - c_k^z \\ \psi - c_k^\psi \end{bmatrix}, \quad \mathcal{A}_2 = \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ 0 & M \end{bmatrix}, \quad \mathcal{C}_2 = \begin{bmatrix} \mathcal{C} & 0 \end{bmatrix}, \quad \mathcal{E}_2 = \begin{bmatrix} \mathcal{E} \\ 0 \end{bmatrix}. \tag{4.36}$$

Consequently, if the evolution of the closed-loop system (4.28) and the new estimation error (4.35) are combined in a single dynamical system, this leads to

$$\begin{bmatrix} z_{k+1} \\ \tilde{e}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathcal{A} & \mathcal{B}N \\ 0 & \mathcal{A}_2 - \mathcal{G}^*\mathcal{C}_2 \end{bmatrix} \begin{bmatrix} z_k \\ \tilde{e}_k \end{bmatrix} + \begin{bmatrix} \mathcal{E} \\ \mathcal{E}_2 \end{bmatrix}\alpha_k + \begin{bmatrix} 0 \\ -\mathcal{G}^*F \end{bmatrix} v_k, \tag{4.37}$$

and thus its state matrix is an upper triangular matrix. Therefore, the stability of the watermarked system (4.37) is related with the eigenvalues of $\mathcal{A}$, which is associated to an asymptotically stable system, and the eigenvalues of $\mathcal{A}_2 - \mathcal{G}^*\mathcal{C}_2$.

It must be highlighted that the capability to design an observer gain such that $\mathcal{A}_2 - \mathcal{G}^*\mathcal{C}_2$ is asymptotically stable depends on the detectability of the pair $(\mathcal{A}_2, \mathcal{C}_2)$. In this regard, the following mild sufficient condition can be stated.

Proposition 4.7. Given the pair (A, C) detectable, a sufficient condition for the detectability of the pair $(\mathcal{A}_2, \mathcal{C}_2)$, is that there is a column of matrix $\mathcal{B}$, denoted as $\mathcal{B}_i$, that satisfies

$$\mathcal{C}[\mathcal{B}_i,\ \mathcal{A}\mathcal{B}_i,\ \ldots,\ \mathcal{A}^{n-2}\mathcal{B}_i] \neq 0,$$

where n denotes the dimension of the square matrix $\mathcal{A}$.

Proof. Let $\mathcal{B}_i$ denote the i-th column of matrix $\mathcal{B}$, and assume that signal ξ_k is only introduced through $\mathcal{B}_i$. Then, from the fact that $\mathcal{A}$ is an asymptotically stable matrix, if follows that in order to guarantee the detectability of the pair $(\mathcal{A}_2, \mathcal{C}_2)$, with

$$\mathcal{A}_2 = \begin{bmatrix} \mathcal{A} & \mathcal{B}_i \\ 0 & M \end{bmatrix}, \quad \mathcal{C}_2 = \begin{bmatrix} \mathcal{C} & 0 \end{bmatrix},$$

and $M \in \mathbb{R}$ is a scalar, only the observability of the extended state must be ensured.

Accordingly, writing the observability matrix $\mathcal{O}_2(\mathcal{A}_2, \mathcal{C}_2)$ as

$$\mathcal{O}_2 = \begin{bmatrix} \mathcal{O}(\mathcal{A}, \mathcal{C}) & O_{12} \\ \mathcal{C}\mathcal{A}^n & O_{22} \end{bmatrix} \in \mathbb{R}^{n_y(n+1)\times(n+1)},$$

with

$$O_{22} = \mathcal{C}\mathcal{A}^{n-1}\mathcal{B}_i + \mathcal{C}\sum_{j=1}^{n-1} \mathcal{A}^{n-1-j}\mathcal{B}_i M^j \in \mathbb{R}^{n_y},$$

it follows that if $rank\big(\mathcal{C}[\mathcal{B}_i \ \mathcal{A}\mathcal{B}_i, \ ..., \ \mathcal{A}^{n-2}\mathcal{B}_i]\big) \geq 1$ is satisfied, then the term O_{22} can be modified by means of the parameters M, allowing to impose the independence of the last column of $\mathcal{O}_2$ and thus the observability of the extended state. □

Note that Proposition 4.7 provides a sufficient condition concerning the applicability of the proposed watermarking method.

4.5.3 Attack detection

Here the detectability of the replay attack under the proposed watermark signal is analysed. To that end, let Ψ denote the mRPI set for the estimation error generated by the extended observer (4.35), and let $\langle 0, \tilde{\mathcal{H}} \rangle \supset \Psi$ denote an RPI zonotopic outer-approximation to Ψ. Furthermore, note that, if the matrix $\mathcal{G}^*$ is designed according to Proposition 4.6, then from Proposition 4.3 it follows that $\|\Psi\|_F^2 = Tr[P_\infty^\Psi]$, with P_∞^Ψ the solution of the corresponding DARE.

Besides, taking advantage of the fact that the constant vector ψ is known by the defender, the watermark signal ξ_k can also be used in order to assess the system operation. Hence, the following can be stated for the steady-state

$$\begin{cases} \xi_k \in \langle 0, N\tilde{\mathcal{H}} \rangle & \Longrightarrow \text{Healthy operation,} \\ \text{otherwise} & \Longrightarrow \text{Something is wrong.} \end{cases} \tag{4.38}$$

The following shows that the signal ξ_k is sensitive to replay attacks. Using the notation presented in Section 2.3, during the replay phase of the attack the watermark signal ξ_k^a can be written as

$$\xi_k^a = \psi - c_k^{\psi,a} = \psi - c_k^{\psi,r} + (c_k^{\psi,r} - c_k^{\psi,a}) = \xi_k^r + (c_k^{\psi,r} - c_k^{\psi,a}), \tag{4.39}$$

where $c_k^{\psi,r}$ and $c_k^{\psi,a}$ are the center of the estimation of ψ during the record and replay phases, respectively.

Moreover, from the steady-state operation of the detector introduced in Assumption 3.4, the signal ξ_k^r satisfies

$$\xi_k^r = \psi - c_k^{\psi,r} \in \langle 0, N\tilde{\mathcal{H}} \rangle, \quad \forall k \in \mathcal{K}_{REP}. \tag{4.40}$$

Therefore, taking into consideration (4.40), the evolution of ξ_k^a depends on the dynamics of the difference $c_k^{\psi,r} - c_k^{\psi,a}$. In this regard, the evolution of the new estimator during the record and replay phases is governed by the equations

$$\begin{bmatrix} c_{k+1}^{z,r} \\ c_{k+1}^{\psi,r} \end{bmatrix} = \left(\begin{bmatrix} \mathcal{A} & 0 \\ 0 & M \end{bmatrix} - \mathcal{G}^* \begin{bmatrix} \mathcal{C} & 0 \end{bmatrix} \right) \begin{bmatrix} c_k^{z,r} \\ c_k^{\psi,r} \end{bmatrix} - \begin{bmatrix} 0 \\ (M-I)\psi \end{bmatrix} + \mathcal{G}^* y_k^r, \tag{4.41a}$$

$$\begin{bmatrix} c_{k+1}^{z,a} \\ c_{k+1}^{\psi,a} \end{bmatrix} = \left(\begin{bmatrix} \mathcal{A} & 0 \\ 0 & M \end{bmatrix} - \mathcal{G}^* \begin{bmatrix} \mathcal{C} & 0 \end{bmatrix} \right) \begin{bmatrix} c_k^{z,a} \\ c_k^{\psi,a} \end{bmatrix} - \begin{bmatrix} 0 \\ (M-I)\psi \end{bmatrix} + \mathcal{G}^* y_k^r, \tag{4.41b}$$

starting at

$$\begin{bmatrix} c_{k_1}^{z,r} \\ c_{k_1}^{\psi,r} \end{bmatrix} = \begin{bmatrix} c_{k_0}^{z} \\ c_{k_0}^{\psi} \end{bmatrix}, \quad \begin{bmatrix} c_{k_1}^{z,a} \\ c_{k_1}^{\psi,a} \end{bmatrix} = \begin{bmatrix} c_{k_1}^{z} \\ c_{k_1}^{\psi} \end{bmatrix}.$$

Hence, splitting the matrix $\mathcal{G}^*$ into the blocks $\mathcal{G}^* = \begin{bmatrix} \mathcal{G}_1 \\ \mathcal{G}_2 \end{bmatrix}$, then from the comparison of (4.41a) and (4.41b) it follows that the center difference between phases evolves according to

$$\begin{bmatrix} c_{k+1}^{z,r} - c_{k+1}^{z,a} \\ c_{k+1}^{\psi,r} - c_{k+1}^{\psi,a} \end{bmatrix} = \begin{bmatrix} \mathcal{A} - \mathcal{G}_1\mathcal{C} & 0 \\ -\mathcal{G}_2\mathcal{C} & M \end{bmatrix} \begin{bmatrix} c_k^{z,r} - c_k^{z,a} \\ c_k^{\psi,r} - c_k^{\psi,a} \end{bmatrix}. \tag{4.42}$$

Consequently, since the resulting state matrix is a lower-triangular matrix, it follows that the dynamics of (4.42) are directly related with the eigenvalues of the tuning matrix M. That is, if one of the eigenvalues of matrix M is set outside the unit disk, it is guaranteed that the signal ξ_k^a will ultimately come out of the associated healthy set $\langle 0, N\tilde{\mathcal{H}} \rangle$, which guarantees the detection of the attack.

Moreover, if an analysis similar to the one carried out in Section 4.4 is performed, it can be seen that with injection of the watermark signal ξ_k the evolution of the residual signal is given by

$$r_k^a = r_k^r + C\tilde{A}^{k-k_1}(c_{k_1}^r - c_{k_1}^a) + C\sum_{i=1}^{k-k_1} \tilde{A}^{i-1}B(\xi_k^r - \xi_k^a). \tag{4.43}$$

Since the last term of (4.43) can be rewritten as

$$C\sum_{i=1}^{k-k_1} \tilde{A}^{i-1}B(\xi_k^r - \xi_k^a) = -C\sum_{i=1}^{k-k_1} \tilde{A}^{i-1}B(c_k^{\psi,r} - c_k^{\psi,a}),$$

it follows that, if the dynamics of $c_k^{\psi,r} - c_k^{\psi,a}$ in (4.42) are destabilized by means of matrix M, then the residual signal r_k^a will also exit the healthy residual set $\mathcal{R}^H$, triggering thus the alarm.

Remark 4.3. In the moment that the attack is detected, the signal ξ_k^a should stop being injected so as not to affect the system with its exponential growth.

Remark 4.4. From the structure of $\mathcal{A}_2$ (with $M > 1$) it follows that the pair $(\mathcal{A}_2, Q_2^{\frac{1}{2}})$, with $Q_2 = \mathcal{E}_2\mathcal{E}_2^T$, does not have unobservable eigenvalues on the unit disks, and thus the corresponding DARE has a unique, maximal, positive definite solution [Bitmead and Gevers, 1991, Theorem 10.3]. However, in order to impose sufficient conditions for the stability of the closed-loop form in Proposition 4.6, then $(\mathcal{A}_2, Q_2^{\frac{1}{2}})$ must be stabilizable. Note that this can be easily attained by extending the uncertain matrix in (4.30a) with the introduction of an epsilon term in the positions of the zero-valued elements.

4.6 Case study

Below, the four-tank process described in Section B.1 of Appendix B is used in order to validate the results presented in this chapter. In this regard, it is considered that the discrete-time LTI model under study presents the distribution matrices

$$E = \begin{bmatrix} 0.05 & 0 & 0 & 0 \\ 0.05 & 0.01 & 0 & 0 \\ 0.05 & 0 & 0.02 & 0 \\ 0.05 & 0 & 0 & 0.02 \end{bmatrix}, \qquad F = \begin{bmatrix} 0.03 & 0 \\ 0 & 0.01 \end{bmatrix}.$$

In addition, during all the simulations that have been carried out, it has been considered that system disturbances w_k and measurement noise v_k are random Gaussian white noise bounded within the zonotopes $w_k \in \langle 0, I_4 \rangle$ and $v_k \in \langle 0, I_2 \rangle$, for all $k \in \mathbb{N}$.

The process is controlled by means of an optimal LQR controller designed for the weighting matrices $W = 10 \cdot I_4$ and $Q = I_2$. The fixed steady-state gain of the controller is

$$L^* = \begin{bmatrix} 2.2171 & 0.0004 & 0.1650 & 0.0744 \\ -0.0016 & 2.4915 & 1.0060 & 0.4850 \end{bmatrix}.$$

Besides, by taking into consideration the covariation matrices $Q_w = EE^T$ and $Q_v = FF^T$, the obtained steady-state gain of the ZKF is

$$G^* = \begin{bmatrix} 0.5106 & 1.2686 \\ 0.1399 & 1.6852 \\ 0.5472 & 1.1097 \\ 0.1060 & 1.6209 \end{bmatrix}.$$

At this point, it must be highlighted that the eigenvalues of matrix $\tilde{A} = A - G^*C - BL^*$ are $\lambda(\tilde{A}) = \{-0.1739, 0.4005, 0.9513 \pm 0.0278i\}$, that is, $\tilde{A}$ is asymptotically stable, and thus there are no guarantees that the replay attack can be detected (cf. (4.25)). In other words, in the remainder of this section, Assumption 4.2 concerning the attack undetectability prevails motivating the injection of a watermark signal in the control loop.

4.6.1 Performance assessment

Concerning the estimation error $e_k = x_k - c_k$ generated by the observer, from Proposition 4.3 it follows that the size of its mRPI set Φ can be computed directly by solving the corresponding DARE. By doing so, the obtained F-radius is $\|\Phi\|_F = \sqrt{Tr[P_\infty]} = 0.1393$, with P_∞ the unique positive definite solution of the DARE. Consequently, if the mRPI set Φ could be used in order to bound the estimation error in the steady-state, applying (4.19) it follows that the optimal infinite horizon cost function that could be obtained would be $J_\infty = 0.9275$.

Due to the difficulties of computing an exact representation of the Φ, different RPI outer-approximation of such mRPI set will be computed below. In this regard, an initial low order RPI zonotopic set $\tilde{Z}_0 = \langle 0, \tilde{H}_0 \rangle$ is computed as presented in Section A.3.1 of Appendix A. The

Iteration	Zonotope order	Est. error F-radius	Perf. loss $\tilde{J}_\infty$
$\tilde{Z}_0$	1	1.2544	19.5451
$\tilde{Z}_5$	7	0.8034	7.7927
$\tilde{Z}_{10}$	14.5	0.4994	3.3956
$\tilde{Z}_{20}$	29.5	0.2368	1.3144
$\tilde{Z}_{30}$	44.5	0.1610	0.9958
$\tilde{Z}_{50}$	74.5	0.1405	0.9307
$\tilde{Z}_{75}$	112	0.1394	0.9276
$\tilde{Z}_{100}$	149.5	0.1393	0.9275
Φ	∞	0.1393	0.9275

Table 4.1: Zonotopic RPI outer approximations of the mRPI set Φ.

obtained set is a parallelotope, that is, a first order zonotope with the following generators matrix

$$\tilde{H}_0 = \begin{bmatrix} 0.0894 & 0.0256 & -0.3773 & 0.3080 \\ 0.0928 & -0.0238 & 0.0413 & -0.0335 \\ 0.0898 & -0.2360 & -0.9798 & -0.0529 \\ 0.0935 & -0.5073 & 0.1513 & -0.0000 \end{bmatrix}.$$

The initial RPI set $\tilde{Z}_0$ is now forward propagated, cf. Section A.3.2 of Appendix A, obtaining successively tighter RPI over-approximations of the mRPI set Φ. In this regard, Table 4.1 shows the zonotope order and the F-radius of the over-approximations obtained for different iteration values. Besides, the value of the cost function for the infinite horizon problem is also shown.

After analysing the data reported in Table 4.1, it can be seen how the F-radius of the successive over-approximations converges to the previously computed $\|\Phi\|_F = 0.1393$ value (shown in red). However, this is achieved at the cost of increasing the complexity of the zonotope as reflected in the second column. Note that, the bad performance index obtained with the initial set motivates its further propagation in search of sets whose F-radius, and hence the imposed performance loss, is closer to the optimal value.

Furthermore, it must be remarked that the computation of the sets shown in Table 4.1 is done offline. Then, the only online computation is the evaluation of the condition $r_k \in \mathcal{R}^H$, with $\mathcal{R}^H = C\mathcal{Z} \oplus F\mathcal{V} = \langle 0, [C\tilde{H},\ F] \rangle$. Testing whether or not a point belongs to a zonotopic set can be efficiently done by solving the constraint satisfaction problem expressed in Property A.10 of Appendix A.

4.6.2 Watermark signal design

In order to design a watermark signal like the one presented in Section 4.5, the tuning matrix M must be defined. Remember that to guarantee the effectiveness of the scheme, the matrix M must present at least one eigenvalue outside the unit disk. In consequence, in the simulations performed below M has been set to

$$M = \begin{bmatrix} 1.05 & 0 \\ 0 & 1.05 \end{bmatrix}.$$

Iteration	**Zonotope**	**New est. error**	**Perf. loss**
	order	**F-radius**	**$\Delta\tilde{J}$**
$\tilde{\mathcal{Z}}_0$	1	10.9688	127.7466
$\tilde{\mathcal{Z}}_5$	4.2	9.0130	86.0918
$\tilde{\mathcal{Z}}_{10}$	8.2	7.1512	54.0752
$\tilde{\mathcal{Z}}_{20}$	16.2	4.5690	22.0604
$\tilde{\mathcal{Z}}_{30}$	24.2	2.9530	9.2396
$\tilde{\mathcal{Z}}_{50}$	40.2	1.2805	1.7299
$\tilde{\mathcal{Z}}_{75}$	60.2	0.6131	0.3662
$\tilde{\mathcal{Z}}_{100}$	80.2	0.5059	0.2361
Ψ	∞	0.4929	0.2223

Table 4.2: Watermark signal design.

Then, following the development presented in Section 4.5, the optimal steady-state value of the gain $\mathcal{G}^*$ of the new observer is

$$\mathcal{G}^* = \begin{bmatrix} 0.6203 & 0.1388 & 0.1039 & 0.0202 & 0.1204 & 0.0037 & 0.1951 & -0.014 & 0.8998 & -0.130 \\ 1.2684 & 1.8383 & 1.8712 & -0.238 & 0.0427 & 0.1566 & -0.030 & 0.1739 & 0.2774 & 4.1170 \end{bmatrix}^T .$$

Concerning the new estimation error, from Proposition 4.3 it follows that the F-radius of its mRPI set Ψ can be computed by solving the corresponding DARE. In this case, the obtained value is $\|\Psi\|_F = \sqrt{Tr[\mathcal{P}_\infty]} = 0.4929$, with $\mathcal{P}_\infty$ the unique positive definite solution of the DARE. At this point, the projection matrix $N = [0_{2\times 8},\ I_2]$ is defined since the watermark signal only requires the terms of the estimation error that relate with the new states defined by the constant ψ. Hence, according to the developments presented in Section 4.4.1, the performance loss induced by a watermark signal confined within a zonotope with covariance $P_\xi = N\mathcal{P}_\infty N^T$ is given by

$$\Delta J = Tr\big[(U + B^T S_\infty B)P_\xi\big] = 0.2223.$$

Next, similar to Section 4.6.1 several RPI outer-approximations to Ψ are computed starting from an initial first order zonotope $\tilde{\mathcal{Z}}_0 = \langle 0, \tilde{\mathcal{H}}_0\rangle$. In this regard, Table 4.2 presents the zonotope order and the F-radius of the over-approximations obtained for different iteration values. Note that given the set RPI set $\tilde{\mathcal{Z}}_i = \langle 0, \tilde{\mathcal{H}}_i\rangle$, the healthy watermark signal ξ_k is guaranteed to converge within $\langle 0, N\tilde{\mathcal{H}}_i\rangle$ and once it is inside, it will only exit that set in case the system in under attack. Consequently, the fourth column of Table 4.2 shows for each iteration the performance loss ΔJ induced by a watermark signal constrained to $\xi_k \in \langle 0, N\tilde{\mathcal{H}}_i\rangle$ and thus with the covariation matrix $P_\xi = N\mathcal{H}_i\mathcal{H}_i^T N^T$.

4.6.3 Attack simulation

A replay attack has been simulated with the following time windows

$$\mathcal{K}_{REC} = [100,\ 300],$$
$$\mathcal{K}_{REC} = [700,\ 900].$$

Furthermore, the selected RPI over-approximation for the ZKF is $\tilde{Z}_{20}$ while for the state estimator used in the watermarking generation the selected RPI over-approximation is $\tilde{\mathcal{Z}}_{50}$. In this regard, the temporal evolution of the watermark signal during the attack scenario is represented in Figure 4.3. In this figure, it can be seen how after the start of the replay phase at $k = 700$, the watermark signal starts growing exponentially until the time instant $k = 737$, when the condition $\xi_k^a \in N\tilde{\mathcal{Z}}_{50}$ is no longer satisfied and thus the attack is detected. Note that the watermark signal stops being injected after the attack is detected.

On the other hand, Figure 4.4 shows the effect caused on the system outputs by the injection of the watermark signal. It can be seen how the performance of the system is hardly degraded before the replay phase and how after the start of this phase, the effect of the increasing watermark signal becomes more evident. It should be highlighted that since the control loop is affected by the data replay, the stability of the plant after the attack detection, at $k = 737$, is due solely to the fact that the open-loop system is stable, i.e., the state matrix A is asymptotically stable.

Moreover, Figure 4.5 also shows how the injection of the watermark signal causes the

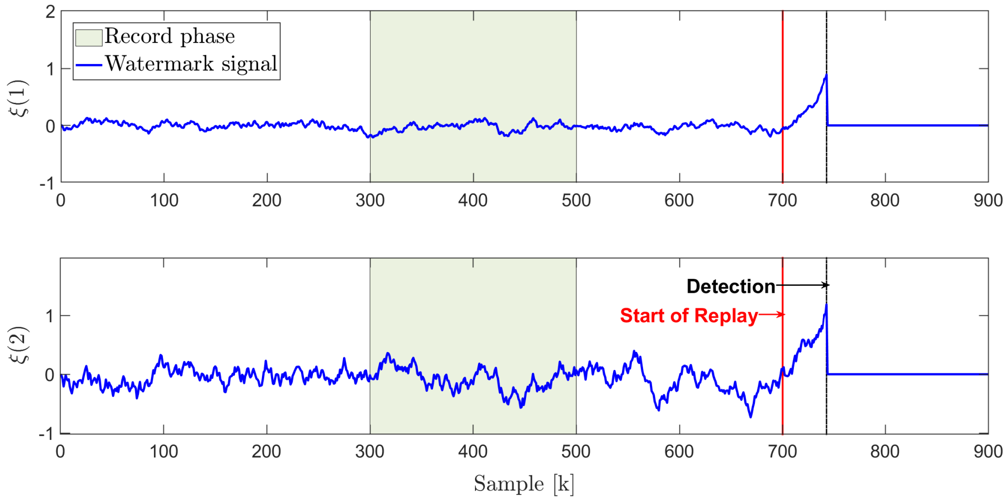


Figure 4.3: Temporal evolution of the watermark signal.

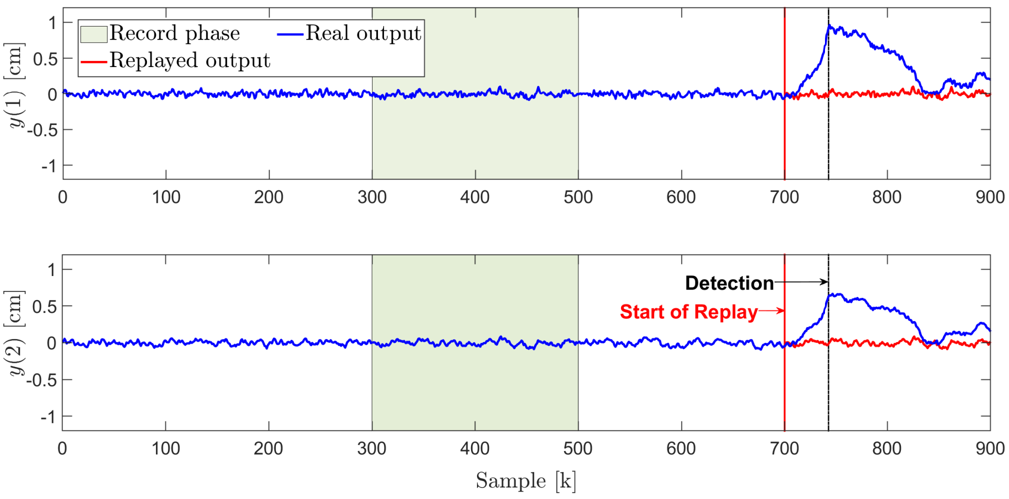


Figure 4.4: Temporal evolution of the system outputs with watermark.

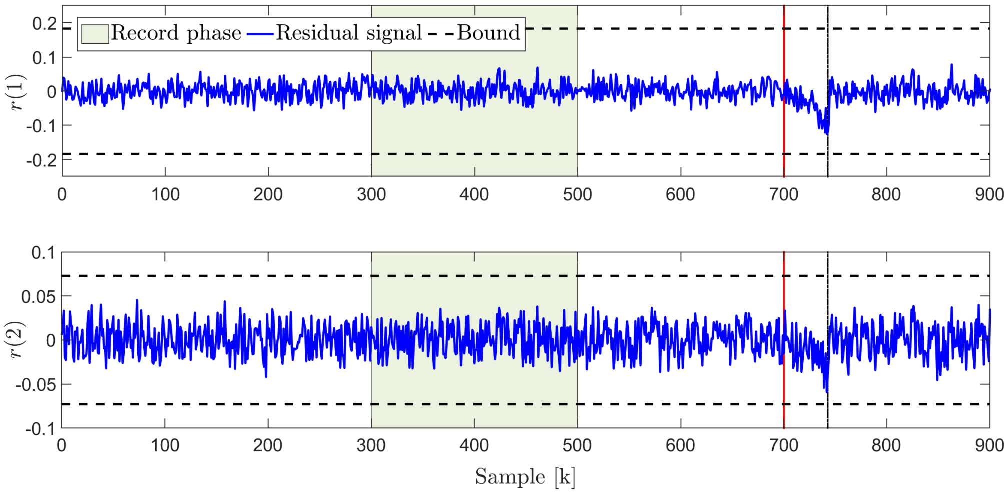


Figure 4.5: Temporal evolution of the residual signal with watermark.

divergence of the residual signal (cf. (4.43)), in such a way that it is guaranteed that at some time instant $r_k \in \mathcal{R}^H = \langle 0, [C\tilde{H}_{20},\ F] \rangle$ will no longer be satisfied. Therefore, this test can also be used to detect the replay attack. The bounds shown in this figure have been obtained through the computation of the interval hull (cf. Definition A.20) of the set $\mathcal{R}^H$. In addition, in order to compare the effect of the watermark signal, Figures 4.6 and 4.7 show the evolution of the system outputs and of the residual signal for the same attack scenario but without the injection of the watermark signal ξ_k.

The mean times required in the simulations to monitor the operation of the system online are[1]:

– Evaluation of $\xi_k \in N\tilde{\mathcal{Z}}_{50} = \langle 0, N\tilde{\mathcal{H}}_{50} \rangle$, mean time $t_m = 8.9$ms.

– Evaluation of $r_k \in \mathcal{R}^H = \langle 0, [C\tilde{H}_{20},\ F] \rangle$, mean time $t_m = 6.3$ms.

and thus well below the $T_s = 1$s sampling time of the discrete-time LTI model of the four-tank process.

Finally, Figure 4.8 compares the optimal performance loss ΔJ_{opt}, the performance loss ΔJ induced by the RPI approximation $\tilde{\mathcal{H}}_{50}$ and the mean detection time for two different matrix M parametrizations

$$M_1 = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \text{(Fig. 4.8a)}, \qquad M_2 = \begin{bmatrix} m & 0 \\ 0 & 0.5 \end{bmatrix} \text{(Fig. 4.8b)},$$

and the values of m varying in the interval $m \in [1.05,\ 1.5]$. In this regard, the mean detection time was obtained running 50 simulations for each value of the parameter m. The simulations show the existing trade-off between performance-loss and detection speed.

[1] Laptop (Intel i7 1.8 GHz, 16 GB RAM) running Windows 10; optimization using Cplex 12.8 ILOG [2018]

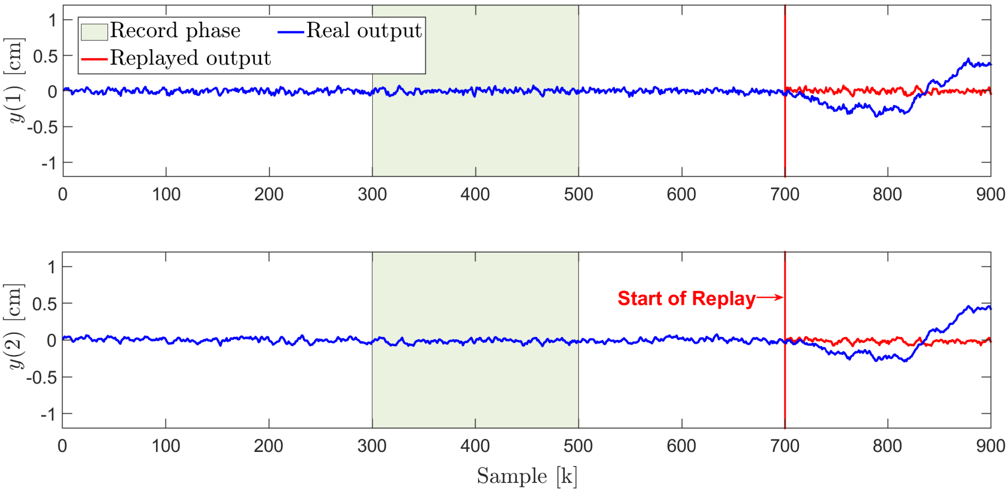


Figure 4.6: Temporal evolution of the system outputs without watermark.

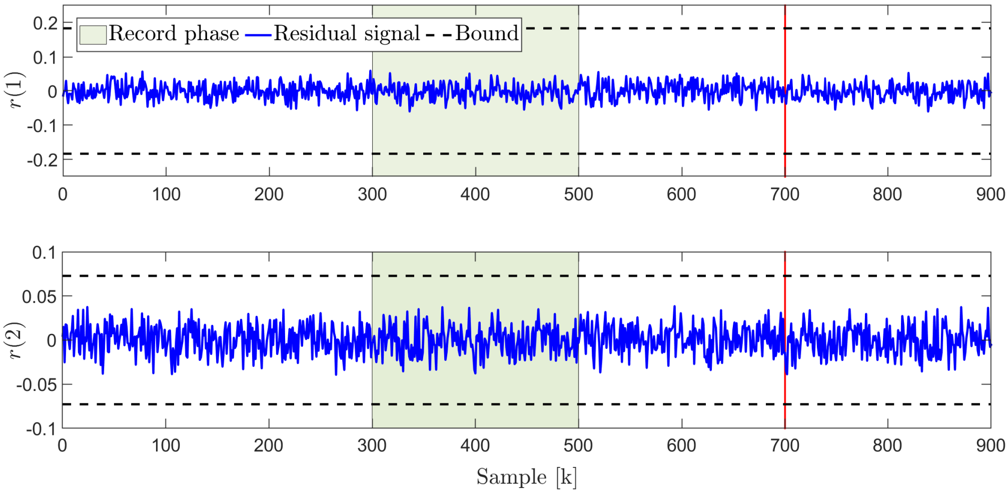
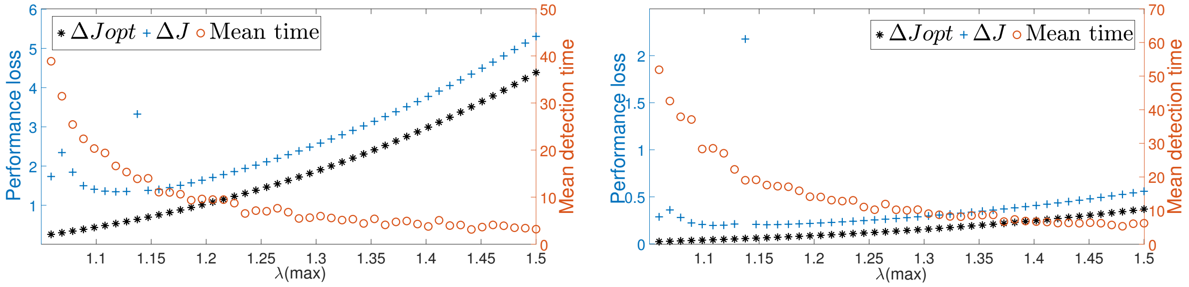


Figure 4.7: Temporal evolution of the residual signal without watermark.



(a) Matrix M_1

(b) Matrix M_2

Figure 4.8: Comparison of performance loss vs mean detection time.

4.7 Summary

This chapter has delved into the analogy between stochastic and deterministic approaches applied to replay attacks. By means of this analogy, despite working in a set-based framework, the system performance can be assessed and its steady-state operation may be related with the size of the mPRI set of the estimation error. Furthermore, similar expressions regarding the impact of a Gaussian watermark signal on the performance of an LQG system and the impact of a zonotopically bounded signal on the weighted Frobenius norm of the states, were obtained. These analogies motivate the approach of security-related problems also from a set-based perspective, as well as impel further studies in the relationship between both fields. Moreover, some unresolved issues like the effect of the reduction operator in the convergence of the Ricatti difference equations, appear as appealing problems that clearly deserve in-depth study.

Another aspect of vital importance is the scalability of the proposed techniques to large-scale systems. This is certainly an open issue in set-based techniques due to the fact that the number of vertices of a bounding set scales exponentially with the number of dimensions. Nevertheless, this pathological effect can be somewhat palliated by means of the reduction operator (which in an extreme case can generate an interval hull approximation, imposing thus the independence of the components). When it comes to the application of the proposed techniques to large-scale real-world systems, investigate the joint design of watermarking techniques with decoupling strategies that split the system into smaller-scale systems is presented as an appealing approach. A smart combination of both approaches may allow to secure the operation of the overall system by designing appropriate watermark signals for each decoupled subsystem.

