# Sequential Detection and Isolation of Cyber-physical Attacks on SCADA Systems

van Long Do

Thèse
de doctorat
de l'UTT

Van Long DO

# Sequential Detection and Isolation of Cyber-physical Attacks on SCADA Systems

**Spécialité :**
Optimisation et Sûreté des Systèmes

2015TROY0032

Année 2015

utt
université de technologie
Troyes

# THESE

*pour l'obtention du grade de*

## DOCTEUR de l'UNIVERSITE DE TECHNOLOGIE DE TROYES

**Spécialité : OPTIMISATION ET SURETE DES SYSTEMES**

*présentée et soutenue par*

**Van Long DO**

*le 17 novembre 2015*

# Sequential Detection and Isolation of Cyber-physical Attacks on SCADA Systems

## JURY

| | | |
|---|---|---|
| M. D. BRIE | PROFESSEUR DES UNIVERSITES | Président |
| M. L. FILLATRE | PROFESSEUR DES UNIVERSITES | Directeur de thèse |
| M. M. KINNAERT | PROFESSEUR ORDINAIRE | Rapporteur |
| M. R. LENGELLÉ | PROFESSEUR DES UNIVERSITES | Examinateur |
| M. I. NIKIFOROV | PROFESSEUR DES UNIVERSITES | Directeur de thèse |
| M. P. WILLETT | PROFESSOR | Rapporteur |

## Personnalités invitées

| | |
|---|---|
| M. F. CAMPAN | CHEF DE PROJET ONDEO Systems |
| M. M. VUILLAUME | INGENIEUR D'ETUDES ONDEO Systems |

# Acknowledgments

This PhD thesis has been carried out within the Laboratory of Systems Modeling and Dependability (LM2S) at the University of Technology of Troyes (UTT) under the co-supervision of Professor Igor NIKIFOROV and Professor Lionel FILLATRE.

First of all, I would like to express my deepest gratitude to my supervisors, Professor Igor NIKIFOROV and Professor Lionel FILLATRE, for their highly professional guidance, unlimited support and unceasing encouragement. Their expertise in science and mathematics as well as their valuable remarks have contributed the most part to the success of my PhD thesis. I am extremely thankful and indebted to them for that.

I would like to express my sincere thanks to Professor Peter WILLETT and Professor Michel KINNAERT for accepting to review my PhD thesis. I am also grateful to Professor Régis LENGELLÉ and Professor David BRIE for agreeing to examine this thesis. I wish also to thank M. Francis CAMPAN and M. Martin VUILLAUME for their respectful presence in the committee. The valuable remarks provided by the respectful experts in both academy and industry have helped in improving the quality of this manuscript.

I gratefully acknowledge the French National Research Agency (ANR), the Suez environment, the Ondeo Systems, and the University of Technology of Troyes for providing me with financial and technical support through the project SCALA.

I would like to extend my thanks to secretaries of doctoral school of UTT, Isabelle LECLERCQ, Thésèse KAZARIAN and Pascale DENIS, for their availability and understanding. I also thank to secretaries of ROSAS department, Bernadette ANDRÉ and Véronique BANSE, for their support and availability. I would like to thank all members of LM2S team for a friendly and adorable environment, as well as for providing me with all necessary facilities for doing research.

A special word of thank should go to M. Phuc DO for offering me an opportunity to study in France. Also, I would like to express my gratitude to M. Noël PHAM and his wife, Ms. Huong BUI, for their unlimited support during the last four years in France. I would like thank all Vietnamese students and friends at Troyes for sharing with me unforgettable moments.

I would like to take this opportunity to thank my colleague, Patric, and his wife, Sandy, for sharing with me the moments of fraternal exchanges and friendliness during these three years.

I am very grateful to my mother for her understanding, unlimited support and encouragement, and to my little brother for taking care of my mother during the last four years.

Finally, I am greatly indebted to my wife, Hong Nhung NGUYEN, for her love, understanding and encouragement, and to my little daughter to be born in several months. You have made my life more meaningful. I love you both!

– Van Long DO –

*To my Mom, my Dad, and my little brother,*

*To Hong Nhung, my wife,*

*for their unlimited support, encouragement, and love.*

## Abstract

This PhD thesis is registered in the framework of the project "SCALA" which received financial support through the program ANR-11-SECU-0005. Its ultimate objective involves the on-line monitoring of Supervisory Control And Data Acquisition (SCADA) systems against cyber-physical attacks. The problem is formulated as the sequential detection and isolation of transient signals in stochastic-dynamical systems in the presence of unknown system states and random noises. It is solved by using the analytical redundancy approach consisting of two steps: residual generation and residual evaluation. The residuals are firstly generated by both Kalman filter and parity space approaches. They are then evaluated by using sequential analysis techniques taking into account certain criteria of optimality. However, these classical criteria are not adequate for the surveillance of safety-critical infrastructures. For such applications, it is suggested to minimize the worst-case probability of missed detection subject to acceptable levels on the worst-case probability of false alarm and false isolation. For the detection task, the optimization problem is formulated and solved in both scenarios: exactly and partially known parameters. The sub-optimal tests are obtained and their statistical properties are investigated. Preliminary results for the isolation task are also obtained. The proposed algorithms are applied to the detection and isolation of malicious attacks on a simple SCADA water network.



**Keywords:** Sequential analysis, Signal detection, Change-point problems, Linear models (Statistics), Computer crimes.



## Résumé

Cette thèse s'inscrit dans le cadre du projet "SCALA" financé par l'ANR à travers le programme ANR-11-SECU-0005. Son objectif consiste à surveiller des systèmes de contrôle et d'acquisition de données (SCADA) contre des attaques cyber-physiques. Il s'agit de résoudre un problème de détection-localisation séquentielle de signaux transitoires dans des systèmes stochastiques et dynamiques en présence d'états inconnus et de bruits aléatoires. La solution proposée s'appuie sur une approche par redondance analytique composée de deux étapes : la génération de résidus, puis leur évaluation. Les résidus sont générés de deux façons distinctes, avec le filtre de Kalman ou par projection sur l'espace de parité. Ils sont ensuite évalués par des méthodes d'analyse séquentielle de rupture selon de nouveaux critères d'optimalité adaptés à la surveillance des systèmes à sécurité critique. Il s'agit donc de minimiser la pire probabilité de détection manquée sous la contrainte de niveaux acceptables pour la pire probabilité de fausse alarme et la pire probabilité de fausse localisation. Pour la tâche de détection, le problème d'optimisation est résolu dans deux cas : les paramètres du signal transitoire sont complètement connus ou seulement partiellement connus. Les propriétés statistiques des tests sous-optimaux obtenus sont analysées. Des résultats préliminaires pour la tâche de localisation sont également proposés. Les algorithmes développés sont appliqués à la détection et à la localisation d'actes malveillants dans un réseau d'eau potable.



**Mots-clés:** Analyse séquentielle, Détection du signal, Rupture (statistique), Modéles linéaires (statistique), Criminalité informatique.

# Contents

**Appendix A**
**Proofs of Lemmas, Theorems and Propositions**

**Appendix B**
**Résumé en Français**

# List of Figures

# List of Abbreviations

| | |
|---|---|
| c.d.f. | cumulative distribution function |
| i.i.d. | independent identically distributed |
| p.d.f. | probability density function |
| w.r.t. | with respect to |
| AC | Alternative Current |
| ADI | Attack Detection and Isolation |
| ARL | Average Run Length |
| BDD | Bad Data Detector |
| CUSUM | CUmulative SUM |
| DC | Direct Current |
| DMZ | Demilitarized Zone |
| FDI | Fault Detection and Isolation |
| FMA GLR | Finite Moving Average Generalized Likelihood Ratio |
| FMA WLR | Finite Moving Average Weighted Likelihood Ratio |
| GCUSUM | Generalized CUSUM |
| GLR | Generalized Likelihood Ratio |
| GWL | Generalized Window Limited CUSUM |
| HMI | Human Machine Interface |
| IED | Intelligent Electronic Device |
| IDS | Intrusion Detection System |
| K-L | Kullback-Leibler distance |
| LLR | Log Likelihood Ratio |
| LQG | Linear Quadratic Gaussian |
| LQR | Linear Quadratic Regulator |
| MC | Monte Carlo |
| MP | Most Powerful |
| MCUSUM | Matrix CUSUM |
| MTU | Master Terminal Unit |
| MWL | Matrix Window Limited CUSUM |
| PCA | Principle Component Analysis |
| PI | Proportional-Integral |
| PID | Proportional-Integral-Derivative |
| PLC | Programmable Logic Controller |
| PSSE | Power System State Estimator |
| RTU | Remote Terminal Unit |
| SCADA | Supervisory Control And Data Acquisition |
| SPRT | Sequential Probability Ratio Test |

| | |
|---|---|
| SVM | Support Vector Machine |
| UIO | Unknown Input Observer |
| UMP | Uniformly Most Powerful |
| VPN | Virtual Private Network |
| VCUSUM | Vector CUSUM |
| VTWL | Variable Threshold Window Limited |
| VTWL CUSUM | Variable Threshold Window Limited CUSUM |
| VTWL GLR | Variable Threshold Window Limited Generalized Likelihood Ratio |
| VTWL WLR | Variable Threshold Window Limited Weighted Likelihood Ratio |
| VWL | Vector Window Limited CUSUM |
| WL CUSUM | Window Limited CUSUM |
| WLR | Weighted Likelihood Ratio |

# Glossary of Notations

| | |
|---|---|
| $h$ | Threshold of a test |
| $h_1, \cdots, h_L$ | Variable thresholds of the VTWL CUSUM test |
| $h_1^*, \cdots, h_L^*$ | Optimal thresholds of the VTWL CUSUM test |
| $\tilde{h}_L$ | Threshold of the FMA test |
| $k$ | Discrete time instant |
| $k_0$ | Attack instant or change-point |
| $m$ | Dimension of vector of control signals |
| $m_\alpha$ | Time window for measuring the probability of false alarm |
| $n$ | Dimension of vector of system states |
| $p$ | Dimension of vector of sensor measurements |
| $q$ | Dimension of vector of disturbances |
| $r$ | Dimension of attack vector on system states |
| $s$ | Dimension of attack vector |
| $a_k$ | Vector of attack signals |
| $d_k$ | Vector of disturbances |
| $r_k$ | Vector of residuals |
| $u_k$ | Vector of control signals |
| $v_k$ | Vector of sensor noises |
| $x_k$ | Vector of system states |
| $y_k$ | Vector of sensor measurements |
| $w_k$ | Vector of process noises |
| $r_{k-L+1}^k$ | Concatenated vector of residuals |
| $u_{k-L+1}^k$ | Concatenated vector of control signals |
| $v_{k-L+1}^k$ | Concatenated vector of sensor noises |
| $y_{k-L+1}^k$ | Concatenated vector of sensor measurements |
| $w_{k-L+1}^k$ | Concatenated vector of process noises |
| $A$ | Matrix $A$ in state space model |
| $B$ | Matrix $B$ in state space model |
| $B_a$ | Matrix $B_a$ in state space model |
| $C$ | Matrix $C$ in state space model |
| $C_\alpha$ | Class of statistical hypothesis tests |
| $D$ | Matrix $D$ in state space model |
| $D_a$ | Matrix $D_a$ in state space model |
| $F$ | Matrix $F$ in state space model |
| $G$ | Matrix $G$ in state space model |
| $L$ | Transient change duration and/or attack duration |
| $Q$ | Covariance matrix of process noises |

| | |
|---|---|
| $R$ | Covariance matrix of sensor noises |
| $S_i^k$ | Logarithm of the likelihood ratio |
| $T$ | Stopping time of a detection-isolation procedure |
| $T_{\mathrm{FMA}}$ | Stopping time of the FMA test |
| $T_{\mathrm{GWL}}$ | Stopping time of the generalized WL CUSUM test |
| $T_{\mathrm{MWL}}$ | Stopping time of the matrix WL CUSUM test |
| $T_{\mathrm{VTWL}}$ | Stopping time of the VTWL CUSUM test |
| $T_{\mathrm{VWL}}$ | Stopping time of the vector WL CUSUM test |
| $T_{\mathrm{WL}}$ | Stopping time of the WL CUSUM test |
| $\hat{T}_{\mathrm{FMA}}$ | Stopping time of the FMA GLR test |
| $\hat{T}_{\mathrm{GLR}}$ | Stopping time of the VTWL GLR test |
| $\check{T}_{\mathrm{FMA}}$ | Stopping time of the FMA WLR test |
| $\check{T}_{\mathrm{WLR}}$ | Stopping time of the VTWL WLR test |
| $\phi_{k-L+1}^k(k_0)$ | Concatenated vector of transient profiles |
| $\psi_{k-L+1}^k(k_0)$ | Concatenated vector of transient profiles by Kalman filter approach |
| $\varphi_{k-L+1}^k(k_0)$ | Concatenated vector of transient profiles by parity space approach |
| $\xi_{k-L+1}^k(k_0)$ | Concatenated vector of random noises |
| $\varrho_{k-L+1}^k(k_0)$ | Concatenated vector of random noises by Kalman filter approach |
| $\varsigma_{k-L+1}^k(k_0)$ | Concatenated vector of random noises by parity space approach |
| $\rho_{01}$ | Kullback-Leibler distance between $\mathcal{P}_0$ and $\mathcal{P}_1$ |
| $\rho_{\mathrm{KF}}$ | Kullback-Leibler distance of the residuals generated by Kalman filter approach |
| $\rho_{\mathrm{PS}}$ | Kullback-Leibler distance of the residuals generated by parity space approach |
| $\tau_a$ | Attack period |
| $\theta$ | Parameter of a test |
| $\theta_1, \theta_2, \cdots, \theta_L$ | Vector of transient change profiles (or attack profiles) |
| $\theta_{k-L+1}^k(k_0)$ | Concatenated vector of transient profiles |
| $\Lambda$ | Likelihood ratio |
| $\Phi(\cdot)$ | Standard normal cumulative distribution function |
| $\Sigma$ | Covariance matrix of the random noise vector $\xi_{k-L+1}^k$ |
| $\mathbb{E}$ | Mathematical expectation |
| $\mathbb{E}_0$, $\mathbb{E}_\infty$, $\mathbb{E}_\infty^l$ | Mathematical expectation w.r.t. the pre-change mode |
| $\mathbb{E}_{k_0}$, $\mathbb{E}_{k_0}^l$ | Mathematical expectation w.r.t. the change-point $k_0$ and the change-type $l$ |
| $\mathbb{P}$ | Probability measure |
| $\overline{\mathbb{P}}_{\mathrm{fa}}$ | Worst-case probability of false alarm |
| $\tilde{\mathbb{P}}_{\mathrm{fa}}$ | Upper bound for the worst-case probability of false alarm |
| $\overline{\mathbb{P}}_{\mathrm{fi}}$ | Worst-case probability of false isolation |
| $\tilde{\mathbb{P}}_{\mathrm{fi}}$ | Upper bound for the worst-case probability of false isolation |
| $\overline{\mathbb{P}}_{\mathrm{md}}$ | Worst-case probability of false missed detection |
| $\tilde{\mathbb{P}}_{\mathrm{md}}$ | Upper bound for the worst-case probability of false missed detection |
| $\mathcal{N}(\theta, \Sigma)$ | Gaussian distribution with mean vector $\theta$ and covariance matrix $\Sigma$ |
| $\mathcal{P}_0$, $\mathcal{P}_\infty$, $\mathcal{P}_\infty^l$ | Joint distribution of random variables w.r.t. the pre-change mode |
| $\mathcal{P}_{k_0}$, $\mathcal{P}_{k_0}^l$ | Joint distribution of random variables w.r.t. the change-point $k_0$ and the change-type $l$ |

# General Introduction

## Context and Motivation

This manuscript addresses the problem of sequential detection and isolation of cyber-physical attacks on Supervisory Control And Data Acquisition (SCADA) systems. The SCADA systems are large-scale industrial control systems designed for controlling and monitoring geographically dispersed assets such as electric power grids, gas pipelines and water distribution networks. The rapid development in information and communication technology renders modern SCADA systems more and more susceptible to cyber-physical attacks, not only on physical elements but also on cyber infrastructures. The security of SCADA systems against malicious attacks has been receiving a great deal of research attention over the past few years, especially after the Stuxnet incident in 2010 [47]. Methods proposed for improving the security of safety-critical infrastructures can be broadly classified into two main categories: protection and surveillance. The protection of SCADA systems focuses mainly on the confidentiality, the integrity and the availability of data by information security measures [16]. The surveillance of SCADA systems, on the other hand, consists in distinguishing their nominal operation from their abnormal behavior and identifying between different types (or locations, sources) of malicious attacks.

The system surveillance can be globally divided into two smaller classes: parametric approach and non-parametric approach. The parametric approach consists in determining a set of mathematical equations governing the operation of the system under normal operation as well as under abnormal behavior. The system is said to operate normally if its outputs correspond to those generated from the parametric model under normal operation. On the other hand, if the outputs of the system are consistent with one abnormal mode of the parametric model, the system is said to be in that abnormal behavior. The parametric model of the system is sometimes difficult to obtain in many practical situations. Hence, the non-parametric approach is generally considered as an alternative solution to the parametric approach in such circumstances. The non-parametric approach, which does not require the parametric model, focuses mainly on analyzing the relationship of observed data (i.e., system outputs). The system is said to be in abnormal behavior if the observations are sufficiently scattered from those obtained during the normal operation.

This PhD thesis is registered in the framework of the project “SCALA” (i.e., Surveillance Continue d’Activité et Localisation d’Agression), received financial support from the “Agence Nationale de la Recherche” through the program “Concepts, Systèmes et Outils pour la Securité Globale”, i.e., ANR-CSOSG, Project ANR-11-SECU-0005). The ultimate target of this project is to develop monitoring schemes for detecting and isolating cyber attacks on SCADA systems. In the project SCALA, there are two PhD theses focusing on two aforementioned methods, i.e., parametric approach and non-parametric approach. This PhD thesis follows the parametric

setting where it is required to develop the models of SCADA systems under normal operation as well as under different attack scenarios.

The physical layer of most SCADA systems can be described by a set of partial differential equations (PDEs). Sometimes, it is more convenient to describe the SCADA systems in the discrete-time state space model by linearizing the PDEs around the operating point. Generally, the discrete-time state space model is infected by random noises, i.e., process noises and sensor noises. The process noises are injected into the state evolution equation for reflecting some non-modeled phenomena or model uncertainties. The sensor noises are added to the sensor measurement equation for describing the inaccuracy of measurement instruments. In this manuscript, the process noises and the sensor noises are assumed to be independent identically distributed (i.i.d.) zero-mean multivariate normal random vectors. The cyber-physical attacks, on both the physical layer and the cyber layer, are modeled as additive signals of short duration on both system equations. For this reason, the on-line monitoring of SCADA systems against malicious attacks is transformed into the sequential detection and isolation of transient changes in stochastic-dynamical systems in the presence of unknown system states (often regarded as the nuisance parameter) and Gaussian random noises.

The monitoring of safety-critical applications against cyber-physical attacks is closely related to the fault detection and isolation (FDI) problem in the fault diagnosis community. The ultimate objective of a statistical FDI problem consists in deciding whether something has gone wrong or everything is fine and then determining the location as well as nature of the fault [206]. Generally, the fault diagnosis problem is solved by the analytical redundancy approach which is comprised of two steps: residual generation and residual evaluation. The negative impact of unknown system states is eliminated by utilizing the residual generation techniques in the fault diagnosis literature and the negative effect of random noises is reduced by exploiting well-known methods in statistical decision theory. This manuscript focuses mainly on the sequential detection and isolation of anomalies in the sequence of residuals.

The sequential change detection and isolation techniques are suitable to the on-line monitoring of SCADA systems against cyber-physical attacks due to their ability to process observed data in real time. The operation of a SCADA system is assumed to be initially in normal behavior and, at an unknown time instant (i.e., the change-point $k_0$), it may unexpectedly undergo an abrupt (or a gradual, an incipient) change-of-state from normal to abnormal because of the malicious attacks. The problem of interest is to design detection-isolation algorithms being capable of detecting the change-point and identifying the change-type subject to certain criteria of optimality [175].

The criteria of optimality for the classical quickest change detection problem, which deals with the change of infinitely long duration, should attain a trade-off between the risk associated with raising the false alarm and the risk related to the detection delay. Globally, the optimality criteria should be in favor of minimizing the "worst-case" average detection delay subject to an acceptable value on the false alarm rate. The "worst-case" operation is imposed on all possible values of the change-point $k_0$ since it is generally unknown. The false alarm rate can be measured by either the average run length (ARL) to false alarm or the (conditional) probability of false alarm within any time window of predefined length. Taking into account such criteria, several optimal or asymptotically optimal detection algorithms have been proposed for both Bayesian approach (i.e., where the change-point $k_0$ is considered as unknown and random) and non-Bayesian approach (i.e., where the change-point $k_0$ is considered as unknown but non-random).

The sequential change detection-isolation problem is considered as a generalization of the quick-

est change detection problem where there are several change types (i.e., multiple hypotheses on the change types). The criteria of optimality for the joint detection-isolation problem must take into consideration the risk associated with the false isolation. Classical optimality criteria for the joint detection-isolation problem aim at minimizing the worst-case average delay for detection-isolation subject to acceptable levels on the false alarm and false isolation rates. Similar to the detection problem, the false alarm rate can be measured by either the ARL to false alarm or the probability of false alarm within any time window of predefined length. The false isolation rate, on the other hand, can be evaluated by multiple indexes, including the ARL to false isolation, the (worst-case, conditional) probability of false isolation and the probability of false isolation within any time window of given length. Asymptotically optimal procedures with respect to various detection-isolation criteria have been proposed under both Bayesian and non-Bayesian settings.

The classical quickest change detection-isolation problem posits that the post-change period is infinitely long. The average delay for detection-isolation is, therefore, the only quantity of interest for evaluating the risk associated with the detection of abrupt changes. Recently, special attention has been paid to the problem of detecting transient changes, i.e., the changes of short period. The traditional quickest change detection criterion minimizing the average detection delay subject to an acceptable level of false alarms is not adequate for the detection of short-duration signals. In such circumstances, the criteria of optimality should be favorable of maximizing the "worst-case" probability of detection (or minimizing the "worst-case" probability of missed detection) subject to an acceptable level of false alarms.

In addition, for safety-critical infrastructures such as electric power grids, water distribution networks, or gas pipelines, a hard limit $L$ is generally imposed on the detection delay since the detection of signals with the delay greater than $L$ may cause catastrophic damage. The acceptable delay $L$ represents the "point of no return" since it is impossible to bring the system back to normal operation after being compromised for a period greater than $L$. This value $L$ can be calculated *a priori* from the gravity of the changes (i.e., the magnitude of the changes) and the permitted consequence of the changes. Any detection of the changes with detection delay greater than the predefined value $L$ is considered as missed. Hence, the optimality criteria for safety-critical applications aim also at maximizing the "worst-case" probability of detection (or minimizing the "worst-case" probability of missed detection) subject to an acceptable level of false alarms.

The on-line monitoring of SCADA systems against cyber-physical attacks considered in this manuscript includes both aforementioned types of transient changes. The malicious attacks are generally performed within a short period due to the resource limits of the attackers. Moreover, it is needless to say that the SCADA systems have been playing an extremely important role in almost safety-critical infrastructures, including electric power grids, gas pipelines, water networks or industrial processes. For these reasons, it is extremely suitable to formulate the attack detection-isolation problem as the problem of detecting and identifying transient signals in stochastic-dynamical systems. The optimality criteria involves the minimization of the worst-case probability of missed detection subject to acceptable levels of false alarm and false isolation rates.

## Structure of the PhD Thesis

This manuscript is organized as follows. The security of SCADA systems against cyber-physical attacks is introduced in chapter 1. The rest of the manuscript is split into two parts, consisting of five chapters. The first part, which includes chapter 2, chapter 3 and chapter 4, focus mainly on the sequential detection and isolation of transient signals on stochastic-dynamical systems. In chapter 2, we recount recent results on the statistical decision theory, including non-sequential hypothesis testing, sequential hypothesis testing, sequential change detection and isolation, and sequential detection of transient signals. Chapter 3 and chapter 4, which are the principal contribution of this thesis, are reserved for designing suboptimal algorithms for detecting and isolating additive signals of short duration in the discrete-time state space model driven by Gaussian noises. The second part of this manuscript, which is comprised of chapter 5 and chapter 6, is dedicated to applying theoretical results obtained in the first part to the detection and isolation of cyber-physical attacks on two SCADA systems, including a simple SCADA gas pipeline and a simple water distribution network. The models of two aforementioned SCADA systems as well as cyber-physical attacks are developed in chapter 5. In chapter 6, the detection-isolation schemes designed in chapter 3 and chapter 4 are applied to the detection and isolation of several attack scenarios. Several concluding remarks are drawn on the basis of the numerical examples. The details of each chapter are presented in the following.

Chapter 1 is dedicated to studying the security of SCADA systems against cyber-physical attacks. Firstly, we study the architecture of modern SCADA systems and investigate system vulnerabilities as well as susceptible points which could be exploited by adversaries for performing malicious attacks. Secondly, we resume various approaches for improving the security of SCADA systems, including the information security approach, the secure control theory approach and the fault detection and isolation (FDI) approach. Following the FDI approach, the SCADA systems are described as the discrete-time state space model with Gaussian noises and the cyber-physical attacks are modeled as additive signals of short duration of both system equations. The on-line monitoring of safety-critical infrastructures is formulated as the sequential detection and isolation of transient signals on stochastic-dynamical systems.

The state-of-the-art of statistical decision theory is reviewed in chapter 2. In this chapter, we present essential methods for dealing with random noises in a stochastic system. The statistical decision theory considered in this chapter is split into four main sub-classes. The first sub-class is the non-sequential hypothesis testing which deals with the choice between two or more hypotheses on the basis of the fixed number of observations generated from random variables. The second sub-class is concerned with the sequential hypothesis testing problem where the sample size is not *a priori* fixed but depends on the observations themselves. The sequential detection and isolation of abrupt changes (i.e., changes of infinitely long duration) in a stochastic system are classified into the third sub-class. Various optimal or asymptotically optimal detection-isolation algorithms with respect to different criteria of optimality are considered. The results of the third sub-class is closely related to the final sub-class, i.e., the sequential detection of transient signals (i.e., changes of short duration) in a stochastic system. Similar to the third sub-class, several criteria for the transient change detection problem as well as optimal (and sub-optimal) algorithms are also reviewed. Up to our best knowledge, the joint detection-isolation of transient signals has not been considered.

Chapter 3 presents the main contribution of this PhD thesis. The on-line monitoring of SCADA systems against cyber-physical attacks is officially formulated as the detection of additive signals

of short duration on both equations of the discrete-time state space model in the presence of unknown system states (i.e., the nuisance parameter) and Gaussian random noises. The criterion for the transient change detection problem, minimizing the worst-case probability of missed detection for a given value on the worst-case probability of false alarm within any time window of predefined length, is utilized through this chapter. The nuisance parameter is eliminated by exploiting classical techniques in fault diagnosis community, i.e., the steady-state Kalman filter and the fixed-size parity space approaches. The unified statistical model of residuals generated by both aforementioned techniques is developed. The Variable Threshold Window Limited (VTWL) CUSUM algorithm, which was first introduced in [67, 69] for independent Gaussian observations, is adapted to the unified statistical model. The optimal choice of thresholds with respect to (w.r.t.) the transient change detection criterion is solved and it is shown that the optimized VTWL CUSUM test is equivalent to the simple Finite Moving Average (FMA) test. In addition, a numerical method is introduced for investigating th statistical performance of these detection rules. Furthermore, the proposed numerical method is exploited for analyzing the sensibility of the sub-optimal FMA test w.r.t. several operational parameters. Finally, a more practical scenario where the transient change parameter is partially known (i.e., the change profiles are assumed to be known but the change magnitude is unknown) is considered. Sub-optimal detection procedures are also proposed for such circumstances.

Chapter 4 generalizes the results obtained in chapter 3 to the joint detection-isolation of transient signals on the discrete-time state space model. The unified statistical model is revised so as to adapt to the detection-isolation problem where there are multiple transient change hypotheses. A completely novel criterion of optimality for the transient change detection-isolation problem is introduced. The criterion involves the minimization of the worst-case probability of missed detection subject to acceptable levels on the worst-case probability of false alarm within any time window of predefined length and the worst-case probability of false isolation during the transient change period. Traditional algorithms for the quickest change detection-isolation problem are adapted to the transient change scenario, including the generalized WL CUSUM test, the matrix WL CUSUM test and the vector WL CUSUM test. Especially, we propose the FMA version for the transient change detection-isolation problem. Upper bounds on the error probabilities of the FMA test are also calculated. Though no optimality result is obtained, the FMA detection rule is shown to offer better statistical performance than traditional detection rules by simulation results.

In order to demonstrate theoretical results obtained in chapter 3 and chapter 4, we develop in chapter 5 the models of two typical SCADA systems, including a simple SCADA gas pipeline and a simple SCADA water distribution network under normal operation as well as under cyber-physical attacks. By linearizing a set of PDEs around the operating point, the physical layer of both systems can be described in the discrete-time state space model driven by Gaussian random noises. The cyber-physical attacks on both physical layer and cyber layer can be modeled as additive signals of short duration on both state evolution and sensor measurement equations.

In chapter 6, we apply the theoretical results obtained in chapter 3 and chapter 4 to the detection and isolation of cyber-physical attacks on the SCADA gas pipeline and the SCADA water network developed in chapter 5. This chapter is organized as follows. Firstly, the negative impact of several types of cyber-physical attacks on closed-loop control systems is demonstrated by performing different attack scenarios on the simple SCADA gas pipeline. Secondly, theoretical findings in chapter 3 are applied to the detection of cyber-physical attacks on the simple SCADA water network. The statistical performance of several detection procedures is investigated by both Monte Carlo simulation and numerical method. It is shown that the FMA test

performs better than traditional detection rules, including non-parametric $\chi^2$ detector, CUSUM detector and WL CUSUM detector, for both the steady-state Kalman filter approach and fixed-size parity space approach. The sensitivity analysis of the FMA test w.r.t. several operational parameters is carried out by both numerical method and Monte Carlo simulation. The comparison between two residual-generation methods, i.e., Kalman filter and parity space, is also performed. The statistical performance of several detection algorithms under partially known transient change parameters is also examined by the Monte Carlo simulation. Finally, a more complex SCADA water network is utilized for investigating the statistical performance of several detection-isolation algorithms proposed in chapter 4. The proposed FMA test is compared with the generalized WL CUSUM, the matrix WL CUSUM and the vector WL CUSUM under different scenarios. It is shown that the FMA detection-isolation rule offers better statistical performance than traditional methods.

## Contribution of the PhD Thesis

The main results of this PhD thesis have been reported in the following papers.

| Citation | Publication in referenced international journals |
|---|---|
| [40] | Van Long Do, Lionel Fillatre, and Igor Nikiforov. "Statistical approaches for detecting cyber-physical attacks on SCADA systems". In preparation to submit to the IEEE Transactions on Control Systems Technology, 2015. |

| Citation | Publication in referenced national journals |
|---|---|
| [38] | Van Long Do, Lionel Fillatre, and Igor Nikiforov. "Sequential detection of transient changes in stochastic-dynamical systems". In Journal de la Société Française de Statistique (J-SFdS), pages 60–97, Vol. 156, No. 4, 2015. |

| Citation | Publication in international conferences with full papers |
|---|---|
| [36] | Van Long Do, Lionel Fillatre, and Igor Nikiforov. "A statistical method for detecting cyber/physical attacks on SCADA systems". In 2014 IEEE Conference on Control Applications (CCA), pages 364– 369. IEEE, 2014. |
| [41] | Van Long Do, Lionel Fillatre, and Igor Nikiforov. "Two sub-optimal algorithms for detecting cyber/physical attacks on SCADA systems". In Proceedings of the X International Conference on System Identification and Control Problems (SICPRO'15), 2015. |
| [39] | Van Long Do, Lionel Fillatre, and Igor Nikiforov. "Sequential monitoring of SCADA systems against cyber/physical attacks". In 9th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS 2015), Paris, France, September 2015. |
| [37] | Van Long Do, Lionel Fillatre, and Igor Nikiforov. "Sensitivity analysis of the sequential test for detecting cyber-physical attacks". In 23rd European Signal Processing Conference (EUSIPCO 2015), September 2015. |

| Citation | Publication in national conferences with full papers |
|---|---|
| [24] | Francis Campan, Van Long Do, Patric Nader, Paul Honeine, Pierre Beauseroy, Lionel Fillatre, Philippe Cornu, Igor Nikiforov, Guillaume Prigent, Jérôme Rouxel. "SCALA - Surveillance continue d'activité et localisation d'agressions". In Workshop Interdisciplinaire sur la sécurité Globale, WISG 2013, pages 1–7, 2013. |

# Chapter 1

# Security of SCADA Systems against Cyber-physical Attacks

**Contents**



## 1.1 Introduction to SCADA Systems

As defined in [172], Supervisory Control And Data Acquisition (SCADA) systems[1] are highly distributed control systems used to control geographically dispersed assets, often scattered over thousands of square kilometers, where centralized data acquisition and control are critical to system operation. These large-scale industrial control systems (i.e., SCADA systems) have been playing an extremely important role in almost safety-critical infrastructures [98] such as electric power grids, transportation systems, communication networks, oil and gas pipelines, water distribution and irrigation networks and multiple facilities, including heating, ventilation and air conditioning (HVAC) systems for buildings, or traffic control systems for airports, etc. These safety-critical assets, however, are becoming more and more susceptible to cyber-physical

[1]SCADA systems are closely related to several types of control systems, including Distributed Control Systems (DCS) [172], Networked Control Systems (NCS) [71,77], Process Control Systems (PCS) [172], Industrial Control Systems (ICS) [172], and Cyber-Physical Systems (CPS) [108,140]. In order to avoid confusion, these terms are utilized interchangeably in this manuscript.

attacks[2], not only on the physical infrastructures but also on the communication network and the control center.

The typical architecture of a modern SCADA system, as shown in figure 1.1, consists of three layers: supervisory control layer, automatic control layer and physical layer. The exchange of data among elements in the system is carried out through the communication network [61].

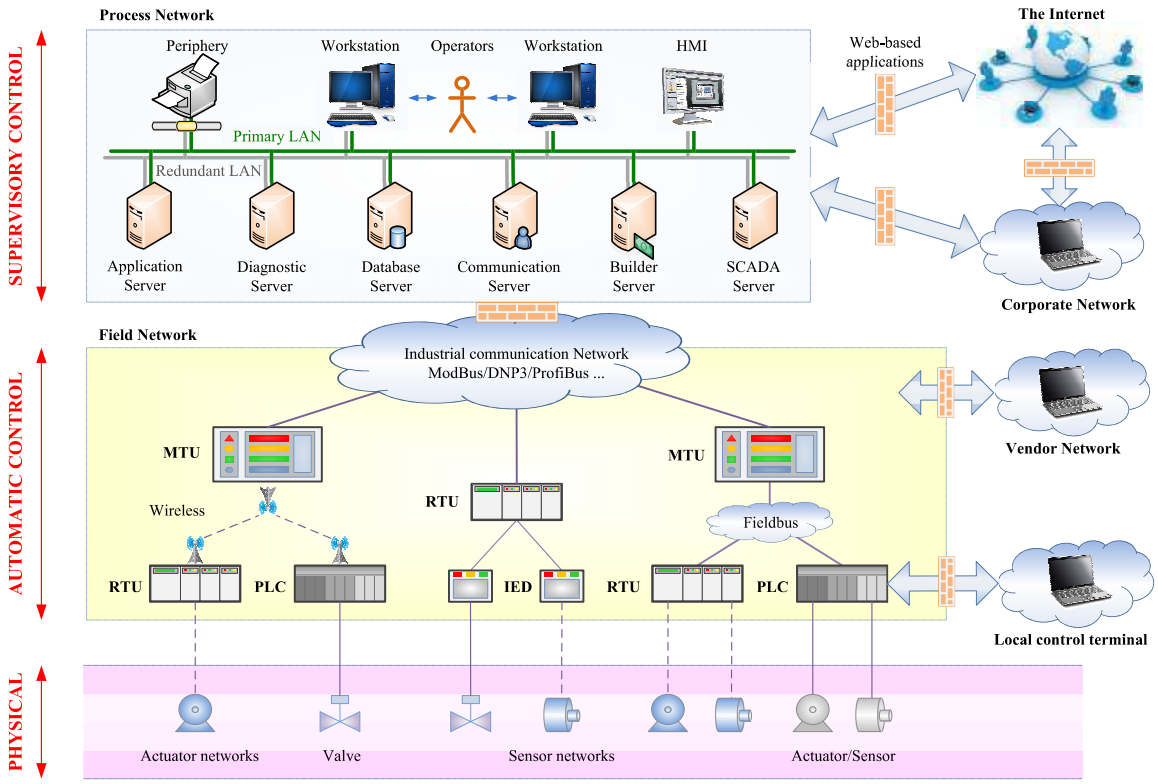



Figure 1.1 – Typical architecture of a modern SCADA system.

### Supervisory control layer

The supervisory control layer (or the control center) is responsible for controlling, monitoring and supervising the operation of a SCADA system by gathering data from field devices, performing supervisory tasks, and sending control commands to field controllers through the communication network. The control center of a typical SCADA system consists of following elements:

- *SCADA server*: Being considered as the heart of the control center, the SCADA server is in charge of controlling and supervising the operation of the system.
- *Communication server*: The communication sever, as its name implies, enables the data exchange between the control center and lower-level layers. The OLE (Object Linking and Embedding) for Process Control (or OPC server) is an example of communication server,

[2]In this manuscript, we use the term "*cyber-physical attack(s)*" instead of "*cyber attack(s)*" for describing the coordination of both cyber and physical activities into the malicious attack(s).

acting as an interface for different software packages to access data from field devices such as Master Terminal Units (MTUs), Remote Terminal Units (RTUs) or Programmable Logic Controllers (PLCs).

- *Builder server*: The builder server is used to load, unload or re-program field devices such as PLCs or RTUs through Ethernet and/or serial cables. An example of a builder server is the software package WinCC/STEP7 of Siemens.
- *Diagnostic server*: The diagnostic server is equipped with intrusion detection systems (IDSs) to detect and identify any abnormal situations, including faults and attacks, occurring to the system.
- *Application server*: The application server is any software framework that helps in developing and implementing complementary applications to the operation of SCADA systems. For instance, the optimal power flow or the electric price policy in power grids are generally located in the application server.
- *Human Machine Interface*: The Human Machine Interface (HMI) is an application that allows system operators to graphically interact with SCADA systems, enabling them to modify control commands and to monitor system variables.
- *Database server*: The database server (or data historian) is a centralized database for logging all process information. This information is then used by the diagnostic server for detecting and identifying any abnormal situations occurring to the system. It can be used also for data analysis, varying from process control analysis to company's plan level.
- *Operators*: The operators working at the control center are in charge of monitoring and supervising the operation of the system and taking action in case of abnormal situations such as faults, failures or even cyber-physical attacks.

#### Automatic control layer

The automatic control layer (or regulatory control layer) is responsible for regulating the operation of physical processes based on the control commands transmitted from the control center and the sensor measurements received from field devices. The control signals, which are the outputs of the controllers, are then sent to the actuators through the communication network. System variables, including control commands, sensor measurements, and control signals, are gathered to the control center for supervisory and management purposes. In large-scale SCADA systems, the automatic control layer is often divided into sub-stations (or sub-systems), whose center is Master Terminal Units (MTUs), and field devices such as Remote Terminal Units (RTUs), Programmable Logic Controllers (PLCs) or Intelligent Electronic Devices (IEDs).

- *Master Terminal Unit*: The MTU, the center of a sub-station, is in charge of exchanging information between the control center and field devices (i.e., RTUs, PLCs or IEDs). The MTU can be regarded as the control center of a small part of a large-scale SCADA system.
- *Remote Terminal Unit*: The RTU is a standalone, special-purpose control and data acquisition unit designed to monitor and control equipments at remote locations from the central station (MTU). Modern RTUs are often equipped with wireless communication such as radio or satellite for exchanging information with the MTU.

- *Programmable Logic Controller*: The PLC is a small industrial computer originally designed for performing logic functions. Nowadays, modern PLCs are developed with the capability of controlling complex processes such as Proportional-Integral-Derivative (PID) control algorithms or file manipulations. In modern SCADA systems, PLCs are used substantially as field devices because they are more economical, more flexible and more configurable than special-purpose RTUs.

- *Intelligent Electronic Device*: The IEDs are smart sensors/actuators which can perform simple control algorithms and data-processing methods. Modern IEDs are generally equipped with wireless technology for communicating with other field devices such as RTUs, PLCs or even MTUs.

### Physical layer

The physical processes, including electric power grids, gas pipelines or water networks, are equipped with actuators (e.g., motors, compressors, pumps, valves), sensors (e.g., temperature sensors, pressure sensors, flow sensors, level sensors, speed sensors) and other protection devices (e.g., circuit breakers, protective relays) to realize technological processes. The physical elements are controlled and monitored by the control center through the automatic control layer and the communication network.

The physical layer of most SCADA systems can be described by a set of partial differential equations (PDEs). These PDEs are generally linearized around the operating point for obtaining the continuous-time state space model. Sometimes, it is preferable to transform the continuous-time state space model into the discrete-time counterpart for exploiting precious results in the digital control theory domain. This task can be realized by utilizing either the zero-order hold method, the first-order hold method, or the Tustin's approximation method [56] with the sample time $T_S$. For this reason, we employ throughout this manuscript the following discrete-time state space model for describing the physical layer of a SCADA system:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + v_k \end{cases}; \quad x_0 = \overline{x}_0, \tag{1.1}$$

where $x_k \in \mathbb{R}^n$ is the vector of system states with unknown initial values $\overline{x}_0 \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$ is the vector of control signals, $d_k \in \mathbb{R}^q$ is the vector of disturbances, $y_k \in \mathbb{R}^p$ is the vector of sensor measurements, $w_k \in \mathbb{R}^n$ is the vector of process noises and $v_k \in \mathbb{R}^p$ is the vector of sensor noises; the matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $F \in \mathbb{R}^{n\times q}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, $G \in \mathbb{R}^{p\times q}$ are assumed to be completely known. The components $Du_k$ and $Gd_k$ stand for the feed-through effect from the control signals $u_k$ and the disturbances $d_k$ to the sensor measurements $y_k$, respectively.

### Communication network

The communication network plays an extremely important role in the operation of a modern SCADA system. Hence, a profound understanding about the communication network will help in analyzing SCADA vulnerabilities. The communication network in a SCADA system can be classified into the corporate network, the process network, the field network and the vendor network [172] (see also figure 1.1).

- *Corporate network*: The corporate network is a group of computers linked together in a particular area, allowing personnel in a company to work collaboratively. Nowadays, the enterprise network is connected to the process network of a SCADA system, enabling the management board to access the process information always and everywhere.
- *Process network*: The process network is a set of servers connected together in the control center of a SCADA system. The cooperation of the servers via the process network helps in monitoring and supervising the operation of the system. The process network is connected to the field network that is responsible for controlling field devices. It is also linked to the business network for sharing process information with the management board.
- *Field network*: The field network connects local controllers (MTUs, RTUs, PLCs, or IEDs) together and links the controllers with actuators/sensors for realizing technological processes. For the maintenance purpose, modern SCADA systems allow to access to the field controllers directly from local access points. This convenience may expose the systems to cyber-physical attacks.
- *Vendor network*: The majority of modern SCADA systems are connected to the vendor network for the purpose of maintenance or technical support. This fact renders modern SCADA systems susceptible to cyber attacks because malicious agents may get access to the SCADA network from the vendor network [213].

The evolution of industrial communication networks has undergone three distinct generations [61,162], from the traditional serial-based fieldbus protocols (e.g., Modbus, Profibus or DNP3) to the industrial Ethernet-based networks (e.g., Modbus-TCP/IP, Ethernet/IP) and the wireless-based communication technologies (e.g., WLAN, WiMAX or Blue-tooth). The standardization of communication protocols renders modern SCADA systems more vulnerable to cyber attacks. More precisely, powerful attackers can break into the communication channels, enabling them to modify the command signals, control signals or sensor measurements for disrupting the systems.

## 1.2 Security of SCADA Systems

The evolution of the SCADA architecture and the communication technology makes modern SCADA systems more and more susceptible to cyber-physical attacks, not only on the physical infrastructures but also on the communication network and the control center [53]. In addition, cyber attacks have become an attractive choice of malicious adversaries to sabotage critical infrastructures since they are cheaper, less risky and easier to execute in comparison with traditional physical methods. Sometimes, malicious adversaries integrate both cyber and physical activities in a coordinated manner for causing more catastrophic damage. A great deal of research effort has been devoted to improve the security of SCADA systems against cyber attacks. For example, the National Institute of Standards and Technology (NIST) in the U.S. has issued even a guide to industrial control systems security [172].

In order to improve the security of SCADA systems and protect safety-critical infrastructures, it is required to investigate system vulnerabilities and to review previous cyber incidents. The vulnerability analysis helps in understanding the susceptible points of the systems and how they might be exploited to launch malicious attacks. The survey of cyber incidents, on the other hand, provides us with a general idea of how the attacks have been carried out in the past so that protection measures can be implemented for avoiding future attacks [118].

### 1.2.1 SCADA cyber incidents

Numerous cyber incidents involving safety-critical infrastructures have been documented over the last decades. Though the attacks might cause huge damage or not, they have raised a big concern about the security of SCADA systems, especially after the Stuxnet incident in 2010. In the following, we present outstanding cyber incidents occurred to ICSs in chronological order.

**Siberian pipeline explosion (1982).** The first cyber incident involving safety-critical infrastructures might be counted as the explosion of the gas pipeline in Siberia in 1982 [156]. It was believed that a Trojan horse had been planted in the SCADA system that controls the Siberian gas pipeline. By changing the cooperation of pumps, turbines and valves, the malicious program caused the pressure in gas pipelines to increase far beyond the acceptable level, leading to an explosion with the power of three kilotons of TNT [189].

**Salt river project hack (1994).** Between July 8th and August 31st, 1994, Mr. Lane Jarrett Davis gained unauthorized access to the computer network of the Salt River Project via a dial-up modem, enabling him to steal and alter essential information such as the water and power monitoring and delivery, customer information, or computer system log files [189]. The hacker installed also a back door to the system so that he could access to the system later.

**Russian gas pipelines (1999).** In 1999, hackers broke into Gazprom, the Russian biggest gas company, through the collaboration with a disgruntled employee [118]. It was believed that the attacker had used Trojan horse to gain control of the central switchboard which controls gas flow through the pipelines. This incident was reported in 2000 by the Interior Ministry of Russia [26, 152, 188].

**Maroochy water breach (2000).** In 2000, Mr. Boden, a disgruntled ex-employee, used a laptop computer and a radio transmitter to take control of 150 sewage pumping stations in Maroochy Shire, Queensland, Australia [168]. Over a three-month period, he released one million liters of untreated sewage into a storm-water drain from where it flowed to local waterways. The attack was motivated by his revenge after he failed to obtain a job at the Maroochy Shire Council.

**Slammer worm crashed Ohio nuke plant network (2003).** In January 2003, a Slammer worm penetrated into a private computer network at Ohio's Davis-Besse nuclear power plant and disabled a safety monitoring system for nearly five hours, despite a belief by plant personnel that the network was protected by a firewall [84, 149]. The Slammer worm spread from the enterprise network to the SCADA systems controlling the nuclear power plant by exploiting the vulnerabilities of the MS-SQL. It was reported that the HMI and the plant process computers had crashed for hours, causing big trouble to system operators.

**Taum Sauk hydroelectric power station failure (2005).** The Taum Sauk incident in December 14, 2005 [159] was not an attack but a failure of a hydroelectric power station. Various reasons, including design/construction flaws, instrumentation errors, and human errors, have been attributed to the catastrophic failure of an upper reservoir. It was reported in [159] that the sensors failed to indicate that the reservoir was full and the pumps were not shut down until the water overflowed for about 5-6 minutes. This overflow undermined the parapet wall, resulting in the collapse of the reservoir. Though this incident was not an attack, the idea behind it can be exploited to perform undetectable attacks in safety-critical infrastructures. For example, the authors in [7] have designed stealthy attacks on a SCADA water irrigation canal by sending compromised feedback signals (i.e., false sensor measurements) to the control center.

**Cyber incident blamed for nuclear power plant shutdown (2008).** In March 2008, a nuclear power plant in Georgia was forced into an emergency shutdown for 48 hours because a

computer used to monitor chemical and diagnostic data from the corporate network rebooted after a software update [97]. For more details, when the updated computer restarted, it reset the data on the control system. The safety systems interpreted the lack of data as the reduction in water reservoirs that cool the plant's radioactive nuclear fuel rods, triggering a system shutdown. Though this cyber incident was not an attack, it has raised a big concern about the security of industrial control systems that operate safety-critical infrastructures.