Part II

Reconfiguration with back-up components

Chapter 5

System reconfiguration with back-up components

This part of the thesis is concerned with the problem of, given a system with back-up components, select the (in some sense) optimal actuator configuration after a fault occurrence. In addition, the consideration of constrained large-scale systems has been established as an initial premise in the study of the problem. This premise holds in the different chapters that make up this part of the dissertation. Consequently, this chapter is intended to motivate, introduce and formulate the problem, establishing thus the basis on which the following chapters will be developed.

The remainder of this chapter is structured as follows: Section 5.1 presents a general introduction to the topic. Section 5.2 discusses the role of the reconfiguration with back-up components problem within the FTC field and presents a literature review on it. Finally, in Section 5.3, some key concepts for the development of the following chapters are introduced, and the problem under consideration is formulated.

5.1 Introduction

The design of systems capable to handling unexpected events is a prevailing need that has motivated the development of the currently mature fields of fault diagnosis (FD) and fault tolerant control (FTC) [Hwang et al., 2009]. To this end, the evolution of complex systems tends to be continuously monitored aiming at detecting inconsistencies with respect to the nominal operation of the system. In this context, fault diagnosis methods can be categorized as either *passive* or *active*, depending on whether the diagnosis is attempted by comparing the input-output data of the system to model or historical data [Gertler, 1998, Isermann, 2005], or by injecting a signal into the system to improve the fault detectability [Niemann, 2006, Scott et al., 2014].

Whenever a reasonable diagnosis of the fault is achieved, *active* FTC schemes are intended to reconfigure the control loop by adapting the controller to the new faulty model [Blanke et al., 2006]. In this regard, the majority of the research work on active FTC has concentrated on achieving fault accommodation by exploiting the analytical redundancy of the system variables for a fixed sensors-actuators structure, whereas the use of the physical redundancy has been mainly relegated to consider the duplication of hardware components, and thus reducing the FTC problem to the switch of the faulty component by a healthy one [Lunze and Richter, 2008].

Nevertheless, the trend towards highly connected systems has led up to networked systems which may present alternative connections between components that differ from the nominal ones. The existence of these alternatives, together with the presence of back-up elements, provides certain physical redundancy that extends the FTC capabilities of the system, allowing thus to replace the effect of a faulty element by modifying the system configuration. This situation appears, for example, in water distribution networks (WDNs), where, apart from the nominal set of valves and pumps that control the water tanks levels, there also exist a set of alternative valves/pumps (normally off) which allow to redistribute the water through secondary pipes in case of emergency. Hence, by selecting the appropriate alternative hardware components, the reconfigured system may present an admissible performance in a situation where the performance of the nominal configuration becomes inadmissible.

Accordingly, the consideration of these highly connected systems modifies the traditional system reconfiguration problem. In this regard, after a fault occurrence, system reconfiguration is normally understood as seeking for an admissible control law using the remaining healthy components. However, in large-scale systems with back-up elements, the reconfiguration problem is extended to find the (in some sense) new optimal configuration that yields an admissible operation. Nonetheless, this extension introduces several points that deserve a detailed study, namely: to assess the admissibility of a configuration in a constrained networked system with algebraic equations that describe possible static relations in the network; make an efficient configuration selection among the large number of possibilities that arise in large-scale systems; provide stability guarantees in the selection of a new configuration, etc. Therefore, this part of the thesis is intended to analyse all the problems stated above (and some others that will appear), and propose an engineering solution that allows to address the reconfiguration with back-up elements in an efficient manner.

At this point, it should be noted that the developments presented hereafter consider that the anomaly that gives rise to considering the reconfiguration of the system is a components fault. This particularization of the anomalies as faults, has been made in order to be able to frame the studied problem within the fault tolerant control techniques. Nevertheless, the developments presented throughout this part of the thesis are also valid if the anomaly that leads to reconfigure the system is the detection of an attack that compromises certain system components.

5.2 System reconfiguration with back-up components in FTC

Below, the system reconfiguration problem is revisited with the aim of highlighting the challenges faced when dealing with large-scale hardware-redundant systems subject to the occurrence of faults. Some of the concepts that are referenced below, are an adaptation of the definitions presented in [Blanke et al., 2006, Chapter 8] to the current problem of reconfiguration with back-up components.

5.2.1 The system reconfiguration problem

The general control problem can be posed as finding a control law in a given set of control laws U, such that the controlled system achieves the control objectives O, while its behaviour satisfies a set of constraints C. Thus, the solution of the problem is defined by the triple $< O, C, U >$. The occurrence of a discrete fault f will indeed modify the system constraints to C_f (modifying the system parameters or even changing the constraints). Fault occurrence may also restrict the admissible control laws, yielding the new set U_f. Therefore, the existence of a solution to the control problem $< O, C_f, U_f >$ determines the system tolerance to fault f.

Active fault-tolerant control tackles the control problem $< O, C_f, U_f >$ by adapting the control law to each faulty situation, that is, each of the different problems $< O, C, U >$ and $< O, C_f, U_f >$ for all $f \in \mathcal{F}$ ($\mathcal{F}$ indexes the set of all considered faults), has its own specific solution. To that end, active FTC strategies rely on the information provided by a fault diagnosis block. In the case where fault diagnosis algorithms are able to provide a reasonable estimation of the fault impact $(\hat{C}_f, \hat{U}_f)$, the problem becomes the so-called *fault accommodation* problem, which is formulated through the triple $< O, \hat{C}_f, \hat{U}_f >$.

Nevertheless, in many other cases, the diagnosis block is only capable of providing fault detection and isolation information, i.e., information that the system is no longer behaving as in healthy mode, and a division in the set of constraints such that $C_f = C_h \cup C'_f$, where C_h (respectively C'_f) denotes the subset of constraint associated with the healthy (faulty) part of the system. The same applies for the set of control laws $U_f = U_h \cup U'_f$. The lack of further information regarding the impact of the fault, suggests as the most reasonable approach to address the control of the faulty plant to reconfigure the system in order to get rid of the faulty components. In this context, the system *reconfiguration problem* is defined as:

Definition 5.1 (Reconfiguration problem [Blanke et al., 2006]). Find a new set of system constraints C^*, and admissible control laws U^*, such that the control problem $< O, C^*, U^* >$ has a solution, find and activate this solution.

Traditionally, depending on whether or not there are back-up elements characterized by the new constraints C_{off}, the reconfiguration problem is addressed by

- **No back-up components:** faulty components are turned off, that is, $C^* = C_h$ and $U^* = U_h$.
- **Back-up components:** the back-up elements are activated, i.e., $C^* = C_h \cup C_{off}$. The activation of this components may also extend the set of control laws such that $U^* \supseteq U_h$.

However, large-scale systems tend to have high intrinsic redundancy, with a large number of alternative actuators that are not used in nominal operation but could be activated if necessary. Hence, the consideration of all possible alternative actuators for solving the control problem may yield, although mathematically feasible, impractical solutions for real-life implementations. Consider for example to open all the possible alternative valves in a city water distribution system with the aim to palliate a local fault.

Therefore, it seems necessary to establish some optimality criterion in the selection of the alternative constraints. Consequently, Definition 5.1 is nuanced by emphasizing the optimal selection of a subset $C_a \subseteq C_{off}$, such that the resulting control problem $< O^*, C^*, U^* >$ has a solution for $C^* = C_h \cup C_a$. Note that, due to the modification of the system input-output structure, the control objectives O^* may differ from the nominal ones. Normally, high impact decisions like the reconfiguration of system structure and/or the modification of the control objectives are framed within the supervision problem, and usually require the approval of a human operator.

5.2.2 Literature review

Below, an analysis of the main approaches that have been taken to face the problem of reconfiguring a system with redundant hardware are presented.

On the one hand, the search for an optimal network configuration is a well-known problem in power networks, where the so-called power distribution network reconfiguration (PDNR) problem studies the system switches disposition in order to reduce power loss, increase system security and/or enhance power quality [Mishra et al., 2017]. Nevertheless, the binary structure of the switches has led to specific algorithms for PDNR, which seem hard to extrapolate to more general cases. In this regard, in the bibliographical survey Kalambe and Agnihotri [2014], the network reconfiguration approaches that have been proposed in the last decades in order to minimize the power losses are discussed.

Specifically within the automatic control field, the problem of system reconfiguration taking into consideration redundant hardware attracted attention at the end of the last century, mainly through the three-tank benchmark [Heiming and Lunze, 1999]. In order to address that problem, different solutions arose: the use of hybrid systems [Tsuda et al., 2001], qualitative solutions [Askari et al., 1999, Lunze and Schröder, 1999], multi-model switching control task [Rato and Lemos, 1999], or neural networks [Marcu et al., 1999]. Besides, in [Mignone, 2002, Chapter 4], the reconfiguration problem is addressed within a mixed logical dynamical (MLD) model framework: I) together with the control using a predictive control scheme; II) with a two decision level procedure, a first level that selects the inputs configuration, and a second level where a receding control strategy including the new inputs is applied. Some more recent practical examples of reconfiguration in network systems can be found for WDNs. In this sense, in Mahmoud et al. [2018] the authors provide a response methodology for reducing the impact of failure situations, where the selection among a possible set of interventions rely on genetic algorithms; in Vamvakeridou-Lyroudia et al. [2010] a hierarchical algorithm is proposed to palliate the effect of a single pipe burst.

Moreover, although not addressing directly the existence of back-up components, a great deal of effort has been put in the evaluation of the admissibility of a system configuration, as well as the control reconfiguration after a fault occurrence. On this subject, in the work of Staroswiecki [2002], the admissibility of an actuators configuration in an LTI system is assessed by checking a combination of structural and performance properties, namely: I) the controllability of the system; II) the controllability Gramian. These concepts are analysed in more depth in [Blanke et al., 2006, Chapter 8]. Besides, in Staroswiecki et al. [2004] the dual case of sensor configurations is studied. Additionally, the potential and the key properties of the lattice of configurations are discussed in Staroswiecki [2010]. In Veillette [1995] the idea of reliable control for the Linear Quadratic problem, that is dedicated to several actuator configurations, was introduced. In Staroswiecki and Berdjag [2010], a general framework that combines passive and active optimal control strategies is proposed. Furthermore, a formal approach for system reconfigurability is formulated in Gehin and Staroswiecki [1999, 2008] based on the different versions of the same service that a system may offer.

At this point, it must be highlighted that the presence of state and input constraints modifies the admissibility criterion used in the majority of the works presented in the preceding paragraph. In this scenario, the admissibility criterion must be extended to also assess the existence of a feasible control law, that is, the presence of constraints imposes the dependence of the admissibility assessment on the system state at fault diagnosis time. Thus, the selection of a new system configuration in a constrained system must deal with the online evaluation of the admissibility of each candidate configuration. Furthermore, another underlying challenge not treated in the aforementioned works is the system extension. Large-scale systems may present a high number of possible alternatives, where the selection of the optimal configuration results in a complex problem whose resolution may introduce latency in the control loop, or even become intractable.

5.3 Problem statement

The objective of this section is twofold: I) the introduction of some key concepts and definitions which will be used recursively in the following chapters; II) establish a general formulation of the problem that is addressed in this part of the thesis.

5.3.1 Actuator configurations

Let $C_0 = \{a_1, a_2, ..., a_n\}$ denote the full set of actuators (nominal plus back-up actuators), where a_i denotes the i^{th} actuator. Furthermore, let $C_i \subseteq C_0,\ i \in \{0, 1, ..., 2^n - 1\}$, denote the possible actuator configurations that arise from the selection of a subset of the n actuators. In the sequel, the cardinality of configuration C_i, i.e., the number of elements that it contains, is denoted as $|C_i|$. In addition, let 2^{C_0} denote the power set $2^{C_0} = \{C_i : C_i \subseteq C_0\}$ of all system configurations.

Below, some useful definitions associated with the usual set-inclusion based partial ordering of the actuator configurations are introduced.

Definition 5.2 (Predecessors / Successors [Staroswiecki, 2010])**.** Given the set of actuators C_0 and a configuration $C_i \subseteq C_0$, the predecessors $\mathbb{P}(C_i)$ and successors $\mathbb{S}(C_i)$ of C_i are defined as

$$\begin{aligned}\mathbb{P}(C_i) &= \{C_j : C_i \subseteq C_j \subseteq C_0\},\\ \mathbb{S}(C_i) &= \{C_j : C_j \subseteq C_i\}.\end{aligned}$$

Definition 5.3 (Predecessors after fault)**.** Given the set of actuators C_0 and a configuration $C_i \subseteq C_0$, the set of remaining predecessors $\mathbb{P}_{\tilde{C}_f}(C_i)$ of configuration C_i, after an outage of the actuators in $\tilde{C}_f$, is defined as

$$\mathbb{P}_{\tilde{C}_f}(C_i) = \{C_j : C_i \subseteq C_j \wedge C_j \subseteq (C_0 \setminus \tilde{C}_f)\}.$$

Furthermore, given a generic property $\mathcal{P}$, which can be either structural or non-structural, the assessment of $\mathcal{P}$ on the configuration C_i is formulated as

$$\begin{cases}\mathcal{P}(C_i) & \implies & \mathcal{P} \text{ is satisfied on } C_i,\\ \bar{\mathcal{P}}(C_i) & \implies & \mathcal{P} \text{ is not satisfied on } C_i.\end{cases}$$

Definition 5.4 (Span)**.** The span of $\mathcal{P}$, is the set $S(\mathcal{P})$ of all configurations in 2^{C_0} that satisfy $\mathcal{P}$, that is

$$S(\mathcal{P}) = \{C_i \in 2^{C_0} : \mathcal{P}(C_i)\}.$$

Definition 5.5 (Bottom-up Monotonicity)**.** A property $\mathcal{P}$ is bottom-up monotonous (BUM) if

$$\forall C_i \in 2^{C_0} : \mathcal{P}(C_i) \implies \mathcal{P}(C_j),\ \forall C_j \in \mathbb{P}(C_i).$$

Definition 5.6 (Minimal configuration)**.** The set of minimal elements of a subset of configurations $Q \subseteq 2^{C_0}$ is defined by

$$m(Q) = \{C_i \in Q : C_j \subset C_i \implies C_j \notin Q\}.$$

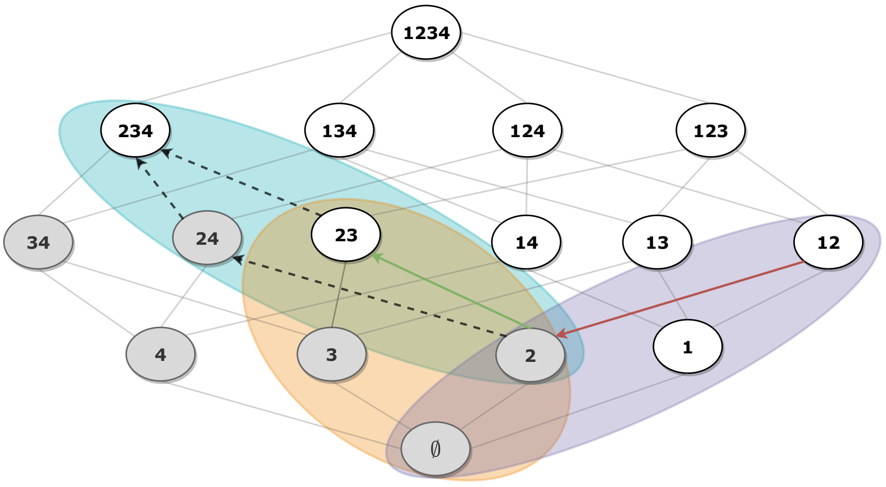


Figure 5.1: Lattice of actuators. $C_N = \{1,2\}$ - $\mathbb{S}(C_N)$ purple ellipse; $C_F = \{2\}$ - $\mathbb{P}_{\tilde{C}_f}(C_F)$ blue ellipse; $C_{new} = \{2,3\}$ - $\mathbb{S}(C_{new})$ orange ellipse.

5.3.2 Selection of an admissible configuration

In the sequel, the admissibility of an actuator configuration C_i at a given time instant is characterized by means of the property $\mathcal{A}$. In this regard, different definitions of admissibility are used in Chapter 6 and Chapter 7 according to distinct performance or stability requirements. Consequently, at each time instant, the power set 2^{C_0} of all system configurations can be partitioned as

$$2^{C_0} = S(\mathcal{A}) \cup S(\bar{\mathcal{A}}).$$

Throughout the remainder of this part of the thesis, it is considered that the system operates on an admissible nominal configuration C_N that only employs a subset of the actuators in C_0, i.e., $C_N \subset C_0$ (with $C_N \in S(\mathcal{A})$), whereas the remaining $C_A = C_0 \setminus C_N$ are considered as healthy back-up actuators. In this scenario, the occurrence of a fault on a subset of the nominal actuators $\tilde{C}_f \subseteq C_N$, drives the system to a faulty configuration denoted as $C_F \in \mathbb{S}(C_N)$.

If the configuration C_F is not admissible, that is, $C_F \in S(\bar{\mathcal{A}})$, then the available back-up components can be brought into play seeking for a new admissible configuration denoted as C_{new}. Therefore, the search of a new system configuration poses the following optimization problem

$$\begin{aligned} C_{new} = \quad & \underset{C_j}{\arg\min.}\, J(C_j), \\ \text{s.t.} \quad & C_j \in \mathbb{P}_{\tilde{C}_f}(C_F), \\ & C_j \in S(\mathcal{A}), \end{aligned} \tag{5.1}$$

where $J(C_j)$ stands for the overall cost of configuration C_j.

With illustrative purposes, a system with four different actuators indexed with numbers from 1 to 4, i.e., $C_0 = \{1,2,3,4\}$, is taken into consideration. In this regard, Figure 5.1, represents the lattice associated with the power set 2^{C_0} of all actuator configurations [Staroswiecki, 2010]. Besides, at a given time instant, it is considered that the set of non-admissible configurations are $S(\bar{\mathcal{A}}) = \big\{\{3,4\}, \{2,4\}, \ \{4\}, \ \{3\}, \ \{2\}, \ \emptyset\big\}$ (grey nodes in Figure 5.1), whereas the set $S(\mathcal{A}) = 2^{C_0} \setminus S(\bar{\mathcal{A}})$ is represented by the white nodes.

In addition, it is considered that the nominal configuration is $C_N = \{1, 2\}$, and thus the back-up elements are $C_A = \{3, 4\}$. Under this elements partition, if a fault occurs in the first actuator, that is, $\tilde{C}_f = \{1\}$, translating the system to $C_F = \{2\} \in S(\bar{\mathcal{A}})$, then the optimization problem (5.1) is solved for the available predecessors of C_F generated by the back-up elements (blue ellipse in Figure 5.1). Note that in Figure 5.1 the configuration $C_{new} = \{2, 3\}$ is highlighted as the solution of (5.1).

Therefore, given a nominal configuration C_N, its successors $\mathbb{S}(C_N)$ can be divided into

$$2^{C_N} = \mathcal{R} \cup \bar{\mathcal{R}},$$

with

- **Recoverable configurations**:

$$\mathcal{R} = \{C_i \in \mathbb{S}(C_N) : C_i \in S(\mathcal{A}) \vee \exists C_j \in \mathbb{P}_{C_N \setminus C_i}(C_i) \text{ such that } C_j \in S(\mathcal{A})\}.$$

- **Non-recoverable configurations**:

$$\bar{\mathcal{R}} = 2^{C_N} \setminus \mathcal{R}.$$

It must be highlighted that the optimization problem (5.1) depends on the capability to assess whether the different configurations considered are admissible. In this regard, the consideration of constrained systems entails that in order to consider that a configuration is admissible, it is not enough to satisfy some performance requirements, but it is also necessary to check the feasibility of each configuration. Unfortunately, testing the feasible evolution of the system depends on the system state at the fault diagnosis time, and thus it must be done online.

Note that this thesis only focuses on actuator faults. Nevertheless, the reconfiguration problem (5.1) can be solved for other types of faults. In this regard, a fault affecting a non-directly controllable element may provoke that the current system configuration is not admissible, motivating the quest for suitable actuator configurations.

5.3.3 Specific notation

Throughout this part of the thesis, bold letters are used to denote a sequence related with the time instants expressed in the subscript, i.e., $\boldsymbol{x}_{k_1:k_2} = \{x_{k_1}, x_{k_1+1}, ..., x_{k_2-1}\}$ (with $k_1 < k_2$). On the other hand, in the optimization problems, a sequence of T variables is denoted as $\boldsymbol{z}_T = \{z(0), ..., z(T-1)\}$. Finally, the set of positive integer numbers including the origin is termed as $\mathbb{I}_a = \{0, 1, ..., a\}$.

Chapter 6

Reconfiguration with back-up components - Planning

Large-scale control systems tend to present a large number of alternative and back-up elements that, although not used in nominal operation, could be brought into play if necessary. For systems with this structure, the occurrence of a fault poses the optimization problem of selecting the (in some sense) optimal configuration that yields an admissible operation. In this chapter, the admissibility of an actuator configuration is evaluated according to its ability to, at fault detection time, generate a feasible, and with an acceptable performance, system trajectory for a horizon that relates with the time interval by which the system is expected to have restored its nominal service. The configuration selection is posed as a multi-objective mixed-integer program (MIP) solved using a lexicographic approach. Aiming at reducing the worst-case execution time, the analysis of necessary properties for the existence of an admissible solution, as well as how to manage the information obtained by evaluating such properties to be included in the MIP, are investigated. A portion of a water transport network is used in order to validate the proposed solution.

6.1 Introduction

The main aim of this chapter is the development of a software tool that, located in the supervisory layer of large-scale systems, would help the system operator to manage critical situations by reconfiguring the system structure. In a fault scenario, the tool must select the (in some sense) optimal back-up elements such that the resulting configuration provides an admissible performance during the time interval up to the system restoration to its nominal operation. Throughout the rest of this chapter, the reconfiguration problem is particularized for flow-based networks, like for example water transport networks (WTNs), characterized by means of discrete-time linear models. This system selection is motivated by the inherent physical redundancy of networked systems, and supported by the interest arisen within the collaboration in external projects that have been carried out during the course of this Ph.D. thesis.

The hybrid nature of the configuration selection leads to representing the system with back-up elements as an MLD model, for which the optimal configuration is obtained by solving a multi-objective mixed integer program online. In this regard, motivated by the difficulty of assessing non-directly related objectives, the multi-objective MIP optimization is addressed using a lexicographic approach. Nevertheless, MIPs are known to be NP-complete with a worst-case optimization time that scales exponentially with the problem size. Aiming at reducing the worst-case execution time, necessary properties for the existence of an admissible solution are evaluated offline using structural and graph-theoretic techniques. Besides, techniques for feeding the online optimization with the information obtained from the offline tests are also studied. The aggregated water transport network presented in Section B.2 of Appendix B is used in order to exemplify the problem and validate the proposed solution.

It should be highlighted that the developments presented below are framed within the general formulation introduced in Chapter 5, and, in particular, the current chapter is aimed at solving the optimization problem posed in (5.1). To that end, repeated references are made to concepts and definitions related with the actuator configurations that are presented in Section 5.3.

The remainder of the chapter is structured as follows: Section 6.2 presents a general overview of the proposed methodology, whereas Section 6.3 is devoted to present the flow-based systems under study and introduce the admissibility criterion. In Section 6.4, the optimization problem that selects the adequate configuration is formulated. Throughout Section 6.5, graph-theoretic techniques for analysing necessary properties are presented. Section 6.6 proposes different algorithms for extracting valuable information from the evaluation of the previous necessary properties. The considered case study is presented in Section 6.7. Finally, Section 6.8 presents some concluding remarks.

6.2 Methodology overview

This section presents the main points in the design of a reconfiguration block which is intended to solve the optimization problem (5.1). The overall control/monitoring scheme considered throughout this chapter is represented in Figure 6.1. In this regard, the system operation is governed from a supervisory layer that consists of a fault diagnosis block, as well as a control block in charge of setting the required set-points to the low-level controllers placed at the plant's

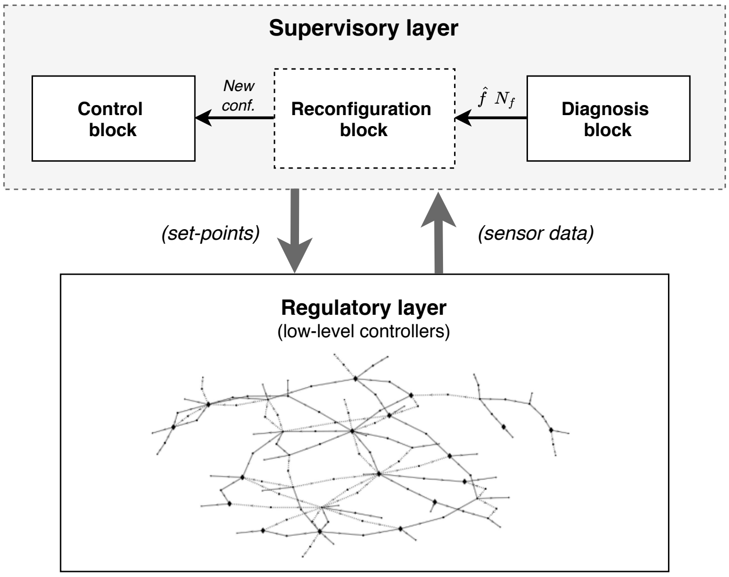


Figure 6.1: Reconfiguration block placement.

side. Note that in the following, no further considerations are made regarding the structure of the control block, which may well be an automatic controller or a human operator.

In addition, a reconfiguration block is added to feed the control block with valuable information regarding the most suitable system configuration after a fault occurrence. As discussed in Section 6.1, the reconfiguration block design is particularized for flow-based networks tracking a periodic reference. Regarding the fault information available, the following assumption is introduced.

Assumption 6.1. The diagnosis block is able to isolate the occurrence of an actuator fault. If an estimation of the fault is available, the model is updated coherently. If only fault isolation information is available, a total component fault is assumed and the corresponding actuator is turned-off.

Below, the temporal evolution of the events related with a fault occurrence defines the following sequence:

1. fault occurs at time k_f,
2. fault is diagnosed at $k_d \geq k_f$,
3. appropriate interventions are immediately carried out

Assumption 6.2. Depending on the nature of the fault, a time horizon N_f is considered such that the normal restoration of the service is expected before $k_s = k_d + N_f$.

According to Assumption 6.2, the reconfiguration block must find an admissible actuator configuration during the horizon N_f, since at k_s the nominal configuration is expected to be fully operative again. Note that Assumption 6.2 characterizes a natural scenario, since it implies that the evaluation of a fault is associated to a worst-case repair time.

Figure 6.2 presents an overview of the proposed methodology for solving the reconfiguration problem (5.1) associated with the time horizon N_f. To this end, the reconfiguration block takes as inputs the configuration C_F, the restoration time N_f and the system state at detection time x_{k_d}. A first step, tests the admissibility of C_F in order to decide whether or not the quest for alternative configurations must be carried out (green block). If the result is negative, the actual reconfiguration problem is solved.

Since the presence of constraints requires to know x_{k_d} in order to elucidate which configuration yields a feasible solution, the yellow block poses the optimization (5.1) as an MIP that is solved online (see Section 6.4). Nevertheless, in order to bound the worst-case execution time of the MIP, the blue block that takes into account the information retrieved from the offline analysis of necessary properties for the existence of a solution, is included. This necessary properties, which are detailed in Section 6.5, allow to either discard the existence of a solution or to filter the nodes explored by the MIP. How to pass the information obtained from the offline analysis in the form of constraints to the MIP optimization is analysed in Section 6.6.

6.3 System admissibility

This section is intended to introduce the systems under study, as well as the definition of admissibility that will be used in the remainder of this chapter.

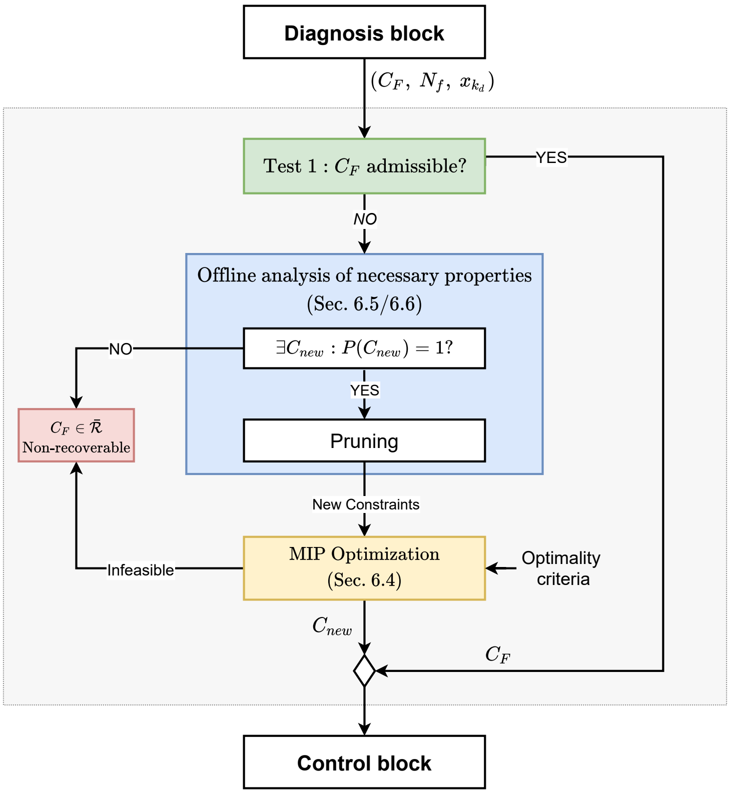


Figure 6.2: Reconfiguration block operation diagram.

6.3.1 System description

Basically, flow-based networks are conformed by a combination of flow sources, flow handling components, sinks, nodes, links and storage elements [Grosso Pérez, 2015]. The control-oriented model of flow-based networks may be described by means of a set of linear (or linearised) discrete difference-algebraic equations (DAE) [Grosso et al., 2014]. Accordingly, the system under study is described as

$$x_{k+1} = Ax_k + B\Sigma\bar{u}_k + B_d d_k, \tag{6.1a}$$
$$0 = E\Sigma\bar{u}_k + E_d d_k, \tag{6.1b}$$

where the difference equations in (6.1a) describe the dynamics of the storage elements, and the algebraic equations in (6.1b) describe the static relations, i.e., mass balance at junction nodes, in the network. For any time instant $k \in \mathbb{N}$, vector $x_k \in \mathbb{R}^{n_x}$ denotes the state variables, $\bar{u}_k \in \mathbb{R}^{\bar{n}_u}$ the control actions and $d_k \in \mathbb{R}^{n_d}$ represents the demanded flow which is modelled as additive measured disturbances. System matrices A, B, B_d, E, E_d are of suitable dimensions set by the network topology, whereas matrix Σ is a diagonal binary matrix that dictates the actuator configuration.

Henceforth, $u_k \in \mathbb{R}^{n_u}$ denotes the control variables used in the nominal configuration C_N (with $n_u = |C_N|$), whereas $v_k \in \mathbb{R}^{n_v}$ denotes the alternative control variables associated with the back-up elements in C_A (with $n_v = |C_A|$). For the sake of simplified notation, in the following it is considered that the control variables have been rearranged in such a way that $\bar{u}_k = [u_k^T,\ v_k^T]^T$. Consequently, the input matrices (B, E) are split into the nominal (B_N, E_N) and alternative (B_A, E_A) input matrices as

$$B = \begin{bmatrix} B_N, & B_A \end{bmatrix}, \qquad E = \begin{bmatrix} E_N, & E_A \end{bmatrix},$$

and the configuration selection matrix Σ is rewritten as

$$\Sigma = \begin{bmatrix} I_{n_u} & 0 \\ 0 & diag(\delta) \end{bmatrix}, \tag{6.2}$$

where δ is a binary vector $\delta = [\delta_1, ..., \delta_{n_v}]^T \in \{0, 1\}^{n_v}$ that rules the selection of the corresponding back-up actuator (see Section 6.4).

Furthermore, for all time instants k, system (6.1) is considered to be subject to hard state and control constraints given by the following polytopic sets

$$\mathcal{X} = \{x_k \in \mathbb{R}^{n_x} \mid Gx_k \leq g\} \subset \mathbb{R}^{n_x}, \tag{6.3a}$$
$$\mathcal{U} = \{u_k \in \mathbb{R}^{n_u} \mid Fu_k \leq f\} \subset \mathbb{R}^{n_u}, \tag{6.3b}$$
$$\mathcal{V} = \{v_k \in \mathbb{R}^{n_v} \mid Hv_k \leq h\} \subset \mathbb{R}^{n_v}, \tag{6.3c}$$

with $G \in \mathbb{R}^{c_x \times n_x}$, $F \in \mathbb{R}^{c_u \times n_u}$, $H \in \mathbb{R}^{c_v \times n_v}$ and $g \in \mathbb{R}^{c_x}$, $f \in \mathbb{R}^{c_u}$, $h \in \mathbb{R}^{c_v}$, being c_x and c_u, c_v the number of state and input constraints, respectively.

According to the above, in the nominal configuration C_N, i.e., with $\delta = 0$, the system dynamics are governed by

$$x_{k+1} = Ax_k + B_N u_k + B_d d_k, \tag{6.4a}$$
$$0 = E_N u_k + E_d d_k. \tag{6.4b}$$

At this point, the following assumption is introduced.

Assumption 6.3. The states in x_k and the demands in d_k are observable at time instant k, and the pair (A, B_N) is stabilizable.

Moreover, after fault occurrence at k_f and diagnosis at k_d, for $k \geq k_d$ the faulty system will evolve according to

$$x_{k+1} = A_F x_k + B_F u_k + B_d d_k, \tag{6.5a}$$
$$0 = E_F u_k + E_d d_k, \tag{6.5b}$$

with A_F, B_F, E_F the estimated faulty matrices and the system state starting in x_{k_d}.