**Electricity grid in U.S. penetrated by spies (2009).** The World Street Journal reported on April 8, 2009 [66] that cyber spies had penetrated into the U.S. electric power grid and left behind a software program that could be used to disrupt the system. Previously, on August 14, 2003 [111], the Northeast and Midwest regions of the United States and some provinces in Canada suffered from a serious blackout due to a software bug. Though there is no connection between two incidents, they have raised a big concern about the security of electric power grids since disrupting power systems might cause catastrophic damage on economic losses and even human life.

**Stuxnet virus (2010).** Stuxnet [20, 47, 48] is a computer worm that was primarily written to target Iranian nuclear centrifuges. Its final goal is to disrupt industrial control systems by modifying programs implemented on PLCs to make them work in a manner that the attacker intended and to hide those changes from system operators.

It is believed that Stuxnet is introduced to a computer network through an infected removable drive. The virus, once penetrated into a Windows computer, installs its own drivers by using stolen certificates from well-known companies, JMicron and Realtek. In order to hide itself while spreading across the network and realizing the final target, the virus installs a Window rootkit by exploiting four zero-day vulnerabilities. The goal of the virus is to search for the WinCC/Step7 software, a typical software of Siemens for programming and monitoring the PLCs. If Stuxnet does not find the software, it does nothing; otherwise, it replaces some *.dll files in WinCC/Step7 folders by infected *.dll files. According to [47], these *.dll files are responsible for loading and unloading PLC programs from Windows computers and the connected PLCs. By this way, the virus is able to infect the PLCs and modify their programs. For hiding itself in the PLC environment, Stuxnet uses the first known PLC rootkit. Interested readers are referred to [48] for more information about how the virus propagates from a Windows computer to the PLC environment.

It has been announced by well-known security companies, including Symantec and Kaspersky, that Stuxnet was the most sophisticated attack at that time [27]. Its sophistication leads to some speculation that Stuxnet was written with state-level financial support. The success of the virus to penetrate into the PLC environment clearly shows that information security-based techniques are not sufficient for the security of safety-critical infrastructures. Therefore, it is required to implement the defense-in-depth strategy [22, 25, 26] for the complete protection of these critical assets.

**Duqu (2011) and Flame (2012).** Duqu and Flame [14] are computer malwares that were discovered in 2011 and 2012, respectively. It has been reported that Duqu is nearly identical to Stuxnet but with completely different purpose. The goal of Duqu is to collect information that could be useful in attacking ICSs later. Similar to Stuxnet and Duqu, Flame uses the rootkit functionality to evade information security methods. Flame is said to be the most sophisticated virus ever found [14]. The virus contains up to 20 megabytes, which is 20 times more powerful than existing computer malwares, including Stuxnet. Unlike Stuxnet, which was designed to sabotage ICSs, the target of Flame is to gather technical diagrams such as AutoCAD drawings,

PDFs and text files. Though Duqu and Flame were not designed to target ICSs directly, the computer worms have raised a big concern about the security of safety-critical infrastructures. Their recent activities, acquiring information about the systems, may be exploited for completely stealthy attacks in the future.

**Pumping station in U.S. (2011).** On November 8, 2011, the SCADA system of the city water utility in Springfield, Illinois, U.S. was hacked [213]. The system kept turning on and off, leading to the burnout of a water pump. The investigation showed that the attackers penetrated into the control system by exploiting the backdoor left by a control system software vendor. In order to provide maintenance and update services, the software vendor used remote access to the SCADA system of its customers. By some methods, the intruders obtained usernames and passwords and gained unauthorized access to the vendor network, providing them with a path to hack into the control system, causing real physical damage.

**Telvent in Canada (2012).** A breach on the internal firewall and security systems of Telvent Canada [154], a company that supplies remote administration and monitoring tools to the energy sector, was discovered on September 10, 2012. After penetrating into the network, the intruders stole project files related to the OASyS SCADA product, a remote administration tool allowing companies to combine older IT equipments with modern "smart grid" technologies. It is very likely that the adversaries gathered information about the novel product in order to find the vulnerabilities of the software and to prepare for future attacks against SCADA systems in energy sector.

**Georgia Water treatment plant (2013).** The incident [33] occurred at the Carters Lake Water Treatment Plant in Murray County, Northwest of Atlanta, U.S. on April 26, 2013. It is believed that someone entered the water treatment plant and tampered with the equipment controlling how much chlorine and fluoride should be added to the water. Though this incident was not a cyber attack, similar attack scenarios may be performed if the water network is connected to the Internet. For example, in stead of entering the plant directly, the intruders can break into the SCADA network and modify the set points of chlorine and fluoride levels.

### 1.2.2 SCADA vulnerabilities

Recent cyber incidents clearly show that the vulnerabilities of modern SCADA systems have been well exploited for performing malicious attacks on safety-critical infrastructures. In order to improve the security of these important assets, it is required to investigate the vulnerabilities of modern SCADA systems so that appropriate protection measures could be taken. The vulnerabilities of modern ICSs can be broadly classified into five categories [53]: architectural vulnerabilities, security policy vulnerabilities, software and hardware vulnerabilities, communication network vulnerabilities and other vulnerabilities.

**Architectural vulnerabilities.** In general, modern SCADA architectures are not so different in principle from the architectures used in the '80s and '90s except the move from an "*isolated environment*" to an "*open environment*". This advanced feature renders modern SCADA systems more and more vulnerable to cyber attacks. Firstly, the majority of SCADA networks are connected to the corporate network for being more flexible in management process. For example, many SCADA systems store process data and process logs in data historian units, enabling the management board to gain access to the information from the business network. This flexibility leaves a backdoor for computer malwares to enter the process network through the

enterprise network [149]. Secondly, a large number of SCADA systems have been using web-based applications for monitoring physical processes and this direct connection to the Internet could be one possible path for hackers to penetrate into the SCADA network. Moreover, local access points to field devices could be another backdoor for malicious agents to get into the field network of the system. Finally, adversaries can break into the SCADA network through their connection with the vendor network which is available in modern SCADA systems [213].

**Security policy vulnerabilities.** Several security policies, such as patching or anti-virus update, might cause negative impact to SCADA systems. The utilization of several patches and anti-virus software often (1) grants the process network access to the Internet, which may addict the systems with malicious agents and (2) requires system reboot, which may lead to the disruption of the systems. An excellent demonstration for this vulnerability is the cyber incident blamed for the shutdown of a nuclear power plant [97] after a software update. Therefore, it is preferable to use software patches and update the anti-virus software rarely so as to keep the process network as isolated as possible.

**Software and hardware vulnerabilities.** In order to respond to industrial requirements, SCADA systems have become more and more complex in both their software and hardware. It is inevitable for modern SCADA systems to contain software bugs and hardware failures [111]. Typical software bugs can be listed as [214]: buffer overflow, SQL-injection, and format string, etc. In fact, the cyber incident [84, 149] was due to the vulnerabilities of the MS-SQL software. Moreover, SCADA systems are real-time operating systems, preventing the systems from implementing traditional encryption algorithms due to the requirement for the availability of data. This real-time demand makes it difficult to implement data encryption algorithms, exposing SCADA systems to integrity attacks.

**Communication protocol vulnerabilities.** Historically, with the idea in mind that SCADA systems would be isolated from other networks, SCADA designers paid little attention to the security problems such as integrity checking mechanism, authentication mechanism, anti-repudiation and anti-replay mechanism. Many SCADA communication protocols, including Modbus, DNP3 and Allen-Bradley Ethernet/IP, lack authentication features to prove the origin or the freshness of network traffic [62]. Hence, these systems are susceptible to Denial-of-Service (DoS) attacks, man-in-the-middle attacks and replay attacks.

Being implemented with proprietary communication protocols, traditional SCADA systems were thought to be secure. However, the "*security through obscurity*" is not obvious in modern world. The information technology has been evolving rapidly, leading to the adoption of common communication protocols such as Ethernet, TCP/IP or wireless networks [155] such as radio frequencies, satellite communication, IEEE 802.x and Bluetooth in the majority of modern SCADA systems. This evolution has reduced the isolation of SCADA systems from outside environment.

**Other vulnerabilities.** The existence of organized cyber-crime groups (terrorists or state-funded groups) enhances the attacker's capabilities to perform powerful attacks on safety-critical infrastructures. There has been speculation that such complex computer worms as Stuxnet, Duqu or Flame received financial support from state-sponsored groups.

### 1.2.3 Possible attack points

The review of cyber incidents and the analysis of system vulnerabilities allow us to figure out the vulnerable points which might be exploited begin malicious attacks. As shown in figure 1.2, the

back-doors to modern SCADA systems can be broadly classified into three categories [7]: cyber attacks on supervisory control layer, cyber attacks on automatic control and communication layer and physical attacks on technical processes.

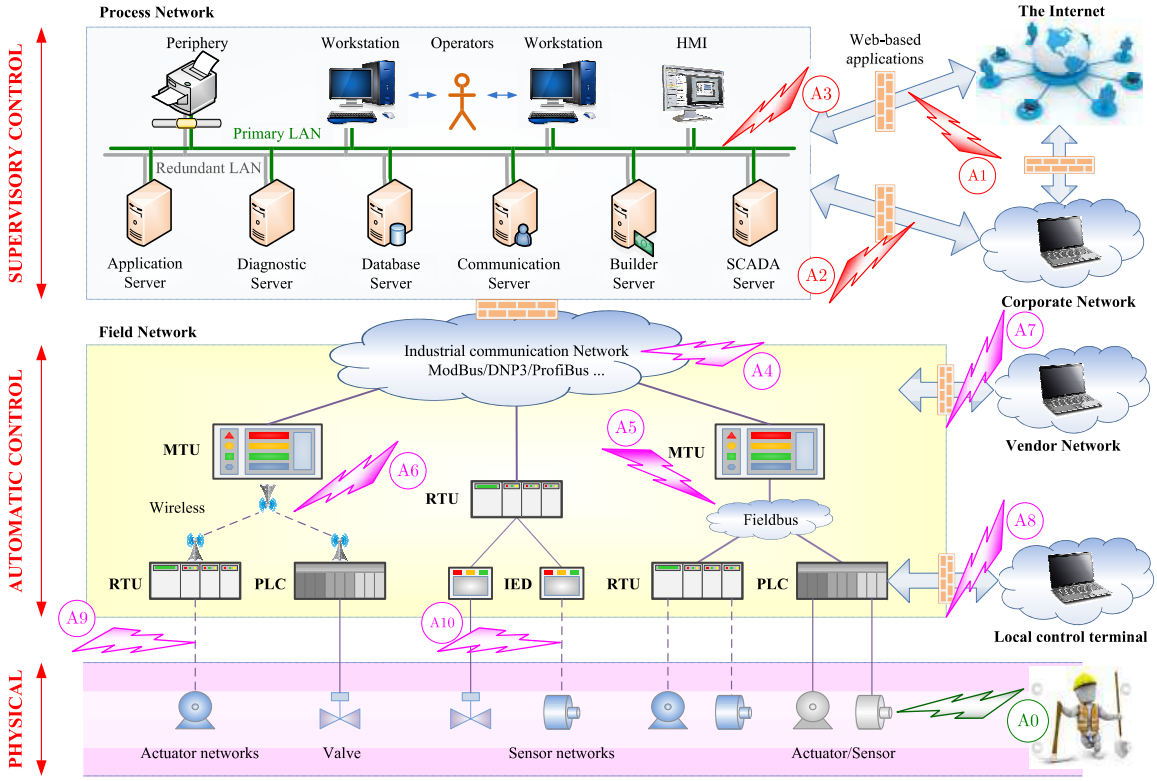



Figure 1.2 – Possible attack points to modern SCADA systems.

### Cyber attacks on supervisory control layer

It is required to discover how an adversary can penetrate into the control center of a SCADA system so that appropriate measures could be taken. According to previous analysis, there are three main back-doors for the attacker to enter the control center:

- *Attack point A1*: Modern SCADA systems use web-based applications for being flexible to management process. In accordance with those advantages, web-based applications also exhibit some inconvenience, especially as regards cybersecurity. An attacker can gain unauthorized access to the control center through those applications.

- *Attack point A2*: A disgruntled employee of a company plugs a USB key containing a virus into a computer in the corporate network. The virus can break through misconfigured firewalls between the business network and the SCADA network and take control of system operation. For example, powerful computer worms such as Stuxnet or Flame are able to bypass traditional IDSs designed by information technology methods.

- *Attack point A3*: In some scenarios, a USB key containing malicious software can be plugged directly into a computer of the control center. Once it enters the process network, the malware can propagate across the network and perform its malicious target.

The control center hosts the SCADA server, the communication server, the builder server, the diagnostic server, the database server, the HMI, and the application server. Since these servers are critical to system operation, an attack on a single element could cause severe consequences.

- *Attack on SCADA server*: Since the SCADA server is responsible for controlling and monitoring the operation of the system, the penetration into the SCADA server may lead to catastrophic damage. For example, the attacker may force the system to stop operating or he may send wrong commands to lower-level stations for disrupting the system.
- *Attack on communication server*: The communication server acts as a bridge to exchange data between the control center and sub-stations. Therefore, the attacker can prevent the data flow between the control center and lower-level devices (DoS attack) or modify the data (integrity attack) when gaining access to the server.
- *Attack on builder server*: The builder server is in charge of loading, unloading or modifying programs from MTUs, RTUs and PLCs. Therefore, if the attacker takes control of builder server, he can re-program the PLCs to disrupt the operation of physical processes[3].
- *Attack on diagnostic server*: The attacker can hack into the diagnostic server to modify the outputs of diagnostic algorithms while conducting other attacks so that system operators are unable to recognize what are wrong with the system[4].
- *Attack on database server*: The database server contains important information such as process data or set-points used for monitoring and controlling physical processes. For example, the diagnostic server uses real-time data from database server to perform the intrusion detection algorithms or the HMI displays process status to system operators. Therefore, by attacking the database server, the attacker can hide other malicious attacks from the operators. In addition, the attacker can steal essential information and then use them for negative purposes.
- *Attack on HMI*: If the attacker can modify some data displayed on the HMI, he can prevent the operators from discovering what is wrong with the system.

### Cyber attacks on automatic control layer

Modern SCADA systems contain numerous vulnerabilities which could be exploited by malicious agents for launching cyber attacks. Hence, it is essential to recognize how the adversaries could penetrate into the automatic control layer and what they would do afterward. The attackers could begin their malevolent activities through following vulnerable points (see also [7]):

[3]This kind of attack is exactly what the virus Stuxnet did when it got access to the control center from a USB key. Stuxnet attacked on the builder server located on computers which had been installed STEP7, a software used for programming PLCs of Siemens. By replacing the file *.dll used by STEP7 to load and unload the programs, the virus could modify the programs loaded into PLCs.

[4]The output of a diagnosis algorithm depends on sensor measurements. Hence, adversaries can modify control/sensor signals for altering the output of the diagnostic server.

- *Attack point A4, A5 and A6*: By exploiting the vulnerabilities of communication protocols such as ModBus, DNP3, Ethernet/IP or wireless-based protocols, the attacker can get access to communication channels between control center and sub-stations (i.e., attack point A4). Once broken into this channel, the intruder may introduce fake control commands to the MTUs, send back false data to the control center, or even jam the communication channels by launching DoS attack. The attack on the communication links between the MTUs and the PLCs/RTUs (i.e., attack points A5 and A6) can be carried out in the same manner.
- *Attack point A7 and A8*: For being flexible in maintenance and update services, modern SCADA systems support communication links between field devices and vendor networks (i.e., attack point A7) or local terminals (i.e., attack point A8). This flexibility leaves a backdoor for malicious hackers to take control of field devices. In fact, an attack has been carried out successfully via the vendor network, causing real physical damage [213].
- *Attack point A9 and point A10*: The communication between local controllers (i.e., RTUs or PLCs) and field devices (actuators or sensors) are sometimes implemented by insecure technologies (i.e., wireless, satellite or radio). As a result, the control signals sent from the controllers to the actuators (i.e., attack point A9) and the feedback signals transmitted from the sensors to the controllers (i.e., attack point A10) are susceptible to cyber attacks. These vulnerabilities may be exploited for designing coordinated attacks, causing catastrophic damage.

#### Physical attacks on technological processes

Due to their geographically dispersed characteristics, it is very difficult to protect SCADA systems from physical attacks (i.e., attack point A0) like cutting the communication cables or compromising sensors and actuators. Sometimes, malicious adversaries integrate both physical and cyber activities into a coordinated attack to cause more catastrophic damage. For these reasons, it is necessary to enforce security measures for protecting physical assets, thus eliminating negative impact of the attack.

## 1.3 Attack Detection and Isolation Methods

SCADA systems are at the core of safety-critical infrastructures, playing a vital role in the development of a nation. Previous analysis has pointed out that these large-scale ICSs are becoming more susceptible to cyber-physical attacks than ever before. It is needless to say that greater concern should be paid for improving the resilience of SCADA systems against cyber-physical attacks so as to avoid physical destruction, economic losses or even human life. There exists a vast literature on the security of SCADA systems against cyber-physical attacks. These methods can be broadly classified into three groups: the information technology (IT) approach, the secure control approach and the fault detection and isolation (FDI) approach.

### 1.3.1 Information technology approach

The information technology (IT) approach focus mainly on ensuring confidentiality, integrity and availability of information [16]. The confidentiality is related to the non-disclosure of information

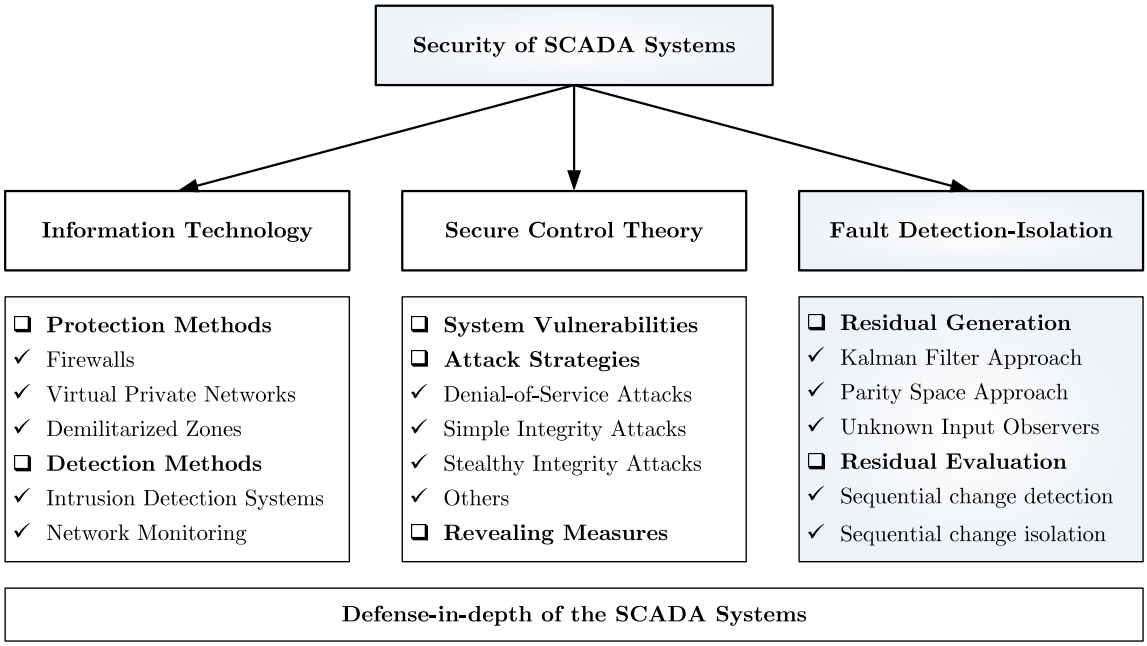



Figure 1.3 – Attack detection and isolation methods.

to unauthorized parties. The confidentiality of data is generally performed by authentication or access control methods. The integrity of data, on the other hand, refers to the trustworthiness of data (i.e., there is no unauthorized modification of data contents or properties). The data integrity is generally realized by both prevention mechanisms (i.e., encryption algorithms, authentication and/or access control) and detection mechanisms (i.e., integrity checking methods). The availability of data is concerned with the utilization of information or resources when needed.

Guidelines and methods [32,42,98,172] have been proposed for improving the security of SCADA systems against cyber-physical attacks. Some examples, among others, include (1) designing specific firewalls between the process network and the corporate network or between the MTUs and RTUs/PLCs, (2) utilizing the Demilitarized Zones (DMZs) for isolating the process network from the corporate network, (3) exploiting Virtual Private Networks (VPNs) for transmitting data over public networks, and (4) developing the Intrusion Detection Systems (IDSs) for SCADA systems [215]. In addition, sequential methods have been proposed in [175,182] for the monitoring of network traffic in computer systems against Denial-of-Service (DoS) attacks.

It is believed that appropriate utilization of aforementioned information security measures may help in reducing the number of cyber incidents as well as their consequences. However, these methods are mainly applicable for protecting SCADA systems from cyber attacks on the control center (i.e., attack points A1, A2 and A3 in figure 1.2) and on the communication layer between the control center and the MTUs (i.e., attack point A4 in figure 1.2). Sometimes, firewalls and VPNs can be utilized for preventing the intrusion into SCADA systems through vendor networks and local terminals (i.e., attack points A7 and A8 in figure 1.2). However, the Stuxnet incident [47,48] and the pumping station incident [213] have given a strong evidence that these IT-based tools can offer only necessary mechanisms for the security of SCADA systems. The complete protection of these large-scale ICSs against cyber-physical attacks requires a defense-in-depth strategy [22,25,26,28], where safety-critical infrastructures are protected by layers of

security.

Moreover, SCADA systems are very different from IT systems in many aspects. Firstly, the requirement of continuous operation prevents SCADA systems from applying IT security solutions like anti-virus software updates. Secondly, it is extremely difficult to implement traditional security solutions to lower layers of SCADA systems. For example, advanced encryption algorithms, which require a huge amount of computational resources, can not be implemented in communication channels between PLCs and sensors/actuators due to the hard real-time requirements [214]. In addition, wireless technologies are often utilized for transmitting data over long distances due to the geographically dispersed characteristics. Finally, the key difference between SCADA systems and IT systems lies in the interaction of the control systems to the physical world. However, traditional IT-based solutions do not exploit the compatibility of the cyber layer (i.e., control algorithms, command signals, control signals and sensor measurements) with the physical layer (i.e., actuators, sensors or physical processes), thus being ineffective against cyber-physical attacks targeting at disrupting the physical processes [141].

### 1.3.2 Secure control theory approach

In contrast to IT methods, the secure control approach, as its name implies, focuses mainly on analyzing the security of networked control systems against cyber attacks. The general approach consists in investigating the negative impact of different types of cyber attacks on particular systems. Especially, a great deal of research effort has been dedicated to investigating the vulnerabilities of networked control systems, designing stealthy/deception attacks which can partially or completely bypass traditional anomaly detectors, and proposing countermeasures for revealing undetectable attacks.

#### Secure control framework

A secure control framework for resource-limited adversaries has been proposed in [184, 187] for studying the cyber security of networked control systems against malicious attacks (see also figure 1.4). The capabilities of attackers are described by an attack space, including the model knowledge (i.e., the information about the system and attack models), the disclosure resources (i.e., the ability to capture control and sensor signals) and the disruption capabilities (i.e., the ability to modify captured signals). It has been shown that this secure control framework can be used for modeling and analyzing various attack scenarios (i.e., attack strategies) found in literature.

The following discrete-time state space model is generally employed for describing the operation of networked control systems under cyber attacks:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + Ka_k^u + w_k \\ y_k &= Cx_k + Du_k + Gd_k + Ha_k^u + Ma_k^y + v_k \end{cases}; \quad x_0 = \overline{x}_0, \tag{1.2}$$

where $x_k \in \mathbb{R}^n$ is the vector of system states with unknown initial value $x_0 = \overline{x}_0$, $u_k \in \mathbb{R}^m$ is the vector of control signals, $d_k \in \mathbb{R}^q$ is the vector of disturbances, $y_k \in \mathbb{R}^p$ is the vector of sensor measurements, $a_k^u \in \mathbb{R}^m$ is the attack vector on control signals, $a_k^y \in \mathbb{R}^p$ is the attack vector on sensor measurements, $w_k \in \mathbb{R}^n$ is the vector of process noises and $v_k \in \mathbb{R}^p$ is the vector of sensor noises; the matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $F \in \mathbb{R}^{p\times q}$, $C \in \mathbb{R}^{p\times p}$, $D \in \mathbb{R}^{p\times m}$, $G \in \mathbb{R}^{p\times q}$, $K \in \mathbb{R}^{n\times m}$, $H \in \mathbb{R}^{p\times m}$ and $M \in \mathbb{R}^{p\times p}$ are assumed to be known to system operators.