6.3.2 Admissibility criterion

At fault diagnosis time, the predicted optimal system evolution for the time interval $[k_d,\ k_s]$ is assumed to be quantified by means of the weighted sum of n_p scalar-valued cost functions $\phi_j(\cdot)$, $\forall j \in \{1, ..., n_p\}$. Thus, the overall system performance index for configuration C_i is defined as

$$\Phi_{C_i} = \sum_{j=1}^{n_p} \phi_j(C_i, N_f, x_{k_d}, \bar{\boldsymbol{u}}_{N_f}^{C_i}, \boldsymbol{d}_{k_d:k_s}, \boldsymbol{x}_{k_d:k_s+1}^r), \tag{6.6}$$

where $\bar{\boldsymbol{u}}_{N_f}^{C_i} = \{\bar{u}(0), ..., \bar{u}(N_f - 1)\}$ is the predicted sequence of optimal control actions for configuration C_i computed at diagnosis time, $\boldsymbol{d}_{k_d:k_s} = \{d_{k_d}, ..., d_{k_s-1}\}$ is the sequence of flow demand predictions which are typically inferred from the periodic operation cycles of flow-based networks [Quevedo et al., 2010] and $\boldsymbol{x}_{k_d:k_s+1}^r$ characterizes a reference trajectory or set-point.

Moreover, let $\boldsymbol{x}^{C_i}_{N_f+1} = \{x(0), ..., x(N_f)\}$ denote the predicted sequence of system states at time k_d obtained by applying the sequences $\bar{\boldsymbol{u}}^{C_i}_{N_f}$, $\boldsymbol{d}_{k_d:k_s}$ and $\boldsymbol{x}^r_{k_d:k_s+1}$. Therefore, following the standard definition of admissibility [Blanke et al., 2006], the admissibility of configuration C_i is assessed according to the following definition.

Definition 6.1 (Admissibility)**.** Configuration C_i is admissible for the time interval $[k_d,\ k_s]$ if and only if:

1. $\boldsymbol{x}_{C_i}$ is a feasible sequence of system states.
2. The performance index satisfies
$$\Phi_{C_i} \leq \beta\Phi_{C_N}, \tag{6.7}$$
where Φ_{C_N} is the cost function associated with the nominal configuration which is obtained by simulating its behaviour for the time horizon N_f, and $\beta \geq 1$ is a predefined scalar set by the user.

Remark 6.1. In the remainder of this chapter, it is implicitly assumed that, starting at x_{k_d}, the nominal configuration C_N is able to obtain a feasible solution in $[k_d,\ k_s]$. That is, it is assumed that $k_d \approx k_f$, and thus the faulty system did not deviate much from the nominal behaviour at the detection time.

6.3.3 Admissibility test for the faulty configuration

Fault occurrence causes constraint modification in the control problem, thus affecting the set of feasible solutions. Hence, the admissibility of C_F is tested (green block in Figure 6.2) to decide whether or not to search for admissible alternative configurations. In this regard, by introducing the performance condition (6.7) as a constraint, the admissibility of C_F can be assessed by checking the existence of a solution to the following problem [Ocampo-Martinez et al., 2007a]

$$\begin{aligned} &\min_{\boldsymbol{u}_{N_f}}.\ h(\cdot), \\ \text{s.t.}\ &\Phi_{C_F} \leq \beta\Phi_{C_N}, \\ &x(j+1) = A_F x(j) + B_F u(j) + B_d d_{k_d+j}, && \forall j \in \mathbb{I}_{N_f-1}, \\ &0 = E_F u(j) + E_d d_{k_d+j}, && \forall j \in \mathbb{I}_{N_f-1}, \\ &x(0) = x_{k_d}, \\ &x(j) \in \mathcal{X}, && \forall j \in \mathbb{I}_{N_f}, \\ &u(j) \in \mathcal{U}, && \forall j \in \mathbb{I}_{N_f-1}, \end{aligned} \tag{6.8}$$

where $x(j)$ and $u(j)$ are the problem variables and $h(\cdot)$ stands for the null function since it is only needed to check if the problem constraints are violated.

6.4 Configuration selection

This section formulates the online selection of a new alternative configuration C_{new} (yellow block in Figure 6.2). In this regard, even though the systems under consideration are modelled by means of discrete difference-algebraic equations, the binary modification of the system configuration confers the problem with a hybrid nature. Each possible actuator configuration represents a different mode, where, contrary to hybrid models where the mode is ruled by the state of the system, in this situation the mode selection becomes an input to the system.

In order to generate such a hybrid model, the logic statements that rule the activation of only one possible configuration can be transformed into linear integer equations by means of auxiliary variables [Williams, 2013]. Hence, following the notation introduced in (6.2), the selection of a subset of alternative actuators is formulated as

$$z_k = diag(\delta) v_k, \tag{6.9}$$

in such a way that, for all $k \in [k_d,\ k_s]$,

$$\begin{aligned} &[\delta_i = 0] \rightarrow [z_k^i = 0], && \text{the } i^{th} \text{ actuator is not used,} \\ &[\delta_i = 1] \rightarrow [z_k^i = v_k^i], && \text{the value of the } i^{th} \text{ actuator is } v_k^i, \end{aligned} \tag{6.10}$$

and thus the binary vector δ characterizes the selected configuration over the 2^{n_v} different possibilities.

Logic statements formulated as the product of continuous and logic variables, can be transformed into equivalent linear integer programs through the following inequalities [Bemporad and Morari, 1999]

$$\begin{aligned} z_k &\leq diag(\delta)\bar{v}, \\ z_k &\geq diag(\delta)\underline{v}, \\ z_k &\leq v_k - \big(I_{n_v} - diag(\delta)\big)\underline{v}, \\ z_k &\geq v_k - \big(I_{n_v} - diag(\delta)\big)\bar{v}, \end{aligned} \tag{6.11}$$

where $\bar{v}$, $\underline{v} \in \mathbb{R}^{n_v}$ represent the upper and lower bounds of the variable v_k, which can be obtained through the interval hull computation of the set $\mathcal{V}$.

6.4.1 MLD formulation

Below, the resulting hybrid system is formalized within an MLD representation [Bemporad and Morari, 1999]. To this end, by taking into consideration the injection of the auxiliary variables z_k through the input matrices $(B_A,\ E_A)$, the dynamics of the resulting system are governed by the model

$$\begin{aligned} x_{k+1} &= A_F x_k + B_F u_k + B_A z_k + B_d d_k, \\ 0 &= E_F u_k + E_A z_k + E_d d_k, \end{aligned} \tag{6.12}$$

subject to the state and input constraints (6.3), as well as the linear integer constraints (6.11).

Hence, recalling the structure of the overall input variables $\bar{u}_k$ and extending the auxiliary variables with the demand flow prediction as follows

$$\bar{u}_k = \begin{bmatrix} u_k^T, & v_k^T \end{bmatrix}^T, \qquad \bar{z}_k = \begin{bmatrix} z_k^T, & d_k^T \end{bmatrix}^T, \tag{6.13}$$

then, after the corresponding mathematical transformations, an MLD formulation for the predicted evolution of system (6.12) can be obtained as

$$\begin{aligned} x_{k+1} &= A_F x_k + \underbrace{\begin{bmatrix} B_F & 0 \end{bmatrix}}_{B_1} \bar{u}_k + \underbrace{\begin{bmatrix} B_A & B_d \end{bmatrix}}_{B_2} \bar{z}_k, \\ E_2\delta + \underbrace{\begin{bmatrix} E_{31} & E_{32} \end{bmatrix}}_{E_3} \bar{z}_k &\leq \underbrace{\begin{bmatrix} E_{11} & E_{12} \end{bmatrix}}_{E_1} \bar{u}_k + E_4 x_k + E_5, \end{aligned} \tag{6.14}$$

with the system starting at x_{k_d} and the matrices

$$\begin{aligned}
E_2 &= \begin{bmatrix}\bar{0}_{1\times 5} & diag(\underline{v}) & -diag(\bar{v}) & diag(\bar{v}) & -diag(\underline{v})\end{bmatrix}^T,\\
E_{31} &= \begin{bmatrix}-E_A^T & E_A^T & \bar{0}_{1\times 3} & -I_{n_v} & I_{n_v} & -I_{n_v} & I_{n_v}\end{bmatrix}^T,\\
E_{32} &= \begin{bmatrix}-E_d^T & E_d^T & \bar{0}_{1\times 7}\end{bmatrix}^T,\\
E_{11} &= \begin{bmatrix}E_F^T & -E_F^T & \bar{0} & -F^T & \bar{0}_{1\times 5}\end{bmatrix}^T,\\
E_{12} &= \begin{bmatrix}\bar{0}_{1\times 4} & -H^T & \bar{0}_{1\times 2} & -I_{n_v} & I_{n_v}\end{bmatrix}^T,\\
E_4 &= \begin{bmatrix}\bar{0}_{1\times 2} & -G^T & \bar{0}_{1\times 6}\end{bmatrix},\\
E_5 &= \begin{bmatrix}\bar{0}_{1\times 2} & g^T & f^T & h^T & \bar{0}_{1\times 2} & \bar{v}^T & -\underline{v}^T\end{bmatrix}^T,
\end{aligned}$$

where $\bar{0}$ denotes a null matrix of appropriate dimensions and $\bar{0}_{1\times m}$ the concatenation of m null matrices (e.g. $\bar{0}_{1\times 3} = [\bar{0}, \bar{0}, \bar{0}]$).

It must be highlighted the fact that, in formulation presented in (6.14), vector δ is time independent, that is, the same configuration maintains throughout the whole time horizon N_f.

Remark 6.2. In order to avoid the procedure of deriving the MLD form by hand, a compiler was developed in Torrisi and Bemporad [2004] to automatically generate the different system matrices.

6.4.2 Lexicographic optimality criterion

The selection of the optimal system configuration must be done following some pre-established criteria. For this purpose, p different objectives that dictate the configuration selection are considered, in such a way that, each one of them is characterized through $f_i(\cdot)$, $i \in \{1, ..., p\}$, scalar-value functions.

Motivated by the difficulty of carrying out an adequate weighting when comparing objectives that are not directly related, particularly in large-scale systems, the remainder of this chapter considers a clear hierarchy among the different optimization objectives. In such a way that the optimization of the first defined objective is infinitely more important than the optimization of the second one, and so on. Previous assumption has led to the consideration of a lexicographic approximation for solving the resulting multi-objective optimization. Specifically, this lexicographic optimization is solved using the sequential solution method detailed in Algorithm A.2 of Appendix A .

Remark 6.3. It must be noted that the proposed methodology would also accommodate the use of other well-known approaches to solve the resulting multi-objective optimization [Marler and Arora, 2004].

6.4.3 MIP optimization

Given the reparation time N_f and the initial state $x(0) = x_{k_d}$, the sequences

$$\boldsymbol{x}_{N_f+1} = \{x(0), ..., x(N_f)\}, \qquad \boldsymbol{z}_{N_f} = \{z(0), ..., z(N_f-1)\},$$

are generated by applying the input sequences

$$\bar{\boldsymbol{u}}_{N_f} = \{\bar{u}(0), ..., \bar{u}(N_f-1)\}, \qquad \boldsymbol{d}_{k_d:k_s} = \{d_{k_d}, ..., d_{k_d+N_f-1}\},$$

to the system configuration imposed by the binary vector δ. Hence, the actuator selection subject to the satisfaction of the admissibility conditions expressed in Definition 6.1, is formulated as the multi-objective optimization problem

$$\begin{aligned}
\delta = \; & \underset{\delta,\bar{\boldsymbol{u}}_{N_f}}{\arg\min.} \; f = [f_1,\; f_2,\; ...,\; f_p]^T, && && (6.15a)\\
\text{s.t.} \; & \Phi_{C_j} \leq \beta \Phi_{C_N}, && && (6.15b)\\
& x(i+1) = A_F x(i) + B_1 \bar{u}(i) + B_2 \bar{z}(i), && \forall i \in \mathbb{I}_{N_f - 1}, && (6.15c)\\
& E_2 \delta + E_3 \bar{z}(i) \leq E_1 \bar{u}(i) + E_4 x(i) + E_5, && \forall i \in \mathbb{I}_{N_f - 1}, && (6.15d)\\
& x(0) = x_{k_d}, && && (6.15e)\\
& x(i) \in \mathcal{X}, && \forall i \in \mathbb{I}_{N_f}, && (6.15f)\\
& \bar{u}(i) \in \mathcal{U} \times \mathcal{V}, && \forall i \in \mathbb{I}_{N_f - 1}, && (6.15g)
\end{aligned}$$

where Φ_{C_j} denotes the performance loss of the configuration imposed by the input variable δ.

Notice the influence of parameter β in the performance of the method. In this regard, large values of β may result in configurations with a significant performance degradation Φ_{C_i}, whereas small values may lead to unfeasible solutions. On this subject, the use of a lexicographic optimization allows to set the performance index Φ_{C_i} as a secondary optimization objective. By doing so, the constraint (6.15b) filters non-admissible configurations in the first iteration of the lexicographic optimization, while the performance index can be minimized within the obtained set of admissible configurations in the following iterations.

6.5 Analysis of necessary properties

This section aims to reduce the computational complexity of the MIP optimization proposed in (6.15). On that subject, MIPs are known to be NP-complete with a worst-case optimization time that scales exponentially with the problem size. This issue becomes critical when facing large-scale systems with a vast number of alternative actuators. Aiming at limiting the worst-case execution time, offline tests may be developed for identifying the necessary components that at least must be brought into play for the existence of an admissible solution.

Note that, on the one hand, the necessary properties analysed must be independent of the system state at detection time x_{k_d}, the restoration horizon N_f or the parameter β. On the other hand, in offline tests the computational burden is not a problem, allowing thus to explore a great number of combinations for different possible faults.

In this regard, valuable information can be retrieved from a structural analysis of the system [Siljak, 2011]. Hence, properties like: connectivity, reachability or structural controllability [Lin, 1974, Reinschke and Wiedemann, 1997], which are necessary for the satisfaction of the conditions stated in Definition 6.1, should be analysed. Furthermore, structural analysis can be easily related with graph-theoretic techniques, for which there exist well-grounded algorithms that suit well with the analysis of large-scale systems.

6.5.1 Flow-based networks

Since the current chapter focuses on the study of flow networks, the periodic patterns on the demand flow, as well as the flow directionality, can be exploited in order to analyse the capacity of a given configuration C_i to satisfy the maximum expected demand on each node of the network during a periodic cycle. Note that this analysis subsumes the information retrieved from other structural properties like the ones discussed above.

To that end, given a flow-based network with a particular actuator configuration, this can be modelled as a directed graph (digraph) $G = (V, E)$, where V is the set of vertices and E the set of edges such that each edge is related with a capacity function c_i [Ford Jr and Fulkerson, 1962]. Besides, denoting as s the set of nodes that correspond to a source and t the set of nodes that correspond to a sink, flow networks can be characterized through (G, c, s, t). On this subject, the procedure followed to obtain the set (G, c, s, t) from the control-oriented model of a constrained flow-based network is detailed below.

- *Graph*: In order to build a directed graph from a model of the form (6.4), the following considerations are taken into account: I) flow handling elements u_k and demand nodes d_k are modelled as positive, and thus the directionality of the flow is reflected by the sign of matrices (B_N, E_N) and (B_d, E_d), respectively; II) the different storage elements are decoupled, i.e., matrix A is diagonal.

  The selected vertices of the graph are: system states $X = \{x_1, ..., x_{n_x}\}$, control inputs $U = \{u_1, ..., u_{n_u}\}$, algebraic nodes $N = \{n_1, ..., n_n\}$ and demand sinks $D = \{d_1, ..., d_{n_d}\}$. Hence, sorting the vertices as $V = X \cup U \cup N \cup D$, the interconnection (adjacency) matrix [Siljak, 2011] of the graph is given by

$$Q = \left[\begin{array}{c|c|c|c} \bar{A} & \bar{B}_N^+ & 0 & \bar{B}_d^+ \\ \hline (\bar{B}_N^-)^T & 0 & (\bar{E}_N^-)^T & 0 \\ \hline 0 & \bar{E}_N^+ & 0 & \bar{E}_d^+ \\ \hline (\bar{B}_d^-)^T & 0 & (\bar{E}_d^-)^T & 0 \end{array}\right], \tag{6.16}$$

  where for each matrix M the superscript $^+$ indicates the elements such that $m_{i,j} \geq 0$ ($m_{i,j} \leq 0$ for superscript $^-$), and $\bar{M}$ denotes the Boolean representation of the matrix such that

$$\bar{m}_{ij} = \begin{cases} 1, & m_{ij} \neq 0, \\ 0, & m_{ij} = 0. \end{cases} \tag{6.17}$$

  Hence, Q is a boolean matrix for which the component $q_{ij} = 1$ dictates that there is a directed edge that goes from vertex v_i to vertex v_j.

- *Edge capacity:* The edges that start or end in an flow-handling vertex $u_i \in U$ are associated with a capacity c_i equal to the maximum value of the corresponding actuation variable. That is, the capacity of each tube passing through an actuator is limited by the maximum value of the actuator itself.

- *Sinks and sources:* The demand nodes D are set as the network sinks $t = (t_1, ..., t_{n_d})$, whereas the storage elements X as sources $s = (s_1, ..., s_{n_x})$.

Note that the previous steps for obtaining the set (G, c, s, t) have been particularized for the nominal configuration C_N. However, a similar procedure can be followed for any arbitrary configuration. Consequently, the capacity of a configuration C_i to satisfy the maximum expected demand in the sinks can be analysed by means of Algorithm 6.1. In this regard, this algorithm tests the capacity of C_i of fulfilling the different demands independently, establishing thus a conservative filter for non admissible configurations. Note that the fifth step solves the max-flow problem of a flow based-network, for which there exist efficient algorithms [Edmonds and Karp, 1972, Boykov and Kolmogorov, 2004].

Algorithm 6.1 Capacity test of C_i

```
1: Compute the maximum expected demand in the sinks $\bar{t} = (\bar{t}_1, ..., \bar{t}_{n_d})$.
2: Connect all the sources to a super-source $S^*$ with infinity capacity.
3: Generate the new graph $G^*_{C_i}$ with the corresponding capacity $c^*$
4: for $j = 1$ to $n_d$ do
5:     $\bar{c}_j$ = max-flow problem $(G^*_{C_i}, c^*, S^*, t_j)$
6: end for
7: if any $(\bar{c} < \bar{t})$ then
8:     $C_i$ is not admissible
9: end if
```

6.6 Pruning

Whereas previous section discussed the offline examination of several necessary properties for the existence of a solution to the reconfiguration problem, this section focuses in the use of such properties to prune the decision tree of the MIP optimization posed in (6.15). Henceforth, a generic necessary property is termed as $\mathcal{P}$.

At this point, it must be highlighted that all necessary properties related with the actuator configurations are bottom-up monotonous, that is, if one configuration of actuators satisfies $\mathcal{P}$ any superset of it will also satisfy $\mathcal{P}$ (cf. Definition 5.5). This fact complicates the pruning of alternative actuators that will not help to obtain an admissible configuration, since any actuator can be part of a configuration that satisfies $\mathcal{P}$, meanwhile it is added to a configuration that already satisfies it.

Hence, in order to extract valuable information from the offline tests, attention will be focused on looking for those configurations of actuators that must be necessarily included instead of those that can be discarded. At this point, two different approaches are proposed:

1. look for the minimal configuration of actuators for which $\mathcal{P}$ is satisfied,

2. look for the minimal actuator configurations without which $\mathcal{P}$ can not be satisfied.

Both approaches will be presented below.

With the purpose of represent graphically the concepts discussed in this section, Figure 6.3 is presented. In this figure, the set of actuators $\{1, 2, 3, 4, 5, 6, 7\}$ are available in order to link the satisfaction of property $\mathcal{P}$ with its requirements. Accordingly, $\mathcal{P}(C_i)$ implies that C_i is a configuration of actuators that connects the lower and upper part of the scheme.

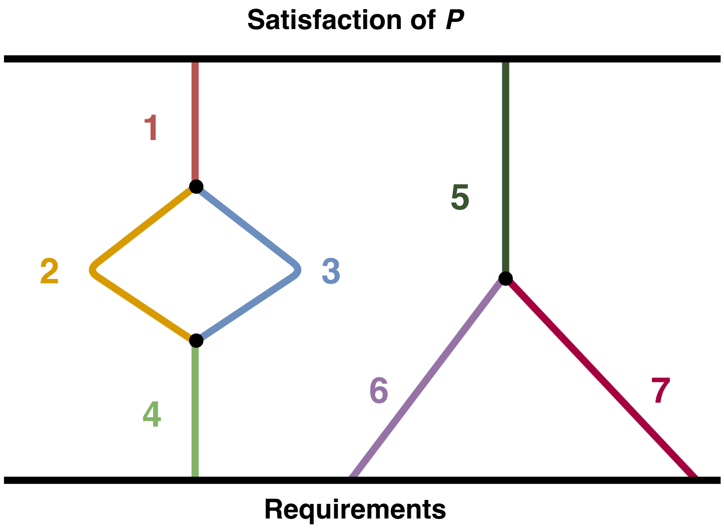


Figure 6.3: Schematic representation of actuator configurations.

6.6.1 Minimal actuator configurations

The first approach is to look for the set of minimal actuator configurations (MACs) that satisfy property $\mathcal{P}$. According to Definition 5.6, this set is defined as

$$\Psi = m\big(S(\mathcal{P})\big) = \{C_i \in S(\mathcal{P}) : C_j \subset C_i \implies C_j \notin S(\mathcal{P})\}.$$

Therefore, for the case of systems with back-up elements, given a faulty configuration C_F such that $\bar{\mathcal{P}}(C_F)$, from the bottom-up monotonicity of the necessary properties it follows that

$$\forall C_i \in \mathbb{P}_{\tilde{C}_f}(C_F) : \mathcal{P}(C_i), \quad \exists C_j \text{ such that } C_i \supseteq C_j \text{ and } C_j \in \Psi,$$

and thus at least one of the MACs must be necessarily activated in order to satisfy $\mathcal{P}$.

Given a configuration C_F, Algorithm 6.2 summarizes the procedure followed in order to look for the configurations of alternative actuators that with its activation the resulting configuration is a MAC for property $\mathcal{P}$. This algorithm takes as inputs a faulty C_F and back-up C_A configurations (with $|C_A| = n_v$). The possible 2^{n_v} configurations of the elements in C_A are sorted in increasing order of cardinality, generating the set of candidate configurations: $\Lambda = [C_1, ..., C_{2^{n_v}}]$ (with $|C_1| = 0$, ..., $|C_{2^{n_v}}| = n_v$). The output of the algorithm it the set of configurations Ψ.

For the example presented in Figure 6.3, the set of MACs is $\Psi = [(1,2,4),\ (1,3,4),\ (5,6),\ (5,7)]$. Note that it is necessary that at least one of these configurations is activated.

6.6.1.1 MACs as constraints

Assume the final set of MACs $\Psi = [C'_1, ..., C'_m]$, such that at least one of those configurations must be activated. In order to feed the MIP solver with that information, the binary vector δ that dictates the use of the alternative elements is taken back into consideration.

Accordingly, the use of at least one configuration from Ψ implies that the following con-

Algorithm 6.2 Search for MACs($\mathcal{P}$)

```
    Input: $\Lambda = [C_1, ..., C_{2^{n_v}}]$, $C_F$
    Output: $\Psi$
Initialize the empty set $\Psi = \emptyset$
while $\Lambda \neq \emptyset$ do
    $C = \Lambda(1)$
    if $\mathcal{P}(C_F \cup C)$ then
        $\Psi \leftarrow C$
        Remove all $C_j : C_j \supseteq C$ from $\Lambda$
    else
        Remove $C$ from $\Lambda$
    end if
end while
```

straint will be satisfied

$$\sum_{i=1}^{m} \left(\frac{\prod_{j=1}^{m} |C'_j|}{|C'_i|} \delta_{C'_i} \right) \geq \prod_{j=1}^{m} |C'_j|, \tag{6.18}$$

where $\delta_{C'_i}$ denotes the elements of the binary vector δ corresponding to the actuators present in configuration C'_i.

For the example introduced above, with $\Psi = [(1,2,4), (1,3,4), (5,6), (5,7)]$, the activation of any of the configurations of Ψ implies the satisfaction of

$$12(\delta_1 + \delta_2 + \delta_4) + 12(\delta_1 + \delta_3 + \delta_4) + 18(\delta_5 + \delta_6) + 18(\delta_5 + \delta_7) \geq 36.$$

Notice that the converse is not necessarily true, i.e., the satisfaction of (6.18) does not necessarily imply that one of the configurations in Ψ is activated.

6.6.2 Minimal necessary actuator configurations

For the case under investigation, more information can be retrieved by searching for the minimal configurations of actuators without which the property cannot be satisfied. Therefore, considering a property for which $\mathcal{P}(C_0)$, the following definition is introduced.

Definition 6.2 (Minimal necessary actuator configurations (MNACs))**.** Given the configuration C_0 with $\mathcal{P}(C_0)$. Then, the set of minimal necessary actuator configurations for property $\mathcal{P}$ is defined by

$$\Gamma = m_n(\mathcal{P}) = \{C_i : \bar{\mathcal{P}}(C_0 \setminus C_i) \implies \mathcal{P}(C_0 \setminus C_j),\ \forall C_j \subset C_i\}.$$

According to Definition 6.2, starting from C_0, configuration C_i is minimally necessary if by removing it the property $\mathcal{P}$ is not satisfied. But, if any single of its elements is not removed then $\mathcal{P}$ is satisfied. Consequently, it follows that at least one element from each MNAC must be activated.

Algorithm 6.3 summarizes the procedure followed in order to look for the set of MNACs by considering the same inputs than in Algorithm 6.2. The output of the algorithm is the set of MNACs Γ.

Algorithm 6.3 Search for MNAC($\mathcal{P}$)

```
    Input: $\Lambda = [C_1, ..., C_{2^{n_v}}]$, $C_F$
    Output: $\Gamma$
Initialize the empty set $\Gamma = \emptyset$
while $\Lambda \neq \emptyset$ do
    $C = C_{2^{n_v}} \setminus \Lambda(1)$
    if $\bar{\mathcal{P}}(C_F \cup C)$ then
        $\Gamma \leftarrow C_{2^{n_v}} \setminus C$
        Remove $C_j \supseteq (C_{2^{n_v}} \setminus C)$ from $\Lambda$
    else
        Remove $C_{2^{n_v}} \setminus C$ from $\Lambda$
    end if
end while
```

6.6.2.1 MNACs as constraints

Assume the final set of MNACs $\Gamma = [C'_1, ..., C'_m]$, such that, according to the stated above, at least one actuator from each MNAC must be necessarily activated. By means of the binary vector δ, this condition can be feed to the MIP solver as the linear constraint

$$\delta_{C'_i} \geq 1, \quad \forall i \in \{1, ..., m\}. \tag{6.19}$$

For the example presented in Figure 6.3, the obtained set of MNACs is given by $\Gamma = [(1,5), (4,5), (1,6,7), (4,6,7), (2,3,5), (2,3,6,7)]$. Thus, the following set of linear constraints to be included in the MIP is generated:

- $\delta_1 + \delta_5 \geq 1$,
- $\delta_1 + \delta_6 + \delta_7 \geq 1$,
- $\delta_2 + \delta_3 + \delta_5 \geq 1$,
- $\delta_4 + \delta_5 \geq 1$,
- $\delta_4 + \delta_6 + \delta_7 \geq 1$,
- $\delta_2 + \delta_3 + \delta_6 + \delta_7 \geq 1$.

Note that all the MACs obtained in Section 6.6.1, i.e., $\Psi = [(1,2,4), (1,3,4), (5,6), (5,7)]$, satisfy the previous set of inequalities, and hence the superset of any of them.

6.7 Case study

The case study is based on the aggregated version of the Barcelona DWTN presented in Section B.2 of Appendix B.

6.7.1 Network model

The considered network consists of 9 water sources (including 5 underground and 4 superficial), 17 water tanks, 61 actuators (37 valves and 24 pumps), 12 nodes and 25 demands. Besides, the following data are available: maximum/minimum levels of tanks, upper bound on the flow passing through the different actuators and a collection of historical water consumption data for each of the 25 demand nodes. In addition, hourly variations in the electricity rate are also

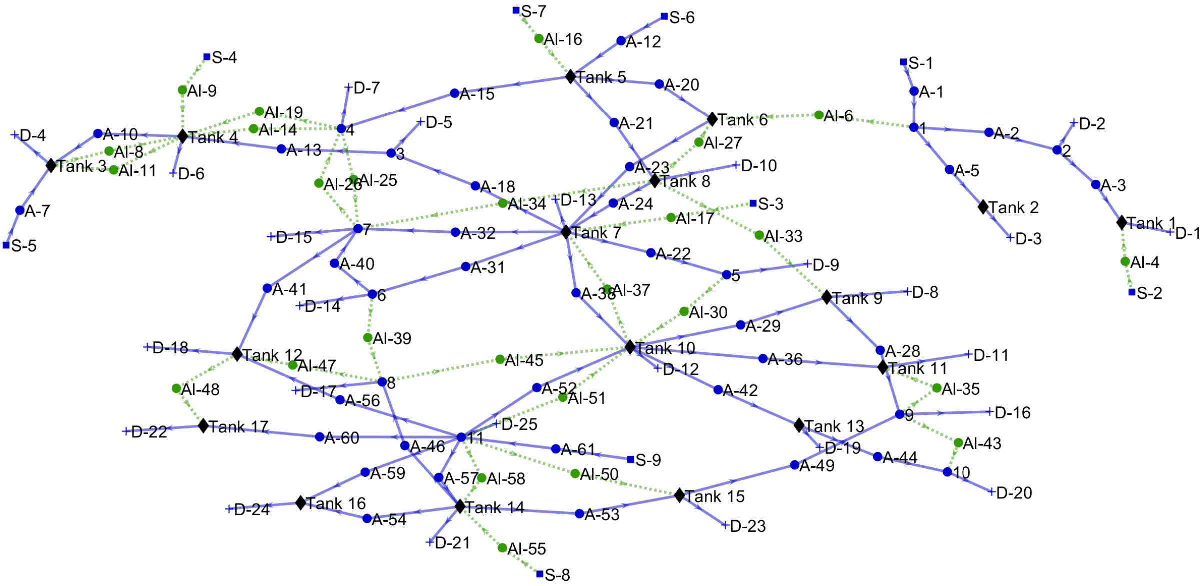


Figure 6.4: DWTN Graph. Nominal set of actuators (blue) alternative set of actuators (green).

available. These variations in the electricity cost affect the pumping stations, and will be taken into consideration in the design of the reference trajectories.

In order to illustrate the problem, it is assumed that only a subset of the 61 possible actuators are used in nominal operation. Consequently, indexing the different actuators with numbers from 1 to 61, the partition expressed in Table 6.1 has been taken into consideration. Hence, out of the 61 actuators, only 35 (14 pumps and 21 valves) will be used in nominal operation, while the other 26 actuators (10 pumps and 16 valves) are considered as back-up elements. As a consequence, given a faulty configuration C_F, this generates $|\mathbb{P}_{\tilde{C}_f}(C_F)| = 2^{26}$ candidate configurations.

Figure 6.4 introduces a graph representation of the system, where the nodes represents the different elements (tanks, actuators, sources, nodes and demands) and the edges describe the directed pipes connecting the network components. Accordingly, elements in blue illustrate the nominal system configuration, whereas green dashed elements are the back-up components through which water does not flow under normal conditions (valves closed and pumps off).

Following the stated in Section B.2 of Appendix B, a mathematical model of the form (6.4) has been obtained for the system in nominal operation. This model considers the $n_x = 17$ tank

	Pumps	**Valves**
Nominal	$3, 5, 10, 15, 20, 21,$ $22, 23, 24, 29, 36,$ $38, 42, 53$	$1, 2, 7, 12, 13, 18,$ $28, 31, 32, 40, 41, 44$ $46, 49, 52, 54, 56, 57,$ $59, 60, 61$
Back-up	$4, 9, 11, 17, 19,$ $27, 33, 34, 48, 55$	$6, 8, 14, 16, 25, 26,$ $30, 35, 37, 39, 43,$ $45, 47, 50, 51, 58$

Table 6.1: Elements partition.

volumes as the system states. The model is discretized with a sampling time $T_s = 1$h.

6.7.2 Reference model

As it will be further discussed in Section 6.7.4, the performance index Φ_{C_i} assesses the system evolution with respect the trajectories generated by a reference model. The considered reference model has the same structure than the nominal system (cf. Table 6.1), and its evolution in time takes into consideration the variations in the electricity rate.

Hence, by identifying periodic patterns in water consumption, and taking into account the hourly cost of electricity, an open-loop optimization is solved for the design of the reference model trajectory such that minimizes the cost of the plant operation [Wang et al., 2017]. The obtained trajectory exhibits a 24h periodic pattern that takes advantage of the lower electricity prices to fill the deposits during the night. In this regard, the blue dashed line in Figure 6.7 exemplifies the temporal evolution obtained for the reference trajectory of the state of the twelfth tank.

6.7.3 Offline tests

Since the case study deals with a flow-based network, the capability of the system to meet the maximum expected demand after a single fault occurrence, has been evaluated offline (see Section 6.5.1). In this regard, the set of minimal necessary alternative actuator configurations MNACs with cardinalities lower or equal than three, were searched by means of Algorithm 6.3. The maximum flow problems were solved using the method in Boykov and Kolmogorov [2004]. The test yielded the following results:

- The faulty configurations C_F resulting after a fault in any of the following nominal components:
$$\{1, 2, 15, 18, 22, 28, 31, 40, 49, 61\},$$
were identified as non-recoverable, i.e., $C_F \in \bar{\mathcal{R}}$. In other words, it does not exist any combination of alternative elements such that the configuration resulting from a fault in any of the above elements can satisfy the maximum expected demand in all the demand nodes. Note that capacity test offers a more detailed information than other test like, for example, connectivity. This can be seen in elements $\{28, 49, 40\}$ (cf. Figure 6.5), which were identified as critical despite the connectivity of all the demand nodes is preserved.

- For faults in the components $\{32, 44, 46\}$, the following MNACs were found

Fault	**MNAC** $(\lvert C_i' \rvert \leq 3)$	**Constraint**
32	$(25, 34)$	$\delta_{25} + \delta_{34} \geq 1$
44	(43)	$\delta_{43} \geq 1$
46	(39)	$\delta_{39} \geq 1$

- For the rest of elements, the faulty configuration satisfied the capacity test, and thus no information could be retrieved.

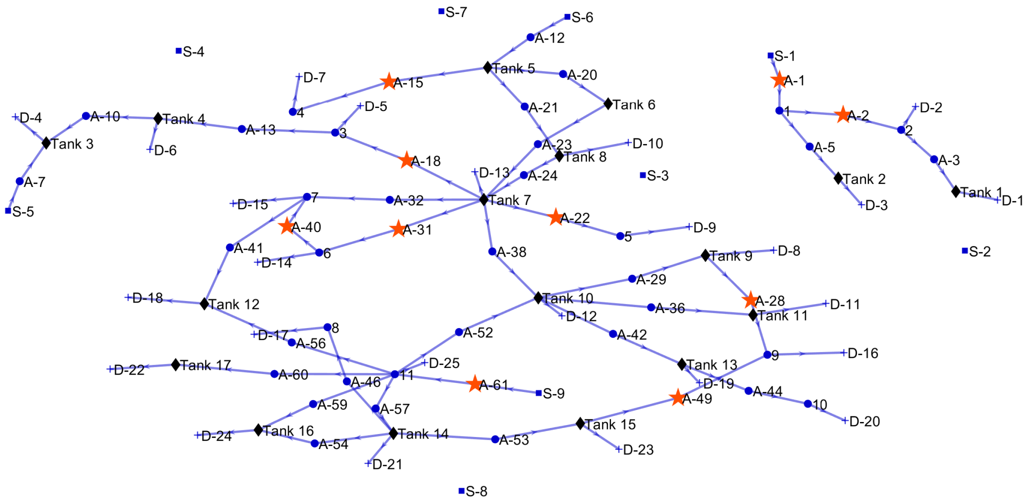


Figure 6.5: Non-recoverable elements (red nodes).

6.7.4 Admissibility criterion

Given a reparation time horizon N_f, the performance index is computed as

$$\Phi_{C_i} = \sum_{j=1}^{N_f} ||Q(x^r_{k_d+j} - x(j))||_2 + \sum_{j=0}^{N_f-1} ||Ru(j))||_2,$$

with x^r the reference model state and matrices $Q = I_{n_x}$ and $R = I_{n_u}$.

The configuration selection problem solved online (yellow block in Figure 6.2) is addressed following the lexicographic approach presented in Section A.4 of Appendix A. For that purpose, the following objectives, which are listed in order of decreasing importance, are taken into consideration:

- *Objective 1:* Minimize the number of back-up actuators used (f_1).
- *Objective 2:* Minimize the performance loss during N_f (f_2).

Previous objectives are formulated mathematically as the minimization of the following scalar-valued functions

$$\begin{cases} f_1 &= ||Q_\delta \delta||_2, \\ f_2 &= \Phi_{C_i}, \end{cases}$$

where δ is the binary vector that dictates the activation of the back-up elements and $Q_\delta = I_{n_v}$.

6.7.5 Results

Figure 6.6 plots the mean and worst-case solver times[1] for the first MIP optimization of the lexicographic approach. Note that this optimization is subject to the highest computational load, since hard constraints on the binary vector are imposed for the second iteration. The plotted data

[1] Laptop (Intel i7 1.8 GHz, 16 GB RAM) running Windows 10; optimization using Cplex 12.8 ILOG [2018].

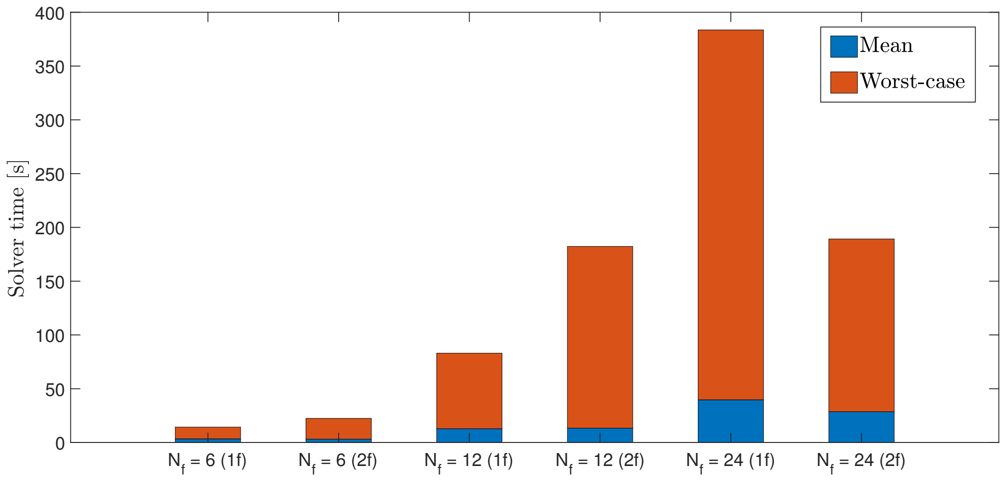


Figure 6.6: Solver time.

represents the values obtained after running 250 successful optimizations for the time horizons $N_f = \{6, 12, 24\}$, and for one single actuator fault (1f) and faults affecting two actuators (2f). It can be seen that the computation times are well below the sampling time $T_s = 1$h used by the control-oriented model of the network. Hence, in this case, the reconfiguration block would not introduce latency in the control loop.

A particular fault scenario is illustrated in Figure 6.7. This scenario simulates a fault blocking actuator $\tilde{C}_f = 44$ (closing the valve) at $k_f = 2$h, and detected at $k_d = 4$h. The red line depicts the faulty evolution of the system state in the time interval between the fault occurrence and the detection (yellow background) for the twelfth tank. Setting an admissible performance degradation of $\beta = 10$, and a reparation horizon $N_f = 24$h, the faulty configuration C_F fails the admissibility test, leading thus to run the online optimization with the inclusion of the linear constraint ($\delta_{43} \geq 1$) obtained in the offline analysis. The obtained solution advises that just opening 43 is enough to obtain an admissible solution for N_f.

In this regard, the first iteration of the lexicographic sequential solution method (see Al-

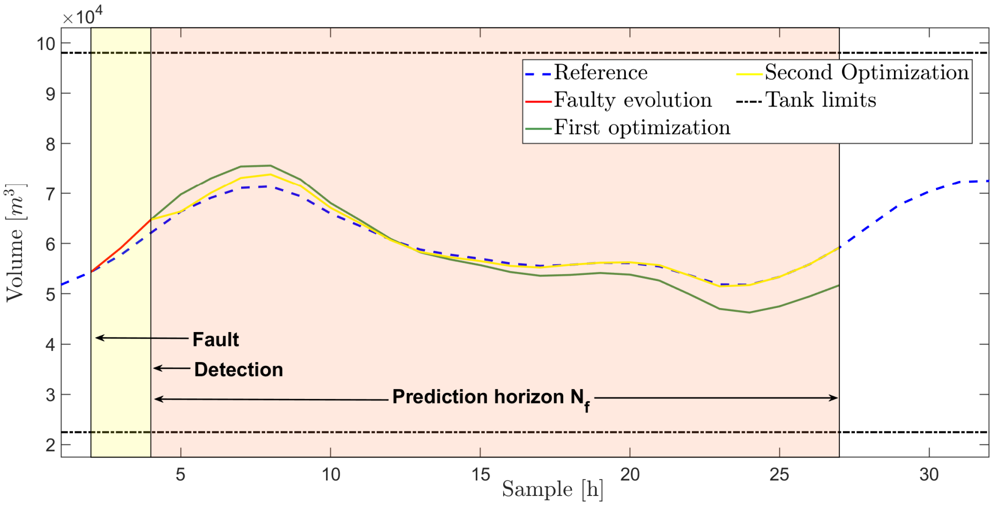


Figure 6.7: Temporal evolution - Tank 12.

gorithm A.2 in Appendix A), provides the information that the minimum number of back-up actuators such that yield an admissible performance is one. Hence, this reduces the candidate configurations used in the second iteration to just 26 candidates. Figure 6.7, shows in green the temporal evolution of the predicted trajectory obtained at $k_d = 4$h in the first iteration of the optimization, and in yellow the predicted trajectory obtained in the second optimization where the cost function f_2 that accounts for the performance loss during N_f is minimized. Note that, for this particular case it would not be necessary to run the second optimization since the first one establishes that the minimum number of back-up actuators required is one, whereas the offline tests set that the inclusion of actuator 43 is necessary.

Besides, Figure 6.8 illustrates the system reconfiguration for a double fault blocking the actuators $\tilde{C}_f = (32, 38)$, highlighted in red, at sampling time $k_f = 10$h. This case assumes an instantaneous detection, i.e., $k_d = k_f = 10$h, and an admissibility threshold of $\beta = 5$. At time 10h, the lexicographic optimization is launched in such a way that the first iteration establishes that the minimum number of back-up elements required is three. This reduces the candidate configurations from 2^{26} to the 2951 possible configurations that use a maximum of three back-up actuators. On the other hand, the solution of the second iteration suggests that the combination of three elements that minimize the performance degradation is $(30, 34, 51)$. This elements appear highlighted in green in Figure 6.8.

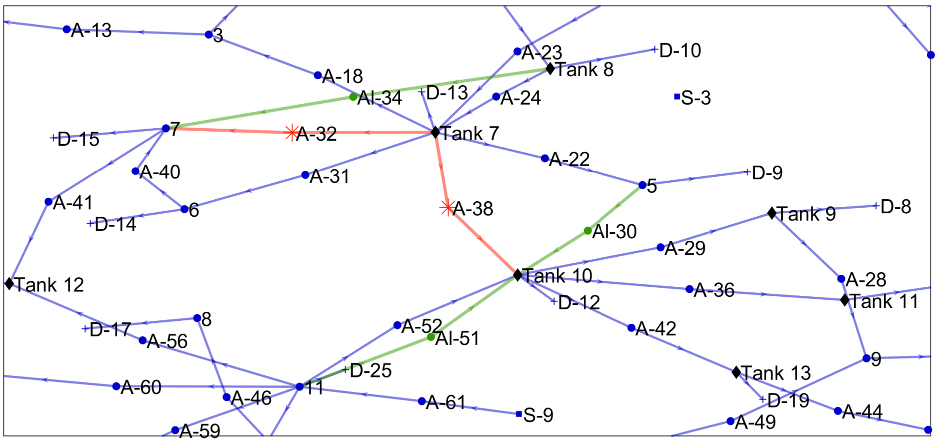


Figure 6.8: Example of system reconfiguration.

6.8 Summary

This chapter addresses the problem of system reconfiguration for systems with back-up components (i.e. hardware redundancy). After a fault occurrence, the selection of the most suitable system configuration is performed solving a mixed-integer program. Several offline necessary properties for the existence of an admissible solution were presented and the procedure for extracting information from those tests was investigated. So far, the configuration selection has been based on the predicted evolution of the system generated at fault diagnosis time. On the other hand, the following chapter studies the coupling between the reconfiguration and the control blocks, in such a way that the configuration selection must be constrained to provide certain stability guarantees.

Moreover, two natural extensions have been identified as future research directions, namely:

the consideration of system non-linearities and model uncertainties. Certainly, none of those issues is a trivial task. For the case of water networks, the non-linearities refer to the consideration of pressure, as well as flow interactions. As regards improving the robustness of the proposed approach, the use of standard robust optimization techniques, which use a local controller to compensate for the effect of possible sources of uncertainties, would generate a coupling between the selection of the actuator configuration and the design of such local controller that hinders the solution of the optimization problem. This latter issue will be also addressed in the following chapter of the thesis.

Chapter 7

Reconfiguration with back-up components - Control

This chapter addresses the robust configuration selection of flow-based networks with back-up components after a component's fault. To that end, the admissibility of each candidate configuration is split into a performance property and a stability property. On the one hand, the performance property assesses the economic cost of the best steady-state cyclic trajectory that can be attained using a robust model predictive control (MPC) policy for a specific actuator configuration. The minimal configurations that span the performance property are computed offline with a double purpose: I) filter out non-admissible configurations by formulating the online configuration selection as a sequential mixed-integer program (MIP) on the superset of each minimal configuration; II) design a local controller to compensate the uncertainty effect in the robust MPC scheme. On the other hand, the stability property is guaranteed imposing, at fault diagnosis time, the satisfaction of the constraints used by a single-layer robust MPC controller. A portion of a water transport network is used in order to validate the proposed solution.

7.1 Introduction

This chapter delves into the reconfiguration problem posed in Chapter 5. In this sense, unlike the previous chapter, the reconfiguration block is integrated in the control loop of a constrained large-scale system. Consequently, the selection of a new system configuration C_{new} must guarantee the satisfaction of the stability requirements in the control of constrained systems, namely: the closed-loop convergence towards a reachable trajectory (or equilibrium point) and the recursive feasibility in the system evolution.

In line with this part of the thesis, the present chapter focuses on the optimal operation of large-scale flow-based networks subject to uncertain periodic disturbances such as exogenous periodic demands and periodically varying power prices. Regarding the operational control of these type of systems: driven by their ability to efficiently control complex processes, a large number of solutions presented in the literature make use of MPC schemes [Kennel et al., 2012, Ocampo-Martinez et al., 2013]. In particular, motivated by the high economic costs associated with the operation of large-scale flow-based networks, the use of economic MPC strategies [Angeli and Müller, 2019] is presented as a convenient approach as evidenced by the good results obtained in the control of different types of flow-based networks [Pereira et al., 2017, Wang et al., 2017]. On this subject, it is worth mentioning that the periodic nature of the disturbances causes that, in some cases, the best way of operating these networks is the imposition of cyclic steady-state operation [Huang et al., 2011, Lee et al., 2001]. To that end, two main economic MPC architectures have been proposed: I) double-layer schemes composed by an upper layer dynamic real time optimizer (RTO) that plans the optimal system steady-state trajectory and a low-level predictive controller that tracks the previous reference [Würth et al., 2011]; II) single-layer schemes where the economic cost function is included in the computation of the control law, allowing thus to assess the economic cost during the transients [Angeli et al., 2011].

According to the previous discussion, an economic criterion is used for addressing the robust version of the configuration selection problem posed in (5.1). This robust selection problem introduces several additional challenges, specifically:

P1 The establishment of the stability guarantees in the configuration selection procedure. This follows from model inconsistencies provoked by a fault, which may cause that the explored candidate configurations are not able to reach the steady-state trajectory set for the nominal configuration. On this subject, if the problem is addressed using steady-state first-principle models [Würth et al., 2011], with the objective of selecting the configuration that yields the closest to the nominal steady-state operation, then feasibility problems may arise at the lower layer in charge of the plant's control (particularly after the transient induced by a fault). This problem is tackled using single-layer MPC schemes.

P2 The uncertainty consideration. In this regard, robust control schemes usually devise a suboptimal control policy through the *a priori* design of a local controller in charge of compensating the effect of uncertainty sources [Mayne et al., 2000, Chisci et al., 2001]. Consequently, the prior selection of the set of actuators that are used (and thus activated) in the design of a local controller, may have a huge impact on the optimality of the configuration selection.

P3 The large-scale of flow-based networks precludes an online evaluation of all the alternative configurations.

P4 The robust control of flow-based networks subject to algebraic equations that describe the static relations in the system, is still a topic under investigation.

Of special relevance for the developments presented in the present chapter is the single-layer MPC scheme proposed for tracking in Limón et al. [2008], and later adapted to the economic MPC in Limon et al. [2014]. In this approach, the system states are extended with the inclusion of a virtual system, which is forced to converge towards the best attainable tracking (or economic) objective. Besides, the capability to set the stability ingredients with respect to the virtual model (which can be updated coherently with the plant model in the configuration selection), allows to tackle (**P1**), and guarantee that, for the selected configuration, there exist a control policy that steers the system towards its best attainable steady-state trajectory. The selection of this particular single-layer control scheme responds to the fact that, despite in this chapter attention is focused on an economic performance, the same methodology holds if the control objective is to track a (possibly non reachable) reference trajectory.

The main contribution of this chapter is to address the configuration selection problem for constrained flow-based networks using robust economic MPC schemes. To that end, a configuration is deemed admissible if it satisfies a performance property and a stability property. On the one hand, the performance property is assessed offline by means of the average cost of the best attainable steady-state cyclic trajectory. In this regard, the monotonicity of the performance property is exploited for conducting an efficient search of the minimal configurations that span such property. These minimal configurations are used online with a double purpose: filter out non-admissible configurations, reducing thus the computational complexity (**P3**); and design a local controller to compensate the effect of the uncertainty (**P2**). On the other hand, the configuration selection is approached by sequentially solving an MIP in the superset of each minimal configuration. The stability property is posed by enforcing that the constraints used by a single-layer robust MPC scheme are satisfied at fault diagnosis time. Additionally, a novel formulation that includes the matrix that distributes the uncertainty in the static equations of the network as an optimization variable in the computation of the best attainable steady-state trajectory (**P4**), is proposed.

The remainder of this chapter is organised as follows: Section 7.2 presents the system under consideration and the adopted admissibility criterion. In Section 7.3, the stability of each candidate configuration is addressed, whereas Section 7.4 analyses the performance assessment of the different configurations. Then, Section 7.5 presents the proposed approach for the robust configuration selection. The considered case study is detailed in Section 7.6. Finally, Section 7.7 draws the main conclusions of the chapter.

7.2 Problem statement

This section presents the system under study, as well as introduces the admissibility criterion upon which the rest of the chapter is based. In addition, the definitions related to the partial ordering of the actuator configurations that are detailed in Section 5.3, will be used extensively throughout this chapter.

This chapter presents an extension of the flow-based network model presented in Section 6.3, by considering uncertainties in the demand flow predictions. Consequently, the model under study has the structure

$$\begin{aligned} x_{k+1} &= Ax_k + B\Sigma\bar{u}_k + B_d d_k + B_w w_k, && (7.1a)\\ 0 &= E\Sigma\bar{u}_k + E_d d_k + E_w w_k, && (7.1b) \end{aligned}$$

where $x_k \in \mathbb{R}^{n_x}$, $\bar{u}_k \in \mathbb{R}^{\bar{n}_u}$, $d_k \in \mathbb{R}^{n_d}$ and $w_k \in \mathbb{R}^{n_w}$ are the state, input, measurable and unknown disturbance vectors of the system at time $k \in \mathbb{N}$, respectively. System matrices $A, B, B_d, B_w, E, E_d, E_w$ are of suitable dimensions dictated by the network topology.

Moreover, similarly to Chapter 6, the input variables are assumed to be rearranged in such a way that $\bar{u}_k = [u_k^T, v_k^T]^T$, where $u_k \in \mathbb{R}^{n_u}$ denotes the control variables used in nominal configuration C_N (with $n_u = |C_N|$), and $v_k \in \mathbb{R}^{n_v}$ denotes the alternative control variables associated with the back-up elements in C_A (with $n_v = |C_A|$). Consequently, the input matrices (B, E) are split into the nominal (B_N, E_N) and alternative (B_A, E_A) input matrices as

$$B = \begin{bmatrix} B_N, & B_A \end{bmatrix}, \qquad E = \begin{bmatrix} E_N, & E_A \end{bmatrix},$$

whereas the configuration selection matrix Σ is rewritten as

$$\Sigma = \begin{bmatrix} I_{n_u} & 0 \\ 0 & diag(\delta) \end{bmatrix}, \tag{7.2}$$

with δ being a binary vector $\delta = [\delta_1, ..., \delta_{n_v}]^T \in \{0, 1\}^{n_v}$ that rules the selection of the corresponding back-up actuator.

Furthermore, for any time instant k, system (7.1) is considered to be subject to hard state and control constraints given by the following polytopic sets

$$\mathcal{X} = \{x_k \in \mathbb{R}^{n_x} \mid Gx_k \leq g\} \subset \mathbb{R}^{n_x}, \tag{7.3a}$$
$$\mathcal{U} = \{u_k \in \mathbb{R}^{n_u} \mid Fu_k \leq f\} \subset \mathbb{R}^{n_u}, \tag{7.3b}$$
$$\mathcal{V} = \{v_k \in \mathbb{R}^{n_v} \mid Hv_k \leq h\} \subset \mathbb{R}^{n_v}, \tag{7.3c}$$

where G, F, H and g, f, h are real matrices and vectors with dimensions consistent with the number of state and input constraints. Besides, the number of equations n_e in (7.1b) satisfies $n_e < n_u$. In addition, for all k, the unknown disturbance satisfies $w_k \in \mathcal{W}$ with

$$\mathcal{W} = \langle 0, I_{n_w} \rangle \subset \mathbb{R}^{n_w}. \tag{7.4}$$

Remark 7.1. Note that, as long as w_k is zonotopically bounded, a zero-centered unitary box representation like (7.4) can be obtained by performing: I) a change of coordinates that shifts the uncertainty center to the zero; II) a coherent modification of the matrices (B_w, E_w).

In the nominal configuration C_N (i.e., with $\delta = 0$), the system dynamics are governed by

$$x_{k+1} = Ax_k + B_N u_k + B_d d_k + B_w w_k, \tag{7.5a}$$
$$0 = E_N u_k + E_d d_k + E_w w_k. \tag{7.5b}$$

Assumption 7.1. The states in x_k are observable at time instant k.

Assumption 7.2. The predicted disturbance signal d_k presents a periodic behaviour with known period T, i.e., $d_k = d_{k+T}$.

Assumption 7.3. The uncertainty affecting the static equations (7.1b) is related with the error on the flow consumption prediction d_k, and thus can be measured at current time instant. That is, a vector $\hat{w}_k \in \mathcal{W} : E_w(w_k - \hat{w}_k) = 0$ can be computed at k.

Assumption 7.2 reflects the expected cyclic behaviour in the flow consumption. Besides, possible uncertainties in the forecast can be embedded into the uncertain variable w_k. Observe that, by means of this assumption, the system expressed in (7.1) becomes a periodic system [Bittanti and Colaneri, 2009]. On the other hand, Assumption 7.3 imposes that the current perturbation affecting the static nodes is known at current k, but unknown at future samples. Note that Assumption 7.3 does not involve the disturbances affecting the system dynamics, where process disturbances can not be known at current time instant.

Furthermore, the same fault scenario and assumption concerning the capabilities of a fault diagnosis block that have been introduced in Chapter 6, are followed here. These are recalled below in order to complement the developments presented in this chapter.

Assumption 7.4. An FDI block is able to detect and isolate the presence of actuator faults. The faulty components are immediately shut down.

Accordingly, an FDI block is able to detect an isolate that a fault occurs in the set of actuators $\tilde{C}_f$. In particular, this chapter focuses on the case where the resulting faulty configuration $C_F = C_N \setminus \tilde{C}_f \in \mathbb{S}(C_N)$ is not admissible, that is, $\bar{\mathcal{A}}(C_F)$.

The temporal evolution of the events related with a fault occurrence defines the following sequence:

1. an actuator fault occurs at time k_f;
2. an FDI block detects and isolates the set of faulty actuators $\tilde{C}_f \subseteq C_N$ at $k_d \geq k_f$, modifying the input matrices to (B_F, E_F).

7.2.1 Performance assessment

Hereafter, it is considered that the closed-loop performance of the system with the different actuator configurations is assessed by means of an economic time varying stage cost function $l : \mathbb{N} \times \mathcal{X} \times \mathcal{U} \times \mathcal{V} \to \mathbb{R}$. As regards the cost function, the following assumption is introduced.

Assumption 7.5. The stage cost function $l(\cdot)$ is assumed to be positive; convex in (x, u, v) for all k; and periodic, i.e., $l(k, x, u, v) = l(k + T, x, u, v)$.

Note that the first two conditions on Assumption 7.5 are introduced for convergence issues, whereas the periodicity constraint in the stage function typically follows from the periodic patterns in electricity pricing.

Therefore, the system performance is assessed as the average of the stage cost function obtained by the closed-loop trajectories. For a configuration C_i, this can be written as

$$\mathcal{L}^{\infty}_{C_i}(x_0, \boldsymbol{u}_\infty, \boldsymbol{v}_\infty) = \lim_{m \to \infty} \frac{1}{m} \sum_{k=0}^{m-1} l(k, x_k, u_k, v_k), \tag{7.6}$$

where x_0 is the initial state and $(\boldsymbol{u}_\infty, \boldsymbol{v}_\infty)$ are the set of corresponding closed-loop input trajectories. In this regard, due to periodic nature of the systems under study, and under the implicit assumption of uniqueness of the solution, it is well-known that the optimal trajectories of the system without unknown disturbances can be obtained by solving a finite horizon open-loop problem that optimizes the average cost over a single period T [Limon et al., 2014, Theorem 1]. For the sake of simplified notation, from this point onward the optimal T-period average stage cost value that can be achieved by the system operating on configuration C_i, is termed as $\mathcal{L}^T_{C_i}$.

Remark 7.2. Note that the computation of the optimal trajectory should be performed by taking into consideration the effect of the uncertainty sources in the cost function, while guaranteeing a robust constraint satisfaction. However, due to the computational complexity of the approaches that consider the uncertainty in the stage cost predictions [Bayer et al., 2016a,b], in the remainder of the chapter the approach considered in Pereira et al. [2017], Wang et al. [2019] will be followed, and only the stage cost of the nominal system is minimized.

Remark 7.3. Some economic MPC related works consider an extra input parameter $p \in \mathbb{R}^{n_p}$ in $l(\cdot)$ that accounts for variations in the parameters of the stage function, e.g., modification of the prices by the electric utility. For simplicity, this has not been taken into account in the remainder of the chapter.

7.2.2 Admissibility

The following definition is introduced concerning the admissibility of an arbitrary configuration.

Definition 7.1 (Admissibility)**.** Given system (7.1) with configuration C_i imposed through matrix Σ. Then, the admissibility of C_i at time k^* is expressed as

$$\mathcal{A}_{k^*}(C_i) = \mathcal{A}^s_{k^*}(C_i) \wedge \mathcal{A}^p(C_i), \tag{7.7}$$

where

i. **Stability property:** $\mathcal{A}^s_{k*}(C_i)$ if starting at x_{k^*}, the closed-loop system with configuration C_i converges asymptotically towards a neighbourhood of the cyclic steady-state trajectory with cost $\mathcal{L}^T_{C_i}$, while guaranteeing the robust constraint satisfaction for all $k \geq k^*$.

ii. **Performance property:** $\mathcal{A}^p(C_i)$ if $\mathcal{L}^T_{C_i} \leq \beta \mathcal{L}^T_{C_N}$, for a given $\beta \in \mathbb{R}_+$.

Note that in the previous definition of admissibility, it is implicitly assumed that there exist a control law for which the nominal configuration C_N converges to a neighbourhood of the cyclic trajectory that yields the cost $\mathcal{L}^T_{C_N}$. In this regard, the following sections are devoted to the analysis of the two conditions that make up the admissibility property. Particularly, Section 7.3 present the robust control law that is used in order to satisfy the stability property $\mathcal{A}^s_k(\cdot)$, whereas Section 7.4 focuses on the analysis of the performance property $\mathcal{A}^p(\cdot)$.

Remark 7.4. Under an MPC control policy, $\mathcal{A}^s_k(C_i)$ is equivalent to verify $x_k \in \mathbb{X}_{C_i}$, where $\mathbb{X}_{C_i}$ is the region of attraction of a robust MPC controller that uses configuration C_i. On this subject, the consideration of large-scale systems precludes the explicit computation of $\mathbb{X}_{C_i}$, and thus $\mathcal{A}^s_k(C_i)$ must be evaluated online.

7.3 Stability property

This section analyses the robust control of a constrained periodic system like (7.1). Motivated by the discussion performed in Section 7.1, the satisfaction of the stability property $\mathcal{A}^s_k(\cdot)$ is addressed using a single-layer economic MPC control scheme like the one proposed in Limon et al. [2014], Pereira et al. [2016b].

7.3.1 Robust planner

Firstly, the optimal T-periodic average stage cost for an N-horizon MPC controller is computed. In this regard, the so-called *robust planner* optimization problem for a system with configuration C_j imposed through matrix Σ, is posed as

$$\begin{aligned}
&\min_{x_0,\bar{\boldsymbol{u}}_T} \ \mathcal{L}^T_{C_j} = \frac{1}{T}\sum_{i=0}^{T-1} l(k+i, x(i), \bar{u}(i)), && \\
&\text{s.t. } x(i+1) = Ax(i) + B\Sigma\bar{u}(i) + B_d d_i, && \forall i \in \mathbb{I}_{T-1}, \\
&0 = E\Sigma\bar{u}(i) + E_d d_i, && \forall i \in \mathbb{I}_{T-1}, \\
&x(T) = x(0) = x_0, && \\
&x(i) \in \mathcal{X}_{C_j}(N), && \forall i \in \mathbb{I}_T, \\
&\bar{u}(i) \in \mathcal{U}_{C_j}(N) \times \mathcal{V}_{C_j}(N), && \forall i \in \mathbb{I}_{T-1},
\end{aligned} \tag{7.8}$$

where the optimal points $x^o_{C_j}, \bar{\boldsymbol{u}}^o_{C_j}$ of (7.8) are denoted as *robust planner trajectories*. Besides, $\mathcal{X}_{C_j}(N)$, $\mathcal{U}_{C_j}(N)$, $\mathcal{V}_{C_j}(N)$ represent a tightened set of state and input constraints for configuration C_j, whose computation will be detailed later.

It is worth mentioning that, due to the periodicity of the known disturbance d_k and the stage function $l(\cdot)$, the obtained robust planner trajectories are independent on the instant in which the optimization problem is formulated. This time independence will be further exploited in Section 7.4 for the offline assessment of the admissible configurations.

7.3.2 Single-layer robust MPC

This section presents the design of an MPC control law that ensures the asymptotic convergence to a neighbourhood of the robust planner trajectories, while guaranteeing the robust constraint satisfaction. For the sake of simplified notation, the developments presented below have been outlined for the nominal configuration C_N. Nevertheless, the same formulation applies for any arbitrary configuration C_i through the introduction of the selection matrix Σ and the consideration of the overall control variables $\bar{u}_k$.

At this point, the linearity of the system is exploited to separate the effect of the uncertainty in the predictions. To that end, let $\tilde{x}_k \in \mathbb{R}^{n_x}$ term the nominal predictions (i.e., without considering the disturbances). Then, the error $e_k = x_k - \tilde{x}_k \in \mathbb{R}^{n_x}$ evolves according to

$$e_{k+1} = Ae_k + B_N e_k^u + B_w w_k, \tag{7.9a}$$

$$0 = E_N e_k^u + E_w w_k, \tag{7.9b}$$

where $e_k^u = u_k - \tilde{u}_k \in \mathbb{R}^{n_u}$ and $\tilde{u}_k$ is the nominal input that yields $\tilde{x}_k$.

Following the proposal of Pereira et al. [2017], an auxiliary input $h_k \in \mathbb{R}^{n_u - n_e}$ is introduced in order to guarantee the satisfaction of (7.5b) for any possible $w_k \in \mathcal{W}$. The auxiliary input h_k is derived from the explicit solution of (7.9b), and thus the input difference e_k^u can rewritten as

$$e_k^u = M_w w_k + M_v h_k, \tag{7.10}$$

for some matrices M_v and M_w. Hence, the robust satisfaction of (7.9b) can be guaranteed by designing the matrices M_v and M_w such that satisfy (the computation of M_v and M_w is addressed in Section 7.3.3)

$$E_N M_v = 0, \tag{7.11a}$$

$$E_w + E_N M_w = 0. \tag{7.11b}$$

From (7.10) and (7.11), system (7.9) can rewritten as the equivalent model

$$e_{k+1} = Ae_k + \hat{B}h_k + \hat{B}_w w_k, \tag{7.12}$$

where $\hat{B} = B_N M_v$ and $\hat{B}_w = B_w + B_N M_w$.

In order to attenuate the uncertainty propagation characterized by (7.12), the standard approach consists on obtaining a suboptimal solution to the control problem through the *a priori* design of a linear control law of the form $h_k = Ke_k$, with matrix K designed such that $A + \hat{B}K$ is an asymptotically stable matrix [Mayne et al., 2005, Alvarado et al., 2010].

At this point, the single-layer robust MPC policy is presented. Following the schemes of Limon et al. [2014], Pereira et al. [2017], the following cost function is introduced

$$V(x_k, x_0^v, \boldsymbol{u}_T^v, \boldsymbol{u}_N, \boldsymbol{d}_{k:k+N}) = V_t(x_k, x_0^v, \boldsymbol{u}_T^v, \boldsymbol{u}_N, \boldsymbol{d}_{k:k+N}) + V_p(x_0^v, \boldsymbol{u}_v^T, \boldsymbol{d}_{k:k+N}), \tag{7.13}$$

where $\boldsymbol{u}_N$ is the N-sequence of control input variables; $\boldsymbol{d}_{k:k+N}$ the sequence of flow demand predictions; while x_0^v and $\boldsymbol{u}_T^v$ are decision variables associated with the introduction of an artificial reference denoted by the superscript ^{v}.

On the one hand, the term $V_t(\cdot)$ penalizes the error between the open-loop trajectories and the artificial reference. On the other hand, $V_p(\cdot)$ penalizes the average stage cost of such artificial reference. These objectives are formulated as

$$V_t(\cdot) = \sum_{i=0}^{N-1} \|x(i) - x^v(i)\|_Q^2 + \|u(i) - u^v(i)\|_R^2, \tag{7.14a}$$

$$V_p(\cdot) = \frac{1}{T} \sum_{i=0}^{T-1} l(k+i, x^v(i), u^v(i)), \tag{7.14b}$$

with $Q = Q^T \succ 0$, $R = R^T \succ 0$ and $N \leq T$.