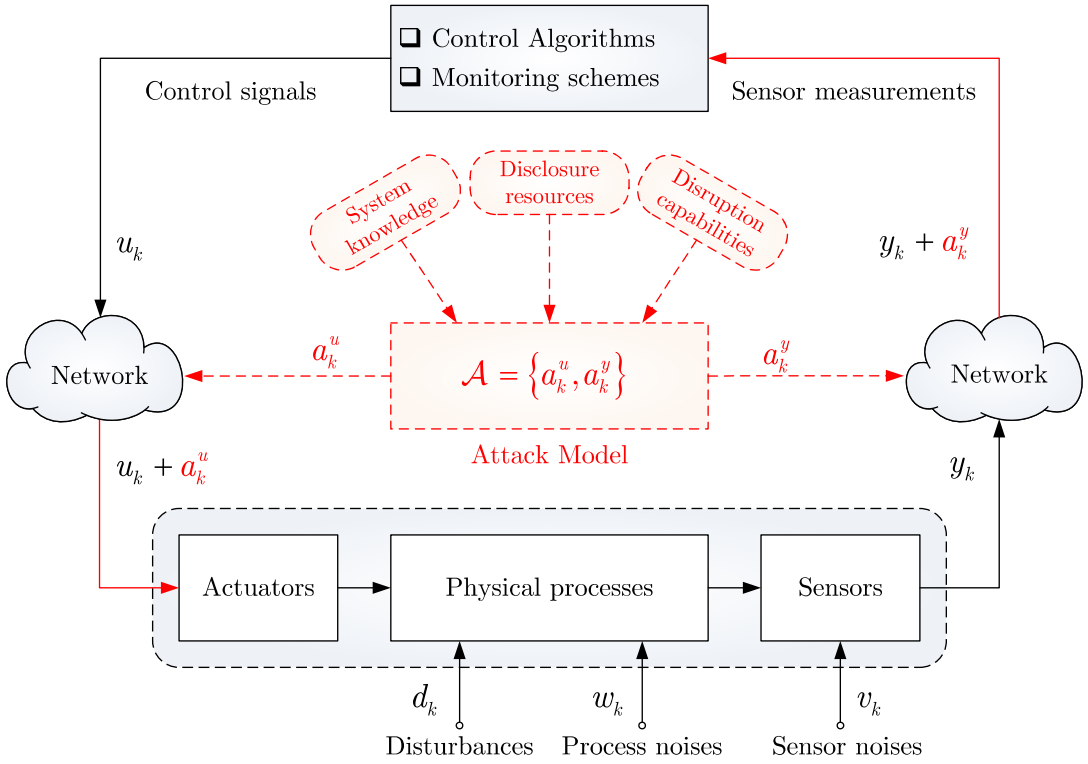



Figure 1.4 – Secure control framework for studying cyber-physical attacks on networked control systems.

**Remark 1.1.** *The system matrices $A$, $B$, $F$, $C$, $D$ and $G$ depend only on the system architecture. On the other hand, the attack matrices $K$, $H$ and $M$ depend not only on the system architecture but also on the capability of malicious adversaries to compromise control and/or sensor signals. As discussed in [99], the attack matrices $K$, $H$ and $M$ should satisfy: $span(K) \subseteq span(B)$, $span(H) \subseteq span(D)$ and $span(M) \subseteq \mathbb{R}^p$, where $span(\Delta)$ denotes the subspace spanned by the columns of matrix $\Delta$. For example, if the attackers are able to gain access to all control and sensor channels, the attack matrices $K$, $H$ and $M$ can be chosen, without loss of generality, as $K \triangleq B$, $H \triangleq D$ and $M \triangleq I_p$ where $I_p \in \mathbb{R}^{p \times p}$ is the identity matrix of size $p$.*

**Remark 1.2.** *The discrete-time state space model (1.2) is more general than those employed in literature for describing networked control systems under cyber attacks. For example, the authors in [184, 187] consider the vector of disturbances $d_k$ as faults (i.e., anomalies) in fault diagnosis literature [30]. Moreover, both the vector of disturbances $d_k$ and the feed-through components $Du_k$ and $Gd_k$ are excluded from the discrete-time state space model used in [88, 99, 120, 122]. Finally, the deterministic state space model (i.e., without random noises $w_k$ and $v_k$) has been used substantially in literature (see, for example [141, 169, 186]).*

**Remark 1.3.** *Let us add some comments on the attack duration $\tau_a$. In the literature, two different approaches have been considered for modeling the attack duration. The first approach posits that the attack duration is infinitely long, i.e., $\tau_a = [k_0, +\infty)$, where $k_0$ is the unknown attack instant (see, for example, in [88, 99, 141]). The second approach assumes that the malicious action is of short duration, i.e, $\tau_a = [k_0, k_0 + L - 1]$, where $k_0$ is the unknown attack instant and $L$ is the attack duration (see, for instance, in [7, 27, 80, 184, 186, 187]).*

The attacks on SCADA systems can be realized by designing the attack vectors $a_k^u$ and $a_k^y$ on, respectively, control signals and sensor measurements in stead of launching physical attacks directly on physical processes. The design of such attack vectors depends heavily on the targets and the capabilities of malicious adversaries. The cyber attacks on SCADA systems can be broadly classified into two main categories [99, 141]: Denial-of-Service (DoS) attacks and integrity attacks, as shown in figure 1.5. DoS attacks refer to such attempts and efforts that aim at disrupting temporarily or indefinitely the exchange of data among entities in the network, for instance, by jamming the communication channels or compromising the routing protocols [99]. The integrity attacks, on the other hand, refer to the possibility of compromising the integrity of data packets (e.g., command signals, control signals or sensor measurements) and they are performed by altering the behavior of actuators and sensors or by breaking into the communication channels between the physical layer and the control center [141]. The integrity attacks can be further divided into two smaller sub-classes: simple integrity attacks and stealthy integrity attacks. The simple integrity attacks include such attack strategies that the modification of data packets is carried out without knowledge about the system models. The stealthy integrity attacks, on the other hand, require the model knowledge, the disclosure resources and the disruption capabilities for bypassing classical detection schemes. Less powerful attackers can choose simple attack strategies such as DoS attacks or simple integrity attacks. However, more powerful adversaries equipped with model knowledge, disclosure resources and disruption capabilities may perform stealthy/deception attacks for partially or completely bypassing traditional anomaly detectors.

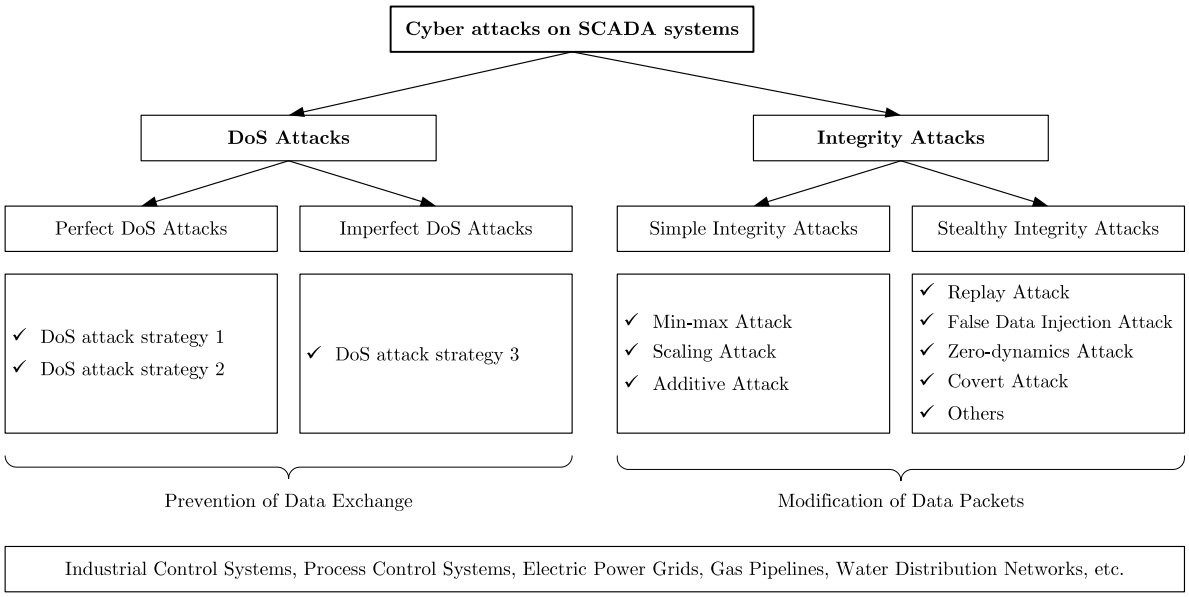



Figure 1.5 – Classification of cyber attacks on SCADA systems.

Consider now the transmission of signals (e.g., command signals, control signals or sensor measurements) from a transmitter to a receiver through a communication channel. Let $z_k$ be the sourced signals sent by the transmitter and $\tilde{z}_k$ be the targeted signals arriving at the receiver. The targeted signals may be different from the original ones due to the malicious attacks on the communication channel (i.e., $\tilde{z}_k \neq z_k$). Let us assume also that the malicious attack is performed within a short period $\tau_a = [k_0, k_0 + L - 1]$, where $k_0$ is the attack instant and $L$ is the attack duration. The mathematical models of several attack strategies are described in the following.

### DoS attacks

A great deal of research effort has been paid to studying the negative impact of DoS attacks on networked control systems over the last few years. For example, the authors in [3] studied the robust feedback control design against DoS attacks; and the impact of random packets drop on controller and estimator performance was investigated in [163, 167]. The first mathematical model of DoS attacks was proposed in [26], where the targeted signals $\tilde{z}_k$ are considered as zero if the sourced signals $z_k$ do not arrive at the receiver. Such an attack strategy can be modeled as

$$\tilde{z}_k = \begin{cases} z_k & \text{if} \quad k \notin \tau_a \\ 0 & \text{if} \quad k \in \tau_a \end{cases}. \tag{1.3}$$

The second mathematical model of DoS attacks was introduced by [80], the received signals $\tilde{z}_k$ are considered as the last arrived signals (i.e., $\tilde{z}_k = \tilde{z}_{k_0-1} = z_{k_0-1}$) if the sourced signals $z_k$ do not arrive at the receiver. The mathematical model of this DoS attack strategy can be described as

$$\tilde{z}_k = \begin{cases} z_k & \text{if} \quad k \notin \tau_a \\ z_{k_0-1} & \text{if} \quad k \in \tau_a \end{cases}. \tag{1.4}$$

The mathematical models (1.3)–(1.4) refer to perfect DoS attacks where powerful attackers are able to completely block the communication channel between the transmitter and the receiver. In practice, malicious adversaries are able to jam the communication link so that data packages are dropped during the transmission process. More precisely, some packages may arrive at the receiver and the others may not [77, 163, 167]. The following model is utilized to describe such realistic scenarios [3]:

$$\tilde{z}_k = \begin{cases} z_k & \text{if} \quad k \notin \tau_a \\ \gamma_k z_k & \text{if} \quad k \in \tau_a \end{cases}, \tag{1.5}$$

where $\gamma_k \in \{0, 1\}$. The authors in [3] has proposed an optimal causal feedback controller (for a discrete-time linear system) that minimizes an objective function subject to safety and power constraints under the assumption that the coefficient $\gamma_k$ follows the Bernoulli distribution.

### Simple integrity attacks

Let $\mathcal{Z} = [z_{\min}, z_{\max}]$ be reasonable union of signals, where $z_{\min}$ and $z_{\max}$ denote, respectively, the minimal and maximal values for both sourced and targeted signals. Performing simple integrity attacks requires no information about the system. For conducting an integrity attack, the attacker captures the sourced signals $z_k$ transmitted over the network, modifies the captured signals, and re-transmits the compromised signals $\tilde{z}_k$ to the receiver. This strategy is often referred to as the "man-in-the-middle" attack. In the following, we introduce some examples of simple integrity attacks that have been introduced in [80], including min-max attack, scaling attack and additive attack.

- *Min, max attack*: Min (resp. max) attack can be carried out simply by returning extremely low (resp. high) values to the receiver. They can be modeled as

$$\tilde{z}_k = \begin{cases} z_k & \text{if} \quad k \notin \tau_a \\ z_{\min} & \text{if} \quad k \in \tau_a \end{cases}, \text{ for min attack;} \quad \tilde{z}_k = \begin{cases} z_k & \text{if} \quad k \notin \tau_a \\ z_{\max} & \text{if} \quad k \in \tau_a \end{cases}, \text{ for max attack.} \tag{1.6}$$

- *Scaling attack*: For the scaling attack, the adversary needs to capture the sourced signals $z_k$ and multiples with a predefined coefficient $\alpha_k$. The model of the scaling attack is described as

$$\tilde{z}_k = \begin{cases} z_k & \text{if} \quad k \notin \tau_a \\ \alpha_k z_k & \text{if} \quad k \in \tau_a \text{ and } \alpha_k z_k \in \mathcal{Z} \\ z_{\min} & \text{if} \quad k \in \tau_a \text{ and } \alpha_k z_k < z_{\min} \\ z_{\max} & \text{if} \quad k \in \tau_a \text{ and } \alpha_k z_k > z_{\max} \end{cases}, \tag{1.7}$$

  where the coefficient $\alpha_k$ is defined by the attacker.

- *Additive attack*: Similar to the scaling attack, the additive attack is performed by adding predefined values $z_k^a$ to the sourced signals $z_k$. The model of the additive attack is written as

$$\tilde{z}_k = \begin{cases} z_k & \text{if} \quad k \notin \tau_a \\ z_k + a_k^z & \text{if} \quad k \in \tau_a \text{ and } z_k + a_k^z \in \mathcal{Z} \\ z_{\min} & \text{if} \quad k \in \tau_a \text{ and } z_k + a_k^z < z_{\min} \\ z_{\max} & \text{if} \quad k \in \tau_a \text{ and } z_k + a_k^z > z_{\max} \end{cases}, \tag{1.8}$$

  where the additive value $a_k^z$ is designed by the attacker.

### Stealthy integrity attacks

It has been shown in literature that powerful adversaries equipped with model knowledge, disclosure resources and disruption capabilities are able to perform stealthy/deception attacks for partially or completely bypassing traditional anomaly detectors. The stealthiness of an attack strategy depends heavily on the capabilities of adversaries to coordinate the attack vectors on control signals and sensor measurements. The characteristic difference between such undetectable attacks lies in how to coordinate the attack vectors $a_k^u$ and $a_k^y$ on control signals and sensor measurements, respectively. In the following, we resume several well-known stealthy integrity attacks on networked control systems.

**Replay attack.** The negative effect of replay attack on a feedback control system has been studied in [120, 122, 187]. The system is described by the discrete-time linear time-invariant Gaussian model driven by an infinite horizon Linear Quadratic Gaussian (LQG) controller, i.e., consisting of the Kalman filter and the Linear Quadratic Regulator (LQR). The $\chi^2$ detector is employed to detect any abnormal behavior occurring to the system. The replay attack strategy is carried out by two steps. In the first step, the attacker records sensor measurements for a certain amount of time before performing the attack. In the second step, he replaces actual sensor measurements by previously recorded signals while performing malicious attacks on control signals for driving system states out of their normal behavior. It has been shown in [120, 122] that the replay attack is able to bypass the $\chi^2$ detector. Two countermeasures have been proposed in [120,122] for revealing the replay attack. It has been also discussed in [140] that the replay attack is not the worst-case stealthy attack since it can be detected by an active monitor (i.e., an anomaly detector that injects unknown auxiliary signals to the control signals).

**Static false data injection attack.** The problem of static false data injection attacks on the Power System State Estimator (PSSE) of DC power models was first considered in [112]. It has been shown that the adversary could launch cyber attacks on sensor measurements with the target of introducing arbitrary errors into certain state variables while bypassing existing bad

data detection (BDD) schemes. Following the work in [112], the authors in [153, 183] studied stealthy/deception attacks on AC power grids based on outdated, inaccurate and incomplete system models. Furthermore, the authors in [210,211] have shown that malicious attackers could modify sensor measurements in order to bias the estimated state variables for profiting in electric prices. In addition, the problem of cyber attacks on PSSE affecting the optimal power flow and load redistribution has been also mentioned in [185] and [212], respectively. Sequential analysis methods have been considered in [93–95] and [79] for detecting cyber attacks on the PSSE instead of using traditional BDDs. Interested readers should refer to [17, 34, 79, 93–95, 160, 174] for other research about deception attacks on the PSSE.

**Dynamic false data injection attack.** In [119, 121], the authors have studied the negative impact of false data injection attack on a discrete-time linear time-invariant Gaussian system. The Kalman filter is used to perform state estimation and a failure detector is employed to detect abnormal situations. The target of the attacker is to fool the state estimator by carefully injecting a certain amount of false data into sensor measurements transmitted to the state estimator over a communication channel. Necessary and sufficient conditions under which the attacker could destabilize the system are also given. According to an analysis in [140, page 46], the false data injection attacks proposed in [119,121] correspond to the output attacks rendering an unstable mode (if any) of the system unobservable. The analysis in [119,121,140] shows that the false data injection attacks are inapplicable if either the system has no unstable pole or some "critical" sensors are protected.

**Zero-dynamics attack.** By utilizing the output-nulling controlled invariant subspace in geometric control theory, the authors in [186] have studied the zero-dynamics attack on networked control systems. The disclosure of the zero-dynamics attack strategy has also been considered, including the modification of the system's structure. Moreover, the authors in [88] have proposed a method to render the attack detectable by triggering data losses on control signals corrupted by the attack. Two observations can be drawn from studying the zero-dynamics attack strategy. Firstly, this attack strategy requires only the modification of control signals for its stealthiness. However, the attack signals added to the control signals can not be chosen freely. These non-zero signals must be designed in such way that their effects to the outputs are null by exploiting the output-nulling problem in the automatic control theory. Simulation results in [186] have shown that there are situations where the attack signals drive the system into a saturation region (i.e., the control signals are greater than the capacity actuators). The zero-dynamics attack strategy reveals itself in such circumstances. Secondly, it has been proved in [186] that the zero-dynamics attack can be revealed by equipping the system with more sensors. It seems that this sensor placement strategy is effective in revealing not only the zero-dynamic attack but also other stealthy attacks.

**Covert attack.** Another kind of stealthy attack, namely the covert attack on networked control systems, has been investigated in [169]. The covert attack strategy consists in coordinating control signals and sensor measurements into a malicious attack. The idea of the covert attack is as follows. Firstly, the state attack vector can be chosen freely based on malicious targets and available resources. Secondly, the sensor attack vector is designed in such a way that it can compensate for the effects of the state attack vector on the sensor measurements. It has been shown that the covert attack is completely stealthy to any anomaly detectors. The covert attack strategy can be considered as the worst-case attack due to its ability to completely bypass traditional anomaly detectors. The disadvantage of this strategy, however, lies in the strategy itself. More precisely, the covert attack requires to compromise enough number of sensors for assuring its stealthiness. By exploiting this inconvenience, defenders can reveal the covert attack

by protecting some important sensors or even equipping new secure sensors.

**Surge attack, bias attack and geometric attack.** While studying the security of process control systems against cyber attacks, the authors in [27] have designed three types of stealthy attacks, named as the surge attack, the bias attack and the geometric attack. The surge attack seeks to maximize damage as soon as possible while the bias attack tries to modify the system by small perturbations over a long period of time. Finally, the geometric attack integrates the surge attack and the bias attack by shifting the system behavior gradually at the beginning and maximizing the damage at the end.

### Covert attack strategy and sensor protection framework

Consider the discrete-time state space model under cyber attacks (1.2). The attack vectors $a_k^u$ and $a_k^y$ can be designed by the covert attack strategy as follows:

- The attack vector $a_k^u$ on control signals can be chosen arbitrarily based on the target and available disruption resources of the attacker.
- The attack vector $a_k^y$ on sensor measurements is calculated by the following equation:

$$\begin{cases} x_{a,k+1} &= Ax_{a,k} + Ka_k^u \\ a_k^y &= -Cx_{a,k} - Ha_k^u \end{cases}; \quad \{x_{a,k}\}_{k \leq k_0} = 0, \tag{1.9}$$

  where $x_{a,k} \in \mathbb{R}^n$ denotes the "attacked" states, reflecting the difference between the system states under attack and those under normal operation.

The covert attack strategy has been shown to be undetectable to any anomaly detectors if the attackers are able to compromise enough number of sensor measurements. In addition, it has been discussed in [140,141] that an attack is undetectable if and only if it excites the system zero-dynamics. In order to reveal stealthy attacks, it is required to reduce the disruption capabilities of the attackers.

In this manuscript, we propose to utilize the sensor protection framework for rendering the covert attack detectable. This framework includes the sensor protection scheme or the sensor placement strategy. The sensor protection scheme consists in implementing some protection measures so that the measurements of some "critical" sensors can not be modified by the attackers. These critical sensors should be chosen such that their sensor measurements are suffered from abrupt/recipient changes under attack conditions. The sensor placement strategy focus on equipping the system with more secure sensors for creating physical redundancy. Similar to the sensor protection scheme, it is required that the effects of the attacks are reflected in the changes in measurements of these new equipped sensors.

The sensor protection scheme is reflected in the protection matrix $M$. Without loss of generality, it can be assumed that the matrix $M$ is diagonal, i.e., $M = \text{diag}(\gamma_j)$ such that $\gamma_j = 0$ if the sensor $S_j$ is protected and $\gamma_j = 1$ if the sensor $S_j$ is vulnerable. The sensor placement strategy can be modeled in the same manner. It has been also shown that the sensor placement strategy can be utilized for rendering different types of stealthy attacks (see, for example, in [121], [186] or [99]).

### 1.3.3 Fault detection and isolation approach

It has been discussed in [26] that the distinct difference between SCADA systems and IT systems lies in the interaction of the former with the physical world. The information security approach is dedicated to improving the security of SCADA systems by protection measures. The compatibility between the cyber layer and the physical infrastructure has not been considered. The secure control theory approach focuses mainly on investigating system vulnerabilities, designing stealthy/deception attacks, and proposing countermeasures for revealing undetectable attacks. The joint detection and isolation of attacks have not been considered seriously. Fortunately, the fault diagnosis community has contributed with methodologies for dealing with abnormal situations occurring to stochastic-dynamical systems under the model uncertainties, disturbances and random noises [30]. Recently, the fault detection and isolation (FDI) techniques have been adapting to the detection and isolation of cyber attacks on SCADA systems.

#### Fault diagnosis

Fault diagnosis, together with fault-tolerant control, is an active domain of control theory. Fault diagnosis is concerned with the detection, isolation and identification of faults. According to [206], the fault detection and isolation (FDI) problem consists in "making a binary decision - either that something has gone wrong or everything is fine, and of determining the location as well as nature of the fault" while the purpose of fault identification is to estimate the size, type or nature of the fault. The target of fault-tolerant control, on the other hand, is to ensure the normal operation of the system under faulty condition by reconfiguration mechanisms.

There has been an extremely vast literature on the fault diagnosis of stochastic-dynamical systems, see, for example, in [10,30,35,65,81,83]. The main purpose of a statistical FDI algorithm is to decide whether the fault has occurred and then to identify the types of the fault with respect to random noises and unknown system states (often regarded as the nuisance parameter). This task consists in calculating a pair $(T, \nu)$, where $T$ is the *stopping time* at which the *final decision* $\nu$, i.e., the change type, is decided. The fault diagnosis problem is generally solved by utilizing the analytical redundancy approach, which is comprised of two steps: residual generation and residual evaluation. The so-called residuals are first generated by employing traditional techniques such as the state observer approach [55,92,190,209], the Kalman filter approach [10,116,205,207,208], the parity space approach [31,54,64,73] or the parameter estimation approach [82], etc. They are then evaluated by utilizing statistical decision theory [10,109,175], including non-sequential hypothesis testing [19,49,109], sequential hypothesis testing [105,175,194,195] or sequential change-point detection and isolation [10,105,175].

The preliminary step of model-based FDI methods, however, is to develop the mathematical model of SCADA systems. By linearizing the PDEs around the operating point, the majority of SCADA systems can be described in the time-invariant state space model (1.1). The derivation of system model under faulty condition depends on the type and the position of the fault. Generally, faults occurring to a dynamical system can be classified into three main categories: component fault, actuator fault and sensor fault. In FDI literature, these faults are assumed to be non-colluding, i.e., there is only one fault occurring at any time instant.