In consequence, the optimal nominal trajectories of the robust economic MPC at time k are obtained from the solution of the following finite horizon control problem

$$\begin{aligned}
&\min_{x_0^v, \boldsymbol{u}_T^v, \boldsymbol{u}_N} \; V(x_k, x_0^v, \boldsymbol{u}_T^v, \boldsymbol{u}_N, \boldsymbol{d}_{k:k+N}), && && (7.15a)\\
&\text{s.t.} \;\; x(0) = x_k, && && (7.15b)\\
&\quad x(i+1) = Ax(i) + B_N u(i) + B_d d_{k+i}, && \forall i \in \mathbb{I}_{N-1}, && (7.15c)\\
&\quad 0 = E_N u(i) + E_d d_{k+i}, && \forall i \in \mathbb{I}_{N-1}, && (7.15d)\\
&\quad x(N) = x^v(N), && && (7.15e)\\
&\quad x^v(i+1) = Ax^v(i) + B_N u^v(i) + B_d d_{k+i}, && \forall i \in \mathbb{I}_{T-1}, && (7.15f)\\
&\quad 0 = E_N u^v(i) + E_d d_{k+i}, && \forall i \in \mathbb{I}_{T-1}, && (7.15g)\\
&\quad x(i) \in \mathcal{X}(i), && \forall i \in \mathbb{I}_{N}, && (7.15h)\\
&\quad x^v(i) \in \mathcal{X}(N), && \forall i \in \mathbb{I}_{T}, && (7.15i)\\
&\quad u(i) \in \mathcal{U}(i), && \forall i \in \mathbb{I}_{N-1}, && (7.15j)\\
&\quad u^v(i) \in \mathcal{U}(N), && \forall i \in \mathbb{I}_{T-1}, && (7.15k)\\
&\quad x^v(T) = x^v(0) = x_0^v, && && (7.15l)
\end{aligned}$$

with the tightened set of constraints

$$\begin{aligned}
\mathcal{X}(0) &= \mathcal{X}, & \mathcal{U}(0) &= \mathcal{U} \ominus M_w \mathcal{W},\\
\mathcal{X}(i) &= \mathcal{X} \ominus \mathcal{R}(i), & \mathcal{U}(i) &= \mathcal{U} \ominus M_w \mathcal{W} \ominus M_v K \mathcal{R}(i),\\
\mathcal{R}(i) &= \bigoplus_{0}^{i-1} \mathcal{Q}(i), & \mathcal{Q}(j) &= (A + \hat{B}K)^j \hat{B}_w \mathcal{W}.
\end{aligned} \tag{7.16}$$

Below, the following assumption regarding the local controller K design are introduced.

Assumption 7.6. The closed-loop matrix $A + \hat{B}K$ satisfies $(A + \hat{B}K)^{N-1} \hat{B}_w w = 0, \; \forall w \in \mathcal{W}$; the sets $\mathcal{X}(i)$ and $\mathcal{U}(i)$ are non-empty for $i \in \mathbb{I}_N$.

Denoting as $u^*(0|k)$ the first optimum value of (7.15) computed at k, then, from [Pereira et al., 2017, Theorem 1], it follows that the system (7.5) controlled by means of the control law

$$u_k = u^*(0|k) + M_w \hat{w}_k,$$

is recursively feasible and converges asymptotically to a neighbourhood of the robust planner trajectories that yield the cost $\mathcal{L}_{C_N}^T$.

Remark 7.5. Through the dead-beat controller introduced in Assumption 7.6, Eq. (7.15e) is a terminal ingredient that allows to ensure the closed-loop stability without the need of computing a robust terminal invariant set [Pereira et al., 2016b]. Note that this set computation may be intractable for large-scale systems. Assumption 7.6 can be relaxed by imposing that $\max_{w\in\mathcal{W}} \|(A+\hat{B}K)^{N-1}\hat{B}_w w\|_\infty = \|(A+\hat{B}K)^{N-1}\hat{B}_w\|_\infty$ is below a pre-specified threshold that relates with the numerical precision of the computer (the unitary zonotope constraint (7.4) has been used in the derivation of previous equality).

7.3.3 Parametrized solution of the robust planner

Regarding the constraints introduced in (7.11): matrix M_v can be designed as an orthonormal basis to the null space of the $n_e \times n_u$ matrix E_N (with $n_e < n_u$). However, on the other hand, an inappropriate selection of matrix M_w (which is in charge of distributing among the different actuators the compensations of the uncertainty in the static nodes (see (7.16))), may have a harmful effect on the optimal operation of the system or even generate an infeasible solution region. In this regard, in Pereira et al. [2017], Wang et al. [2018a] the Moore-Penrose pseudo-inverse is used in order to compute matrix M_w, whereas in Grosso et al. [2014], Nassourou et al. [2020] the matrix M_w results from an actuators permutation selected by the user. Nevertheless, none of the above designs takes into account the state and input constraints.

Aiming at addressing this problem, the fact that matrix M_w is affine to the tube of uncertain trajectories expressed by (7.12) is exploited and its design is now included as an optimization variable in the robust planner trajectories computation.

At this point, Lemma 1 in Scott et al. [2014] is rewritten as:

Lemma 7.1 (P-difference)**.** Given the zonotope $Z = \langle c, H\rangle \subset \mathbb{R}^n$, with $c \in \mathbb{R}^n$ and $H \in \mathbb{R}^{n\times z}$, and the polyhedron $S = \{x \in \mathbb{R}^n : Lx \leq l\} \subseteq \mathbb{R}^n$, with $l \in \mathbb{R}^m$ and $L \in \mathbb{R}^{m\times n}$. Then, $S \ominus Z = \{x \in \mathbb{R}^n : Lx \leq l - Lc - |LH|\mathbf{1}_z\}$.

Therefore, the new robust planner formulation is introduced in the following mathematical proposition.

Proposition 7.1. Under Assumption 7.6, the convex optimization problem (7.17) is a robust planner for configuration C_N, where M_w is an optimization variable.

$$\min_{M_w, x_0, \boldsymbol{u}_T, \Omega_x, \Omega_u} \ \mathcal{L}^T_{C_N} = \frac{1}{T}\sum_{i=0}^{T-1} l(k+i, x(i), u(i)),$$

$$\text{s. t.} \quad x(i+1) = Ax(i) + B_N u(i) + B_d d_i, \qquad \forall i \in \mathbb{I}_{T-1} \tag{7.17a}$$

$$0 = E_N u(i) + E_d d_i, \qquad \forall i \in \mathbb{I}_{T-1} \tag{7.17b}$$

$$x(T) = x(0) = x_0, \tag{7.17c}$$

$$E_N M_w = -E_w, \tag{7.17d}$$

$$Gx(i) \leq g - \Gamma - \Omega_x \mathbf{1}_{Nn_w}, \qquad \forall i \in \mathbb{I}_T \tag{7.17e}$$

$$Fu(i) \leq f - \Delta - \Omega_u \mathbf{1}_{Nn_w}, \qquad \forall i \in \mathbb{I}_{T-1} \tag{7.17f}$$

$$\Lambda(I_N \otimes M_w) \leq \Omega_x, \quad -\Lambda(I_N \otimes M_w) \leq \Omega_x, \tag{7.17g}$$

$$\Theta(I_N \otimes M_w) \leq \Omega_u, \quad -\Theta(I_N \otimes M_w) \leq \Omega_u, \tag{7.17h}$$

with

$$\begin{aligned}
&\Gamma = |GH_a(N)|\mathbf{1}_{Nn_w}, && \Delta = |FM_vKH_a(N)|\mathbf{1}_{Nn_w},\\
&\Lambda = GH_b(N), && \Theta = F(\tilde{I} + M_vKH_b(N)),\\
&H_a(N) = \left[(A+\hat{B}K)^{N-1}B_w \ \dots \ B_w\right], && H_b(N) = \left[(A+\hat{B}K)^{N-1}B \ \dots \ B\right],\\
&\tilde{I} = [I_{n_u}\, 0_{n_u\times n_u(N-1)}].
\end{aligned}$$

Proof. Starting from the robust planner (7.8) for the nominal configuration C_N. Matrix M_w affects the sets

$$\mathcal{X}(N) = \mathcal{X} \ominus \mathcal{R}(N), \qquad \mathcal{U}(N) = \mathcal{U} \ominus \mathcal{R}_u(N), \tag{7.18}$$

with $\mathcal{R}(N)$ and $\mathcal{R}_u(N)$ the N^{th} iteration of

$$\begin{aligned}
&\mathcal{R}(i) = \bigoplus_{0}^{i-1} \mathcal{Q}(i), \qquad \mathcal{Q}(j) = (A+\hat{B}K)^j\hat{B}_w\mathcal{W},\\
&\mathcal{R}_u(i) = M_w\mathcal{W} \oplus M_vK\mathcal{R}(i).
\end{aligned}$$

By recalling that $\mathcal{W}$ is a unitary zonotope and that $\hat{B}_w = B_w + BM_w$, it follows that $\mathcal{R}(i)$ and $\mathcal{R}_u(i)$ are also zonotopic sets which can be rewritten as

$$\begin{aligned}
\mathcal{R}(i) &= \langle 0, H_a(i) + H_b(i)(I_i \otimes M_w)\rangle,\\
\mathcal{R}_u(i) &= \langle 0, [M_w \ \ M_vK\big(H_a(i) + H_b(i)(I_i \otimes M_w)\big)]\rangle,
\end{aligned} \tag{7.19}$$

with $H_a(i)$ and $H_b(i)$ the i^{th} elements of the recursion

$$\begin{aligned}
&H_a(i+1) = [(A+\hat{B}K)H_a(i),\ B_w], && H_a(0) = 0,\\
&H_b(i+1) = [(A+\hat{B}K)H_b(i),\ B], && H_b(0) = 0.
\end{aligned}$$

Therefore, from $\mathcal{X}$ and $\mathcal{U}$ in (7.3) and Lemma 7.1, the sets in (7.18) are rewritten as

$$\begin{aligned}
\mathcal{X}(N) &= \{x_k \in \mathbb{R}^{n_x} : Gx_k \leq \tilde{g}\},\\
\mathcal{U}(N) &= \{u_k \in \mathbb{R}^{n_u} : Fu_k \leq \tilde{f}\},
\end{aligned} \tag{7.20}$$

where

$$\begin{aligned}
\tilde{g} &= g - |GH_a(N)|\mathbf{1}_{Nn_w} - |GH_b(N)(I_N \otimes M_w)|\mathbf{1}_{Nn_w},\\
\tilde{f} &= f - |FM_vKH_a(N)|\mathbf{1}_{Nn_w} - |F(\tilde{I} + M_vKH_b(N))(I_N \otimes M_w)|\mathbf{1}_{Nn_w},
\end{aligned} \tag{7.21}$$

and $\tilde{I} = [I_{n_u}\ 0_{n_u\times n_u(N-1)}]$.

In addition, (7.21) can be reformulated as linear constraints in M_w by bounding the absolute value of the matrices from above [Lofberg, 2003]. To this end, the variable matrices Ω_x and Ω_u are introduced as

$$\begin{aligned}
\tilde{g} &= g - |GH_a(N)|\mathbf{1}_{Nn_w} - \Omega_x\mathbf{1}_{Nn_w},\\
\tilde{f} &= f - |FM_vKH_a(N)|\mathbf{1}_{Nn_w} - \Omega_u\mathbf{1}_{Nn_w},
\end{aligned} \tag{7.22}$$

altogether with the set of constraints

$$\begin{aligned}
GH_b(N)(I_N \otimes M_w) &\leq \Omega_x,\\
-GH_b(N)(I_N \otimes M_w) &\leq \Omega_x,\\
F(\tilde{I} + M_vKH_b(N))(I_N \otimes M_w) &\leq \Omega_u,\\
-F(\tilde{I} + M_vKH_b(N))(I_N \otimes M_w) &\leq \Omega_u.
\end{aligned} \tag{7.23}$$

Hence, by means of (7.20)-(7.23) and imposing the satisfaction of (7.11b), then M_w can be set as an optimization variable in the robust-planner for the nominal configuration while preserving the convexity of the optimization problem. □

It must be pointed out that in single-layer economic MPC controllers [Huang et al., 2011, Limon et al., 2014], the robust planner is only introduced in order to proof stability properties, and thus not required to be solved. Conversely, here, the optimization (7.17) is solved with a double purpose: I) computing M_w; II) computing the optimal average stage cost $\mathcal{L}^T_{C_i}$ that can be attained by a specific system configuration. Observe that (7.17) is solved offline, and therefore the computational complexity added with the parametrization of M_w is not a problem.

7.4 Performance property

The convergence of the closed-loop system towards a cyclic steady-state trajectory motivates the assessment of the performance of a configuration C_i based on the average state cost of such steady-state trajectory. Besides, thanks to the time independence of the trajectories retrieved from the robust planner, that information can be used to filter out the configurations C_i that yield a non-admissible steady-state behaviour.

Note that the average stage cost $\mathcal{L}^T_{C_i}$ depends on the set of tightened constraints $\mathcal{X}_{C_i}(N)$, $\mathcal{U}_{C_i}(N)$, $\mathcal{V}_{C_i}(N)$ computed for a local controller. In this regard, consider two configurations C_i and C_j such that $C_i \subset C_j$, and denote as $\mathcal{L}^T_{C_j|C_i}$ the average stage cost obtained by solving the robust planner optimization for C_j and the set of tightened constraints computed using C_i.

Lemma 7.2. The performance property $\mathcal{A}^p(\cdot)$ constitutes a bottom-up monotonous property, that is,

$$\mathcal{A}^p(C_i) \implies \mathcal{A}^p(C_j), \ \forall C_j \in \mathbb{P}(C_i).$$

Proof. The proof follows from the fact that, for any $C_j \supset C_i$, the robust planner optimization can be formulated in such a way that yields the cost $\mathcal{L}^T_{C_j|C_i}$. On the other hand, the robust planner optimization for C_i can be posed similarly to the one that yields $\mathcal{L}^T_{C_j|C_i}$ plus a constraints that sets to zero the elements in $C_j \setminus C_i$. Therefore, from optimality it follows that $\mathcal{L}^T_{C_i} \geq \mathcal{L}^T_{C_j|C_i}$, and thus $\mathcal{A}^p(C_i) \implies \mathcal{A}^p(C_j)$. □

It is worth mentioning that the satisfaction of $\mathcal{A}^p$ subsumes the satisfaction of structural properties like: connectivity ($\mathcal{P}_1$), controllability ($\mathcal{P}_2$), capacity ($\mathcal{P}_3$), etc., which are necessary for the satisfaction of $\mathcal{A}^p$. That is,

$$S_p(\mathcal{A}^p) \subseteq S_p(\mathcal{P}_1) \cap \ldots \cap S_p(\mathcal{P}_l),$$

and thus $\mathcal{A}^p(\cdot)$ stablish a tighter filter than structural properties.

7.4.1 Search for minimal configurations

Below, a new algorithm for the search of minimal elements in a subset of configurations satisfying a monotonic property is presented. In particular, given a fault in $\tilde{C}_f$, this algorithm is used for the offline computation of the set of minimal candidate configurations that satisfy the performance property in Definition 7.1, that is,

$$\Theta_{\tilde{C}_f} = m(\mathcal{S}_p(\mathcal{A}^p) \cap \mathbb{P}_{\tilde{C}_f}(C_F)),$$

this set of minimal configurations plays a key role in the configuration selection addressed in Section 7.5.

Even though the search for minimal configurations is run offline. The assessment of complex properties like $\mathcal{A}^p$ can be computationally demanding, specially if the computation of $\Theta_{\tilde{C}_f}$ requires the evaluation of a large number of configurations. Furthermore, non-structural properties may be subject to seasonal variations in their parameters that impose a recalculation of the minimal set, e.g. monthly or seasonal variations on the parameters of the stage function.

On this subject, Algorithm 7.1 presents a depth-first alike strategy in order to exploit the following facts:

- If $\mathcal{A}^p(C_i)$, then all $C_j \in \mathbb{P}(C_i)$ can be removed from the search since, by definition, they can not be minimum configurations.
- If $\bar{\mathcal{A}}^p(C_i)$, then all $C_j \in \mathbb{S}(C_i)$ can be removed from the search since, from the bottom-up monotonicity of $\mathcal{A}^p(\cdot)$, if follows that they cannot satisfy it.

The input of the algorithm is the set $\Lambda = [C_1, ..., C_{2^{n_v}-1}]$ of $2^{n_v} - 1$ candidate configurations that have been sorted following a cardinality ordering (i.e., $|C_i| \geq |C_{i+1}|$). In addition, the operator $\bar{\mathbb{S}}(C|\Lambda)$ used in line 7 of Algorithm 7.1, returns the first highest cardinality strict successor of C in Λ, that is,

$$\bar{\mathbb{S}}(C|\Lambda) = \Big\{ C_i \subset C : \; i \leq j, \; \forall (C_i, C_j) \in \big(\mathbb{S}(C) \setminus C) \cap \Lambda\big) \Big\}.$$

7.5 Solution of the reconfiguration problem

This section aims at solving the reconfiguration problem (5.1) subject to the admissibility criterion presented in Definition 7.1. On this subject, the satisfaction of the robust stability property $\mathcal{A}_k^s(\cdot)$ is addressed using the robust MPC scheme presented in Section 7.3. Nevertheless, these robust control schemes are designed as a two-step procedure with the following causal relationship among them:

1. An *a priori* state feedback control law is designed for compensating the effect of the uncertainty. This is done by means of configuration C_i.
2. The evolution of the nominal (i.e., no uncertainty) trajectory is optimized subject to a tightened set of constraints. This is done for a configuration $C_j \supseteq C_i$.

Consequently, the *a priori* selection of the actuator set used by the local controller may have a high impact on the optimality of the solution of (5.1). In order to highlight this causal dependence, hereafter, the stability property is characterized by $\mathcal{A}_k^s(C_j|C_i)$, where C_i is the

Algorithm 7.1 Search for minimal configurations over a BUM property.

```
Input: $\Lambda = [C_1, ..., C_{n_a-1}]$
Output: $\Theta$
Initialize the empty set $\Theta$
if $\bar{\mathcal{A}}^p(C_1)$ then
    Stop (No solution)
end if
$C_x = \Lambda(1)$
while $\Lambda \neq \emptyset$ do
    $C_y = \bar{\mathbb{S}}(C_x|\Lambda)$
    if $C_y = \emptyset$ then
        $\Theta \leftarrow C_x$ ($C_x$ is minimal)
        Remove $\mathbb{P}(C_x)$ from $\Lambda$
        $C_x = \Lambda(1)$
    else
        if $\mathcal{A}^p(C_y)$ then
            $C_x = C_y$
        else
            Remove $\mathbb{S}(C_y)$ from $\Lambda$.
        end if
    end if
end while
```

configuration used for designing the local control law, and $C_j \supseteq C_i$ is the configuration used in the simulation of the nominal evolution.

At this point, solving (5.1) would imply to assess the stability property of each candidate configuration using the same configuration for both: the local controller and the nominal predictions, that is, such that $\mathcal{A}_k^s(C_i|C_i)$. Note that, for large-scale systems the online assessment of this stability property becomes intractable. As a consequence, the configuration selection (5.1) is approximated basing the uncertainty compensation on the minimal configurations (termed as C_i^m) that yield an admissible performance.

7.5.1 Solution based on minimal configurations

The idea is exemplified in Figure 7.1, where the lattice of candidate configurations $\mathbb{P}_{\tilde{C}_f}(C_F)$ after a fault is schematically represented. Therefore:

- *Offline*: the performance property (white nodes) is used to prune the candidate configurations. However, by means of Algorithm 7.1 is not needed to assess $\mathcal{A}^p(\cdot)$ in all the nodes, but look for the set of minimal configurations $C_i^m \in \Theta_{C_F}$ since from Definition 5.6, it follows that
$$\mathcal{S}_p(\mathcal{A}^p) = \mathbb{P}(C_1^m) \cap ... \cap \mathbb{P}(C_l^m),$$
where l is the number of elements in Θ_{C_F}.

- *Online*: Instead of designing a local controller and evaluating the stability property for each white node. A local controller is designed only for the minimal configurations and (5.1) is solved locally in the predecessors of each minimal configuration.

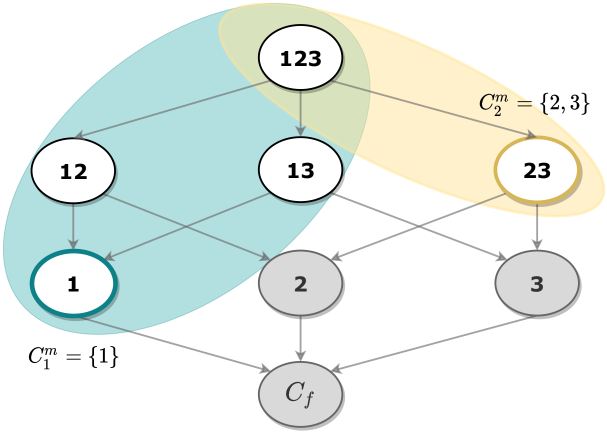


Figure 7.1: Lattice of $\mathbb{P}_{\tilde{C}_f}(C_F)$: $\mathcal{S}_p(\mathcal{A}^p)$ white nodes; $\mathbb{P}(C_1^m)$ blue ellipse; $\mathbb{P}(C_2^m)$ yellow ellipse.

Hence, if the configuration $C_j \in \mathbb{P}(C_i^m)$ is such that $\mathcal{A}_k^s(C_j|C_i^m)$, then, by means of the MPC controller presented in Section 7.3, the convergence of the closed-loop trajectories to a neighbourhood of the robust planner trajectories that yields and average cost $\mathcal{L}_{C_j|C_i^m}^T$ is ensured. Besides, since $\mathcal{A}^p(C_i^m)$, from Lemma 7.2 it follows that $\mathcal{A}^p(C_j)$, and thus C_j is admissible according to Definition 7.1.

7.5.2 MIP optimization

The configuration selection (5.1) in the predecessors of a minimal configuration C_i^m can be posed as an MIP. In this regard, let $\mathcal{X}_{C_i^m}(j), \mathcal{U}_{C_i^m}(j), \mathcal{V}_{C_i^m}(j)$ represent the tightened set of state and input constraints obtained for a local controller designed using C_i^m. Moreover, let $\delta_{C_i^m}$ represent the binary vector with the elements in C_i^m activated (and thus not considered as variables in the optimization problem). Therefore the selection of a new configuration $C_{new}^{\phi_i}$ restricted to the set $\mathbb{P}(C_i^m)$ can be posed as the following MIP

$$C_{new}^{\phi_i} = \underset{\delta_{C_i^m}, x_0^v, \boldsymbol{u}_T^v, \boldsymbol{u}_N}{\arg\min.} \quad J(C_i),$$

$$\begin{aligned}
\text{s.t.}\ \ & x(0) = x_k, && && (7.24a)\\
& x(i+1) = Ax(i) + B_F u(i) + B_A z(i) + B_d d_{k+i}, && i \in \mathbb{I}_{N-1}, && (7.24b)\\
& 0 = E_F u(i) + E_A z(i) + E_d d_{k+i}, && i \in \mathbb{I}_{N-1}, && (7.24c)\\
& x(N) = x^v(N), && && (7.24d)\\
& x^v(i+1) = Ax^v(i) + B_F u^v(i) + B_A z^v(i) + B_d d_{k+i}, && i \in \mathbb{I}_{T-1}, && (7.24e)\\
& 0 = E_F u^v(i) + E_A z^v(i) + E_d d_{k+i}, && i \in \mathbb{I}_{T-1}, && (7.24f)\\
& x^v(T) = x^v(0) = x_0^v, && && (7.24g)\\
& x(i) \in \mathcal{X}_{C_i^m}(i), && i \in \mathbb{I}_N, && (7.24h)\\
& x^v(i) \in \mathcal{X}_{C_i^m}(N), && i \in \mathbb{I}_T, && (7.24i)\\
& \bar{u}(i) \in \mathcal{U}_{C_i^m}(j) \times \mathcal{V}_{C_i^m}(j), && i \in \mathbb{I}_{N-1}, && (7.24j)\\
& \bar{u}^v(i) \in \mathcal{U}_{C_i^m}(N) \times \mathcal{V}_{C_i^m}(N), && i \in \mathbb{I}_{T-1}, && (7.24k)\\
& z(i) = diag(\delta_{C_i^m}) v(i), && i \in \mathbb{I}_{N-1}, && (7.24l)\\
& z^v(i) = diag(\delta_{C_i^m}) v^v(i), && i \in \mathbb{I}_{T-1}, && (7.24m)
\end{aligned}$$

where the product of continuous and logic variables in (7.24l)-(7.24m) can be transformed into equivalent linear integer inequalities [Bemporad and Morari, 1999]. Recall that $J(C_j)$ stands for the overall cost of configuration C_j according to some pre-established criteria that rule the configuration selection (cf. Section 5.3).

Observe that, by means of the single-layer MPC scheme, the integer program (7.24) updates coherently the model used for control and the model used for the virtual planner. This ensures the generation of a reachable trajectory for the new configuration (retrieved from the binary vector $\delta_{C_i^m}$), since the terminal ingredient used for stability (7.24d) is also modified coherently with the virtual planner.

7.5.3 Decision between sets

Section 7.5.2 formulates an MIP for obtaining (if it exists) a new configuration $C_{new}^{\phi_i}$ for each one of the sets $\mathbb{P}(C_i^m)$. Below, a sequential method is followed in order to decide how to conduct the search among the different sets and select a final C_{new}. In this sequential approach, the information retrieved from solving the configuration selection problem in one set is used for limiting the search space in the remaining sets.

For simplicity, the optimization problem (7.24) solved in the set $\mathbb{P}(C_i^m)$ is characterized as the following optimization problem

$$\min_{\gamma_i \in \Gamma_i} . \, J(\gamma_i), \tag{7.25}$$

where the vector γ_i encompasses the different decision variables and Γ_i is the feasibility set obtained for configuration C_i^m. Hence, using the notation in (7.25), Algorithm 7.2 reflects the steps followed in order to the final new configuration C_{new}.

Algorithm 7.2 Sequential solution method for configuration selection

```
$J_1^* = \min._{\gamma_1 \in \Gamma_1} J(\gamma_1)$
for $i = 2$ to $l$ do
    $J_i^* = \min._{\gamma_i \in \Gamma_i} \{J(\gamma_i) \mid J(\gamma_i) \leq J_j^*, \, j = 1, ..., i-1\}$
end for
Determine $C_{new}$ as:
$C_{new} = \arg\min. J(C_i) \,|\, J(C_i) \leq J_j^*, \; j = 1, ..., l.$
```

The performance of Algorithm 7.2 improves when the cost function $J(\cdot)$ is ruled by some criteria for which the minimal configuration yields the optimal value within the search set, e.g., the activation of the minimum number of back-up actuators. In that circumstances, by sorting the search sets $\mathbb{P}(C_i^m)$ in a coherent way, the solution obtained in the first iterations of Algorithm 7.2 can be claimed as the optimal C_{new}, and thus there will be no need to continue the search in the remaining steps. As an example, consider the lattice in Figure 7.1 and a minimum cardinality criterion. Hence, if running (7.24) on $\mathbb{P}(C_1^m)$ yields as optimum points any of the configurations: $C_{new}^{\phi_1} = \{1\}$, $C_{new}^{\phi_1} = \{1, 2\}$, or $C_{new}^{\phi_1} = \{1, 3\}$. Then, there is no need to run (7.24) on $\mathbb{P}(C_2^m)$ since no better solution can be obtained and thus $C_{new} = C_{new}^{\phi_1}$.

7.6 Case study

The case study used in order to illustrate the proposals presented in this chapter is the aggregated version of the drinking water transport network (DWTN) of the city of Barcelona that is detailed in Section B.2 of Appendix B.

In addition, the robustness of the network model is enhanced by considering uncertainties in the water demand predictions. In this regard, similar to Pereira et al. [2016a], the prediction error is assumed to be bounded in the set

$$\mathcal{W} = \{w_k \in \mathbb{R}^{25} : \ |w_k| \leq \bar{w}\},$$

where the maximum prediction error $\bar{w} \in \mathbb{R}^{25}$ is set as the 5% of the maximum expected demand during the tests, i.e., $\bar{w}^i = 0.05 \max_k d_k^i$. Throughout all the simulation presented in this section, the values of the uncertain variable w_k have been randomly generated following a uniform distribution bounded within $\mathcal{W}$.

Akin to the case study presented in Chapter 6, the actuators of the network have been partitioned into nominal and back-up components as reflected in Table 7.1 (with $|C_N| = 46$ and $|C_A| = 15$). Besides, in Figure 7.2 the back-up elements appear highlighted in green. Note that, the consideration of uncertainty provokes that the number of components that are considered as back-up elements is lower than in Section 6.7, that is, the uncertainty causes that more actuators are required for controlling the nominal configuration of the network.

At this point, it must be highlighted the fact that the element partition discussed above is completely artificial, and it has been carried out for illustrative purposes. On this subject, the network is thought to be operated using all the existing components. This particularity of the case study hinders the control of the system, and thus the robust network control becomes a challenging problem even in nominal configuration.

	Pumps	**Valves**
Nominal	$3, 5, 9, 10, 11, 15, 20,$ $21, 22, 23, 24, 29, 33,$ $34, 36, 38, 42, 48, 53$	$1, 2, 7, 8, 12, 13, 18, 28,$ $31, 32, 35, 39, 40, 41, 43,$ $44, 45, 46, 47, 49, 51, 52,$ $54, 56, 57, 59, 60, 61$
Back-up	$4, 17, 19, 27, 55$	$6, 14, 16, 25, 26, 30,$ $32, 37, 50, 58$

Table 7.1: Elements partition.

7.6.1 Management criteria

The stage cost uses a weighted sum of the following management criteria, as described in Ocampo-Martinez et al. [2009, 2012].

1. **Minimising water production and transport costs:** A first term accounts for the

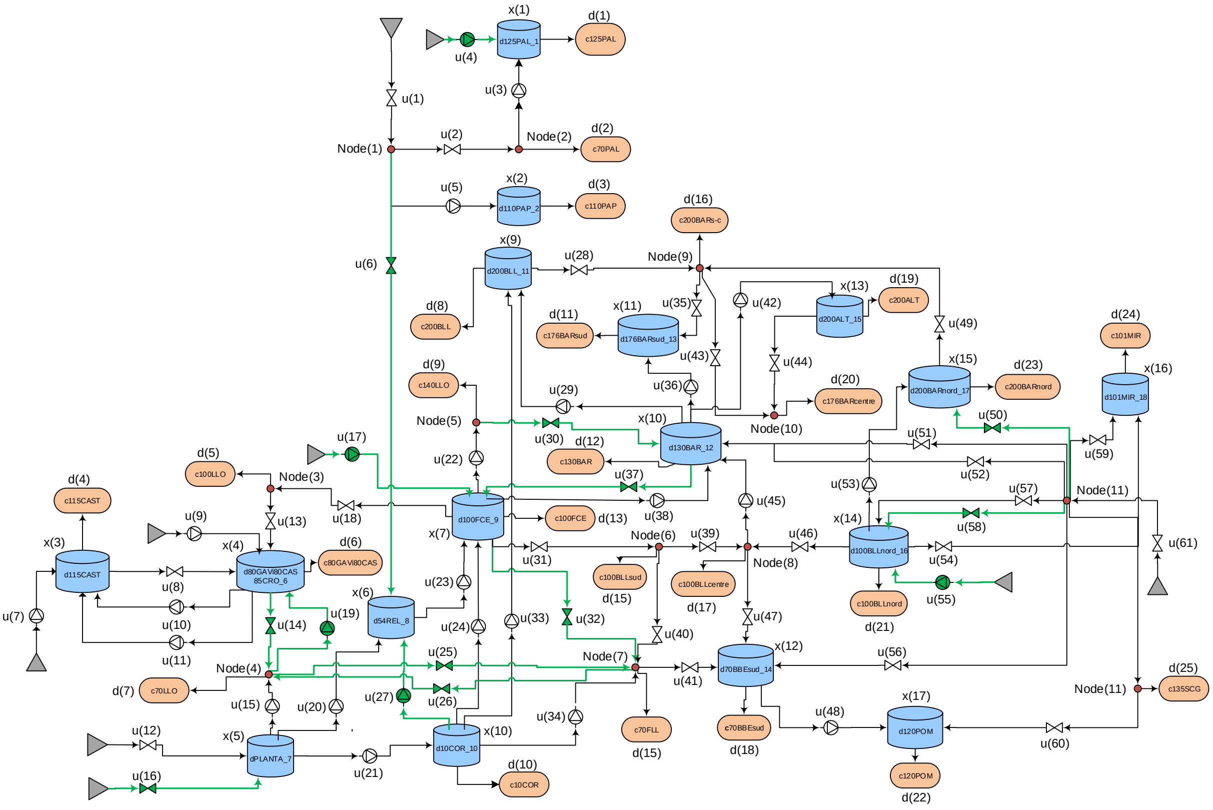


Figure 7.2: Nominal / back-up components.

economic costs associated with the drinking water production (water treatment) and transporting (pumping). The performance index to be minimized is described by the expression

$$f_{1,k} = (\alpha_1 + \alpha_{2,k})u_k,$$

where α_1 is an $1 \times n_u$ dimensional vector that accounts for the fixed economic cost of the water according to its source (treatment plant, dwell, etc.) and $\alpha_{2,k}$ is a vector of dimension $1 \times n_u$ associated with the economic cost of pumping the water. This vector $\alpha_{2,k}$, presents a $T = 24$h cyclic pattern that relates with the daily variations in the electricity rate.

2. **Safety storage term:** The satisfaction of water demands is imposed as a hard constraint in the network model, that should be fulfilled at every time instant. As a consequence, the stored water volume is preferably maintained around a given safety value as a risk prevention mechanism. This concept is formulated as

$$f_{2,k} = (x_k - x_{sf})^T W_x (x_k - x_{sf}),$$

where $x_k \in \mathbb{R}^{n_x}$ denotes the water volume in the tanks and $x_{sf} \in \mathbb{R}^{n_x}$ denotes the safety storage volume. Particularly, the safety volume has been designed as $x_{sf} = 0.75(\bar{x} - \underline{x})$, where $\bar{x}$ and $\underline{x}$ denote the maximum and minimum accepted tank volumes, respectively. Moreover, the selected weighting matrix is set to $W_x = diag(1/(\bar{x} - \underline{x}))$, in order to penalize the deviation from the safety volume proportionally to the size of each one of the tanks.

3. **Smoothness of the control actions:** The variations of the control signal between consecutive sampling intervals is also penalised. By denoting $\Delta u_k = \bar{u}_k - \bar{u}_{k-1}$, this concept is formulated as

$$f_{3,k} = \Delta u_k^T W_u \Delta u_k,$$

where the weighting matrix has been defined as $W_u = I_{\bar{n}_u}$.