The system model under faulty condition (i.e., actuator faults, component faults and sensor

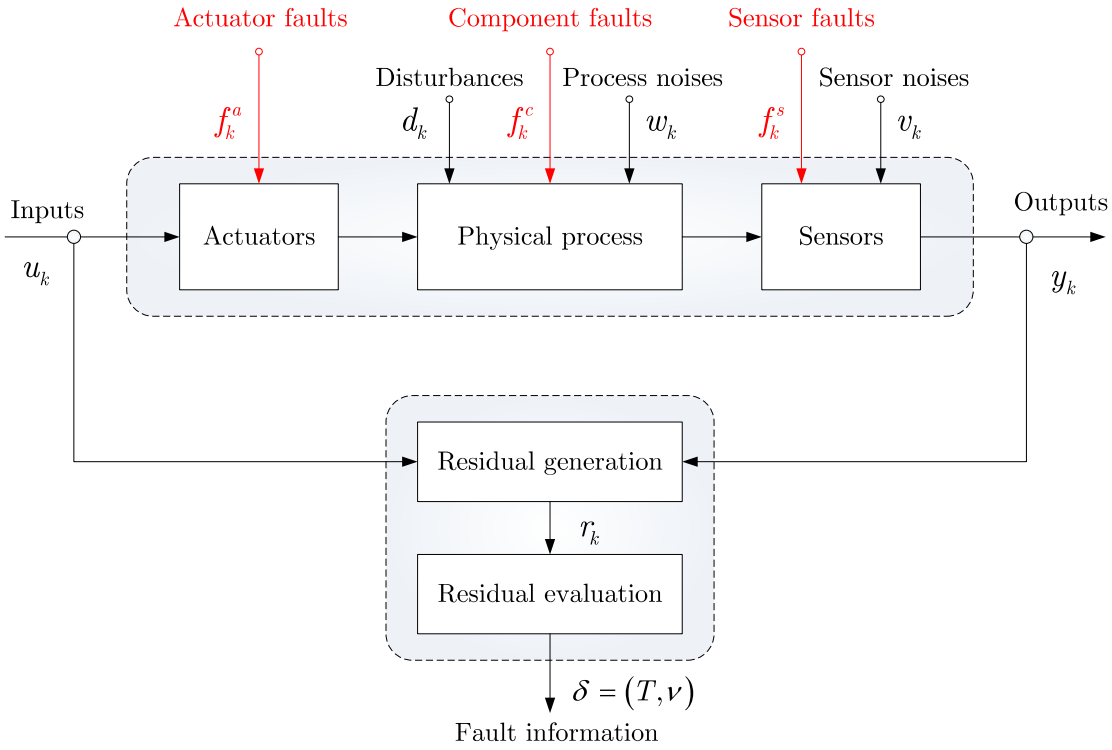



Figure 1.6 – Structure of model-based fault diagnosis: residual generation and residual evaluation.

faults) can be described as

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + K_1 f_k^c + K_2 f_k^a + w_k \\ y_k &= Cx_k + Du_k + Gd_k + H_1 f_k^c + H_2 f_k^a + M f_k^s + v_k \end{cases}, \quad x_0 = \overline{x}_0, \tag{1.10}$$

where $f_k^c \in \mathbb{R}^n$ is the vector of component faults, $f_k^a \in \mathbb{R}^m$ is the vector of actuator faults, $f_k^s \in \mathbb{R}^p$ is the vector of sensor faults; the matrices $K_1 \in \mathbb{R}^{n\times n}$, $K_2 \in \mathbb{R}^{n\times m}$, $H_1 \in \mathbb{R}^{p\times n}$, $H_2 \in \mathbb{R}^{p\times m}$, and $M \in \mathbb{R}^{p\times p}$ are assumed to be known.

**Remark 1.4.** *Let us discuss the system model under a faulty condition (1.10). Generally, faults are assumed to be non-colluding, i.e., there is only one fault occurring at any time instant. For example, if there is a fault occurring at the actuator $j$-th (i.e., modeled by $f_k^a(j) \neq 0$ for $1 \leq j \leq m$ and $f_k^a(l) = 0$ for $1 \leq l \neq j \leq m$, where $f_k^a(j)$ denotes the fault of actuator $j$-th), the component faults and the sensor faults must not occur (i.e., $f_k^c = 0$ and $f_k^s = 0$) and vice versa.*

### Faults and attacks

It has been shown that FDI tools could be used for detecting and mitigating the negative impact of cyber attacks on networked control systems [187]. However, these tools might be exploited more successfully if we could figure out the similarities and the differences between faults and attacks. Both faults and attacks occur at an unknown time instant and they cause unpredictable changes in the behavior of physical systems. Moreover, it follows from (1.2)–(1.10) that both

faults and attacks can be modeled as additive signals on both equations of the discrete-time state space model. Faults and attacks, however, possess inherently distinct features, making it difficult for traditional FDI techniques to be directly applied to detect cyber-physical attacks.

Firstly, the most significant difference between a fault and an attack lies in that the fault is considered as a phenomenon occurring randomly in each component (such as actuators, sensors, or communication channels, etc...) of a system while the attack is an intentional action performed by malicious adversaries. In addition, simultaneous faults are generally assumed to be non-colluding while cyber attacks could be performed in a coordinated way. For these reasons, cyber-physical attacks may cause more catastrophic damage to the system than faults do.

Secondly, in comparison with faults, cyber-physical attacks are much more difficult to detect/isolate since they can be performed in a coordinated way. It has been shown that attack vectors can be manipulated for partially or completely bypassing traditional anomaly detectors (e.g., replay attack [120, 122], false data injection attack [119, 121], zero-dynamics attack [186], or covert attack [169]). Hence, it is required to implement some *a priori* methods for rendering the attacks detectable/identifiable before applying detection/isolation techniques. Fortunately, revealing methods may be available from the security analysis process by the secure control theory approach.

Finally, faults often occur for a long time until they are detected/isolated and repaired while malicious attacks may be performed within a short period due to the limited resources of the adversaries [7, 26, 27, 80]. From the other hand, for safety-critical applications, it is required to detect the attacks with the detection delay upper bounded by a certain prescribed value [9, 68–70]. For these reasons, the detection and identification of attacks should be formulated as the sequential detection and isolation of transient changes in stochastic-dynamical systems.

### Related works

The application of traditional FDI techniques to the detection and isolation of cyber attacks has received considerable amount of research effort. For instance, the authors in [27] have formulated the problem of detecting cyber attacks on process control systems as the fault diagnosis problem. The process control systems are described as a discrete-time linear time-invariant system. The estimated outputs are compared to the received measurements, which are probably compromised, to generate the sequence of residuals. The residuals are then evaluated by using either sequential hypothesis testing or sequential change-point detection techniques. In order to circumvent the unknown parameters, the authors propose using the non-parametric CUSUM algorithm to detect the attack. The disadvantages of this work, however, lie in that it has not considered neither the effects of random noises nor the isolation problem.

Moreover, the security problem of SCADA water irrigation canals against cyber-physical attacks has been treated in [4–7]. The SCADA architecture for water irrigation networks is proposed. The system architecture consists of three layers: supervisory control layer, regulatory control layer and physical layer. The physical layer is modeled by the discrete time-delay state space model [92]. This model is obtained by solving a set of partial differential equations. The automatic control layer contains PI controllers to regulate water flow in the network while the supervisory control layer is equipped with a model-based diagnosis scheme. The diagnosis scheme is designed by utilizing a set of Unknown Input Observers (UIO) [30] adapted to the time-delay system [92]. It has been shown that the UIO-based diagnosis scheme can detect and isolate only the random faults in sensors or actuators. However, this UIO-based scheme can not diagnose

malicious attacks from intelligent adversaries who have knowledge about the system's model, diagnosis scheme and have capability to compromise control signals and sensor measurements.

A comprehensive framework has been proposed in [140,141] for detecting and identifying attacks on cyber-physical systems. The following discrete-time state space model[5] has been considered

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + B_a a_k \\ y_k &= Cx_k + Du_k + Gd_k + D_a a_k \end{cases}; \quad x_0 = \overline{x}_0, \tag{1.11}$$

where $a_k \in \mathbb{R}^{n+p}$ is the attack vector, the attack matrices $B_a \in \mathbb{R}^{n\times(n+p)}$ and $D_a \in \mathbb{R}^{p\times(n+p)}$. It has been shown in [140, 141] that the model (1.11) can be utilized for describing various types of cyber attacks found in literature, including the replay attack, the false data injection attack and the covert attack.

The necessary and sufficient conditions for constructing undetectable and unidentifiable attacks are introduced. Moreover, centralized and distributed algorithms are proposed to detect and isolate the detectable and identifiable attacks. Similar to the work in [6,7], the isolation problem is solved by exploiting the UIO techniques. It has been shown that these algorithms are optimal in the sense that they can detect (resp. identify) any detectable (resp. identifiable) attacks. The main drawback of this work is that it has been formulated in deterministic framework (without random noises).

### 1.3.4 Discussion

The defense-in-depth of SCADA systems against cyber attacks requires the integration of both information security approach, secure control theory approach and fault detection and isolation approach. The IT-based methods provide us with countermeasures for protecting safety-critical infrastructures from cyber attacks on the control center. The secure control theory-based methods focus mainly on (1) investigating the vulnerabilities of networked control systems modeled by discrete-time state space form, (2) designing stealthy/deception attacks for partially or completely bypassing traditional anomaly detectors, and (3) proposing countermeasures for revealing such undetectable attacks. The FDI approach, on the other hand, deals with the detection and identification of cyber attacks by adapting traditional FDI techniques to attack detection-isolation scenarios.

The state space model has often been employed for describing the operation of SCADA systems under normal operation as well as under cyber attacks. Specially, cyber attacks are modeled as additive signals impacting both state evolution and sensor measurement equations. The secure control framework (1.1)–(1.2) proposed in [187] can be utilizing for analyzing the security of networked control systems against various types of cyber attacks. However, this framework can be even improved by integrating both cyber and physical activities into a coordinated attack.

Let us consider the following discrete-time state space model under cyber-physical attacks:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + K_1 a_k^p + K_2 a_k^u + w_k \\ y_k &= Cx_k + Du_k + Gd_k + H_1 a_k^p + H_2 a_k^u + M a_k^y + v_k \end{cases}; \quad x_0 = \overline{x}_0, \tag{1.12}$$

[5] Originally, the continuous-time descriptor system without the vector of control signals $u_k$ and the vector of disturbances $d_k$ is utilized in [140, 141] for describing cyber-physical systems under attack. However, it has been discussed that similar results in [140, 141] can be applied directly to the discrete-time descriptor systems and/or non-singular systems with known inputs (e.g., the control signals $u_k$ and the disturbances $d_k$). For these reasons, the discrete-time state space model with both control signals $u_k$ and disturbances $d_k$ is written here for being consistent with previously used models.

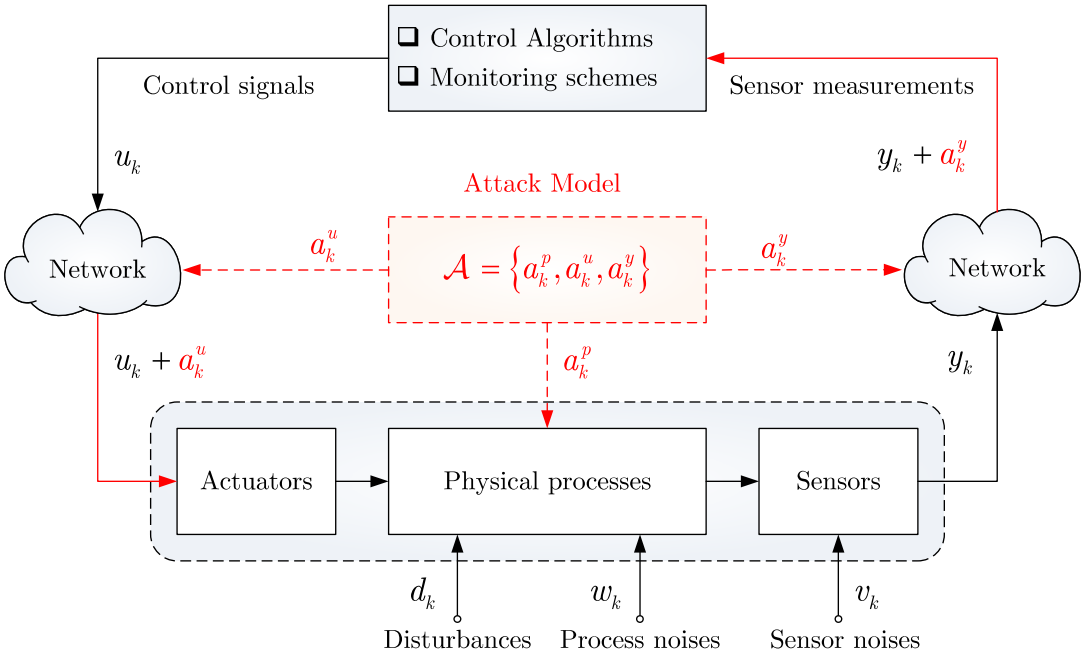



Figure 1.7 – Cyber-physical attacks on SCADA systems: physical attacks on processes (i.e., modeled by physical attack vector $a_k^p$), cyber attacks on control signals (i.e., modeled by control attack vector $a_k^u$), and on sensor measurements (i.e., modeled by sensor attack vector $a_k^u$).

where $a_k^p \in \mathbb{R}^n$ is the attack vector on physical processes, $a_k^u \in \mathbb{R}^m$ is the attack vector on control signals, $a_k^y \in \mathbb{R}^p$ is the attack vector on sensor measurements; the matrices $K_1 \in \mathbb{R}^{n\times n}$, $K_2 \in \mathbb{R}^{n\times m}$, $H_1 \in \mathbb{R}^{p\times n}$, $H_2 \in \mathbb{R}^{p\times m}$ and $M \in \mathbb{R}^{p\times p}$.

For simplifying the model (1.12), the attack vectors $a_k^p$ and $a_k^u$ are grouped into the state attack vector $a_k^x = \left[(a_k^p)^T, (a_k^u)^T\right]^T \in \mathbb{R}^r$, where $r = n + m$. The corresponding attack matrices are $K = [K_1, K_2] \in \mathbb{R}^{n\times r}$ and $H = [H_1, H_2] \in \mathbb{R}^{p\times r}$. The discrete-time state space model (1.12) can be rewritten as

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + Ka_k^x + w_k \\ y_k &= Cx_k + Du_k + Gd_k + Ha_k^x + Ma_k^y + v_k \end{cases}; \quad x_0 = \overline{x}_0. \tag{1.13}$$

For simplifying the model (1.13), let us define the attack matrices $B_a = [K, 0] \in \mathbb{R}^{n\times s}$ and $D_a = [H, M] \in \mathbb{R}^{p\times s}$ and the attack vector $a_k = \left[(a_k^x)^T, (a_k^y)^T\right]^T \in \mathbb{R}^s$, where $s = r + p$. Finally, the discrete-time state space model (1.13) can be simplified as

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + B_a a_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + D_a a_k + v_k \end{cases}; \quad x_0 = \overline{x}_0, \tag{1.14}$$

where the attack vector $a_k$ is designed by the attacker and the attack matrices $B_a$ and $D_a$ are decided by system operators.

The reaction of SCADA systems to cyber-physical attacks is determined by the attack components $B_a a_k$ and $D_a a_k$. The attack matrices $B_a$ and $D_a$ depend on the system architecture, i.e., system operators, while the attack vector $a_k$ is designed by the attacker. Sometimes, the

adversary may be forced to perform his malevolent action within a short period due to the resource limit. In addition, the malicious attack should be terminated once the attacker achieves his target. For these reasons, the attack vector $a_k$ should be designed such that $a_k \neq 0$ if $k \in \tau_a$ and $a_k = 0$ otherwise, where $\tau_a = [k_0, k_0 + L - 1]$ is the attack duration.

**Remark 1.5.** *Let us add some comments on several key differences between the proposed discrete-time state space model (1.12)–(1.14) and various models used in the literature. Firstly, it can be noticed that the system model under the faulty condition (1.10) and the system model under cyber-physical attack (1.12) have the same structure, i.e., the attack vector on physical processes $a_k^p$ corresponds to the component fault vector $f_k^c$, the attack vector on control signals $a_k^u$ corresponds to the actuator fault vector $f_k^a$, and the attack vector on sensor measurements $a_k^y$ corresponds to the sensor fault vector $f_k^s$. The principal difference between the models (1.10) and (1.12) lies in the assumption on the faulty vectors (i.e., $f_k^c$, $f_k^a$ and $f_k^s$) and the attack vectors (i.e., $a_k^p$, $a_k^u$ and $a_k^y$) themselves. In fault diagnosis literature, the faults are generally assumed to be non-colluding, i.e., there is always at most one fault occurring at any time instant, while cyber-physical attacks can be performed at the same time. Secondly, the secure control framework (1.2) proposed in [187] has not taken into consideration the physical attack on processes. In fact, the authors in [187] have considered the physical attack vector $a_k^p$ as the unknown signals representing the effects of faults. Moreover, the attack vector on control signals $a_k^u$ and the attack vector on sensor measurements $a_k^y$ have not been integrated into a single vector, thus making difficult for the design of detection-isolation schemes. Finally, the modeling framework (1.11) proposed in [140, 141] has not taken into consideration the random noises. For these reasons, it can be concluded that the proposed models (1.12)–(1.14) are the generalization of those found in literature. These models will be utilized throughout this manuscript for analyzing the security of SCADA systems against cyber-physical attacks as well as for designing detection-isolation schemes.*

## 1.4 Conclusion

Industrial control systems in general and SCADA systems in particular have been playing a vital role in safety-critical infrastructures of a nation, including transportation systems, electric power grids, gas pipelines, water distribution networks, etc. Along with the development in information and communication technology, modern SCADA systems are becoming more and more vulnerable to cyber-physical attacks, not only on physical infrastructures, but also on the communication network and the control center. Due to their essential role, the security of SCADA systems against malicious attacks has received significant research attention over the last few years.

Though information security approach may provide some protection methods that help in improving the security of SCADA systems, these methods appear to be not sufficient for the defense-in-depth of the systems against malicious attacks being capable of bypassing information security layers, as in the case of Stuxnet incident in 2010. Hence, the secure control approach is considered as a complementary partner to IT-based methods in protecting large-scale ICSs against cyber attacks. However, secure control methods have focused mainly on investigating the vulnerabilities of networked control systems, designing stealthy/deception attacks on the systems and then proposing some countermeasures for rendering these attacks detectable.

The FDI approach, on the other hand, concentrates on the detection and identification of detectable and identifiable attacks. The statistical FDI problem was generally solved by the

analytical redundancy approach, which is composed of residual generation and residual evaluation approach. The residuals are first generated by utilizing traditional methods such as the Kalman filter approach, the parity space approach or the parameter estimation approach. Due to the irreducible effects of random noises, the residuals must be evaluated by using statistical hypothesis testing or the change-point detection/isolation methods.

Sequential analysis seems to be the most suitable tool to the monitoring of SCADA systems against cyber-physical attacks due to inevitable effects of random noises. Based on the idea introduced in [140,141,187], we utilize the discrete-time linear state space model with Gaussian noises impacting both equations to describe SCADA systems. The cyber-physical attacks are modeled as additive signals of short duration on both state evolution and sensor measurement equations. As an extension to the modeling framework in [140,141,187], our framework contains almost recent cyber-physical attack strategies, including both DoS attacks, simple integrity attacks and stealthy/deception attacks. The attack signals are modeled as additive changes of short period to reflect the resource limits (if any) of the attacker. Moreover, for safety-critical applications, it is required to detect the attacks with the detection delay upper bounded by a certain prescribed value [9, 67, 69]. For these reasons, it is more adequate to formulate the detection and identification of cyber-physical attacks on SCADA systems as the sequential detection and isolation of transient changes in stochastic-dynamical systems.

# Part I

# Sequential Detection and Isolation of Transient Signals in Stochastic-dynamical Systems

In the first chapter, we have introduced the problem of detecting and isolating cyber-physical attacks on SCADA systems. It has been pointed out that current tools in fault diagnosis community should be revised so as to adapt to the attack scenarios. Due to the inevitable effects of random noises, statistical tools must be considered in the decision-making processes. This part consists of developing some detection and isolation algorithms that are appropriate for the on-line monitoring of safety-critical infrastructures against malicious attacks by integrating current tools in both fault diagnosis and statistics.

In chapter 2, we describe the state of the art in statistical decision theory, including the classical hypothesis testing, sequential hypothesis testing, sequential change-point detection and isolation, and sequential detection of transient changes. The sequential detection of transient signals, integrated with some residual-generation methods from the FDI community, will be shown to be the most appropriate approach for the on-line surveillance of SCADA systems against cyber-physical attacks. For this reason, we formulate the attack detection problem as the problem of detecting transient changes in stochastic-dynamical systems. This problem will be considered in chapter 3, where several sub-optimal detection algorithms are proposed and their statistical properties are investigated. Finally, some preliminary results on problem of jointly detecting and isolating transient signals in stochastic-dynamical systems are considered in chapter 4.

# Chapter 2

# Statistical Decision Theory

## Contents

## 2.1 Introduction

The security of SCADA systems against cyber-physical attacks has been introduced in chapter 1. It has been discussed that the problem of attack detection and identification could be formulated as the problem of detecting and isolating transient changes (i.e., the changes of short duration) in stochastic-dynamical systems. The attack detection and isolation problem is, therefore, closely related to the fault diagnosis problem. The target of a statistical FDI problem is to distinguish the "normal" operation from the "abnormal" behavior under the effects of model uncertainties, disturbances and random noises. The model uncertainties and disturbances can be reduced or even eliminated by utilizing robust model-based fault detection techniques [30]. The effects of random noises, on the other hand, must be treated by exploiting results from the statistical decision theory. The statistical decision theory, which is concerned with the decision-making process in the presence of random variables, includes three sub-domains: the classical (non-sequential) hypothesis testing problem, the sequential hypothesis testing problem and the sequential change-point detection-isolation problem.

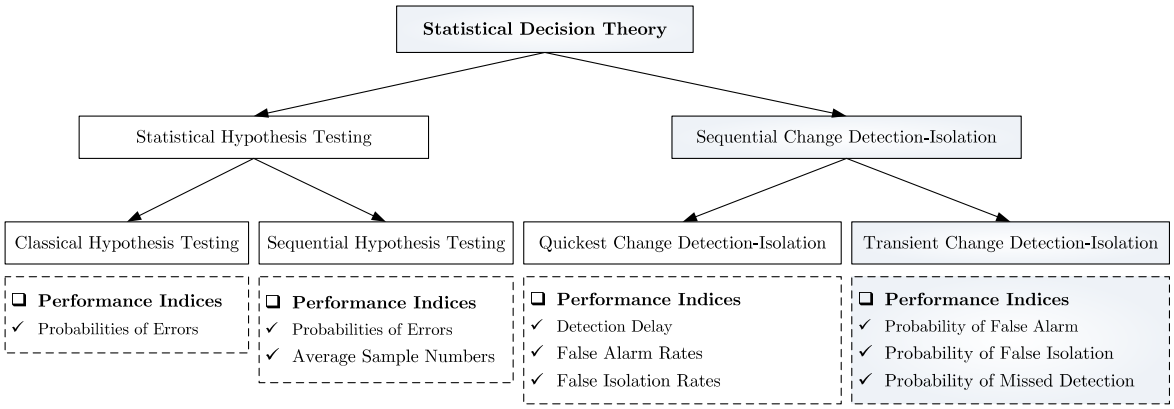



Figure 2.1 – Sub-domains in statistical decision theory.

The classical hypothesis testing theory, whose main results are given in section 2.2, is concerned with the choice between two or multiple conjectures (or hypotheses) based on the set of fixed-size samples $Y_n = (y_1, y_2, \cdots, y_n)$. The sequential hypothesis testing problem, on the other hand, deals with any statistical tests where the number of samples is not *a priori* fixed but it depends on the observations themselves. The sequential change-point detection-isolation problem addresses the detection and identification of abrupt changes (i.e., the changes of infinitely long duration) in stochastic processes. Recent results on the sequential analysis domain, which includes the sequential hypothesis testing and quickest change-point detection-isolation, are introduced in section 2.3 and section 2.4, respectively. The sequential detection of transient changes (i.e., the short-duration signals) is resumed in section 2.5.

## 2.2 Non-sequential Hypothesis Testing

The classical (non-sequential) hypothesis testing problem consists of deciding one of multiple hypotheses $\mathcal{H}_0, \mathcal{H}_1, \cdots, \mathcal{H}_K$ based on the observations $Y_n = (y_1, y_2, \cdots, y_n)$ of fixed size $n$ generated from a parametric family of distributions $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ depending on the parameter $\theta$.