Accordingly, the stage cost is made up of a weighted sum of the previous terms

$$l(k, x_k, u_k) = \lambda_1 f_{1,k} + \lambda_2 f_{2,k} + \lambda_3 f_{3,k},$$

with $\lambda_1 = 1$, $\lambda_2 = 0.05$ and $\lambda_3 = 0.01$.

7.6.2 Robust controller design for the nominal configuration

Here, the computation of the tuning parameters required for the robust MPC of configuration C_N is addressed. When required, a similar procedure has been followed for an arbitrary configuration C_i.

The time horizon of the MPC controller is set to $N = T = 24$h. Matrix M_v is designed as an orthonormal basis to the null space of matrix E_N. A controller gain K is designed such that $A + B_N M_v K$ is an asymptotically stable matrix. On this subject, it must be pointed out the difference in the actuators limits existing in the network: actuator 50 has a maximum value of 15m^3/s, whereas the maximum value of actuator 7 is of 10^{-5}m^3/s. Therefore, the local controller gain K is designed following an LQR design that weights the input power inversely to the flow handling capacity of each actuator. Hence, the matrices $Q = I_{n_x}$ and $R = M_v^T diag(1/u_{max}) M_v$ are used for designing the controller gain, where u_{max} denotes the maximum flow handling capacity of the actuators.

At this point, the solution of (7.17) yields the matrix M_w (and thus the set of tightened constraints), as well as the average stage cost of the robust planner which for the nominal configuration is $\mathcal{L}^T_{C_N} = 2.2518 \cdot 10^3$.

7.6.3 Minimal configurations that satisfy the performance condition

Here an offline analysis of the average stage cost that can be attained by the different candidate configurations is presented. Below, it has been considered that a configuration C_i satisfies the performance property if its associated average stage cost is below $\beta = 1.25$ times the cost of the nominal configuration $\mathcal{L}^T_{C_N}$. The offline tests yielded the following results:

- A single fault in one of the following components has been identified as critical

  $$\{1, 2, 5, 12, 13, 15, 18, 21, 22, 23, 28, 29, 31, 35, 36, 38, 40, 41, 42, 44, 49, 54, 56, 59, 61\}.$$

  That is, either there are no combination of back-up elements for with robust planner (7.17) generates an feasible trajectory, or the attained trajectory has associated cost lower than the threshold.

- For fault in the following components

  $$\{7, 8, 9, 10, 11, 16, 33, 39, 43, 45, 46, 48, 52, 60\}$$

  the robust planner (7.17) is able to generate an admissible trajectory for the faulty configuration C_F.

- For a fault on the elements reported in Table 7.2, the search for minimal configurations that satisfy the performance property has been carried out. In this regard, Table 7.2 shows the number of minimal configurations found for an outage on the actuator, as well as the number (and percentage) of configurations explored by Algorithm 7.1 out of the possible $2^{15} - 1 = 32.767$ candidate configurations. Note the good behaviour of the Algorithm 7.1.

Fault in	**Number of minimal configurations**	**Explored configurations**	**Percentage**
3	1	16	0.049%
20	17	200	0.610%
24	11	144	0.439%
34	1	17	0.052%
47	2	34	0.104%
51	1	17	0.052%
53	1	17	0.052%
57	1	16	0.049%

Table 7.2: Minimal configurations.

7.6.4 Criteria for configuration selection

In the sequel, the selection among the different candidate configurations is performed by minimizing the following criteria

1. Minimize number of back-up actuators activated (h_1).
2. Minimize the expected performance-loss during the transient (h_2).

Furthermore, a lexicographic ordering is assumed among the previous objectives, i.e., the optimization of the first objective is infinitely more important than the optimization of the second one (cf. Section A.4 of Appendix A).

7.6.5 Fault scenario - Fault in 24

Here a fault in actuator 24 is simulated. For this case, the maximum performance degradation is set at 20% of the nominal performance, that is, $\beta = 1.2$. Hence, a configuration C_i satisfies the performance property if its associated stage cost is lower than $\beta \cdot \mathcal{L}^T_{C_N} = 2.702 \cdot 10^3$.

For the above threshold, the set of minimal configurations that have been obtained is presented in Table 7.3. These configurations were computed using Algorithm 7.1, which examined 109 configurations out of the $2^{15} - 1 = 32.767$ candidate ones. This represents the exploration of the 0.3326% of the candidate configurations. Notice that in Table 7.3 the configurations have been sorted by taking into account its cardinality.

Fault scenario: starting with the system operating with configuration C_N, a fault in actuator 24 appears at $k_f = 39$h. This fault causes a performance loss of 25% of the actuator capabilities.

Ordering	$\boldsymbol{C_i^m}$	$\boldsymbol{\mathcal{L}^T_{C_i^m}}$
1	$[6,\ 17,\ 27]$	$2.668 \cdot 10^3$
2	$[17,\ 27,\ 50]$	$2.678 \cdot 10^3$
3	$[14,\ 17,\ 25,\ 27]$	$2.671 \cdot 10^3$
4	$[17,\ 25,\ 26,\ 27]$	$2.671 \cdot 10^3$
5	$[17,\ 27,\ 37,\ 55]$	$2.697 \cdot 10^3$
6	$[17,\ 27,\ 55,\ 58]$	$2.698 \cdot 10^3$
7	$[6,\ 14,\ 25,\ 27,\ 50]$	$2.680 \cdot 10^3$
8	$[6,\ 25,\ 26,\ 27,\ 50]$	$2.681 \cdot 10^3$

Table 7.3: Set of minimal configurations - Fault 24, $\beta = 1.2$.

Moreover, it is assumed that an FDI block detects the fault at $k_d = 43$h and that the actuator 24 is turned-off.

The use of Algorithm 7.2 yields the following results:

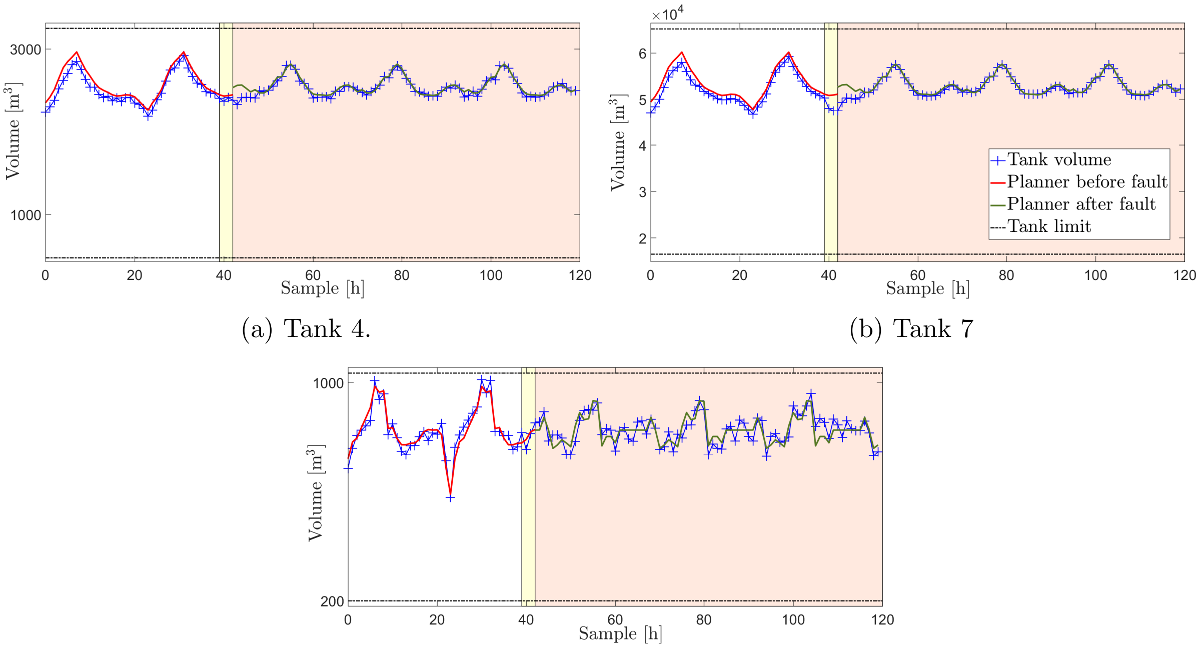
- **First optimization:** The first optimization is launched for the set $\mathbb{P}(C_1^m)$, corresponding with the back-up actuators in $C_1^m = [6,\ 17,\ 27]$ activated. The obtained solution is $C_{new}^{\phi_1} = C_1^m$ $(h_1 = 3)$. Hence, since for all $j \geq 3$, $|C_j^m| \geq 3$ (cf. Table 7.3), then the search must only be run for the set $\mathbb{P}(C_2^m)$. Besides, with the back-up actuators C_1^m activated, the best expected stage cost obtained during the transient is $h_2 = 2.2443 \cdot 10^3$.
- **Second optimization:** A second optimization for the set $\mathbb{P}(C_2^m)$ is run subject to the constraint of a maximum activation of three back-up components. The second optimization yields $C_{new}^{\phi_2} = C_2^m$ $(h_1 = 3)$. Besides, the best expected cost during the transient is $h_2 = 2.2407 \cdot 10^3$.

Accordingly, $C_{new} = C_{new}^{\phi_2}$, since, for the same number of back-up actuators, it produces a better expected average cost during the transient.

Throughout the remaining figures: the yellow background highlights the system evolution in the time interval between the fault occurrence and its diagnosis. On the other hand, the orange background highlights the system evolution after the detection/reconfiguration. Figure 7.3 shows in blue the temporal evolution of different tank volumes during the fault scenario described above. In these figures, it is displayed in red the reference that generates the robust planner for configuration C_N, and in green the new reference towards which the system in the configuration C_{new} will converge. Note the control capabilities of the MPC scheme, which is able to robustly stabilize the system.

Besides, Figure 7.4 shows the evolution of several actuators of the network. In particular, Figure 7.4a shows the evolution of the back-up actuator 17, which it is turned-on. Besides, in Figure 7.4b it can be seen how the pump 23 has to increase its power with the new configuration, what yields a worse economic performance.

Additionally, Figure 7.5 depicts in red the stage cost of the nominal configuration, in green the cost of the new configuration and in purple the maximum stage cost set by the admissibility condition. The blue line represents the evolution of the average stage cost during the fault

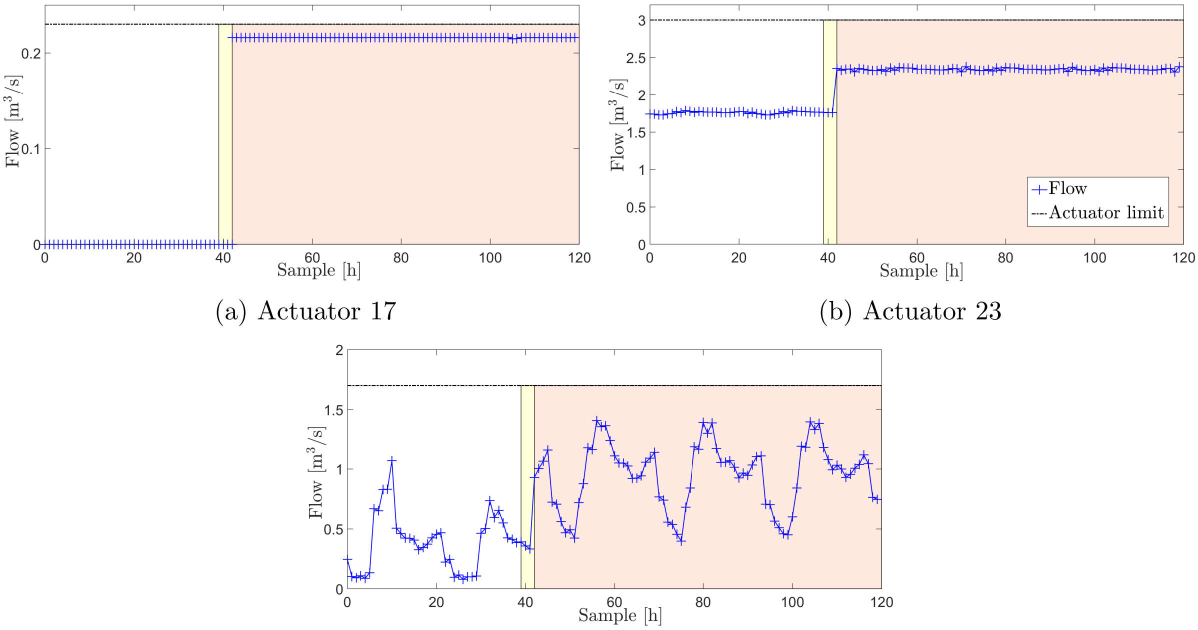


(a) Tank 4.

(b) Tank 7

(c) Tank 11

Figure 7.3: Tank volumes evolution - Fault in 24.



(a) Actuator 17

(b) Actuator 23

(c) Actuator 51

Figure 7.4: Actuators evolution - Fault in 24.

scenario. This cost has been computed at time k by averaging the cost obtained in the previous T samples, that is, for the time interval $[k-T+1,k]$. Besides, $\mathcal{L}^T_{C_N}$ and $\mathcal{L}^T_{C_2^m}$ are also displayed. In this figure, it can be seen how firstly the cost stabilizes at $\mathcal{L}^T_{C_m}$ and how, after the transient induced by the fault and reconfiguration, the stage cost stabilizes at $\mathcal{L}^T_{C_2^m}$. Note the mismatch between the cost values obtained by the real system trajectory and the ones computed by the planner. This is a consequence of the fact that the robust planner is computed for the nominal

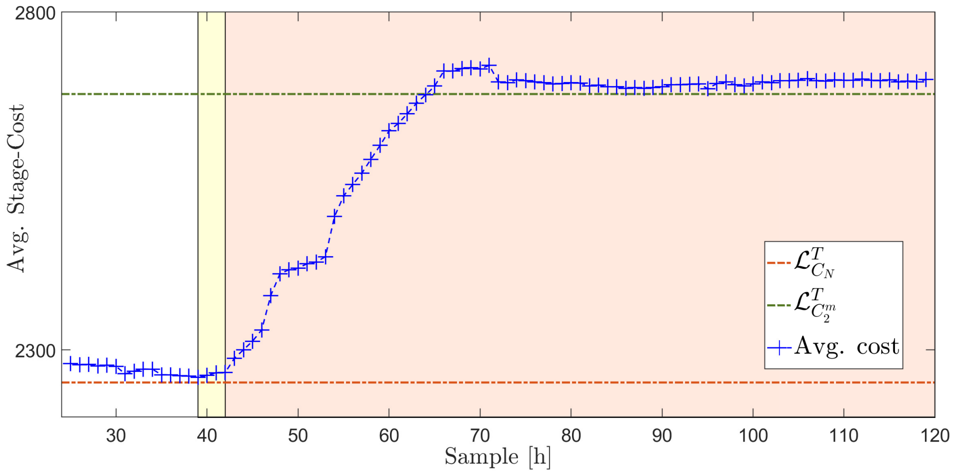


Figure 7.5: Average stage cost - Fault 24.

(non-disturbed) system.

7.6.6 Fault scenario - Tank isolation.

Here, a multiple fault is simulated. Thus, in this case, the maximum performance degradation is set to 40% of the nominal performance ($\beta = 1.4$), in such a way that the threshold is $\beta \cdot \mathcal{L}^T_{C_N} = 3.152 \cdot 10^3$. For that threshold, the obtained set of minimal configurations is presented in Table 7.4. Those configurations were computed using Algorithm 7.1, which explored 45 out of $2^{14} - 1 = 16.383$ candidate configurations. This represents a 0.27% of the total.

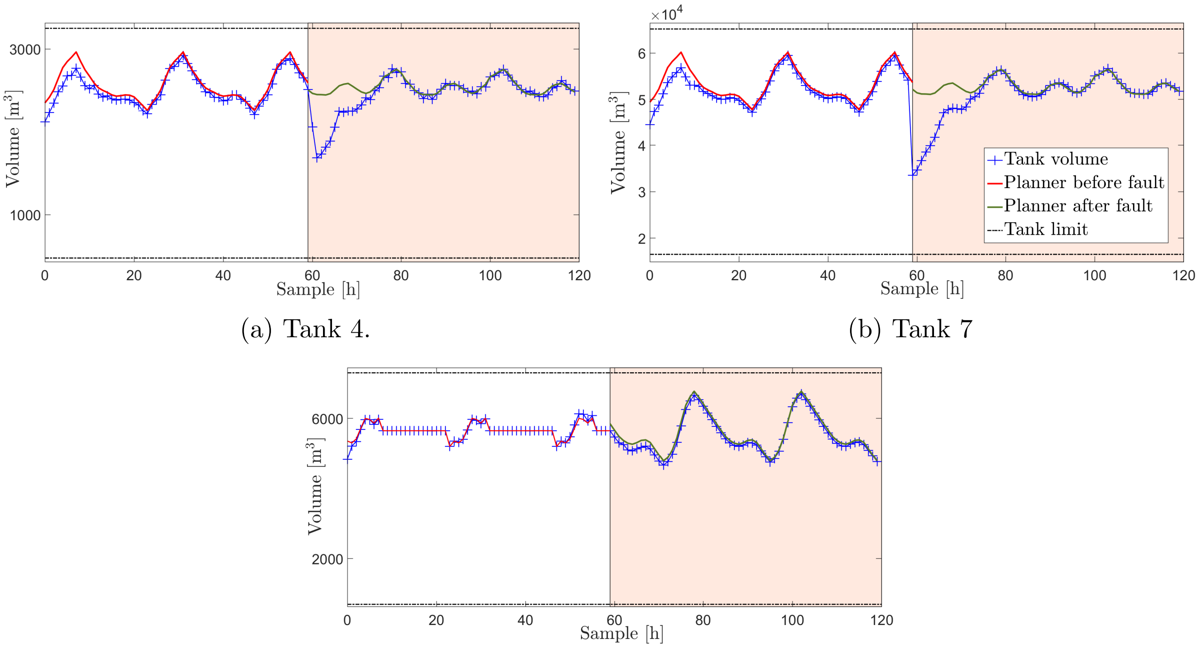
Ordering	C_i^m	$\mathcal{L}^T_{C_i^m}$
1	$[6,\ 32]$	$3.127 \cdot 10^3$
2	$[14,\ 17,\ 25,\ 32,\ 37]$	$3.123 \cdot 10^3$
3	$[17,\ 25,\ 26,\ 32,\ 37]$	$3.123 \cdot 10^3$

Table 7.4: Set of minimal configurations - Tank isolation, $\beta = 1.4$.

Below, a new fault scenario is simulated. In this case, it is considered that at $k_f = k_d =$ 59h no more water can be extracted from Tank 8 (cf. Figure 7.2). This emulates a fault in $\tilde{C}_f = [24,\ 33,\ 34]$, and causes that actuator 27 is removed from the list of candidate back-up actuators. Besides, at k_f a perturbation is introduced in the state of the Tank 7, causing an unexpected emptying of the tank. In this regard, the online implementation of Algorithm 7.2, yields the following results:

- **First optimization:** The first optimization is launched for the set $\mathbb{P}(C_1^m)$. The obtained solution is $C_{new}^{\phi_1} = [6,\ 17,\ 32]$. At this point, since the cardinality of $C_{new}^{\phi_1}$ is lower than the cardinality of the remaining minimal configurations the algorithms stops the search, and thus $C_{new} = C_{new}^{\Phi_1}$.

Hence, apart from the actuators $C_1^m = [6,\ 32]$, which guarantee an admissible steady-state performance, in this scenario it is also required to activate the actuator 17 in order to cope

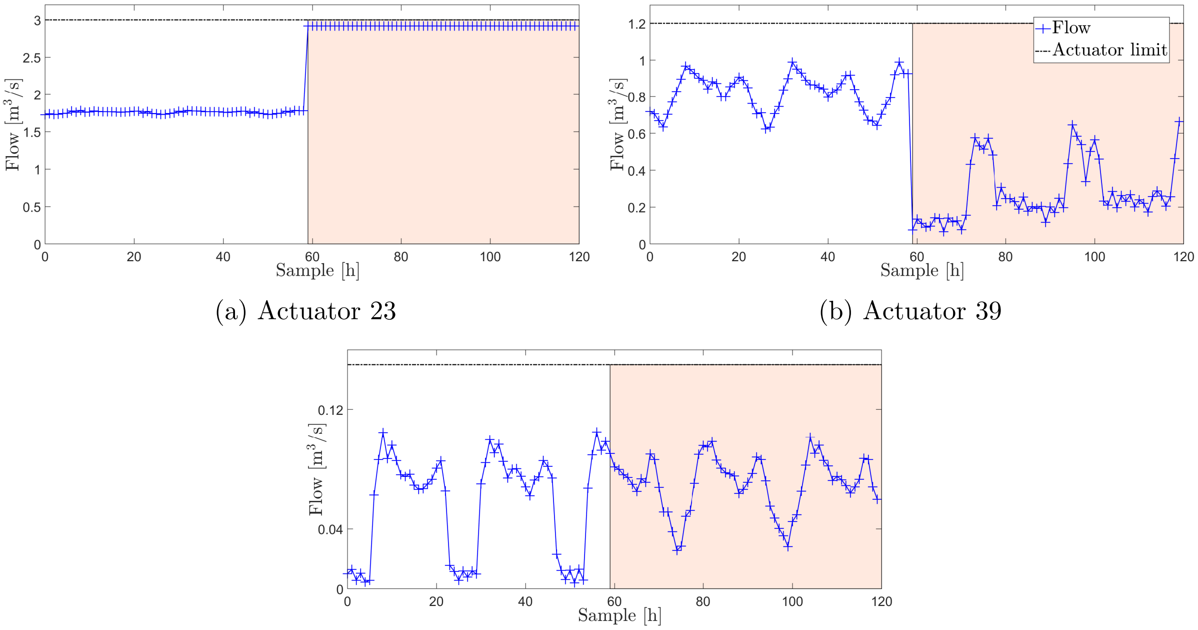


(a) Tank 4.

(b) Tank 7

(c) Tank 9

Figure 7.6: Tank volumes evolution - Tank isolation.



(a) Actuator 23

(b) Actuator 39

(c) Actuator 44

Figure 7.7: Actuators evolution - Tank isolation.

with the transient induced by the fault. Note that the resulting controller optimizes the nominal trajectory (i.e. without disturbances) using C_{new}, while the uncertainty compensation is done by means of $C_1^m = [6,\ 32]$. This yields a new robust planner cost of $\mathcal{L}^T_{C_{new}|C_1^m} = 2.853 \cdot 10^3 \leq \mathcal{L}^T_{C_1^m}$.

Figure 7.6 displays the evolution of the tanks volumes, and the modification of the robust planner trajectories before and after the reconfiguration. In particular, Figure 7.6b shows the effect of the injected disturbance on Tank 7, which also affects Tank 4 in Figure 7.6a. In those figures, it can be seen how the system recovers from the effect of the fault, and how it is able to stabilize over the new reference. On the other hand, Figure 7.7 shows the evolution of different actuators for the fault scenario described above. Finally, Figure 7.8 presents the stage cost during the fault scenario. In particular, it can be seen how, after the reconfiguration, the average stage cost ends up stabilizing in $\mathcal{L}^T_{C_{new}|C_1^m}$, which, by monotonicity, is below the stage cost obtained for the minimal configuration and thus below the admissibility threshold.

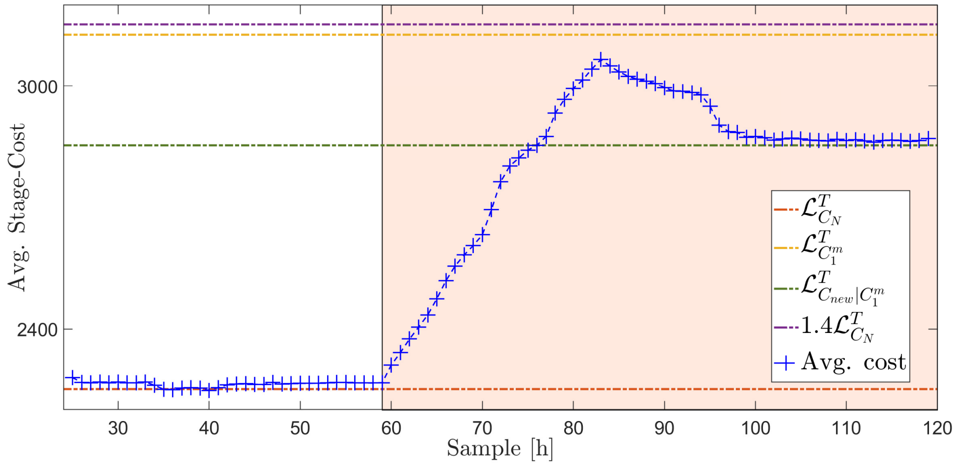


Figure 7.8: Average stage cost - Tank isolation.

7.7 Summary

This chapter presents a robust solution to the system reconfiguration with back-up components problem. The proposed solution uses a single-layer MPC scheme, which allows to embed the coherent modification of the controller and planner into an MIP optimization. Furthermore, the coupling between the configuration of actuators used in the design of a local control law and the one used for simulating the evolution of the nominal system, has been addressed basing the design of the local controller on those minimal configurations which, due to the monotonicity of the performance property, guarantee that any superset of them yields an admissible performance in the steady-state.

One of the greatest difficulties of the problem under study, lies in the parameters on which the different elements that make up the control scheme depend. In such a way that small modifications in these parameters can significantly affect the existence of a control law that is admissible for a given actuators configuration. In this regard, the proposed solution seeks a trade-off between the optimality in the new configuration selection, and the computational complexity of the approach. Finally, the next natural step in the development of these reconfiguration techniques would be to consider possible non-linearities in the system model.

Part III

Conclusions and future research

Chapter 8

Concluding remarks and further extensions

As concluding remarks, it should be pointed out that the objectives set at the beginning of the thesis have been satisfied for the most part. Additionally, during the course of this thesis, new ideas and challenges have appeared that have enriched the results presented in this document. On this subject, it is worth mentioning that the different contributions were reported at each corresponding chapter. For completeness, these contributions are collected below. Furthermore, the proposal of future ways to continue the research developed in the thesis will also be pointed out in this chapter.

8.1 Contributions

Hereafter, the contributions presented in this thesis are summarized.

Contributions presented in *Part I - Replay attack detection*

- The detectability of the replay attack, as well as the effect that a watermark signal causes on the residual signal generated by an anomaly detector, have been characterized using zonotopic sets.
- It has been investigated two different replay attack scenarios that can be launched against the supervisory layer of a control system, and that depend on the attacker's capabilities to access the data sent by the different components of the system.
- Analytical expressions that relate the steady-state replay attack detectability with the output set-point reference set from the supervisory layer have been obtained for both attack scenarios.
- The offline design of finite watermark sequences that asynchronously added in the control loop guarantee the replay attack detection has been proposed.
- A novel zonotope-based metric to assess the performance of an optimal control loop make up of an LQR controller and a ZKF state estimator, has been presented.

- It has been shown the relationship between the F-radius of the mRPI set for the estimation error generated by a ZKF, and the solution of the corresponding DARE. Besides, the F-radius has been proposed to evaluate the performance loss induced by working on over-approximations of mRPI set.
- Analogous expressions regarding performance loss induced by a Gaussian/zonotopically-bounded watermark signal in an optimal control scheme have been obtained.
- A novel zonotopically-bounded watermark signal that exploit the loss of feedback that causes a replay attack in order to guarantee the attack detection, has been proposed.

Contributions presented in ***Part II - Reconfiguration with back-up components***

- A general methodology has been proposed to address the problem of selecting the optimal actuator configuration in a system with back-up components.
- The formulation of the configuration selection as a multi-objective MIP, and its solution using a lexicographic approach, has been investigated.
- The management of the information retrieved from the analysis of necessary properties to improve the worst-case execution of the MIP optimization has been explored.
- The robust configuration selection has been solved using a single-layer robust MPC scheme, where the local controller is designed using the minimal configurations that guarantee an admissible performance.
- It has been proposed to use the robust planner periodic trajectories in order to assess the performance admissibility of an actuator configuration in flow-based systems.
- The matrix that distributes the uncertainty compensation between the different actuators in the robust control of flow-based networks has been designed in an optimal manner.
- Different algorithms to look for minimal configurations in a lattice of configurations have been proposed.

8.2 Directions for future research

This section collects some of the open issues and new ideas that could be addressed in future work.

On the detection of stealthy attacks

- The zonotope-based techniques used in Chapters 3 and 4 can be extended to characterize the effect of other attack policies like, for example, bias injection attacks, as well as for the design of active methods that guarantee the attack detection.
- The effect of the gain of the observer located in the supervision center in the active detection of replay attacks can be studied. Investigating if an adequate tuning of the observer can reduce the performance loss that must be imposed by the watermark signal in order to guarantee the attack detection.

- With respect to the finite watermark sequences presented in Chapter 3, it can be analysed which injection directions yield the minimal performance degradation for a given security index. In addition, a smart injection schedule can be studied that grants a good trade-off between performance loss and detection capabilities, under certain assumptions on the probability of the system to be under attack at a specific time instant.
- The effect of the zonotope reduction operator in the convergence of the Ricatti difference equations, and hence in the convergence of the optimal observer gain, can be analysed.
- The scalability of the proposed approaches can be studied. Particularly, the joint design of watermark signals and decoupling strategies can be explored in order to attain easily scalable detection schemes.
- The proposed zonotope-based detection techniques can be extended to deal with guaranteed detection in the context of distributed systems.
- The zonotopically-bounded watermarking schemes can be easily coupled with robust control schemes for constrained systems. Then, the design of watermarked/attack-resilient control schemes can be further studied.

On the reconfiguration with back-up components

- The results presented in Chapters 6 and 7 can be extended by considering possible non-linearities in the system dynamics. In particular, for DWTN, it can be studied how to include the non-linearities that arise when taking into account the pressure difference between the storage elements and the consumption nodes.
- The modularity and low computation advantages of decentralized control schemes can be exploited to solve the reconfiguration with back-up components locally, while guaranteeing the closed-loop stability.
- Partitioning approaches can be used to split the network into physically redundant areas, in such a way that the search for hardware-redundant components can be conducted in each area independently.
- It can be studied the development of metrics that assess the hardware redundancy and the average running cost of the different configurations, in such a way that, at a network design stage, shed some light on the selection of a secure optimal nominal configuration.

Part IV

Appendices

Appendix A

Mathematical background

A.1 Matrix definitions and properties

Some basic definitions and matrix properties are introduced below [Golub and Van Loan, 2013].

Definition A.1. The symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called *positive definite* if for all $x \neq 0$, $x^T A x > 0$.

Definition A.2. The symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called *negative definite* if $-A$ is positive definite.

Definition A.3. The symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called *positive semidefinite* (or nonnegative definite) if for all $x \neq 0$, $x^T A x \geq 0$.

Definition A.4. The symmetric matrix $A \in \mathbb{R}^{n \times n}$ is *negative semidefinite* (or nonpositive definite) if $-A$ is positive semidefinite.

Definition A.5. The *trace* of the square matrix $A \in \mathbb{R}^{n \times n}$ is defined to be the sum of the elements of its main diagonal, i.e., $Tr[A] = \sum_{i=1}^{n} a_{ii}$. Moreover, for matrices X, A, B, C, D and vectors x, y of appropriate size, and a scalar α, the following holds

$$\begin{aligned}
Tr[A] &= Tr[A^T], \\
Tr[A+B] &= Tr[A] + Tr[B], \\
Tr[Ayx^T] &= x^T A y, \\
Tr[ABCD] = Tr[BCDA] &= Tr[CDAB] = Tr[DABC], \\
\partial_X Tr[AX^T B] &= A^T B^T, \\
\partial_X Tr[AXBX^T C] &= BX^T CA + B^T X^T A^T C^T,
\end{aligned}$$

where $\partial_X f(\cdot)$ is a short notation for $\partial f(\cdot)/\partial X$.

Definition A.6. The *p-norm* of matrix a A is defined for a real number $1 \leq p \leq \infty$ by

$$\|A\|_p = \max_{\|x\|_p = 1} |Ax|_p, \tag{A.1}$$

where $|x|_p$ is a vector norm.

Definition A.7. The *Frobenius norm* of a matrix $A \in \mathbb{R}^{m\times n}$ is given by

$$\|A\|_F = \sqrt{Tr[AA^T]} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}, \tag{A.2}$$

where a_{ij} are the elements of A.

Property A.1. For matrices A and B it follows that $rank(AB) \leq min\{rank(A), rank(B)\}$.

Property A.2. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. If A is invertible then $rank(AB) = rank(B)$.

Property A.3. Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times n}$. If B is invertible then $rank(AB) = rank(A)$.

Property A.4. Let $A \in \mathbb{R}^{k\times m}$, $B \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{n\times l}$. If ABC exists then $rank(AB) + rank(BC) \leq rank(ABC) + rank(B)$.

Property A.5. Let $A \in \mathbb{R}^{n\times p}$ have column rank r. The set $\mathcal{N}(A) = \{x \in \mathbb{R}^p : Ax = 0\}$ is a vector space of dimension equal to $p - r$ and it is called the null space of A.

A.2 Set definitions and properties

Set based-approaches play a fundamental role in several chapters of this thesis. Therefore, it is necessary to present their main definitions and properties. In this context, among the different set descriptions typically employed in the automatic control field (intervals, polyhedrals, zonotopes, ellipsoids, etc), this dissertation makes use of polyhedral and zonotopic sets. Consequently, some basic definitions, properties and operations regarding polyhedrals and zonotopes are introduced in Section A.2.1 and Section A.2.2, respectively.

Prior to introduce the polyhedral and zonotopic sets, the following definitions and basic operations between sets are presented [Le et al., 2013].

Definition A.8. A set $\mathcal{X} \subset \mathbb{R}^n$ is called *convex*, if for any $x_1, x_2, ..., x_k \in \mathcal{X}$ with $k \geq 2$, and any $\alpha_1, \alpha_2, ..., \alpha_k \in \mathbb{N}$ such that $\sum_{i=1}^{k}\alpha_i = 1$, the element $\sum_{i=1}^{k}\alpha_i x_i$ is in $\mathcal{X}$.

Definition A.9. A *convex hull* of a given set $\mathcal{X}$, denoted $conv(\mathcal{X})$, is the smallest convex set containing $\mathcal{X}$.

Definition A.10. The *inclusion operator* between two sets is defined by $\mathcal{X} \subseteq \mathcal{Y}$, if and only if $\forall x \in \mathcal{X}$, then $x \in \mathcal{Y}$. This means that $\mathcal{X}$ is a subset of $\mathcal{Y}$.

Definition A.11. The *intersection operator* of two sets $\mathcal{X}$ and $\mathcal{Y}$ is defined by $\mathcal{X} \cap \mathcal{Y} = \{x : x \in \mathcal{X} \,\text{and}\, x \in \mathcal{Y}\}$.

Definition A.12. The image of a set $\mathcal{X}$ under a *map* (projection) $\mathcal{M}$ is the set $\mathcal{M}(\mathcal{X}) = \{y : y = \mathcal{M}(\mathcal{X}), x \in \mathcal{X}\}$.

Definition A.13. The *Minkowski sum* of two sets $\mathcal{X}$ and $\mathcal{Y}$ is defined by $\mathcal{X} \oplus \mathcal{Y} = \{x + y : x \in \mathcal{X},\ y \in \mathcal{Y}\}$.

Definition A.14. The *Pontryagin difference* (P-difference) of two sets $\mathcal{X}$ and $\mathcal{Y}$ is defined by $\mathcal{X} \ominus \mathcal{Y} = \{z : z + y \in \mathcal{X},\ \forall y \in \mathcal{Y}\}$.

Furthermore, several definitions related to the set invariance of a discrete-time linear time-invariant (LTI) system are introduced. To that end, consider the system

$$x_{k+1} = Ax_k + w_k, \tag{A.3}$$

where $x_k \in \mathbb{R}^n$ is the state, $w_k \in \mathbb{R}^n$ is an unknown disturbance, and k is the time instant. Below, w_k is assumed to be contained at all time instants in a convex and compact set $\mathcal{W} \subset \mathbb{R}^n$ that contains the origin, and that $A \in \mathbb{R}^{n\times n}$ is an asymptotically stable matrix (all the eigenvalues of A are strictly inside the unit disk).

Definition A.15. The set $\Omega \subset \mathbb{R}^n$ is said to be *robustly positively invariant* (RPI) for the system (A.3), if for all $x_0 \in \Omega$ and all $w_k \in \mathcal{W}$ the solution is such that $x_k \in \Omega$ for $k > 0$. Equivalently, Ω is RPI if and only if $A\Omega \oplus \mathcal{W} \subseteq \Omega$.

Definition A.16. The *minimal RPI* (mRPI) set of (A.3) is the RPI set in $\mathbb{R}^n$ that is contained in every closed RPI set of (A.3).

For an LTI asymptotically stable system like (A.3), the mRPI set exists, is unique, compact and contains the origin of the state space if $\mathcal{W}$ contains the origin of the disturbance space [Kolmanovsky and Gilbert, 1998, Sec. IV]. Furthermore, from the linearity and asymptotic stability of (A.3), it follows that the mRPI set is the limit set of all trajectories of system (A.3). Additionally, in set-theoretic notation the mRPI set Φ_∞ is formulated as

$$\Phi_\infty = \bigoplus_{i=0}^{\infty} A^i \mathcal{W}. \tag{A.4}$$

For a comprehensive analysis on set invariance please refer to Kolmanovsky and Gilbert [1998], Blanchini and Miani [2008].

A.2.1 Polyhedral sets

A polyhedral set in a finite-dimensional Euclidean space is the intersection of a finite amount of closed half-spaces. A bounded polyhedral is denoted as a *polytope*. Polytopes can be expressed in any of its dual representations, i.e., half-space representation (H-polytopes) and vertex representation (V-polytopes), which are known to be mathematically equivalent allowing to transform one representation into the other [Ziegler, 2012]. The definition of the dual representations are presented below.

Definition A.17 (Half-space representation)**.** A polyhedral set $\mathcal{P} \subset \mathbb{R}^n$ can be defined as the intersection of a finite number of closed half-spaces

$$\mathcal{P} = \{x \in \mathbb{R}^n : Hx \leq k,\ H \in \mathbb{R}^{m\times n},\ k \in \mathbb{R}^m\}.$$

If $\mathcal{P}$ is bounded, then $\mathcal{P}$ is a polytope.

Definition A.18 (Vertex representation)**.** For a finite set of points $\mathcal{V} \subset \mathbb{R}^n$, with $\mathcal{V} = \{v_1, v_2, ..., v_m\}$, a polytope $\mathcal{P}$ can be defined as the convex hull of the set $\mathcal{V}$

$$\mathcal{V} = conv(\mathcal{V}) = \big\{\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_m v_m :\ \alpha_i \in \mathbb{R}^+,\ \sum_{i=1}^{m} \alpha_i = 1,\ v_i \in \mathbb{R}^n\big\}.$$

A.2.2 Zonotopic sets

Zonotopes are increasingly being used in the field of systems engineering [Kühn, 1998, Combastel, 2003, Alamo et al., 2005]. This is mainly due to their flexibility, reduced complexity and because they allow the efficient computation of linear transformations and Minkowski sums [Le et al., 2013]. Zonotopes are a particular class of convex polytopes, which are symmetric with respect to their center. Accordingly, zonotopes can be represented as H-polytopes or V-polytopes. Nevertheless, the main advantages in the use of zonotopes are associated with alternative representations which allow to implicitly represent the zonotope shape by a rectangular matrix.

Definition A.19. A *zonotope* $\langle c, H\rangle \subset \mathbb{R}^n$ is the affine transformation of a unitary hypercube $\mathbf{B}^m = [-1,1]^m$:

$$\langle c, H\rangle = c \oplus H\mathbf{B}^m = \{c + Hz : z \in \mathbf{B}^m\},$$

where $c \in \mathbb{R}^n$ is the center and $H \in \mathbb{R}^{n\times m}$ the generators matrix. The order of $\langle c, H\rangle$ is m/n.

Some basic definitions and properties used throughout this dissertation are stated below.

Definition A.20. The *interval hull* (or aligned box) of a given zonotope $\mathcal{Z} = \langle c, H\rangle \subset \mathbb{R}^n$ with $H \in \mathbb{R}^{n\times m}$, denoted by $\square\mathcal{Z} = \langle c, b(H)\rangle$, is the smallest interval box that contains $\mathcal{Z}$ and can be computed as

$$b(H) = diag(|H|\mathbf{1}_m) \in \mathbb{R}^{n\times n}, \tag{A.5}$$

where $\mathbf{1}$ is a column vector of ones.

Property A.6. The *permutation* of the generator matrix columns of a zonotope does not modify the zonotope.

Property A.7. The *Minkowski sum* of the zonotopes $\mathcal{Z}_1 = \langle c_1, H_1\rangle$ and $\mathcal{Z}_2 = \langle c_2, H_2\rangle$ is $\mathcal{Z}_1 \oplus \mathcal{Z}_2 = \langle c_1 + c_2, [H_1,\ H_2]\rangle$.

Property A.8. The *linear image* of a zonotope $\mathcal{Z} = \langle c, H\rangle$ by a compatible matrix L is $L\mathcal{Z} = L\langle c, H\rangle = \langle Lc, LH\rangle$.

Definition A.21. Given the zonotope $\mathcal{Z} = \langle c, H\rangle$, the *covariance* of the zonotope is defined as $cov(\mathcal{Z}) = HH^T$.

Definition A.22. The *Frobenius radius* (F-radius) of a given zonotope $\mathcal{Z} = \langle c, H\rangle$ is the Frobenius norm of the generator matrix, i.e., $\|\mathcal{Z}\|_F = \|H\|_F$.

Definition A.23. The *weighted Frobenius radius* (F_W-radius) of a given zonotope $\mathcal{Z} = \langle c, H\rangle$ is the weighted Frobenius norm of the generator matrix, i.e., $\|\mathcal{Z}\|_{F,W} = \|H\|_{F,W}$.

A reduction operator, denoted as $\downarrow_q$, allows to reduce the number of generators of a zonotope $\langle c, H\rangle$ to a fixed number $q \in \mathbb{N}$ while preserving the inclusion property, i.e., $\langle c, H\rangle \subseteq \langle c, \downarrow_q(H)\rangle$. Among the different order reduction techniques existing in the literature [Yang and Scott, 2018], this thesis mainly uses the reduction operator formulated in Combastel [2005] and recalled below.

Property A.9. Given the zonotope $\langle c, H\rangle \subset \mathbb{R}^n$ with $c \in \mathbb{R}^n$ and $H \in \mathbb{R}^{n\times m}$, sort the columns of H on decreasing weighted norm $\|\cdot\|_W$, that is, $H = [h_1, ..., h_j, ..., h_m]$ with $\|h_j\|_W^2 \geq \|h_{j+1}\|_W^2$. Then, by computing $\downarrow_{q,W}(H) = [H^a,\ b(H^b)] \in \mathbb{R}^{n\times q}$ with $H^a = [h_1, ..., h_{q-n}]$ and $H^b = [h_{q-n+1}, ..., h_m]$, it follows that $\langle c, H\rangle \subseteq \langle c, \downarrow_{q,W}(H)\rangle$.

Property A.10. Given the vector $p \in \mathbb{R}^n$ and the zonotope $\langle c, H \rangle \subset \mathbb{R}^n$, with $c \in \mathbb{R}^n$ and $H \in \mathbb{R}^{n \times m}$. The problem of determining whether a point belongs to a zonotope can be formulated as the following constraint satisfaction problem

$$\begin{aligned} \min_{z} .\ & h(\cdot), \\ \text{s.t.}\ \ & p = c + Hz, \\ & \|z\|_\infty \leq 1, \end{aligned}$$

where $h(\cdot)$ denotes the null function.

A.3 Zonotopic ϵ-approximation of the mRPI set

Below, it is presented the procedure that it used in order to compute an outer zonotopic ϵ-approximation of the mRPI set for a perturbed discrete-time LTI system of the form

$$x_{k+1} = Ax_k + Bw_k, \tag{A.6}$$

where $x_k \in \mathbb{R}^n$ is the state and $w_k \in \mathbb{R}^n$ is an unknown disturbance bounded in the zonotope $w_k \in \mathcal{W} = \langle c_w, H_w \rangle$. Matrices A and B are of suitable dimensions with A being asymptotically stable. Besides, let $\bar{\mathcal{W}}$ denote the zero centered zonotope $\bar{\mathcal{W}} = \langle 0, H_w \rangle$.

The computation of a zonotopic RPI outer approximation of the mRPI set follows the iterative procedures used in Rakovic et al. [2005], Olaru et al. [2010]. This approaches are based on computing an initial RPI, and then recursively propagate it obtaining at each iteration a tighter RPI outer-approximation of the mRPI. Accordingly, below it is discussed the computation of an initial zonotopic RPI set in generators representation, and the computation of a bound in the maximum number of required iterations such that the ∞-norm error is bounded. Note that the forward propagation of zonotopic sets can can be computed trivially by means of Property A.7 and Property A.8.

A.3.1 Initial RPI set

Whenever is possible, the so-called Ultimate Bound (UB) method is used for computing an initial RPI set. Under certain conditions on the closed-loop matrix A, the UB allows to directly compute a first-order zonotopic RPI set. In the following, it is considered that the zonotopic set $\bar{\mathcal{W}}$ is over-approximated by an aligned box, i.e., $|w_k| \leq \bar{w}$ for some nonnegative vector $\bar{w}$. Therefore, the following lemma presented in Seron and De Doná [2015] is introduced.

Lemma A.1. Suppose an invertible (complex) transformation $V \in \mathbb{C}^{n_x \times n_x}$ exists such that the matrix $\Lambda = |V^{-1}AV|$ is strictly stable. Then, the set

$$\mathcal{S} = \{x \in \mathbb{R}^{n_x} : |V^{-1}x| \leq (I_{n_x} - \Lambda)^{-1}|V^{-1}B|\bar{w}\}, \tag{A.7}$$

is an invariant set for (A.6).

From (A.7), it follows that the resulting set $\mathcal{S}$ is a parallelotope, and thus a first-order zonotope, if V is a real matrix $V \in \mathbb{R}^{n_x \times n_x}$. Note that if matrix A has real eigenvalues, by performing the Jordan decomposition $\Lambda = V^{-1}AV$, then matrix V presents real entries.

Furthermore, if A presents complex conjugate pairs of eigenvalues $\lambda_{i,j} = a \pm bi$ that satisfy (A.8), then the following steps can be performed in order to obtain a matrix $\bar{V}$ that satisfies Lemma A.1.

1. Compute Jordan decomposition $\Lambda = V^{-1}AV$, such that Λ is an upper triangular matrix with the eigenvalues of A in the main diagonal, and V presents complex pair of columns $x \pm yi$ associated which each pair of complex conjugate eigenvalues $\lambda_{i,j} = a \pm bi$.
2. Substitute each pair of complex columns $x \pm yi$ by the pair of real vectors (x, y) that span the corresponding 2-dimensional invariant subspace of A [Golub and Van Loan, 2013]. This generates a new non-singular matrix $\bar{V}$ such that $\bar{\Lambda} = \bar{V}^{-1}A\bar{V}$.

Note that matrix $\bar{\Lambda}$ results in an upper triangular matrix whose elements in the main diagonal are the real eigenvalues λ_r of A or 2-by-2 block matrices B_i associated with the complex conjugate pairs $\lambda_{i,j} = a \pm bi$ with the form

$$B_i = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

Lemma A.1 requires that the eigenvalues of $|\bar{\Lambda}|$ satisfy $|\lambda(|\bar{\Lambda}|)| < 1$. From the asymptotic stability of A, it follows that the real eigenvalues λ_r already satisfy that$|\lambda_r| < 1$. On the other hand, for the complex eigenvalues it must be satisfied $|\lambda(|B_i|)| < 1$ with $\lambda(|B_i|)_{1,2} = |a| \pm |b|$. Therefore, $\bar{V}$ is a real matrix satisfying $|\lambda(|\bar{\Lambda}|)| < 1$, and thus $\mathcal{S}$ a zonotopic RPI set, for those complex eigenvalues that satisfy

$$||a| - |b|| \leq ||a| + |b|| = |a| + |b| < 1. \tag{A.8}$$

The parallelotope (A.7), can be easily transformed into a first-order zonotope in generator representation by means of the relationships presented in Althoff et al. [2010].

The procedure presented above offers an straightforward method for obtaining a first-order zonotopic RPI set. For the cases were the eigenvalues of matrix A does not satisfy the conditions aforementioned, the following condition presented in Rakovic et al. [2005] can be employed. From [Rakovic et al., 2005, Theorem 1], it follows that for a given scalar $\alpha \in [0, 1)$ there exists a finite $s \in \mathbb{N}_+$ that satisfies

$$A^s B\bar{\mathcal{W}} \subseteq \alpha B\bar{\mathcal{W}}, \tag{A.9}$$

then, if (A.9) is satisfied, the zonotope $\langle 0, H_0 \rangle$ with

$$H_0 = (1 - \alpha)^{-1}[BH_w,\ ABH_w\ ...\ A^{s-1}BH_w], \tag{A.10}$$

is an RPI set for (A.6). The evaluation of the inclusion of one zonotope in generator representation into another, can be formulated as a convex problem as proposed in [Caro, 2004, Appendix B].

A.3.2 Forward propagation and stopping criterion

This section addresses the computation of a stopping criterion in the recursive propagation of an initial RPI set such that guarantees an ∞-norm ϵ-approximation of the mRPI. In this regard, let us recall the following result presented in Olaru et al. [2010].

Proposition A.1. Consider (A.6) and denote as Φ_0 an RPI initial set for (A.6). Each of the set iterations

$$\Phi_{j+1} = A\Phi_j \oplus B\mathcal{W},$$

where $j \in \mathbb{N}$ denotes the j^{th} element of the sequence, is an RPI approximation of the mRPI set. Moreover, as j tends to infinity, the set sequence converges to the mRPI set.

By means of the previous recursion, a certified outer ϵ-approximation of the mRPI set Φ_m can be obtained by means of [Olaru et al., 2010, Theorem 3.5].

Theorem A.1. For all $\epsilon > 0$ there exists an $l \in \mathbb{N}_+$ such that the following RPI outer ϵ-approximation exists

$$\Phi_m \subset \Phi_l \subset \Phi_m \oplus \mathbb{B}_p^n(\epsilon),$$

where $\mathbb{B}_p^n(\epsilon) = \{x \in \mathbb{R}^n : ||x||_p \leq \epsilon\}$ and $||x||_p$ is the p-norm.

Accordingly, from Appendix A of Olaru et al. [2010], by choosing an $l \in \mathbb{N}$ such that

$$A^l\Phi_0 \subset \mathbb{B}_\infty^n(\epsilon/2), \tag{A.11}$$

it is guaranteed that $\Phi_m \subset \Phi_l \subset \Phi_m \oplus \mathbb{B}_\infty^n(\epsilon)$. Hence, given an $\epsilon > 0$ and particularizing for the ∞-norm, the following holds

$$||A^l\Phi_0||_\infty \leq ||A^l||_\infty ||\Phi_0||_\infty < \epsilon/2 \implies A^l\Phi_0 \subset \mathbb{B}_\infty^n(\epsilon/2). \tag{A.12}$$

By eigendecomposing matrix A as $A = T\Psi T^{-1}$, and expressing its spectral radius as $\rho(A) = ||\Psi||_\infty$, then $||A^l||_\infty$ can be bounded as

$$||A^l||_\infty = ||T\Psi^s T^{-1}||_\infty \leq ||T||_\infty ||T^{-1}||_\infty \rho(A)^l. \tag{A.13}$$

Consequently, replacing (A.13) in (A.12), and computing $\phi = ||\Phi_0||_\infty$ and $\kappa = ||T||_\infty ||T^{-1}||_\infty$, an ϵ-approximation to the mRPI set is guaranteed by choosing

$$l > \frac{log(\epsilon/2) - log(\kappa\phi)}{log(\rho(A))}, \quad l \in \mathbb{N}_+. \tag{A.14}$$

Algorithm A.1 summarizes the procedure for obtaining a zonotopic ϵ-approximation of the mRPI for the system (A.6).

Algorithm A.1 Zonotopic ϵ-approximation of the mRPI set

```
Input: Pair $(A, B)$, parameter $\epsilon > 0$ and the zonotopic disturbance set $\mathcal{W} = \langle c_w, H_w \rangle$
Output: Zonotopic RPI approximation $\mathcal{X}$ of the mRPI
Compute $H_0$ either using (A.7) or by means of (A.10)
Compute $\rho(A)$, $\kappa$ and $\phi = ||H_0||_\infty$
Compute the minimum $l \in \mathbb{N}_+$ such that $l > \big(log(\epsilon/2) - log(\kappa\phi)\big)/log\big(\rho(A)\big)$
for $j = 0$ to $j = l - 1$ do
    Propagate $H_{j+1} = [AH_j \; BH_w]$
end for
Compute the RPI set $\mathcal{X} = \langle c_x, 0 \rangle \oplus \langle 0, H_l \rangle$ with $c_x = (I - A)^{-1}Bc_w$
```

A.4 Lexicographic multi-objective optimization

This section is devoted to present a general overview of a lexicographic multi-objective optimization, and of the approach followed in this thesis for searching a lexicographic minimum. In this regard, the following multi-objective problem $\mathcal{Q}$ is introduced as

$$\mathcal{Q}: \quad \min_{\theta \in \Theta} f(\theta), \tag{A.15}$$

where the vector θ encompasses the different decision variables and $\Theta \subset \mathbb{R}^d$ is its admissible set. The vector-valued objective function $f : \Theta \to \mathbb{R}^p$, is such that $f(\theta) := [f_1(\theta), ..., f_p(\theta)]^T$ being $f_i : \Theta \to \mathbb{R}$ scalar-valued objective functions which are assumed to attain their minima inside $\Theta \neq \emptyset$.

According to Kerrigan and Maciejowski [2002], a given $\theta^* \in \Theta$ is a *lexicographic minimizer* and $f(\theta^*)$ is the *lexicographic minimum* of $\mathcal{Q}$, if and only if there does not exist a $\theta \in \Theta$ and an i^* satisfying $f_{i^*}(\theta) < f_{i^*}(\theta^*)$ and $f_i(\theta) = f_i(\theta^*)$, $i = \{1, ..., i^* - 1\}$.

An interpretation of the above definition is that a solution is a lexicographic minimum if and only if an objective f_i can be reduced only at the expense of increasing at least one of the higher-prioritized objectives $\{f_1, ..., f_{i-1}\}$. For a problem $\mathcal{Q}$ a lexicographic minimizer exists and the lexicographic minimum is unique [Kerrigan and Maciejowski, 2002].

Among the different approaches to find the lexicographic minimum, throughout this dissertation, the sequential solution method (see Ocampo-Martinez et al. [2007b]), which is summarized in Algorithm A.2, is employed.

Algorithm A.2 Lexicographic multi-objective optimization using the sequential solution method

```
$f_1^* = min_{\theta\in\Theta} f_1(\theta)$
for $i = 2$ to $r$ do
    $f_i^* = min_{\theta\in\Theta}\{f_i(\theta) \mid f_j(\theta) \leq f_j^*,\ j = 1, ..., i-1\}$
end for
Determine the lexicographic minimizer set as:
$\theta^* \in \{\theta \in \Theta \mid f_j(\theta) \leq f_j(\theta^*), j = 1, ..., r\}$
```

Note from step 5 of Algorithm A.2 that the lexicographic minimizer is not guaranteed to be unique. A sufficient condition for guaranteeing the uniqueness of the lexicographic minimizer is that at least one of the cost functions f_i is strictly convex and attains its minima inside $\Theta \neq \emptyset$. For this particular case, the minimizer is obtained as

$$\theta^* = \arg\min_{\theta \in \Theta}\{f_i(\theta) \mid f_j(\theta) \leq f_j^*,\ j = 1, ..., i-1\}, \tag{A.16}$$

and therefore there will be no need for continue Algorithm A.2 for the remaining steps $\{f_{i+1}^*, ..., f_p^*\}$.

Appendix B

Case study description

The analysis and the algorithms developed in this thesis have been mainly validated by means of two different case studies. The first one is the four-tank system proposed in Johansson [2000] which constitutes a well-known benchmark used to evaluate control and supervision strategies. Recently, this system also has been employed to assess cyber-attack detection techniques [Navarro, 2011, Teixeira et al., 2015a, Sánchez et al., 2019b]. The second case study is an aggregated version of the Barcelona drinking water network (DWN). This network represents a constrained large-scale system, that has been extensively used in recent years in order to evaluate different control strategies [Ocampo-Martinez et al., 2012, 2013, Pereira et al., 2016a].

B.1 Four-tank system

The four-tank is a multi-input/multi-output process proposed by Johansson [2000] as a control benchmark which consists of four interconnected tanks, two pumps and two valves whose value is set prior to the experiment. Figure B.1 represents a schematic diagram of the system setup.

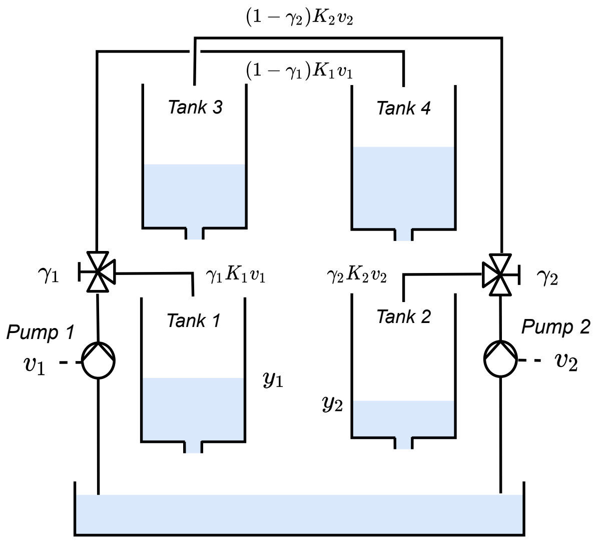


Figure B.1: Schematic representation of the four-tank process.

The four-tank process operates as follows: tank levels are manipulated by means of the voltages v_1 and v_2 applied to Pumps 1 and 2, respectively. These pumps extract water from the basin and pour it into the different tanks. Furthermore, the upper tanks (named Tank 3 and Tank 4) release water into the lower ones (Tank 1 and Tank 2), contributing to coupling the system dynamics. The available measurements are the water levels in Tank 1 and Tank 2 denoted as y_1 and y_2 respectively, which are obtained as voltages from the sensors. Moreover, the amount of water pumped into the different tanks depends on the degree of opening of the two valves, which can vary between 0 and 1, i.e., $0 \leq \gamma_i \leq 1$ with $i = 1, 2$.

Regarding the physical features, the input voltages v_1 and v_2 vary between 0V and 10 V, the height of each tank is 20 cm and the connection of the tank and the pumps is done using a piper with a diameter equal to 6 mm.

From a mathematical point of view, mass balances and Bernoulli's law yield the following set of non-linear differential equations

$$\begin{aligned}
\frac{dh_1(t)}{dt} &= -\frac{a_1}{A_1}\sqrt{2gh_1(t)} + \frac{a_3}{A_1}\sqrt{2gh_3(t)} + \frac{\gamma_1 K_1}{A_1}v_1(t),\\
\frac{dh_2(t)}{dt} &= -\frac{a_2}{A_2}\sqrt{2gh_2(t)} + \frac{a_4}{A_2}\sqrt{2gh_4(t)} + \frac{\gamma_2 K_2}{A_2}v_2(t),\\
\frac{dh_3(t)}{dt} &= -\frac{a_3}{A_3}\sqrt{2gh_3(t)} + \frac{(1-\gamma_2)\,K_2}{A_3}v_2(t),\\
\frac{dh_4(t)}{dt} &= -\frac{a_4}{A_4}\sqrt{2gh_4(t)} + \frac{(1-\gamma_1)\,K_1}{A_4}v_1(t),\\
y_1(t) &= k_c h_1(t),\\
y_2(t) &= k_c h_2(t),
\end{aligned} \tag{B.1}$$

where

- h_i is the water level in the tank,
- A_i is the cross-section of the tank,
- a_i is the cross-section of the outlet hole,
- K_i is the constant of the pump,
- y_i is the measured signal,
- g is the acceleration due to gravity.

Parameter	**Value**	**Unit**
$A_1 = A_3$	28	cm^2
$A_2 = A_4$	32	cm^2
$a_1 = a_3$	0.071	cm^2
$a_2 = a_4$	0.057	cm^2
k_c	0.5	V/cm
g	981	cm/s^2

Table B.1: Four-tank system parameters.

Variable	**Value**	**Unit**
(h_1^*, h_2^*)	$(12.4, 12.7)$	cm
(h_3^*, h_4^*)	$(1.8, 1.4)$	cm
(v_1^*, v_2^*)	$(3, 3)$	V
(k_1^*, k_2^*)	$(3.33, 3.35)$	cm^3/Vs
(γ_1^*, γ_2^*)	$(0.7, 0, 6)$	

Table B.2: Four-tank operating point.

The parameter values of model (B.1) are given in Table B.1. Furthermore, the non-linear model (B.1) is linearized around the operating point that is shown in Table B.2 and denoted by means of the superscript *. Therefore, by considering the incremental state $\Delta h_i(t) = h_i(t) - h_i^*$ and control variable $\Delta u_i(t) = v_i(t) - v_i^*$, the linearized state-space equation is then given by

$$\begin{aligned}\Delta \dot{h}(t) &= \begin{bmatrix} -\frac{1}{T_1} & 0 & \frac{A_3}{A_1 T_3} & 0 \\ 0 & -\frac{1}{T_2} & 0 & \frac{A_4}{A_2 T_4} \\ 0 & 0 & -\frac{1}{T_3} & 0 \\ 0 & 0 & 0 & -\frac{1}{T_4} \end{bmatrix} \Delta h(t) + \begin{bmatrix} \frac{\gamma_1 k_1}{A_1} & 0 \\ 0 & \frac{\gamma_2 k_2}{A_2} \\ 0 & \frac{(1-\gamma_1)k_2}{A_3} \\ \frac{(1-\gamma_1)k_1}{A_4} & 0 \end{bmatrix} \Delta u(t), \\ y(t) &= \begin{bmatrix} k_c & 0 & 0 & 0 \\ 0 & k_c & 0 & 0 \end{bmatrix} \Delta h(t), \end{aligned} \tag{B.2}$$

with $T_i = \frac{A_i}{a_i}\sqrt{\frac{2h_i^*}{g}}$ for $i = 1, 2, 3, 4$.

By discretizing the continuous-time model (B.2) using Euler discretization with a sampling time of 1 second, a discrete-time linear time-invariant system is obtained as

$$\Delta h_{k+1} = A\Delta h_k + B\Delta u_k + E w_k, \tag{B.3a}$$

$$y_k = C\Delta h_k + D\Delta u_k + F v_k, \tag{B.3b}$$

where

$$A = \begin{bmatrix} 0.9841 & 0 & 0.0419 & 0 \\ 0 & 0.9889 & 0 & 0.0333 \\ 0 & 0 & 0.9581 & 0 \\ 0 & 0 & 0 & 0.9667 \end{bmatrix}, \quad B = \begin{bmatrix} 0.2102 & 0 \\ 0 & 0.0628 \\ 0 & 0.0479 \\ 0.0094 & 0 \end{bmatrix},$$

$$C = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

whereas w_k and v_k represent the presence of unknown but bounded process disturbances and measurements noise, respectively. The effect of these uncertainty sources in the evolution of the states and in the outputs signal is modelled by means of the associated distribution matrices E and F with appropriate dimensions.

B.2 The Barcelona Drinking Water Transport Network

According to the 2009 technical report presented in Fambrini and Ocampo-Martinez [2009], the drinking water network of Barcelona covers a territorial extension of 425km^2, with a total pipe length of 4.470km. Every year, the network supplies 237.7hm^3 of drinking water to a population over 2.8 millions of inhabitants. The Barcelona DWN is structurally organised in two layers. The upper layer, termed as *transport network*, links the water treatment plants with the reservoirs distributed all over the city. The lower layer, named *distribution network*, is sectorised in subnetworks. Each of these subnetworks links a reservoir with each consumer. The case study used in this Ph.D. dissertation focuses on the drinking water transport network (DWTN). Hence, each one of the subnetworks that constitute the distribution network, is modelled as a demand sector. Moreover, the water demand of each sector follows a cyclic demand pattern, which can be predicted by using a time-series model [Quevedo et al., 2010].

The water sources of the network are the Ter and Llobregat rivers, which are regulated at their head by some dams with an overall capacity of 600 cubic hectometres. There are four drinking water treatment plants: the Abrera and Sant Joan Despí plants, which extract water from the Llobregat river, the Cardedeu plant, which extracts water from Ter river, and the Besòs plant, which treats the underground flows from the aquifer of the Besòs river. Additionally, there are also several underground sources (wells) that can provide water through pumping stations. Those different water sources currently provide a flow of around 7m^3/s.

The control system of the transport network presents a two-level architecture. At the upper layer, a supervisory control system is in charge of the global control of the network. This upper layer controller establishes the set-points of the regulatory controllers allocated at the lower layer. Regulatory controllers are of PID type, and optimize the pressure profile to minimize losses due leakage and to provide sufficient water pressure [Ocampo-Martinez et al., 2009]. Besides, low-level controllers hide the non-linear behaviour of the network from the supervisory layer, allowing the supervisory controller to make use of the control-oriented linear model described in the following section.

B.2.1 Control-oriented model of the DWTN

Below, the constitutive elements, basic relationships and main steps followed in order to obtain the control-oriented model of a DWTN, are introduced. The reader is referred to Ocampo-Martinez et al. [2009], Caini et al. [2009], Fambrini and Ocampo-Martinez [2009] for further details on DWTN modelling and specific insights related to the aggregated network case study.

B.2.1.1 Tanks

The stored water volume in the tanks represents the system state x_k in a state-space model of the network. On this subject, the dynamics of the network are governed by the mass balance expressions that relate the stored volume in tanks with the manipulated tank inflows/outflows u_k, and the water demands d_k. This can be written for the i-th tanks as

$$x_{k+1}^i = x_k^i + \Delta t\left(\sum_i q_k^{in,i} - \sum_j q_k^{out,j}\right), \tag{B.4}$$

where $q_k^{in,i}$ and $q_k^{out,j}$ correspond to the i-th inflow and the j-th outflow, given in m^3/s. The physical constraints that limit the range of tank volume capacities are expressed as

$$\underline{x} \leq x_k \leq \bar{x},$$

where $\underline{x}$ and $\bar{x}$ denote the minimum and maximum volume capacity, respectively, given in m^3. Note that, this represents a physical constraint, in such a way that it is impossible to send more water to a tank than it can store, or draw more water than the stored amount.

B.2.1.2 Actuators

The control actuators can be either pumps or valves. The flow manipulated by means of the actuators represents the control input variables of the model u_k. Both, pumps and valves, present lower and upper physical limits which are modelled as the constraints

$$\underline{u} \leq u_k \leq \bar{u},$$

where $\underline{u}$ and $\bar{u}$ denote the minimum and maximum flow capacity, respectively, given in m^3/s. Notice that the pumping station flow is modelled as a continuous variable capable to adopt any value within a certain range. In this regard, the pumping stations, which have an ON-OFF behaviour, should implement an additional scheduling procedure for individual pump operation in order to produce the desired flow. This pumping scheduling is not addressed here, since it can be placed in the regulatory level.

B.2.1.3 Nodes

These elements correspond to the network points where water flows are merged or split. The static equation that reflects the mass conservation in these elements can be written as

$$\sum_i q_k^{in,i} = \sum_j q_k^{out,j}, \tag{B.5}$$

where $q_k^{in,i}$ and $q_k^{out,j}$ correspond to the i-th inflow to the node and the j-th outflow of the node, in m^3/s. Note that, the node inflows and outflows can be a combination of manipulated flows u_k and demand flow d_k.