Since the sample size $n$ is fixed, non-sequential tests are often called the Fixed-Size Sample (FSS) tests. Denote by $\Theta$ the *parameter space* including all possible values of $\theta$. The parameter space $\Theta$ can be divided into $K+1$ non-empty disjoint sets $\Theta_0, \Theta_1, \cdots, \Theta_K$ satisfying $\Theta_l \cap \Theta_j = \emptyset$ for $l \neq j$ and $\Theta_0 \cup \Theta_1 \cup \cdots \cup \Theta_K = \Theta$. The target of the hypothesis testing problem is to decide one of $K+1$ hypotheses $\mathcal{H}_l = \{\theta \in \Theta_l\}, 0 \leq l \leq K$ based on the observations $Y_n = (y_1, y_2, \cdots, y_n)$. Let $\Omega$ denote the *observation space* which is defined as all possible values of the observations $Y_n = (y_1, y_2, \cdots, y_n)$. The problem of testing $K+1$ hypotheses $\{\mathcal{H}_l\}_{0 \leq l \leq K}$ corresponds to the fragmentation of the observation space $\Omega$ into $K+1$ disjoint regions $\Omega_l$, for $0 \leq l \leq K$, on which the hypotheses $\mathcal{H}_l$, for $0 \leq l \leq K$, are accepted.

Generally, the statistical hypothesis testing problem is asymmetric. The hypothesis $\mathcal{H}_0$ is called the *null hypothesis*, corresponding to the normal operation of a system and other hypotheses $\mathcal{H}_1, \mathcal{H}_2, \cdots, \mathcal{H}_K$ are denoted as the *alternative hypotheses* (or simply as the alternatives), corresponding to the abnormal behavior of the system. When $K = 1$ (i.e., the detection task), the problem is to decide the null hypothesis $\mathcal{H}_0$ against the alternative hypothesis $\mathcal{H}_1$. When $K > 1$ (i.e., the diagnosis task), the problem is to select one of multiple hypotheses $\mathcal{H}_0, \mathcal{H}_1, \cdots, \mathcal{H}_K$.

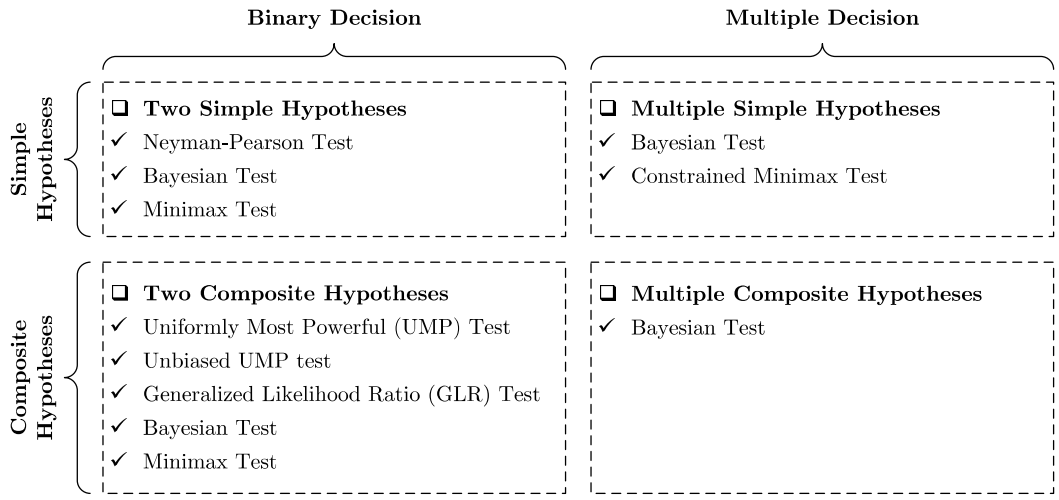



Figure 2.2 – Classical hypothesis testing methods.

This section is organized as follows. In subsection 2.2.1, we give some basic definitions about the statistical hypothesis testing problem. The problem of testing between two simple and composite hypotheses is introduced in subsection 2.2.2 and subsection 2.2.3, respectively. Finally, the multiple hypothesis testing problem is considered in subsection 2.2.4.

### 2.2.1 Basic definitions

In this subsection, we introduce main definitions and optimality criteria of the statistical hypothesis testing framework.

**Definition 2.1.** *(Simple Hypothesis [19, 50, 67, 126]). A simple hypothesis is any assumption which uniquely determines the distribution of the observations* $Y_n = (y_1, y_2, \cdots, y_n)$.

In the parametric framework, a hypothesis $\mathcal{H}_l = \{\theta \in \Theta_l\}$ depending on the parameter $\theta$ is simple if the distribution $P_\theta$ of the observations $Y_n = (y_1, y_2, \cdots, y_n)$ under the hypothesis $\mathcal{H}_l$ is

specified completely. In other words, the subset $\Theta_l$ is reduced to $\Theta_l \equiv \{\theta_l\}$, for $0 \leq l \leq K$, where $\theta_0, \theta_1, \cdots, \theta_K$ are fixed points from $\Theta$. Hence, the problem is to choose one of multiple hypotheses $\mathcal{H}_l = \{\theta = \theta_l\}$ or $\mathcal{H}_l = \{y_1, y_2, \cdots, y_n \sim P_{\theta_l}\}$, where the parameters $\theta_l$, for $l = 0, 1, \cdots, K$, are completely known.

**Definition 2.2.** *(Composite Hypothesis [50, 67, 126, 175]). Any non-simple hypothesis is called a composite hypothesis.*

In the parametric case, a composite hypothesis $\mathcal{H}_l$ can be written as $\mathcal{H}_l = \{\theta \in \Theta_l\}$ or $\mathcal{H}_l = \{y_1, y_2, \cdots, y_n \sim P_\theta | \theta \in \Theta_l\}$, for $0 \leq l \leq K$, where the subsets $\Theta_l \cap \Theta_j = \emptyset$ for $l \neq j$ and $\Theta_l$ is not reduced to a single point $\theta_l$, for $0 \leq l \leq K$.

**Definition 2.3.** *(Statistical Test [19, 50, 67, 126]). A statistical test for testing between $K+1$ hypotheses is any measurable mapping $\delta : \Omega \mapsto \{\mathcal{H}_0, \mathcal{H}_1, \cdots, \mathcal{H}_K\}$ from the observation space $\Omega$ onto the set of hypotheses $\mathcal{H}_0, \mathcal{H}_1, \cdots, \mathcal{H}_K$.*

Statistical tests can be broadly classified into two types: randomized tests and non-randomized tests. In this manuscript, we consider only non-randomized tests. Interested readers are referred to [175] for more details on randomized tests.

The statistical test $\delta(Y_n)$, therefore, can be considered as a random "variable" which takes on the values $\mathcal{H}_0, \mathcal{H}_1, \cdots, \mathcal{H}_K$. If $\delta(Y_n) = \mathcal{H}_l$, then we accept hypothesis $\mathcal{H}_l$, that is, we decide that the parameter $\theta \in \Theta_l$, for $0 \leq l \leq K$.

In hypothesis testing problems, some optimality criteria are often introduced for comparing various statistical tests. For the sake of simplicity, let us consider now the case of testing between multiple simple hypotheses $\mathcal{H}_0, \mathcal{H}_1, \cdots, \mathcal{H}_K$. The quality of a test $\delta$ is generally characterized by the set of probabilities of erroneous decisions:

$$\alpha_{jl}(\delta) = \mathbb{P}_{\theta_j}(Y_n \in \Omega_l) = \mathbb{P}_{\theta_j}(\delta(Y_n) = \mathcal{H}_l), \quad 0 \leq j \neq l \leq K, \tag{2.1}$$

$$\alpha_j(\delta) = \mathbb{P}_{\theta_j}(Y_n \notin \Omega_j) = \mathbb{P}_{\theta_j}(\delta(Y_n) \neq \mathcal{H}_j), \quad 0 \leq j \leq K, \tag{2.2}$$

where $\mathbb{P}_{\theta_j}(\cdot)$ is the probability of an event $(\cdot)$ when the hypothesis $\mathcal{H}_j$ is true (i.e., $\theta = \theta_j$), $\alpha_{jl}(\delta)$ denotes the probability of deciding hypothesis $\mathcal{H}_l$ when hypothesis $\mathcal{H}_j$ is true, and $\alpha_j(\delta)$ stands for the probability of rejecting hypothesis $\mathcal{H}_j$ when it is true. This number $\alpha_j(\delta) = \sum_{l \neq j} \alpha_{jl}(\delta)$ is also denoted as the probability of errors of $j$-th kind for the test $\delta$ [19].

The probability of rejecting the null hypothesis $\mathcal{H}_0$ when it is true is called the *probability of false alarm* and it is defined mathematically as

$$\alpha_0(\delta) = \mathbb{P}_{\theta_0}(\delta(Y_n) \neq \mathcal{H}_0). \tag{2.3}$$

The *power* of the test $\delta$, on the other hand, is defined by the set of probabilities of correct decisions

$$\beta_j(\delta) = \mathbb{P}_{\theta_j}(\delta(Y_n) = \mathcal{H}_j), \quad j = 1, 2, \cdots, K. \tag{2.4}$$

It is clear that $\beta_j(\delta) = 1 - \alpha_j(\delta)$, for all $j = 0, 1, \cdots, K$. It is desirable that the probabilities of errors $\alpha_j(\delta)$, for a given test $\delta$, be as small as possible. However, since the sample size of the observations is fixed at $n$, it is impossible to make all probabilities $\alpha_j(\delta)$ arbitrarily small. Then, the question that arises naturally is how to compare various tests.

**Definition 2.4.** *(Better Test [19]). A test $\delta_1$ is better than $\delta_2$ if, for all $j = 0, 1, \cdots, K$, we have $\alpha_j(\delta_1) \leq \alpha_j(\delta_2)$, and the inequality must be strict for at least one $j$.*

However, it is not always possible to compare two tests $\delta_1$ and $\delta_2$ in this sense. Three possible optimality criteria have been introduced for comparing statistical tests, including the most powerful approach, the Bayesian approach and the minimax approach. Readers are referred to [175, pages 88–90] for the philosophical backgrounds of these criteria.

**Most powerful approach.** Denote by $C_{\alpha_0,\cdots,\alpha_{K-1}}$ a class of tests with $K$ upper bounds for probabilities of errors of rejecting the true hypotheses:

$$C_{\alpha_0,\cdots,\alpha_{K-1}} = \{\delta : \alpha_j(\delta) \leq \alpha_j,\ j = 0, 1, \cdots, K-1\}. \tag{2.5}$$

Within the class $C_{\alpha_0,\cdots,\alpha_{K-1}}$, it is possible to order various tests by the value $\alpha_K(\delta)$: the smaller $\alpha_K(\delta)$, the better the test $\delta$.

**Definition 2.5.** *(Most Powerful Test [19, 175]). A test $\delta^* \in C_{\alpha_0,\cdots,\alpha_{K-1}}$ is the most powerful test in the class $C_{\alpha_0,\cdots,\alpha_{K-1}}$ if for any $\delta \in C_{\alpha_0,\cdots,\alpha_{K-1}}$, we have*

$$\alpha_K(\delta^*) \leq \alpha_K(\delta). \tag{2.6}$$

**Bayesian approach.** This approach assumes the *a priori* distribution $Q = (q_0, q_1, \cdots, q_K)$ on the set of hypotheses $\mathcal{H}_0, \mathcal{H}_1, \cdots, \mathcal{H}_K$, where $q_j = \mathbb{P}(\mathcal{H}_j) \geq 0$, for $0 \leq j \leq K$ and $\sum_{j=0}^{K} q_j = 1$, are *a priori* probabilities of the hypotheses $\mathcal{H}_j$. Let also $L_{jl} = L_j(\delta = \mathcal{H}_l)$ be the loss function associated with the acceptance of hypothesis $\mathcal{H}_l$ when hypothesis $\mathcal{H}_j$ is true. The following average (integrated or weighted) loss can be utilized for comparing tests:

$$J_Q(\delta) = \sum_{j=0}^{K}\sum_{l=0}^{K} L_j(\delta = \mathcal{H}_l)\,\mathbb{P}(\mathcal{H}_j)\,\mathbb{P}_{\theta_j}(\delta(Y_n) = \mathcal{H}_l) = \sum_{j=0}^{K}\sum_{l=0}^{K} L_{jl} q_j \alpha_{jl}(\delta). \tag{2.7}$$

The average loss $J_Q(\delta)$ is sometimes called the *Bayes risk* associated with the loss function $L_{jl}$.

**Definition 2.6.** *(Bayesian Test [19, 175]). A test $\bar{\delta}$ is called the Bayes test if it minimizes the average loss $J_Q(\delta)$, for a given a priori distribution $Q$, i.e,*

$$\bar{\delta} = \arg\inf_{\delta} J_Q(\delta), \tag{2.8}$$

*where the infimum is taken over all FSS tests.*

In the particular case of the $0-1$ loss function, i.e., $L_{jl} = 0$ if $l = j$ and $L_{jl} = 1$ otherwise, the average loss $J_Q(\delta)$ is reduced to the average error probability $\alpha_Q(\delta)$:

$$J_Q(\delta) = \alpha_Q(\delta) = \sum_{j=0}^{K}\sum_{\substack{l=0\\ l\neq j}}^{K} q_j \alpha_{jl}(\delta) = \sum_{j=0}^{K} q_j \left(\sum_{\substack{l=0\\ l\neq j}}^{K} \alpha_{jl}(\delta)\right) = \sum_{j=0}^{K} q_j \alpha_j(\delta). \tag{2.9}$$

In this case, the Bayes test $\bar{\delta}$ minimizes the average error probability $\alpha_Q(\delta) = \sum_{j=0}^{K} q_j \alpha_j(\delta)$ over all FSS tests.

**Minimax approach.** The minimax approach is concerned with the maximum value $\alpha_{\max}(\delta)$ of the probabilities of errors:

$$\alpha_{\max}(\delta) = \max_j \alpha_j(\delta) = \max_Q \alpha_Q(\delta). \tag{2.10}$$

It is clear that the value $\alpha_{\max}(\delta)$ can be utilized for ordering tests.

**Definition 2.7.** *(Minimax Test [19, 175]). A test $\tilde{\delta}$ is called the minimax test if it minimizes the maximum probability of error $\alpha_{\max}(\delta)$, i.e.,*

$$\tilde{\delta} = \arg\inf_{\delta} \alpha_{\max}(\delta), \tag{2.11}$$

*where the infimum is taken over all FSS tests.*

**Remark 2.1.** *The Bayesian tests do have strong connections with both the minimax tests and the most powerful tests. Sometimes, it is possible to find some a priori distribution $Q$ of the hypotheses so that the Bayesian tests become the minimax tests or the most powerful tests. See [19, 50, 109, 175] for further discussion.*

### 2.2.2 Testing between two simple hypotheses

Let $Y_n = (y_1, y_2, \cdots, y_n)$ be the sequence of observations generated from the distribution $P_\theta$ depending on the parameter $\theta$ which may take either $\theta_0$ or $\theta_1$. The problem is to decide between the null hypothesis $\mathcal{H}_0 = \{\theta = \theta_0\}$ and the alternative one $\mathcal{H}_1 = \{\theta = \theta_1\}$ based on the observations $Y_n = (y_1, y_2, \cdots, y_n)$. In this case, the error probability of type I (i.e., the probability of false alarm), $\alpha_0(\delta) = \mathbb{P}_{\theta_0}(\delta \neq \mathcal{H}_0)$, is called the *size* of the test or the *level of significance* of the test. The value $\beta(\delta) = 1 - \alpha_1(\delta) = \mathbb{P}_{\theta_1}(\delta = \mathcal{H}_1)$ is called the *power* of the test, where $\alpha_1(\delta) = \mathbb{P}_{\theta_1}(\delta \neq \mathcal{H}_1)$ is the error probability of type II (i.e., the probability of missed detection). In order to compare various tests, let us define the class

$$C_\alpha = \{\delta : \mathbb{P}_{\theta_0}(\delta(Y_n) \neq \mathcal{H}_0) \leq \alpha\}, \tag{2.12}$$

including all FSS tests with the probability of false alarm upper bounded by a given value $\alpha \in (0, 1)$.

**Definition 2.8.** *(Likelihood Ratio). The Likelihood Ratio (LR) $\Lambda(Y_n)$ is defined as*

$$\Lambda(Y_n) = \frac{p_{\theta_1}(Y_n)}{p_{\theta_0}(Y_n)} = \frac{p_{\theta_1}(y_1, y_2, \cdots, y_n)}{p_{\theta_0}(y_1, y_2, \cdots, y_n)}, \tag{2.13}$$

*where $p_{\theta_j}(Y_n) = p_{\theta_j}(y_1, y_2, \cdots, y_n)$ is the joint probability density function (p.d.f.) of the observations $Y_n = (y_1, y_2, \cdots, y_n)$ under the distribution $P_{\theta_j}$, for $j = 0, 1$.*

When the observations $y_1, y_2, \cdots, y_n$ are independent identically distributed (i.i.d.), the joint p.d.f. of $Y_n$ is calculated as $p_{\theta_j}(Y_n) = \prod_{i=1}^n f_{\theta_j}(y_i)$, where $f_{\theta_j}(y_i)$ is the p.d.f. of the random variable $y_i$. The LR $\Lambda(Y_n)$ plays a critical role in constructing optimal tests, including the most powerful approach, the Bayesian approach and the minimax approach.

**Neyman-Pearson (N-P) Test.** The Neyman-Pearson test (or the most powerful test) for deciding betweentwo simple hypotheses is given in the following theorem. It is based on the fundamental lemma of Neyman-Pearson.

**Theorem 2.1.** *(Neyman-Pearson Test [19, 109]). Suppose that the function* $c \mapsto R(c) = \mathbb{P}_{\theta_0}(\Lambda(Y_n) \geq c)$ *is continuous for all* $c > 0$*. Then, the MP test* $\delta^*(Y_n)$ *in the class* $C_\alpha$ *given in* (2.12) *is defined as*

$$\delta^*(Y_n) = \begin{cases} \mathcal{H}_1 & \text{if} \quad \Lambda(Y_n) \geq h \\ \mathcal{H}_0 & \text{if} \quad \Lambda(Y_n) < h \end{cases}, \tag{2.14}$$

*where the threshold* $h$ *can be found by solving the equation* $\mathbb{P}_{\theta_0}(\Lambda(Y_n) \geq h) = \alpha$*.*

**Bayesian Test.** Let $q_j = \mathbb{P}(\mathcal{H}_j) \geq 0$, for $j = 0, 1$ and $q_0 + q_1 = 1$, be the *a priori* probabilities of the hypothesis $\mathcal{H}_j$. Consider the $0-1$ loss function case. The Bayes risk $J_Q(\delta)$ associated with the *a priori* distribution $Q = (q_0, q_1)$ corresponds to the average error probability $\alpha_Q(\delta)$ and it is written as

$$J_Q(\delta) = \alpha_Q(\delta) = q_0\alpha_0(\delta) + q_1\alpha_1(\delta). \tag{2.15}$$

**Theorem 2.2.** *(Bayesian Test [19, 109]). The Bayesian test* $\overline{\delta}(Y_n)$ *which minimizes the average error probability* $\alpha_Q(\delta)$ *in* (2.15) *is defined as*

$$\overline{\delta}(Y_n) = \begin{cases} \mathcal{H}_1 & \text{if} \quad \Lambda(Y_n) \geq h \\ \mathcal{H}_0 & \text{if} \quad \Lambda(Y_n) < h \end{cases}, \quad \text{where } h = \frac{q_1}{q_0}. \tag{2.16}$$

**Minimax test.** In the case of testing between two simple hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, the maximum value of error probabilities is given as $\alpha_{\max}(\delta) = \max(\alpha_0(\delta), \alpha_1(\delta))$. The minimax test is given in the following theorem.

**Theorem 2.3.** *(Minimax Test [19, 109]). The minimax test* $\tilde{\delta}(Y_n)$ *which minimizes the maximum value* $\alpha_{\max}(\delta)$ *of the error probabilities is defined as*

$$\tilde{\delta}(Y_n) = \begin{cases} \mathcal{H}_1 & \text{if} \quad \Lambda(Y_n) \geq h \\ \mathcal{H}_0 & \text{if} \quad \Lambda(Y_n) < h \end{cases}, \tag{2.17}$$

*where the threshold* $h$ *is chosen such that* $\mathbb{P}_{\theta_0}(\Lambda(Y_n) \geq h) = \mathbb{P}_{\theta_1}(\Lambda(Y_n) < h)$*.*

It is worth noting that the N-P test (2.14), the Bayesian test (2.16) and the minimax test (2.17) are likelihood ratio-based tests, i.e., the decision is made by comparing the LR $\Lambda(Y_n)$ with a threshold which is chosen for assuring an acceptable level of false alarm. Interested readers are referred to [19, 50, 109, 175] for further discussion on the relationship between the most powerful approach, the Bayesian approach and the minimax approach in statistical hypothesis testing.

### 2.2.3 Testing between two composite hypotheses

Consider now the problem of testing between two composite hypotheses $\mathcal{H}_0 = \{\theta \in \Theta_0\}$ and $\mathcal{H}_1 = \{\theta \in \Theta_1\}$, where $\Theta_0 \cap \Theta_1 = \emptyset$. In this case, the fundamental Neyman-Pearson lemma is no longer valid. Hence, optimality criteria used for testing two simple hypotheses need to be revised so as to adapt to composite scenarios.

**Definition 2.9.** *(Size of Composite Test [19, 109, 175]). The size or the error probability of the first kind $\alpha_0(\delta)$ for a composite test $\delta$ is defined as the maximal probability of rejecting the null hypothesis $\mathcal{H}_0$ when it is true, i.e.,*

$$\alpha_0(\delta) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(Y_n \notin \Omega_0) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\delta(Y_n) \neq \mathcal{H}_0). \tag{2.18}$$

**Definition 2.10.** *(Power of Composite Test [19, 109, 175]). The power $\beta(\delta, \theta)$ of a composite test $\delta$ is now a function of parameter $\theta$ and it is defined as the probability of correct acceptance of hypothesis $\mathcal{H}_1$ when the true parameter value is $\theta$, i.e.,*

$$\beta(\delta, \theta) = \mathbb{P}_\theta(Y_n \in \Omega_1) = \mathbb{P}_\theta(\delta(Y_n) = \mathcal{H}_1), \quad \theta \in \Theta_1, \tag{2.19}$$

*where $\beta(\delta, \theta)$ is also called the power function of the test $\delta$ since it depends on the parameter $\theta$.*

Let $C_\alpha$ denote a class of composite tests with the level of significance $\alpha$, for $\alpha \in (0, 1)$, i.e.,

$$C_\alpha = \left\{ \delta : \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\delta(Y_n) \neq \mathcal{H}_0) \leq \alpha \right\}. \tag{2.20}$$

Of course, it is desirable to construct a test $\delta$ in the class $C_\alpha$ given in (2.20) to maximize the power function $\beta(\delta, \theta)$ for all values of $\theta \in \Theta_1$.

**Definition 2.11.** *(UMP Test [19, 109, 175]). A test $\delta^*(Y_n)$ is said to be uniformly most powerful (UMP) test in the class $C_\alpha$ given in* (2.20) *if, for all other tests $\delta \in C_\alpha$, we have*

$$\beta(\delta^*, \theta) \geq \beta(\delta, \theta), \text{ for all } \theta \in \Theta_1. \tag{2.21}$$

Unfortunately, such UMP tests rarely exist in practical situations. Theoretical results on the hypothesis testing between two composite hypotheses have been developed for only some particular cases. We consider in the following two special cases.

**Monotone Likelihood Ratio and UMP test.** Let $Y_n = (y_1, y_2, \cdots, y_n)$ be generated from a parametric family of distributions $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ depending on the scalar parameter $\theta$ and the family $\mathcal{P}$ possesses monotone likelihood ratio. The UMP test exists in the case of testing between two composite hypotheses $\mathcal{H}_0 = \{\theta \leq \theta_0\}$ and $\mathcal{H}_1 = \{\theta > \theta_0\}$. Main results are given in the following.