B.2.1.4 Sector of consume

A sector of consume represents the water demand made by the network users of a specific physical area. Its effect on the network dynamics is modelled as a system disturbance. Since the demand shows a periodic behaviour, with daily and weekly seasonalities, it can be forecasted by using methods based on time series [Quevedo et al., 2010].

B.2.1.5 Network model

Considering the expressions presented above, the control-oriented model of a DWTN in discrete time can be written as

$$x_{k+1} = Ax_k + Bu_k + B_d d_k + B_w w_k, \tag{B.6a}$$
$$0 = Eu_k + E_d d_k + E_w w_k, \tag{B.6b}$$

where $x_k \in \mathbb{R}^{n_x}$ is the state space vector corresponding to the water volumes of the n_x tanks, $u_k \in \mathbb{R}^{n_u}$ represents the vector of manipulated flows through the n_u actuators, $d_k \in \mathbb{R}^{n_d}$ describes the predicted flow demand on the sectors of consume and $w_k \in \mathbb{R}^{n_w}$ characterizes the presence of unknown disturbances. Matrices $A, B, B_d, B_w, E, E_d, E_w$ are of suitable dimensions.

Notice that the difference equation (B.6a) is derived from the mass balance in the tanks presented in (B.4), whereas the algebraic equation (B.6b) results from the static relationships in the nodes described in (B.5). Also note that, for the case where all the network flows are manipulated, then matrix A is an identity matrix with suitable dimensions.

B.2.2 The aggregated case study

The case study used in Chapters 6 and 7 is the *aggregated* version of the Barcelona DWTN [Caini et al., 2009, Fambrini and Ocampo-Martinez, 2009]. This simplified system, constitutes a representative version of the entire network. In order to generate the aggregate model, some of the consumer demand sectors are concentrated in a single point. Similarly, some of the tanks of the overall network are aggregated in a single element, while the respective actuators are considered as a single pumping station or valve.

Figure B.2 displays the aggregated model of the network. This model is made up of 17 tanks, 61 actuators (comprised of 26 pumping stations and 35 valves), 11 nodes and 25 main sectors of water demands. The model has been simulated and compared against the real behaviour of the network, assessing thus its validity. The physical limits of the network elements, as well as the equations that described the system, are reported in the corresponding Appendices of Fambrini and Ocampo-Martinez [2009]. In addition, the predicted water demand that is used in the simulations has been obtained from historic data, and it can be found in the supplementary material of Pereira et al. [2016a].

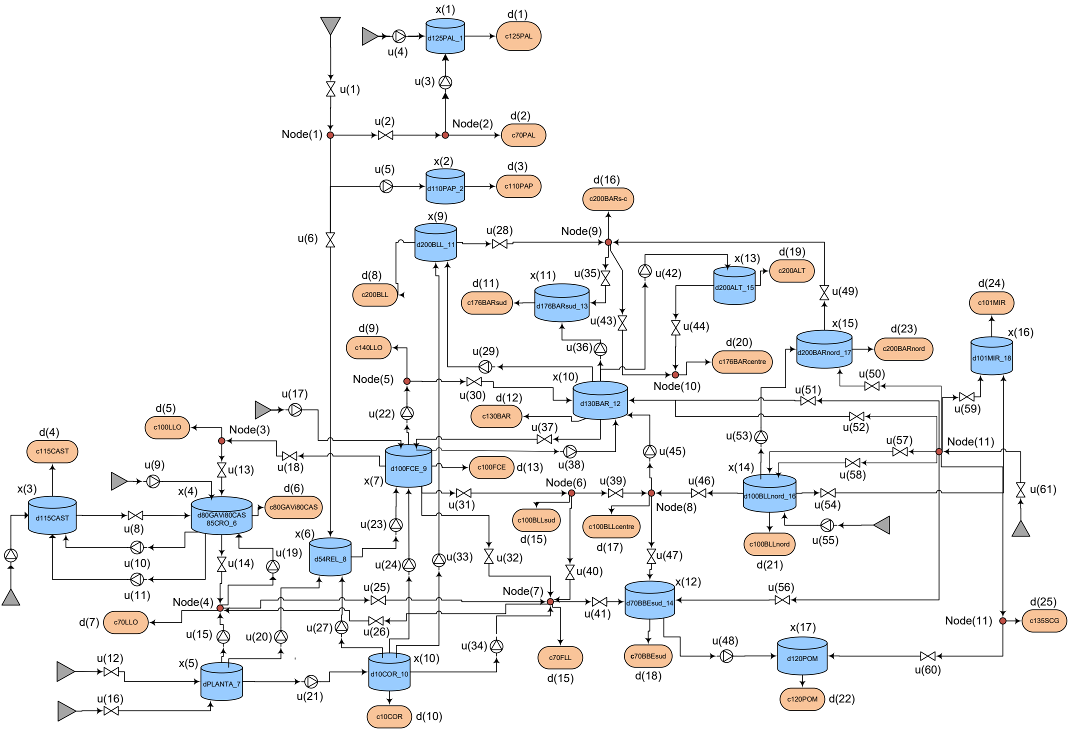


Figure B.2: Aggregated Barcelona DWN. Tanks (blue); Demand Sectors (orange); Sources (grey); Nodes (red).

Bibliography

A. Abdelwahab, W. Lucia, and A. Youssef. Set-theoretic control for active detection of replay attacks with applications to smart grid. In *2020 IEEE Conference on Control Technology and Applications (CCTA)*, pages 1004–1009. IEEE, 2020.

M. Abrams and J. Weiss. Malicious control system cyber security attack case study-maroochy water services, australia. Technical report, MITRE CORP MCLEAN VA MCLEAN, 2008.

T. Alamo, J. M. Bravo, and E. F. Camacho. Guaranteed state estimation by zonotopes. *Automatica*, 41(6):1035–1043, 2005.

M. Althoff, O. Stursberg, and M. Buss. Computing reachable sets of hybrid systems using a combination of zonotopes and polytopes. *Nonlinear analysis: hybrid systems*, 4(2):233–249, 2010.

I. Alvarado, D. Limon, D. M. de la Pena, T. Alamo, and E. Camacho. Enhanced iss nominal mpc based on constraint tightening for constrained linear systems. *UKACC International Conference on Control*, 2010.

D. Angeli and M. A. Müller. Economic model predictive control: Some design tools and analysis techniques. In *Handbook of Model Predictive Control*, pages 145–167. Springer, 2019.

D. Angeli, R. Amrit, and J. B. Rawlings. On average performance and stability of economic model predictive control. *IEEE transactions on automatic control*, 57(7):1615–1626, 2011.

J. Askari, B. Heiming, and J. Lunze. Controller reconfiguration based on a qualitative model: A solution of three-tanks benchmark problem. In *1999 European Control Conference (ECC)*, pages 4035–4040. IEEE, 1999.

C.-Z. Bai, F. Pasqualetti, and V. Gupta. Security in stochastic control systems: Fundamental limitations and performance bounds. In *2015 American Control Conference (ACC)*, pages 195–200. IEEE, 2015.

F. A. Bayer, M. Lorenzen, M. A. Müller, and F. Allgöwer. Robust economic model predictive control using stochastic information. *Automatica*, 74:151–161, 2016a.

F. A. Bayer, M. A. Müller, and F. Allgöwer. Min-max economic model predictive control approaches with guaranteed performance. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 3210–3215. IEEE, 2016b.

A. Bemporad and M. Morari. Control of systems integrating logic, dynamics, and constraints. *Automatica*, 35(3):407–427, 1999.

D. P. Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3): 334–334, 1997.

R. R. Bitmead and M. Gevers. Riccati difference and differential equations: Convergence, monotonicity and stability. In *The Riccati Equation*, pages 263–291. Springer, 1991.

S. Bittanti and P. Colaneri. *Periodic systems: filtering and control*, volume 5108985. Springer Science & Business Media, 2009.

F. Blanchini and S. Miani. *Set-theoretic methods in control*. Springer, 2008.

M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, and J. Schröder. *Diagnosis and fault-tolerant control*, volume 2. Springer, 2006.

Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE transactions on pattern analysis and machine intelligence*, 26(9):1124–1137, 2004.

E. Byres and J. Lowe. The myths and facts behind cyber security risks for industrial control systems. In *Proceedings of the VDE Kongress*, volume 116, pages 213–218. Citeseer, 2004.

E. Caini, V. Puig, and G. Cembrano. Development of a simulation environment for drinking water networks: Application to the validation of a centralized mpc controller for the barcelona case study. *Institut de Robotica i Informatica Industrial (CSIC-UPC), Tech. Rep. IRI-TR-03-09*, 2009.

A. A. Cárdenas, S. Amin, and S. Sastry. Research challenges for the security of control systems. In *HotSec*, 2008.

A. A. Cardenas, S. Amin, and S. Sastry. Secure control: Towards survivable cyber-physical systems. In *2008 The 28th International Conference on Distributed Computing Systems Workshops*, pages 495–500. IEEE, 2008.

J. M. B. Caro. *Control predictivo no lineal robusto basado en técnicas intervalares*. PhD thesis, Universidad de Sevilla, 2004.

D. U. Case. Analysis of the cyber attack on the ukrainian power grid. *Electricity Information Sharing and Analysis Center (E-ISAC)*, 388, 2016.

L. Chisci, J. A. Rossiter, and G. Zappa. Systems with persistent disturbances: predictive control with restricted constraints. *Automatica*, 37(7):1019–1028, 2001.

C. Combastel. A state bounding observer based on zonotopes. In *2003 European Control Conference (ECC)*, pages 2589–2594. IEEE, 2003.

C. Combastel. A state bounding observer for uncertain non-linear continuous-time systems based on zonotopes. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 7228–7234. IEEE, 2005.

C. Combastel. Zonotopes and kalman observers: Gain optimality under distinct uncertainty paradigms and robust convergence. *Automatica*, 55:265–273, 2015.

C. Combastel. An extended zonotopic and gaussian kalman filter (ezgkf) merging set-membership and stochastic paradigms: Toward non-linear filtering and fault detection. *Annual Reviews in Control*, 42:232–243, 2016.

D. Dobkin, J. Hershberger, D. Kirkpatrick, and S. Suri. Computing the intersection-depth of polyhedra. *Algorithmica*, 9(6):518–533, 1993.

J. Edmonds and R. M. Karp. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of the ACM (JACM)*, 19(2):248–264, 1972.

P. M. Esfahani, M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson. Cyber attack in a two-area power system: Impact identification using reachability. In *Proceedings of the 2010 American control conference*, pages 962–967. IEEE, 2010.

V. Fambrini and C. Ocampo-Martinez. Modelling and decentralized model predictive control of drinking water networks. *Institut de Robotica i Informatica Industrial (CSIC-UPC), Tech. Rep. IRI-TR-04-09*, 2009.

C. Fang, Y. Qi, P. Cheng, and W. X. Zheng. Optimal periodic watermarking schedule for replay attack detection in cyber–physical systems. *Automatica*, 112:108698, 2020.

H. Fawzi, P. Tabuada, and S. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic control*, 59(6):1454–1467, 2014.

R. M. Ferrari and A. M. Teixeira. Detection and isolation of replay attacks through sensor watermarking. *IFAC-PapersOnLine*, 50(1):7363–7368, 2017a.

R. M. Ferrari and A. M. Teixeira. Detection and isolation of routing attacks through sensor watermarking. In *2017 American Control Conference (ACC)*, pages 5436–5442. IEEE, 2017b.

R. M. Ferrari and A. M. Teixeira. A switching multiplicative watermarking scheme for detection of stealthy cyber-attacks. *IEEE Transactions on Automatic Control*, 2020.

L. R. Ford Jr and D. R. Fulkerson. *Flows in networks*. Princeton university press, 1962.

J. Fortuny-Amat and B. McCarl. A representation and economic interpretation of a two-level programming problem. *Journal of the operational Research Society*, 32(9):783–792, 1981.

G. F. Franklin, J. D. Powell, A. Emami-Naeini, and J. D. Powell. *Feedback control of dynamic systems*, volume 4. Prentice hall Upper Saddle River, NJ, 2002.

G. Franzè, F. Tedesco, and W. Lucia. Resilient control for cyber-physical systems subject to replay attacks. *IEEE Control Systems Letters*, 3(4):984–989, 2019.

A. Gehin and M. Staroswiecki. A formal approach to reconfigurability analysis application to the three tank benchmark. In *1999 European Control Conference (ECC)*, pages 4041–4046. IEEE, 1999.

A.-L. Gehin and M. Staroswiecki. Reconfiguration analysis using generic component models. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(3): 575–583, 2008.

J. Gertler. *Fault detection and diagnosis in engineering systems*. CRC press, 1998.

M. Ghaderi, K. Gheitasi, and W. Lucia. A novel control architecture for the detection of false data injection attacks in networked control systems. In *2019 American Control Conference (ACC)*, pages 139–144. IEEE, 2019.

J. Giraldo, A. Cardenas, and R. G. Sanfelice. A moving target defense to detect stealthy attacks in cyber-physical systems. In *2019 American Control Conference (ACC)*, pages 391–396. IEEE, 2019.

G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU press, 2013.

P. Griffioen, S. Weerakkody, and B. Sinopoli. A moving target defense for securing cyber-physical systems. *IEEE Transactions on Automatic Control*, 2020.

J. Grosso, C. Ocampo-Martínez, V. Puig, and B. Joseph. Chance-constrained model predictive control for drinking water networks. *Journal of process control*, 24(5):504–516, 2014.

J. M. Grosso Pérez. *On model predictive control for economic and robust operation of generalised flow-based networks*. PhD thesis, Universitat Politècnica de Catalunya, 2015.

J. D. Hamilton. *Time series analysis*. Princeton university press, 2020.

N. Hashemi and J. Ruths. Generalized chi-squared detector for lti systems with non-gaussian noise. In *2019 American Control Conference (ACC)*, pages 404–410. IEEE, 2019.

B. Heiming and J. Lunze. Definition of the three-tank benchmark problem for controller reconfiguration. In *1999 European Control Conference (ECC)*, pages 4030–4034. IEEE, 1999.

P. Hespanhol, M. Porter, R. Vasudevan, and A. Aswani. Dynamic watermarking for general lti systems. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 1834–1839. IEEE, 2017.

R. Huang, E. Harinath, and L. T. Biegler. Lyapunov stability of economically oriented nmpc for cyclic processes. *Journal of Process Control*, 21(4):501–509, 2011.

T. Huang, B. Satchidanandan, P. Kumar, and L. Xie. An online detection framework for cyber attacks on automatic generation control. *IEEE Transactions on Power Systems*, 33(6):6816–6827, 2018.

I. Hwang, S. Kim, Y. Kim, and C. E. Seah. A survey of fault detection, isolation, and reconfiguration methods. *IEEE transactions on control systems technology*, 18(3):636–653, 2009.

I. ILOG. Cplex optimizer 12.8, 2018.

R. Isermann. Model-based fault-detection and diagnosis–status and applications. *Annual Reviews in control*, 29(1):71–85, 2005.

S. Jajodia, A. K. Ghosh, V. Swarup, C. Wang, and X. S. Wang. *Moving target defense: creating asymmetric uncertainty for cyber threats*, volume 54. Springer Science & Business Media, 2011.

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

S. Kalambe and G. Agnihotri. Loss minimization techniques used in distribution network: bibliographical survey. *renewable and sustainable energy reviews*, 29:184–200, 2014.

A. Kanellopoulos and K. G. Vamvoudakis. A moving target defense control framework for cyber-physical systems. *IEEE Transactions on Automatic Control*, 65(3):1029–1043, 2019.

F. Kennel, D. Görges, and S. Liu. Energy management for smart grids with electric vehicles based on hierarchical mpc. *IEEE Transactions on industrial informatics*, 9(3):1528–1537, 2012.

E. C. Kerrigan and J. M. Maciejowski. Designing model predictive controllers with prioritised constraints and objectives. In *Proceedings. IEEE International Symposium on Computer Aided Control System Design*, pages 33–38. IEEE, 2002.

A. Khazraei, H. Kebriaei, and F. R. Salmasi. A new watermarking approach for replay attack detection in lqg systems. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 5143–5148. IEEE, 2017.

W.-H. Ko, B. Satchidanandan, and P. Kumar. Theory and implementation of dynamic watermarking for cybersecurity of advanced transportation systems. In *2016 IEEE Conference on Communications and Network Security (CNS)*, pages 416–420. IEEE, 2016.

I. Kolmanovsky and E. G. Gilbert. Theory and computation of disturbance invariant sets for discrete-time linear systems. *Mathematical problems in engineering*, 4(4):317–367, 1998.

E. Kontouras, A. Tzes, and L. Dritsas. Cyber-attack on a power plant using bias injected measurements. In *2017 American Control Conference (ACC)*, pages 5507–5512. IEEE, 2017.

E. Kontouras, T. Anthony, and L. Dritsas. Set-theoretic detection of data corruption attacks on cyber physical power systems. *Journal of Modern Power Systems and Clean Energy*, 6(5): 872–886, 2018.

W. Kühn. Rigorously computed orbits of dynamical systems without the wrapping effect. *Computing*, 61(1):47–67, 1998.

C. Kwon, W. Liu, and I. Hwang. Security analysis for cyber-physical systems against stealthy deception attacks. In *2013 American control conference*, pages 3344–3349. IEEE, 2013.

A. Lalami and C. Combastel. Generation of set membership tests for fault diagnosis and evaluation of their worst case sensitivity. *IFAC Proceedings Volumes*, 39(13):569–574, 2006.

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

R. Langner. To kill a centrifuge: A technical analysis of what stuxnet's creators tried to achieve. *The Langner Group*, 2013.

V. T. H. Le, C. Stoica, T. Alamo, E. F. Camacho, and D. Dumur. *Zonotopes: From guaranteed state-estimation to control*. John Wiley & Sons, 2013.

J. H. Lee, S. Natarajan, and K. S. Lee. A model-based predictive control approach to repetitive control of continuous processes with periodic operations. *Journal of Process Control*, 11(2): 195–207, 2001.

R. M. Lee, M. J. Assante, and T. Conway. German steel mill cyber attack. *Industrial Control Systems*, 30:62, 2014.

D. Limón, I. Alvarado, T. Alamo, and E. F. Camacho. Mpc for tracking piecewise constant references for constrained linear systems. *Automatica*, 44(9):2382–2387, 2008.

D. Limon, M. Pereira, D. M. De La Peña, T. Alamo, and J. M. Grosso. Single-layer economic model predictive control for periodic operation. *Journal of Process Control*, 24(8):1207–1224, 2014.

C.-T. Lin. Structural controllability. *IEEE Transactions on Automatic Control*, 19(3):201–208, 1974.

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

L. Liu, L. Ma, Y. Wang, J. Zhang, and Y. Bo. Distributed set-membership filtering for time-varying systems under constrained measurements and replay attacks. *Journal of the Franklin Institute*, 357(8):4983–5003, 2020b.

Y. Liu, P. Ning, and M. K. Reiter. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security (TISSEC)*, 14 (1):13, 2011.

J. Lofberg. Approximations of closed-loop minimax mpc. In *42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH37475)*, volume 2, pages 1438–1442. IEEE, 2003.

W. Lucia, B. Sinopoli, and G. Franze. A set-theoretic approach for secure and resilient control of cyber-physical systems subject to false data injection attacks. In *2016 Science of Security for Cyber-Physical Systems Workshop (SOSCYPS)*, pages 1–5. IEEE, 2016.

J. Lunze and J. H. Richter. Reconfigurable fault-tolerant control: a tutorial introduction. *European journal of control*, 14(5):359–386, 2008.

J. Lunze and J. Schröder. Qualitative diagnosis of the 3-tanks system. In *1999 European Control Conference (ECC)*, pages 4450–4455. IEEE, 1999.

H. A. Mahmoud, Z. Kapelan, and D. Savić. Real-time operational response methodology for reducing failure impacts in water distribution systems. *Journal of Water Resources Planning and Management*, 144(7):04018029, 2018.

T. Marcu, M. Matcovschi, and P. Frank. Neural observer-based approach to fault-tolerant control of a three-tank system. In *1999 European Control Conference (ECC)*, pages 4053–4058. IEEE, 1999.

R. T. Marler and J. S. Arora. Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization*, 26(6):369–395, 2004.

D. Q. Mayne and W. Schroeder. Robust time-optimal control of constrained linear systems. *Automatica*, 33(12):2103–2118, 1997.

D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36(6):789–814, 2000.

D. Q. Mayne, M. M. Seron, and S. Raković. Robust model predictive control of constrained linear systems with bounded disturbances. *Automatica*, 41(2):219–224, 2005.

F. Miao, M. Pajic, and G. J. Pappas. Stochastic game approach for replay attack detection. In *52nd IEEE conference on decision and control*, pages 1854–1859. IEEE, 2013.

F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas. Coding sensor outputs for injection attacks detection. In *53rd IEEE Conference on Decision and Control*, pages 5776–5781. IEEE, 2014.

D. Mignone. *Control and estimation of hybrid systems with mathematical optimization.* PhD thesis, ETH Zurich, 2002.

J. Milošević, H. Sandberg, and K. H. Johansson. Estimating the impact of cyber-attack strategies for stochastic networked control systems. *IEEE Transactions on Control of Network Systems*, 7(2):747–757, 2019.

S. Mishra, D. Das, and S. Paul. A comprehensive review on power distribution network reconfiguration. *Energy Systems*, 8(2):227–284, 2017.

Y. Mo and B. Sinopoli. Secure control against replay attacks. In *2009 47th annual Allerton conference on communication, control, and computing (Allerton)*, pages 911–918. IEEE, 2009.

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

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

Y. Mo, S. Weerakkody, and B. Sinopoli. Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Systems Magazine*, 35(1):93–109, 2015.

C. Murguia and J. Ruths. Cusum and chi-squared attack detection of compromised sensors. In *2016 IEEE Conference on Control Applications (CCA)*, pages 474–480. IEEE, 2016.

C. Murguia, I. Shames, J. Ruths, and D. Nešić. Security metrics and synthesis of secure control systems. *Automatica*, 115:108757, 2020.

M. Nassourou, J. Blesa, and V. Puig. Robust economic model predictive control based on a zonotope and local feedback controller for energy dispatch in smart-grids considering demand uncertainty. *Energies*, 13(3):696, 2020.

A. F. Navarro. Security analysis of a wireless quadruple tank control system. Master's thesis, KTH Royal Institute of Technology, 2011.

H. Niemann. A setup for active fault diagnosis. *IEEE Transactions on Automatic Control*, 51 (9):1572–1578, 2006.

C. Ocampo-Martinez, P. Guerra, V. Puig, and J. Quevedo. Actuator fault-tolerance evaluation of linear constrained model predictive control using zonotope-based set computations. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 221(6):915–926, 2007a.

C. Ocampo-Martinez, A. Ingimundarson, V. Puig, and J. Quevedo. Objective prioritization using lexicographic minimizers for mpc of sewer networks. *IEEE Transactions on Control Systems Technology*, 16(1):113–121, 2007b.

C. Ocampo-Martinez, V. Puig, G. Cembrano, R. Creus, and M. Minoves. Improving water management efficiency by using optimization-based control strategies: the barcelona case study. *Water science and technology: water supply*, 9(5):565–575, 2009.

C. Ocampo-Martinez, D. Barcelli, V. Puig, and A. Bemporad. Hierarchical and decentralised model predictive control of drinking water networks: Application to barcelona case study. *IET control theory & applications*, 6(1):62–71, 2012.

C. Ocampo-Martinez, V. Puig, G. Cembrano, and J. Quevedo. Application of predictive control strategies to the management of complex networks in the urban water cycle [applications of control]. *IEEE Control Systems Magazine*, 33(1):15–41, 2013.

S. Olaru, J. A. De Doná, M. M. Seron, and F. Stoican. Positive invariant sets for fault tolerant multisensor control schemes. *International Journal of Control*, 83(12):2622–2640, 2010.

O. Ozel, S. Weerakkody, and B. Sinopoli. Physical watermarking for securing cyber physical systems via packet drop injections. In *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 271–276. IEEE, 2017.

F. Pasqualetti, F. Dörfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE transactions on automatic control*, 58(11):2715–2729, 2013.

M. Pereira, D. M. de la Peña, D. Limon, I. Alvarado, and T. Alamo. Application to a drinking water network of robust periodic mpc. *Control Engineering Practice*, 57:50–60, 2016a.

M. Pereira, D. M. de la Pena, D. Limón, I. Alvarado, and T. Alamo. Robust model predictive controller for tracking changing periodic signals. *IEEE Transactions on Automatic Control*, 62(10):5343–5350, 2016b.

M. Pereira, D. M. De La Pena, and D. Limón. Robust economic model predictive control of a community micro-grid. *Renewable Energy*, 100:3–17, 2017.

M. Porter, A. Joshi, P. Hespanho, A. Aswani, M. Johnson-Roberson, and R. Vasudevan. Simulation and real-world evaluation of attack detection schemes. In *2019 American Control Conference (ACC)*, pages 551–558. IEEE, 2019.

M. Porter, P. Hespanhol, A. Aswani, M. Johnson-Roberson, and R. Vasudevan. Detecting generalized replay attacks via time-varying dynamic watermarking. *IEEE Transactions on Automatic Control*, 2020.

M. Pourasghar, C. Combastel, V. Puig, and C. Ocampo-Martinez. Fd-zkf: A zonotopic kalman filter optimizing fault detection rather than state estimation. *Journal of Process Control*, 73: 89–102, 2019.

M. Pourasghar-Lafmejani. *On the Fault Diagnosis of Dynamic Systems Using Set-based Approaches.* PhD thesis, Universitat Polièctinca de Catalunya (UPC), 2019.

Y. Qu and K. Pang. State estimation for a class of artificial neural networks subject to mixed attacks: A set-membership method. *Neurocomputing*, 411:239–246, 2020.

J. Quevedo, V. Puig, G. Cembrano, J. Blanch, J. Aguilar, D. Saporta, G. Benito, M. Hedo, and A. Molina. Validation and reconstruction of flow meter data in the barcelona water distribution network. *Control Engineering Practice*, 18(6):640–651, 2010.

S. V. Rakovic, E. C. Kerrigan, K. I. Kouramas, and D. Q. Mayne. Invariant approximations of the minimal robust positively invariant set. *IEEE Transactions on automatic control*, 50(3): 406–410, 2005.

L. Rato and J. Lemos. Multimodel based fault tolerant control of the 3-tank system. In *1999 European Control Conference (ECC)*, pages 4047–4052. IEEE, 1999.

K. J. Reinschke and G. Wiedemann. Digraph characterization of structural controllability for linear descriptor systems. *Linear algebra and its applications*, 266:199–217, 1997.

R. Romagnoli, S. Weerakkody, and B. Sinopoli. A model inversion based watermark for replay attack detection with output tracking. In *2019 American Control Conference (ACC)*, pages 384–390. IEEE, 2019.

H. S. Sánchez, D. Rotondo, T. Escobet, V. Puig, and J. Quevedo. Bibliographical review on cyber attacks from a control oriented perspective. *Annual Reviews in Control*, 48:103–128, 2019a.

H. S. Sánchez, D. Rotondo, T. Escobet, V. Puig, J. Saludes, and J. Quevedo. Detection of replay attacks in cyber-physical systems using a frequency-based signature. *Journal of the Franklin Institute*, 356(5):2798–2824, 2019b.

H. Sandberg, S. Amin, and K. H. Johansson. Cyberphysical security in networked control systems: An introduction to the issue. *IEEE Control Systems Magazine*, 35(1):20–23, 2015.

B. Satchidanandan and P. Kumar. On the design of security-guaranteeing dynamic watermarks. *IEEE Control Systems Letters*, 4(2):307–312, 2019.

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

C. Schellenberger and P. Zhang. Detection of covert attacks on cyber-physical systems by extending the system dynamics with an auxiliary system. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 1374–1379. IEEE, 2017.

J. K. Scott, R. Findeisen, R. D. Braatz, and D. M. Raimondo. Input design for guaranteed fault diagnosis using zonotopes. *Automatica*, 50(6):1580–1589, 2014.

M. M. Seron and J. A. De Doná. Robust fault estimation and compensation for lpv systems under actuator and sensor faults. *Automatica*, 52:294–301, 2015.

T. Shinohara and T. Namerikawa. Reach set-based attack resilient state estimation against omniscient adversaries. In *2018 Annual American Control Conference (ACC)*, pages 5813–5818. IEEE, 2018.

D. D. Siljak. *Decentralized control of complex systems*. Courier Corporation, 2011.

R. S. Smith. A decoupled feedback structure for covertly appropriating networked control systems. *IFAC Proceedings Volumes*, 44(1):90–95, 2011.

H. Song, P. Shi, C.-C. Lim, W.-A. Zhang, and L. Yu. Set-membership estimation for complex networks subject to linear and nonlinear bounded attacks. *IEEE transactions on neural networks and learning systems*, 31(1):163–173, 2019.

M. Staroswiecki. On reconfigurability with respect to actuator failures. *IFAC Proceedings Volumes*, 35(1):257–262, 2002.

M. Staroswiecki. On reconfiguration-based fault tolerance. In *18th Mediterranean Conference on Control and Automation, MED'10*, pages 1681–1691. IEEE, 2010.

M. Staroswiecki and D. Berdjag. A general fault tolerant linear quadratic control strategy under actuator outages. *International Journal of Systems Science*, 41(8):971–985, 2010.

M. Staroswiecki, G. Hoblos, and A. Aitouche. Sensor network design for fault tolerant estimation. *International journal of adaptive control and signal processing*, 18(1):55–72, 2004.

F. Stoican. *Fault tolerant control based on set-theoretic methods*. PhD thesis, Supélec, 2011.

F. Stoican, S. Olaru, J. A. De Doná, and M. M. Seron. Zonotopic ultimate bounds for linear systems with bounded disturbances. *IFAC Proceedings Volumes*, 44(1):9224–9229, 2011.

Z. Tang, M. Kuijper, M. S. Chong, I. Mareels, and C. Leckie. Linear system security—detection and correction of adversarial sensor attacks in the noise-free case. *Automatica*, 101:53–59, 2019.

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

A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson. A secure control framework for resource-limited adversaries. *Automatica*, 51:135–148, 2015a.

A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson. Secure control systems: A quantitative risk management approach. *IEEE Control Systems Magazine*, 35(1):24–45, 2015b.

A. M. Teixeira and R. M. Ferrari. Detection of sensor data injection attacks with multiplicative watermarking. In *2018 European Control Conference (ECC)*, pages 338–343. IEEE, 2018.

F. D. Torrisi and A. Bemporad. Hysdel-a tool for generating computational hybrid models for analysis and synthesis problems. *IEEE transactions on control systems technology*, 12(2): 235–249, 2004.

K. Tsuda, D. Mignone, G. Ferrari-Trecate, and M. Morari. Reconfiguration strategies for hybrid systems. In *Proceedings of the 2001 American Control Conference.(Cat. No. 01CH37148)*, volume 2, pages 868–873. IEEE, 2001.

R. Tunga, C. Murguia, and J. Ruths. Tuning windowed chi-squared detectors for sensor attacks. In *2018 Annual American Control Conference (ACC)*, pages 1752–1757. IEEE, 2018.

D. Umsonst and H. Sandberg. Anomaly detector metrics for sensor data attacks in control systems. In *2018 Annual American Control Conference (ACC)*, pages 153–158. IEEE, 2018.

L. Vamvakeridou-Lyroudia, J. Bicik, M. Morley, D. Savic, and Z. Kapelan. A real-time intervention management model for reducing impacts due to pipe isolation in water distribution systems. In *Water Distribution Systems Analysis 2010*, pages 209–221. 2010.

R. J. Veillette. Reliable linear-quadratic state-feedback control. *Automatica*, 31(1):137–143, 1995.

Y. Wang, V. Puig, and G. Cembrano. Non-linear economic model predictive control of water distribution networks. *Journal of Process Control*, 56:23–34, 2017.

Y. Wang, D. M. de la Peña, V. Puig, and G. Cembrano. Robust periodic economic model predictive control using probabilistic set invariance for descriptor systems. *IFAC-PapersOnLine*, 51(20):436–441, 2018a.

Y. Wang, D. Muñoz De la Peña, V. Puig, and G. Cembrano. Robust economic model predictive control based on a periodicity constraint. *International Journal of Robust and Nonlinear Control*, 29(11):3296–3310, 2019.

Z. Wang, C.-C. Lim, and Y. Shen. Interval observer design for uncertain discrete-time linear systems. *Systems & Control Letters*, 116:41–46, 2018b.

S. Weerakkody and B. Sinopoli. Detecting integrity attacks on control systems using a moving target approach. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 5820–5826. IEEE, 2015.

S. Weerakkody, Y. Mo, and B. Sinopoli. Detecting integrity attacks on control systems using robust physical watermarking. In *53rd IEEE Conference on Decision and Control*, pages 3757–3764. IEEE, 2014.

S. Weerakkody, B. Sinopoli, S. Kar, and A. Datta. Information flow for security in control systems. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 5065–5072. IEEE, 2016.

S. Weerakkody, O. Ozel, and B. Sinopoli. A bernoulli-gaussian physical watermark for detecting integrity attacks in control systems. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 966–973. IEEE, 2017.

H. P. Williams. *Model building in mathematical programming.* John Wiley & Sons, 2013.

L. Würth, R. Hannemann, and W. Marquardt. A two-layer architecture for economically optimal process control and operation. *Journal of Process Control*, 21(3):311–321, 2011.

X. Yang and J. K. Scott. A comparison of zonotope order reduction techniques. *Automatica*, 95:378–384, 2018.

X. Zhang and F. Zhu. Observer-based sensor attack diagnosis for cyber-physical systems via zonotope theory. *Asian Journal of Control*, 2020.

Y. Zhang, Y. Zhu, and Q. Fan. A novel set-membership estimation approach for preserving security in networked control systems under deception attacks. *Neurocomputing*, 400:440–449, 2020.

M. Zhu and S. Martinez. On distributed constrained formation control in operator–vehicle adversarial networks. *Automatica*, 49(12):3571–3582, 2013a.

M. Zhu and S. Martinez. On the performance analysis of resilient networked control systems under replay attacks. *IEEE Transactions on Automatic Control*, 59(3):804–808, 2013b.

G. M. Ziegler. *Lectures on polytopes*, volume 152. Springer Science & Business Media, 2012.