**Definition 2.12.** *(Monotone LR [19, 109, 175]). Let $Y_n = (y_1, y_2, \cdots, y_n)$ be a sequence of random samples belonging to a parametric family of distributions $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ with the corresponding densities $p_\theta(Y_n)$, where the parameter $\theta$ is scalar. The family $\mathcal{P}$ is said to be with monotone likelihood ratio (LR) if there exists a function $T(Y_n)$ such that, for all $\theta_1$ and $\theta_0$ satisfying $\theta_1 > \theta_0$, the LR*

$$\Lambda(Y_n) = \frac{p_{\theta_1}(Y_n)}{p_{\theta_0}(Y_n)} = g(T(Y_n)) \tag{2.22}$$

*is a non-decreasing or non-increasing function of $T(Y_n)$.*

**Theorem 2.4.** *(UMP Test [19, 109, 175]). Suppose that the sequence of random samples $Y_n = (y_1, y_2, \cdots, y_n)$ is generated from a parametric family of distributions $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ depending on the scalar parameter $\theta$ and that the family $\mathcal{P}$ admits the monotone LR $\Lambda(Y_n) = g(T(Y_n))$.*

*Let $\theta_0$ be a fixed real number, then the UMP test $\delta^*(Y_n)$ for testing hypothesis $\mathcal{H}_0 = \{\theta \leq \theta_0\}$ against hypothesis $\mathcal{H}_1 = \{\theta > \theta_0\}$ in the class $C_\alpha$ given in (2.20) is defined as*

$$\delta^*(Y_n) = \begin{cases} \mathcal{H}_1 & \text{if} \quad T(Y_n) \geq h \\ \mathcal{H}_0 & \text{if} \quad T(T_n) < h \end{cases}, \tag{2.23}$$

*where the threshold $h$ can be found by solving the equation $\mathbb{P}_{\theta_0}(T(Y_n) \geq h) = \alpha$.*

**Unbiased Test.** So far we have discussed the one-sided alternative hypotheses. In many applications, it is required to consider two-sided alternative hypotheses, for example, to test $\mathcal{H}_0 = \{\theta = \theta_0\}$ against $\mathcal{H}_1 = \{\theta \neq \theta_0\}$. However, no UMP test exists except for particular examples [19,109,175]. Let us introduce, therefore, a subclass $\overline{C}_\alpha$ of the so-called *unbiased* tests in the class of UMP tests.

**Definition 2.13.** *(Unbiased Test [19, 109, 175]). A test $\delta$ for testing hypothesis $\mathcal{H}_0 = \{\theta \in \Theta_0\}$ against hypothesis $\mathcal{H}_1 = \{\theta \in \Theta_1\}$ in the class $C_\alpha$ defined in (2.20) is said to be unbiased if the following condition holds:*

$$\alpha_0(\delta) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\delta \neq \mathcal{H}_0) \leq \inf_{\theta \in \Theta_1} \mathbb{P}_\theta(\delta \neq \mathcal{H}_0) = \inf_{\theta \in \Theta_1} \beta(\delta, \theta). \tag{2.24}$$

It should be noted that this condition is very natural because the probability of rejection of $\mathcal{H}_0$ when it is false (i.e., $\inf_{\theta \in \Theta_1} \beta(\delta, \theta)$) must not be less than the probability of rejection of $\mathcal{H}_0$ when it is true (i.e., $\alpha_0(\delta)$). Before introducing the unbiased UMP test, let us consider the exponential family of distributions.

**Definition 2.14.** *(Exponential family of distributions [19, 109]). Let $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ be a parametric family of distributions depending on the scalar parameter $\theta$. The family $\mathcal{P}$ is said to be exponential if its p.d.f. has the form*

$$p_\theta(X) = c(\theta)\, h(X) \exp\{\nu(X)\, \theta\}, \tag{2.25}$$

*where $X \mapsto h(X)$ and $X \mapsto \nu(X)$ are functions from $\mathbb{R}^n$ to $\mathbb{R}$ and $\theta \mapsto c(\theta)$ is a function from $\mathbb{R}$ to $\mathbb{R}$.*

The problem is to design the unbiased UMP test for testing hypothesis $\mathcal{H}_0 = \{\theta \in [\theta_0, \theta_1]\}$ against hypothesis $\mathcal{H}_1 = \{\theta \notin [\theta_0, \theta_1]\}$, where $\theta_0, \theta_1 \in \mathbb{R}$ and $\theta_0 \leq \theta_1$, based on the observations $Y_n = (y_1, y_2, \cdots, y_n)$ generated from an exponential family of distributions $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ with scalar parameter $\theta$.

**Theorem 2.5.** *(Unbiased UMP Test [19, 109]). Let $Y_n = (y_1, y_2, \cdots, y_n)$ be random samples generated from an exponential family of distributions $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ depending on the scalar parameter $\theta$. The unbiased UMP test $\check{\delta}(Y_n)$ for testing hypothesis $\mathcal{H}_0 = \{\theta \in [\theta_0, \theta_1]\}$ against hypothesis $\mathcal{H}_1 = \{\theta \notin [\theta_0, \theta_1]\}$ in the class $C_\alpha$ given in (2.20) is defined as*

$$\check{\delta}(Y_n) = \begin{cases} \mathcal{H}_1 & \text{if} \quad T(Y_n) \notin [k_0, k_1] \\ \mathcal{H}_0 & \text{if} \quad T(Y_n) \in [k_0, k_1] \end{cases}, \tag{2.26}$$

*where $T(Y_n) = \sum_{i=1}^n \nu(y_i)$ and the thresholds $k_0$ and $k_1$ can be found by solving following equations:*

$$\mathbb{P}_{\theta_0}\left(\check{\delta} = \mathcal{H}_1\right) = \mathbb{P}_{\theta_1}\left(\check{\delta} = \mathcal{H}_1\right) = \alpha. \tag{2.27}$$

**Generalized Likelihood Ratio (GLR) Test.** So far we have seen that the optimal tests exist just in several particular cases. Unfortunately, the state of the art of the statistical theory has shown that it is impossible to define a test that is optimal in all situations (e.g. two or simple hypotheses, simple and composite hypotheses, scalar or vector, etc.). Even giving up the optimality criteria, it is difficult to formulate a test that is similar to the Neyman-Pearson test, i.e., by utilizing the LR, since the parameter $\theta$ is unknown. In order to circumvent this difficulty, it is proposed to utilize the Maximum Likelihood Estimation (MLE) of the parameter $\theta$ in $\Theta_0$ and $\Theta_1$ instead of its exact value for calculating the LR. The LR that uses the MLE of parameter $\theta$ is called the Generalized LR (GLR) and it is defined mathematically as

$$\hat{\Lambda}\left(Y_n\right) = \frac{\sup_{\theta\in\Theta_1} p_\theta\left(Y_n\right)}{\sup_{\theta\in\Theta_0} p_\theta\left(Y_n\right)} = \frac{\sup_{\theta\in\Theta_1} p_\theta\left(y_1, y_2, \cdots, y_n\right)}{\sup_{\theta\in\Theta_0} p_\theta\left(y_1, y_2, \cdots, y_n\right)}. \tag{2.28}$$

**Definition 2.15.** *(GLR Test [19, 109, 175]). The Generalized Likelihood Ratio (GLR) test in the class $C_\alpha$ for testing between*

$$\mathcal{H}_0 = \{\theta \in \Theta_0\} \quad \textit{against} \quad \mathcal{H}_1 = \{\theta \in \Theta_1\}$$

*is defined as*

$$\hat{\delta}\left(Y_n\right) = \begin{cases} \mathcal{H}_1 & \textit{if} \quad \hat{\Lambda}\left(Y_n\right) \geq h \\ \mathcal{H}_0 & \textit{if} \quad \hat{\Lambda}\left(Y_n\right) < h \end{cases}, \tag{2.29}$$

*where the threshold $h$ satisfies the following relation $\sup_{\theta\in\Theta_0} \mathbb{P}_\theta\left(\hat{\Lambda}\left(Y_n\right) \geq h\right) = \alpha$.*

It has been shown that the GLR test $\hat{\delta}\left(Y_n\right)$ is in many situations not optimal [109]. However, in some particular circumstances, it coincides with an optimal test (in the Bayesian approach, for example). Therefore, the GLR test is one of the most popular and important methods for solving the composite hypothesis testing problem.

**Bayesian approach for composite hypotheses.** The Bayesian approach for testing between two composite hypotheses $\mathcal{H}_0 = \{\theta \in \Theta_0\}$ and $\mathcal{H}_1 = \{\theta \in \Theta_1\}$, where $\Theta_0 \cap \Theta_1 = \emptyset$, is based on the *a priori* distribution $Q = (q_0, q_1)$, where $q_0 = \mathbb{P}\left(\mathcal{H}_0\right)$, $q_1 = \mathbb{P}\left(\mathcal{H}_1\right)$ and $q_0 + q_1 = 1$, on the hypotheses and the *a priori* distributions $G_j\left(\theta\right)$, for $j = 0, 1$, on the parameter $\theta$. The idea is to replace the unknown densities $p_\theta\left(Y_n\right)$ under $\mathcal{H}_j$ by following average (or integrated) densities over $\Theta_j$, for $j = 0, 1$, i.e.,

$$p_{G_j}\left(Y_n\right) = \int_{\Theta_j} p_\theta\left(Y_n\right) dG_j\left(\theta\right). \tag{2.30}$$

**Theorem 2.6.** *(Bayesian Test [19, 109, 175]). The Bayesian test which minimizes the average error probability $\alpha_Q\left(\delta\right) = q_0\overline{\alpha}_0\left(\delta\right) + q_1\overline{\alpha}_1\left(\delta\right)$ for testing between two composite hypotheses $\mathcal{H}_0 = \{\theta \in \Theta_0\}$ against $\mathcal{H}_1 = \{\theta \in \Theta_1\}$ is given by*

$$\overline{\delta}\left(Y_n\right) = \begin{cases} \mathcal{H}_1 & \textit{if} \quad \dfrac{p_{G_1}\left(Y_n\right)}{p_{G_0}\left(Y_n\right)} \geq \dfrac{q_1}{q_0} \\ \mathcal{H}_0 & \textit{if} \quad \dfrac{p_{G_1}\left(Y_n\right)}{p_{G_0}\left(Y_n\right)} < \dfrac{q_1}{q_0} \end{cases}, \tag{2.31}$$

*where $\overline{\alpha}_j\left(\delta\right) = \int_{\Theta_j} \mathbb{P}_\theta\left(\delta \neq \mathcal{H}_j\right) dG_j\left(\theta\right)$, for $j = 0, 1$.*

**Minimax approach for composite hypotheses.** Similar to the Bayesian approach, the minimax approach assumes also the *a priori* distributions $G_j(\theta)$ concentrated on $\Theta_j^* \subset \Theta_j$, for $j = 0, 1$. In contrast to the Bayesian approach, the minimax approach does not assume the *a priori* distribution on the hypotheses. Therefore, this approach is sometimes called the partially Bayesian approach.

**Definition 2.16.** *( [19, 109, 175]). A test $\tilde{\delta}(Y_n)$ for deciding between two composite hypotheses $\mathcal{H}_0 = \{\theta \in \Theta_0\}$ and $\mathcal{H}_0 = \{\theta \in \Theta_1\}$ is minimax in the class $C_\alpha$ given in* (2.20) *if, for all tests $\delta \in C_\alpha$, we have*

$$\inf_{\theta \in \Theta_1} \mathbb{P}_\theta \left( \tilde{\delta}(Y_n) = \mathcal{H}_1 \right) \geq \inf_{\theta \in \Theta_1} \mathbb{P}_\theta \left( \delta(Y_n) = \mathcal{H}_1 \right).$$

In order to determine the minimax test $\tilde{\delta}(Y_n)$ between two composite hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, let us define the "auxiliary" Neyman-Pearson test for testing between two simple hypotheses

$$\mathcal{H}_{G_0} = \{(y_1, y_2, \cdots, y_n) \sim P_{G_0}\} \quad \text{against} \quad \mathcal{H}_{G_1} = \{(y_1, y_2, \cdots, y_n) \sim P_{G_1}\},$$

where the distributions $P_{G_j}$ are with the densities $p_{G_j}(Y_n) = \int_{\Theta_j^*} p_\theta(Y_n)\, dG_j(\theta)$, for $j = 0, 1$. The MP test between $\mathcal{H}_{G_0}$ against $\mathcal{H}_{G_1}$ in the class $\tilde{C}_\alpha = \{\delta : \mathbb{P}_{G_0}(\delta(Y_n) \neq \mathcal{H}_0) \leq \alpha\}$ is given by

$$\delta^*_{G_0 G_1}(Y_n) = \begin{cases} \mathcal{H}_1 & \text{if} \quad \dfrac{p_{G_1}(Y_n)}{p_{G_0}(Y_n)} \geq h \\ \mathcal{H}_0 & \text{if} \quad \dfrac{p_{G_1}(Y_n)}{p_{G_0}(Y_n)} < h \end{cases}, \tag{2.32}$$

where the threshold $h$ is such chosen that $\mathbb{P}_{G_0}\left( \dfrac{p_{G_1}(Y_n)}{p_{G_0}(Y_n)} \geq h \right) = \alpha$. The power of the test $\delta^*_{G_0 G_1}(Y_n)$ is then defined as

$$\beta_{G_0 G_1}(\delta^*_{G_0 G_1}) = \mathbb{P}_{G_1}(\delta^*_{G_0 G_1}(Y_n) = \mathcal{H}_1). \tag{2.33}$$

**Theorem 2.7.** *(Minimax Test [19, 109, 175]). Suppose that there exist the a priori distributions $G_0$ and $G_1$ defined on the subsets $\Theta_0^* \subset \Theta_0$ and $\Theta_1^* \subset \Theta_1$, respectively, such that*

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\delta^*_{G_0 G_1}(Y_n) \neq \mathcal{H}_0) \leq \alpha \quad \text{and} \quad \inf_{\theta \in \Theta_1} \mathbb{P}_\theta(\delta^*_{G_0 G_1}(Y_n) = \mathcal{H}_1) = \beta_{G_0 G_1},$$

*then the MP test $\delta^*_{G_0 G_1}$ is minimax in the class $C_\alpha$ given in* (2.20) *for testing hypothesis $\mathcal{H}_0 = \{\theta \in \Theta_0\}$ against hypothesis $\mathcal{H}_1 = \{\theta \in \Theta_1\}$.*

Theoretical results obtained for the hypothesis testing between two composite hypotheses are quite limited. The UMP tests and unbiased UMP tests exist in very limited scenarios. For practical situations, the Bayesian approach and the GLR approach are generally considered. The *a priori* distribution on parameter $\theta$ is required for constructing the (completely or partially) Bayesian tests. Hence, the Bayesian tests are quite sensitive to the choice of the *a priori* distribution. On the other hand, the GLR tests do not require any a priori information on the parameter $\theta$ but their optimality can not be guaranteed.

### 2.2.4 Testing between multiple hypotheses

The problem of hypothesis testing between two (simple and composite) hypotheses has been reviewed in previous subsections. In this subsection, we consider the problem of testing between multiple hypotheses.

**Bayesian approach for multiple simple hypotheses.** Let $Y_n = (y_1, y_2, \cdots, y_n)$ be random samples of size $n$ generated from one of $K+1$ distinct distributions $P_{\theta_j}$, for $j = 0, 1, \cdots, K$. Seeking for simplicity, let us assume that the distributions $P_{\theta_j}$ are absolutely continuous, i.e., the densities $p_{\theta_j}(Y_n)$ are continuous functions w.r.t. the samples $Y_n$. The problem becomes deciding between $K+1$ simple hypotheses, i.e., $\mathcal{H}_j = \{\theta = \theta_j\}$, for $j = 0, 1, \cdots, K$.

Consider now the Bayesian approach. Let $q_j = \mathbb{P}(\mathcal{H}_j) > 0$, with $\sum_{j=0}^{K} q_j = 1$, be the *a priori* probabilities of hypothesis $\mathcal{H}_j$, for $j = 0, 1, \cdots, K$. Consider the case of $0-1$ loss function. In this case, the Bayes risk $J_Q(\delta)$ defined in (2.7) is reduced to the average error probability $\alpha_Q(\delta)$ given in (2.9). The Bayesian test for multiple simple hypotheses are given in the following theorem.

**Theorem 2.8.** *(Bayesian test [49, 175]). Consider the multiple hypothesis testing problem between $K+1$ simple hypotheses $\mathcal{H}_0, \mathcal{H}_1, \cdots, \mathcal{H}_K$ with the loss function $0-1$ and the a priori probabilities $q_0, q_1, \cdots, q_K$. The Bayesian test which minimizes the average error probability $\alpha_Q(\delta)$ given in (2.9) is defined as*

$$\overline{\delta}(Y_n) = \mathcal{H}_l \quad \text{if} \quad l = \arg \max_{0 \le j \le K} q_j p_{\theta_j}(Y_n). \tag{2.34}$$

Under above assumption that the distributions $P_{\theta_j}$ are absolutely continuous, the event $q_l p_{\theta_l}(Y_n) = q_j p_{\theta_j}(Y_n)$ has the $\mu$-measure zero for $l \neq j$; hence the maximum in (2.34) is unique with probability 1. Moreover, the Bayesian test (2.34) coincides with the maximum *a posteriori* (MAP) decision rule: choose index $l$ of hypothesis $\mathcal{H}_l$ that maximizes the posterior probability $q_j p_{\theta_j}(Y_n)$ over $j = 0, 1, \cdots, K$, i.e., $l = \arg\max_{0 \le j \le K} q_j p_{\theta_j}(Y_n)$.

**Constrained minimax approach for multiple simple hypotheses.** Let us introduce a class of tests $C_\alpha$ for deciding between multiple hypotheses as follows:

$$C_\alpha = \{\delta : \mathbb{P}_{\theta_0}(\delta(Y_n) \neq \mathcal{H}_0) \le \alpha\}. \tag{2.35}$$

**Definition 2.17.** *(Constrained minimax test [12, 175]). A test $\tilde{\delta}(Y_n)$ is constrained minimax of level $\alpha$ between the hypotheses $\mathcal{H}_0, \mathcal{H}_1, \cdots, \mathcal{H}_K$ if $\tilde{\delta}(Y_n) \in C_\alpha$ and for any other test $\delta(Y_n) \in C_\alpha$, the following inequality is satisfied*

$$\max_{1 \le l \le K} \alpha_l\left(\tilde{\delta}(Y_n)\right) \le \max_{1 \le l \le K} \alpha_l(\delta(Y_n)), \tag{2.36}$$

*where $\alpha_l(\delta(Y_n))$ is the probability of rejecting hypothesis $\mathcal{H}_l$ when it is true.*

**Theorem 2.9.** *(Constrained minimax test [12, 175]). Let $q_0, q_1, \cdots, q_K \ge 0$ be weighting coefficients satisfying $\sum_{j=0}^{K} q_j = 1$. The following weighted GLR test between the hypotheses $\mathcal{H}_0, \mathcal{H}_1, \cdots, \mathcal{H}_K$*

$$\tilde{\delta}(Y_n) = \begin{cases} \mathcal{H}_l & \text{if} \quad \max_{1 \le j \le K}\left(q_j \dfrac{p_{\theta_j}(Y_n)}{p_{\theta_0}(Y_n)}\right) \ge h \\ \mathcal{H}_0 & \text{if} \quad \max_{1 \le j \le K}\left(q_j \dfrac{p_{\theta_j}(Y_n)}{p_{\theta_0}(Y_n)}\right) < h \end{cases}, \quad l = \arg \max_{1 \le j \le K}\left(q_j \frac{p_{\theta_j}(Y_n)}{p_{\theta_0}(Y_n)}\right) \tag{2.37}$$

*is constrained minimax if the threshold $h$ is selected so that*

$$\mathbb{P}_{\theta_0}\left(\max_{1 \le j \le K}\left(q_j \frac{p_{\theta_j}(Y_n)}{p_{\theta_0}(Y_n)}\right) \ge h\right) = \alpha,$$

*and that the weighted coefficients are chosen so that the probability of false classification*

$$\alpha_l\left(\tilde{\delta}\left(Y_n\right)\right)=\alpha_j\left(\tilde{\delta}\left(Y_n\right)\right),\quad \forall l,j\neq 0$$

*is constant over the set of alternative hypotheses* $\mathcal{H}_1,\mathcal{H}_2,\cdots,\mathcal{H}_K$.

The above theorem allows us to design an "equalizer test" which maximizes the common power

$$\beta=\mathbb{P}_{\theta_l}\left(\tilde{\delta}\left(Y_n\right)=\mathcal{H}_l\right)=\mathbb{P}_{\theta_j}\left(\tilde{\delta}\left(Y_n\right)=\mathcal{H}_j\right),\quad \forall l,j\neq 0 \tag{2.38}$$

in the class $C_\alpha$ defined in (2.35).

**Bayesian approach for multiple composite hypotheses.** It is of practical interest to consider now the problem of testing between multiple composite hypotheses $\mathcal{H}_j=\{\theta\in\Theta_j\}$, for $0\leq j\leq K$. The Bayesian approach for testing multiple hypotheses is based on the *a priori* distribution $Q=(q_0,q_1,\cdots,q_K)$ on the hypotheses, i.e., $q_j=\mathbb{P}\left(\mathcal{H}_j\right)$, for $0\leq j\leq K$, and the *a priori* distributions $G_j\left(\theta\right)$, for $j=0,1,\cdots,K$, on the parameter $\theta$. For the sake of simplicity, let us consider now the case of $0-1$ loss function. In such a case, the Bayes risk is equivalent to the following average error probability:

$$J_Q\left(\delta\right)=\alpha_Q\left(\delta\right)=\sum_{j=0}^{K}q_j\overline{\alpha}_j\left(\delta\right)=\sum_{j=0}^{K}q_j\int\limits_{\Theta_j}\mathbb{P}_\theta\left(\delta\neq\mathcal{H}_j\right)dG_j\left(\theta\right). \tag{2.39}$$

The Bayesian approach for testing multiple composite hypotheses is given in the following theorem.

**Theorem 2.10.** *(Bayesian Test). Consider the problem of multiple hypothesis testing between* $K+1$ *composite hypotheses* $\mathcal{H}_0,\mathcal{H}_1,\cdots,\mathcal{H}_K$ *with the a priori distribution* $Q=(q_0,q_1,\cdots,q_K)$ *on the hypotheses, i.e.,* $q_j=\mathbb{P}\left(\mathcal{H}_j\right)$*, for* $0\leq j\leq K$*, and the a priori distributions* $G_j\left(\theta\right)$*, for* $j=0,1,\cdots,K$*, on the parameter* $\theta$*. Suppose also the* $0-1$ *loss function. The Bayesian test* $\overline{\delta}\left(Y_n\right)$ *which minimizes the Bayes risk* $J_Q\left(\delta\right)$ *defined in* (2.39) *is given by*

$$\overline{\delta}\left(Y_n\right)=\mathcal{H}_l,\quad \textit{if}\quad l=\arg\max_{0\leq j\leq K}\left[q_j\int_{\Theta_j}p_\theta\left(Y_n\right)dG_j\left(\theta\right)\right]. \tag{2.40}$$

Similar to the multiple simple hypothesis testing problem, the Bayesian test (2.40) for deciding between multiple composite hypotheses coincides also to the maximum *a posteriori* (MAP) decision rule: choose index $l$ that maximizes the posterior distribution $q_j\int_{\Theta_j}p_\theta\left(Y_n\right)dG_j\left(\theta\right)$ over all $j=0,1,\cdots,K$.

### 2.2.5 Conclusion

In this section, we have briefly presented basic definitions and different results on the classical (non-sequential) statistical hypothesis testing theory. Several optimality criteria, by the most powerful approach, the Bayesian approach and the minimax approach, have been introduced. It has been shown that optimal (or suboptimal) procedures for testing two (or more) simple (or composite) hypotheses could be designed to attain a given criterion of optimality. Generally,

the statistical performance of a decision rule is proportional to the number of observations. In non-sequential setting, however, the sample size is *a priori* fixed. Therefore, the classical hypothesis testing theory is particularly useful for off-line applications. For on-line monitoring tasks, however, another data-processing method needs to be considered for more efficiently reducing the number of observations. These sequential methods for hypothesis testing and change-point detection-isolation will be reviewed in section 2.3 and section 2.4, respectively.

## 2.3 Sequential Hypothesis Testing

The purpose of this section is to introduce some sequential methods for testing between two or multiple hypotheses. The sequential hypothesis testing problem consists in seeking a detection rule $\delta$ that is carried out in real time $k = 1, 2, \cdots$ based on the observations $y_1, y_2, \cdots y_k$. The decision of stopping the test at time $k$ or continuing the test at time $k+1$ depends on the observed data $y_1, y_2, \cdots, y_k$ itself.

This section is organized as follows. Basic definitions are given in subsection 2.3.1. Several results on the sequential tests between two simple and composite hypotheses are introduced in subsection 2.3.2 and subsection 2.3.3, respectively. Finally, we consider in subsection 2.3.4 the problem of sequential testing between multiple hypotheses.

### 2.3.1 Introduction

In the classical hypothesis testing problem, the sample size $n$ is *a priori* fixed. The problem consists of seeking a detection rule $\delta$ satisfying a given optimality criterion. For example, in the case of testing between two simple hypotheses $\mathcal{H}_0 = \{\theta = \theta_0\}$ and $\mathcal{H}_1 = \{\theta = \theta_1\}$, we wish to maximize the power of the test $\beta(\delta) = \mathbb{P}_{\theta_1}(\delta = \mathcal{H}_1)$ for a given value on the probability of false alarm $\alpha_0(\delta) = \mathbb{P}_{\theta_0}(\delta \neq \mathcal{H}_0)$. The error probabilities, i.e., $\alpha_0(\delta)$ and $\alpha_1(\delta) = 1 - \beta(\delta)$, depend on the sample size $n$ which has not been pointed out explicitly. It is well-known that the N-P test given in (2.14) is the most powerful test in the class $C_\alpha = \{\delta : \alpha_0(\delta) \leq \alpha\}$.

The question arises [19]: "Is it possible to improve this statistical procedure?". Of course, the answer is negative under the above-mentioned criterion. However, if we drop the assumption that the sample size is fixed, that is, make $n$ be a random variable depending on the samples already observed, then improvement is possible [19]. This sequential hypothesis testing method is critical in such applications that require some cost for performing experiments. In fact, the theoretical study of sequential hypothesis testing has been ushered by A. Wald [193, 195] in response to demands for more efficient testing of anti-aircraft gunnery during World War II [105].

For testing between two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, the sequential procedure can be described as follows [193]. At any stage of an experiment, a procedure is given for making one of the following three decisions: (1) to accept hypothesis $\mathcal{H}_0$, (2) to accept hypothesis $\mathcal{H}_1$, or (3) to continue the experiment by making an additional observation. Thus, such a decision rule is performed sequentially. Based on the basis of $k$, for $k \geq 1$, observations, one of the aforementioned three decisions is made. If the first or second decision is made, the test is terminated by accepting either hypothesis $\mathcal{H}_0$ or hypothesis $\mathcal{H}_1$, respectively. If the third decision is made, on the other hand, the experiment is continued taking the $k+1$ observation. The process continues until

either hypothesis $\mathcal{H}_0$ or hypothesis $\mathcal{H}_1$ is accepted. The time instant $N$ at which the process terminates is a random variable since it depends on the outcomes of observed data.

**Definition 2.18.** *(Stopping time [67, 126, 133]). A stopping time with respect to a sequence of random variables $y_1, y_2, \cdots$ is a Markov random variable $N$ with values in $\{1, 2, \cdots\}$ and the property that for each $k \in \{1, 2, \cdots\}$, the occurrence or non-occurrence of the event $\{N = k\}$ depends only on the values of $\{y_1, y_2, \cdots, y_k\}$.*

**Definition 2.19.** *(Sequential test [19, 67, 126]). A sequential test $\delta$ between hypothesis $\mathcal{H}_0$ and hypothesis $\mathcal{H}_1$ is a pair $(N, \nu)$, where $N$ is the stopping time and $\nu$ is the final decision.*

In sequential hypothesis testing, it is desirable to achieve the trade-off between the average sample number (ASN) and the error probabilities. The comparison between various sequential tests can be performed with the aide of following definitions.

**Definition 2.20.** *(Better Sequential Test [68, 193]). Consider the problem of sequential testing between two simple hypotheses: $\mathcal{H}_0 = \{\theta = \theta_0\}$ and $\mathcal{H}_1 = \{\theta = \theta_1\}$. Let $\tilde{\delta}$ and $\delta$ be two sequential procedures with the error probabilities of type I and type II being equal to $\alpha_0$ and $\alpha_1$ and with the stopping times $\tilde{N}(\alpha_0, \alpha_1)$ and $N(\alpha_0, \alpha_1)$, respectively. The test $\tilde{\delta}$ is said to be better than the test $\delta$ if*

$$\mathbb{E}_{\theta_0}\left[\tilde{N}(\alpha_0, \alpha_1)\right] \leq \mathbb{E}_{\theta_0}\left[N(\alpha_0, \alpha_1)\right] \quad \text{and} \quad \mathbb{E}_{\theta_1}\left[\tilde{N}(\alpha_0, \alpha_1)\right] \leq \mathbb{E}_{\theta_1}\left[N(\alpha_0, \alpha_1)\right], \tag{2.41}$$

*where $\mathbb{E}_{\theta_j}[N]$ is the ASN under hypothesis $\mathcal{H}_j$ (i.e., $\theta = \theta_j$), for $j = 0, 1$.*

**Definition 2.21.** *(Class of Sequential Test [68, 193]). Let $\alpha_0, \alpha_1 \in [0, 1]$ be two real numbers. The class of all sequential tests with the error probabilities of type I and type II being smaller than or equal to $\alpha_0$ and $\alpha_1$, respectively, is defined as*

$$C_{\alpha_0,\alpha_1} = \left\{\delta = (N, \nu) : \alpha_j(\delta) \leq \alpha_j \quad \text{and} \quad \mathbb{E}_{\theta_j}[N] < \infty, \quad j = 0, 1\right\}. \tag{2.42}$$

**Definition 2.22.** *(Optimal Sequential Test [68, 193]). Consider the problem of sequential testing between two simple hypotheses: $\mathcal{H}_0 = \{\theta = \theta_0\}$ and $\mathcal{H}_1 = \{\theta = \theta_1\}$. The test $\tilde{\delta} = \left(\tilde{N}, \tilde{\nu}\right)$ is said to be optimal in the class $C_{\alpha_0,\alpha_1}$ if, for all sequential tests $\delta = (N, \nu)$ in the class $C_{\alpha_0,\alpha_1}$, the following conditions are satisfied:*

$$\mathbb{E}_{\theta_0}\left[\tilde{N}\right] \leq \mathbb{E}_{\theta_0}[N] \quad \text{and} \quad \mathbb{E}_{\theta_1}\left[\tilde{N}\right] \leq \mathbb{E}_{\theta_1}[N]. \tag{2.43}$$

### 2.3.2 Sequential testing between two simple hypotheses

In this sub-section, we consider the Sequential Probability Ratio Test (SPRT), which was first introduced in [193], for testing hypothesis $\mathcal{H}_0 = \{\theta = \theta_0\}$ and hypothesis $\mathcal{H}_1 = \{\theta = \theta_1\}$. The Bayesian approach can be found in [19, 196].

**Sequential Probability Ratio Test (SPRT).** Let $y_1, y_2, \cdots$ be i.i.d. random variables which have a common p.d.f. $f_\theta(y)$ with respect to some sigma-finite measure $\mu$. The joint p.d.f. of $Y_k = (y_1, y_2, \cdots, y_k)$ is calculated as

$$p_{\theta_j}(Y_k) = p_{\theta_j}(y_1, y_2, \cdots, y_k) = \prod_{i=1}^{k} f_{\theta_j}(y_i), \quad \text{for } j = 0, 1. \tag{2.44}$$

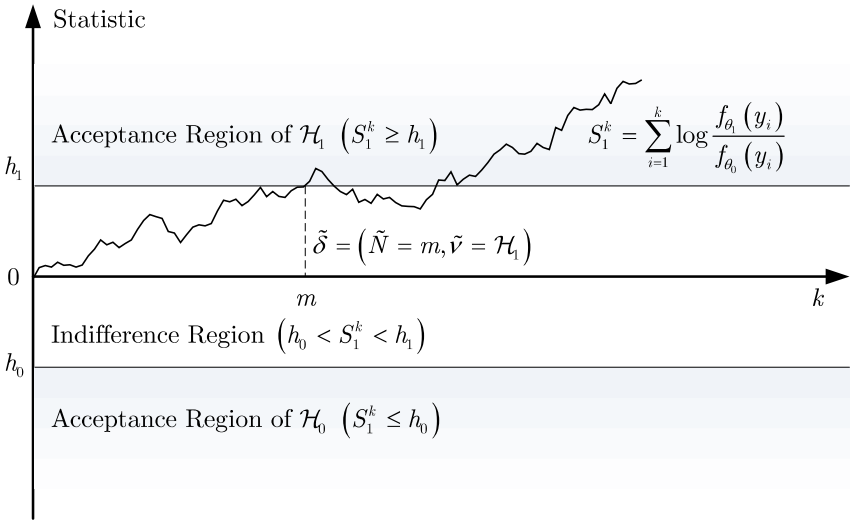



Figure 2.3 – Sequential probability ratio test between two simple hypotheses.

Let $h_0 < 0 < h_1$ be two real numbers (thresholds) and

$$S_1^k = \log \frac{p_{\theta_1}(Y_k)}{p_{\theta_0}(Y_k)} = \log \frac{\prod_{i=1}^{k} f_{\theta_1}(y_i)}{\prod_{i=1}^{k} f_{\theta_0}(y_i)} = \sum_{i=1}^{k} \log \frac{f_{\theta_1}(y_i)}{f_{\theta_0}(y_i)} \tag{2.45}$$

be the log-likelihood ratio (LLR) between hypothesis $\mathcal{H}_1$ and hypothesis $\mathcal{H}_0$ on the basis of the observations $Y_k = (y_1, y_2, \cdots, y_k)$. The sequential procedure $\tilde{\delta} = \left(\tilde{N}, \tilde{\nu}\right)$ introduced by Wald [193] is described as

$$\tilde{N} = \inf\left\{k \geq 1 : S_1^k \notin (h_0, h_1)\right\}, \quad \tilde{\nu} = \begin{cases} \mathcal{H}_1 & \text{if} \quad S_1^{\tilde{N}} \geq h_1 \\ \mathcal{H}_0 & \text{if} \quad S_1^{\tilde{N}} \leq h_0 \end{cases}, \tag{2.46}$$

where the thresholds $h_0$ and $h_1$ are chosen for assuring acceptable levels on the error probabilities of type I and type II.

**Performance of SPRT.** Several properties of the SPRT are given in following theorems. The approximation of the error probabilities and the average sample numbers of the SPRT (2.46) is given in Theorem 2.11 and Theorem 2.12, respectively. Finally, the optimality property of the SPRT (2.46) is shown in Theorem 2.13.

**Theorem 2.11.** *(Error probabilities of SPRT [67, 175, 193]). Consider Wald's SPRT* $\tilde{\delta} = \left(\tilde{N}, \tilde{\nu}\right)$ *given in* (2.46) *with thresholds* $h_0 < 0 < h_1$. *The relations between the error probabilities of type I and type II and the thresholds are described as follows:*

$$\log\left(\frac{\alpha_1\left(\tilde{\delta}\right)}{1-\alpha_0\left(\tilde{\delta}\right)}\right) \leq h_0, \quad \log\left(\frac{1-\alpha_1\left(\tilde{\delta}\right)}{\alpha_0\left(\tilde{\delta}\right)}\right) \geq h_1. \tag{2.47}$$

The exact calculation of thresholds for assuring acceptable values on the error probabilities is elaborate. For this reason, Wald [193] suggested the following approximations:

$$h_0 \simeq \log \frac{\alpha_1\left(\tilde{\delta}\right)}{1-\alpha_0\left(\tilde{\delta}\right)}, \quad h_1 \simeq \frac{1-\alpha_1\left(\tilde{\delta}\right)}{\alpha_0\left(\tilde{\delta}\right)}, \quad \text{when} \quad \alpha_0\left(\tilde{\delta}\right), \alpha_1\left(\tilde{\delta}\right) \to 0. \tag{2.48}$$

**Theorem 2.12.** *(Average Sample Number of SPRT [19, 193, 194]). Consider the Wald's SPRT* $\tilde{\delta} = \left(\tilde{N}, \tilde{\nu}\right)$ *given in* (2.46). *Then, as* $\alpha_0\left(\tilde{\delta}\right), \alpha_1\left(\tilde{\delta}\right) \to 0$, *the average number of samples* $\mathbb{E}_{\theta_0}\left[\tilde{N}\right]$ *and* $\mathbb{E}_{\theta_1}\left[\tilde{N}\right]$ *are given by*

$$\mathbb{E}_{\theta_0}\left[\tilde{N}\right] \simeq \frac{\left(1-\alpha_0\left(\tilde{\delta}\right)\right)\log\left(\frac{1-\alpha_0\left(\tilde{\delta}\right)}{\alpha_1\left(\tilde{\delta}\right)}\right) - \alpha_0\left(\tilde{\delta}\right)\log\left(\frac{1-\alpha_1\left(\tilde{\delta}\right)}{\alpha_0\left(\tilde{\delta}\right)}\right)}{\mathbb{E}_{\theta_0}\left[\log\left(\frac{f_{\theta_0}(y)}{f_{\theta_1}(y)}\right)\right]}, \tag{2.49}$$

$$\mathbb{E}_{\theta_1}\left[\tilde{N}\right] \simeq \frac{\left(1-\alpha_1\left(\tilde{\delta}\right)\right)\log\left(\frac{1-\alpha_1\left(\tilde{\delta}\right)}{\alpha_0\left(\tilde{\delta}\right)}\right) - \alpha_1\left(\tilde{\delta}\right)\log\left(\frac{1-\alpha_0\left(\tilde{\delta}\right)}{\alpha_1\left(\tilde{\delta}\right)}\right)}{\mathbb{E}_{\theta_1}\left[\log\left(\frac{f_{\theta_1}(y)}{f_{\theta_0}(y)}\right)\right]}. \tag{2.50}$$

**Theorem 2.13.** *(Optimality of SPRT [19, 193, 194]). Let* $y_1, y_2, \cdots$ *be the sequence of i.i.d. random observations generated from a parametric family of distributions* $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ *depending on the parameter* $\theta$. *Consider the problem of testing hypothesis* $\mathcal{H}_0 = \{\theta = \theta_0\}$ *against hypothesis* $\mathcal{H}_1 = \{\theta = \theta_1\}$. *Let* $C_{\alpha_0,\alpha_1}$ *given in* (2.42) *be the class of all tests (sequential and non-sequential) with upper bounds on the error probabilities. Then, the Wald's SPRT* $\tilde{\delta} = \left(\tilde{N}, \tilde{\nu}\right)$ *is optimal in the class* $C_{\alpha_0,\alpha_1}$. *In other words, it minimizes the average number of samples* $\mathbb{E}_{\theta_0}\left[\tilde{N}\right]$ *and* $\mathbb{E}_{\theta_1}\left[\tilde{N}\right]$ *among all (sequential and non-sequential) tests* $\delta = (N, \nu)$ *in the class* $C_{\alpha_0,\alpha_1}$, *i.e.,*

$$\mathbb{E}_{\theta_0}\left[\tilde{N}\right] \leq \mathbb{E}_{\theta_0}\left[N\right] \quad \textit{and} \quad \mathbb{E}_{\theta_1}\left[\tilde{N}\right] \leq \mathbb{E}_{\theta_1}\left[N\right]. \tag{2.51}$$

It follows from Theorem 2.13 that the SPRT (2.46) for testing between two simple hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ minimizes the average sample numbers under both hypotheses among all (i.e., sequential and non-sequential) tests in the class $C_{\alpha_0,\alpha_1}$ defined in (2.42). The problem of testing between two composite hypotheses is considered in the following subsection.

### 2.3.3 Sequential testing between two composite hypotheses

It is of practical interest to consider the problem of sequential testing between two composite hypotheses $\mathcal{H}_0 = \{\theta \in \Theta_0\}$ and $\mathcal{H}_1 = \{\theta \in \Theta_1\}$, for $\Theta_0 \cap \Theta_1 = \emptyset$ (see also, for example, in [78, 89, 100, 101, 115, 193, 194, 196]). For solving this problem, Wald [193] suggested to utilize the Weighted Sequential Probability Ratio Test (WSPRT) and the Generalized Sequential Probability Ratio Test (GSPRT).

Let $y_1, y_2, \cdots$ be i.i.d. random variables with a common density $f_\theta(y)$, depending on the parameter $\theta$, with respect to some finite-measure $\mu$. Consider the problem of testing the simple hypothesis $\mathcal{H}_0 = \{\theta = \theta_0\}$ against the composite hypothesis $\mathcal{H}_1 = \{\theta \in \Theta_1\}$, where $\theta_0 \notin \Theta_1$.

The first method is to apply the generalized likelihood ratio (GLR) approach, replacing the unknown LLR $S_1^k$ by the GLR statistic

$$\hat{S}_1^k = \log \sup_{\theta \in \Theta_1} \prod_{i=1}^{k} \left[ f_\theta \left( y_i \right) / f_{\theta_0} \left( y_i \right) \right], \tag{2.52}$$

resulting in the following GSPRT $\hat{\delta} = \left( \hat{N}, \hat{\nu} \right)$:

$$\hat{N} = \inf \left\{ k \geq 1 : \hat{S}_1^k \notin (h_0, h_1) \right\}, \quad \hat{\nu} = \begin{cases} \mathcal{H}_1 & \text{if} \quad \hat{S}_1^{\hat{N}} \geq h_1 \\ \mathcal{H}_0 & \text{if} \quad \hat{S}_1^{\hat{N}} \leq h_0 \end{cases}, \tag{2.53}$$

where the thresholds $h_0$ and $h_1$ are selected for assuring acceptable levels of error probabilities.

The second method consists of replacing the LLR $S_1^k$ by the weighted LLR statistic

$$\overline{S}_1^k = \log \int_{\theta \in \Theta_1} w \left( \theta \right) \prod_{i=1}^{k} \left[ f_\theta \left( y_i \right) / f_{\theta_0} \left( y_i \right) \right] d\theta, \tag{2.54}$$

where $w \left( \theta \right)$ is a suitably selected weighted function on $\Theta_1$, leading to the following WSPRT $\overline{\delta} = \left( \overline{N}, \overline{\nu} \right)$:

$$\overline{N} = \inf \left\{ k \geq 1 : \overline{S}_1^k \notin (h_0, h_1) \right\}, \quad \overline{\nu} = \begin{cases} \mathcal{H}_1 & \text{if} \quad \overline{S}_1^{\overline{N}} \geq h_1 \\ \mathcal{H}_0 & \text{if} \quad \overline{S}_1^{\overline{N}} \leq h_0 \end{cases}, \tag{2.55}$$

where the thresholds $h_0$ and $h_1$ are also chosen for assuring acceptable levels of error probabilities.

In a more general case where the null hypothesis is also composite, i.e., $\mathcal{H}_0 = \{\theta \in \Theta_0\}$, Wald [193] proposed to exploit the WSPRT given in (2.55) with the following weighted LLR

$$\overline{S}_1^k = \log \frac{\int_{\theta \in \Theta_1} w_1 \left( \theta \right) \prod_{i=1}^{k} f_\theta \left( y_i \right) d\theta}{\int_{\theta \in \Theta_0} w_0 \left( \theta \right) \prod_{i=1}^{k} f_\theta \left( y_i \right) d\theta}, \tag{2.56}$$

where $w_0 \left( \theta \right)$ and $w_1 \left( \theta \right)$ are suitably selected weighted functions on $\Theta_0$ and $\Theta_1$, respectively.

By changing the measures and applying the Wald's likelihood ratio identity [175, pages 223–224], the average error probabilities $\overline{\alpha}_0 \left( \overline{\delta} \right)$ and $\overline{\alpha}_1 \left( \overline{\delta} \right)$ are upper bounded by

$$\overline{\alpha}_0 \left( \overline{\delta} \right) = \int_{\Theta_0} \mathbb{P}_\theta \left( \overline{\delta} \neq \mathcal{H}_0 \right) w_0 \left( \theta \right) d\theta \leq e^{-h_1}, \tag{2.57}$$

$$\overline{\alpha}_1 \left( \overline{\delta} \right) = \int_{\Theta_1} \mathbb{P}_\theta \left( \overline{\delta} \neq \mathcal{H}_1 \right) w_1 \left( \theta \right) d\theta \leq e^{h_0}, \tag{2.58}$$

where the thresholds $h_0 \leq 0 < h_1$.

For practical purposes, the upper bounds on the maximal error probabilities of type I and type II would be more preferable than the upper bounds on the average error probabilities, which depend on the choice of weighted functions. Let us introduce the class

$$C_{\alpha_0, \alpha_1} = \left\{ \delta : \sup_{\theta \in \Theta_0} \mathbb{P}_\theta \left( \delta \neq \mathcal{H}_0 \right) \leq \alpha_0, \sup_{\theta \in \Theta_1} \mathbb{P}_\theta \left( \delta \neq \mathcal{H}_1 \right) \leq \alpha_1 \right\}, \quad \alpha_0 + \alpha_1 < 1 \tag{2.59}$$

for testing between two composite hypotheses. The upper bounds on the maximal error probabilities of the WSPRT and the GSPRT have not been obtained in the general case. Interested readers are referred to [175, chapter 5] for more discussion on this topic.

### 2.3.4 Sequential testing between multiple simple hypotheses

Over the last few decades, a great deal of effort has been devoted to study the sequential multihypothesis testing problem. The majority of work has concentrated on proposing suboptimal procedures based on the modification of the sequential probability ratio test for i.i.d. observations. For example, Sobel and Wald [171] considered the problem of sequential testing between three normal distributions. Independently, Armitage [8] proposed a sequential procedure for testing between multiple simple hypotheses. Based on Bayesian framework, the multiple sequential hypothesis testing procedures were introduced in [11, 43, 44, 192]. The multihypothesis testing problem for non-i.i.d. stochastic models has been also considered in [43, 44, 104, 180].

**Definition 2.23.** *(Sequential Multihypothesis Test [43, 175]). A sequential multihypothesis test $\delta = (N, \nu)$ between $K+1$ hypotheses is defined as a pair $(N, \nu)$, where $N \geq 1$ is the Markov stopping time and $\nu \in \{0, 1, \cdots, K\}$ is the final decision. The event $\{\nu = l\}$, for $0 \leq l \leq K$, means that we accept hypothesis $\mathcal{H}_l$ for some stopping time $N < \infty$.*

Let $y_1, y_2, \cdots$ be i.i.d. random variables with a common density $f_\theta(y)$, depending on the parameter $\theta$, with respect to some finite-measure $\mu$. Consider the problem of sequential testing between multiple simple hypotheses $\mathcal{H}_l = \{\theta = \theta_l\}$, for $0 \leq l \leq K$. The MSPRT $\delta = (N, \nu)$ can be defined in the following manner:

$$N_l = \inf\left\{k \geq 1 : \min_{0 \leq j \neq l \leq K} \left\{S_1^k(l, j) - h_{lj}\right\} \geq 0\right\} \tag{2.60}$$

$$N = \min_{l=0,1,\cdots,K} N_l \tag{2.61}$$

$$\nu = \arg \min_{l=0,1,\cdots,K} N_l, \tag{2.62}$$

where

$$S_1^k(l, j) = \log \frac{\prod_{i=1}^k f_{\theta_l}(y_i)}{\prod_{i=1}^k f_{\theta_j}(y_i)} = \sum_{i=1}^k \log \frac{f_{\theta_l}(y_i)}{f_{\theta_j}(y_i)} \tag{2.63}$$

is the log-likelihood ratio between hypothesis $\mathcal{H}_l$ and hypothesis $\mathcal{H}_j$ on the basis of the observations $y_1, y_2, \cdots, y_k$ and $h_{lj}$ are chosen thresholds. Readers are referred to [43, 44, 175, chapter 4] for the asymptotic optimality properties of the MSPRT and also the Bayesian approach for the problem.

### 2.3.5 Conclusion

In this section, we have considered the problem of sequential testing between two simple hypotheses, two composite hypotheses and multiple simple hypotheses. Theoretical results have shown that the sequential tests reduce significantly the number of observations in order to achieve a significant level compared to the non-sequential counterparts. The sequential hypothesis testing introduced in this section is essential in understanding the on-line change-point detection-isolation techniques, which is the subject of the following section.

## 2.4 Sequential Change-point Detection and Isolation

In this section, we focus on the design and analysis of techniques for the quickest change detection and isolation problem. This approach is extremely suitable to surveillance applications, including the monitoring of SCADA systems against cyber-physical attacks.

### 2.4.1 Introduction

The sequential change-point detection deals with the on-line detection of a change in the state of a process, subject to an acceptable level on the risk of false alarms. Specially, the process is assumed to be in a normal state before the surveillance begins and it may unexpectedly undergo an abrupt (or a gradual, an incipient) change-of-state from normal to abnormal. With the arrival of each new observation, the problem is to decide whether the process is in normal behavior or it has been changed to an abnormal state. If the state has become abnormal, we are interested in detecting the change, usually as soon as possible, so that appropriate responses could be provided. The time instant $k_0$ at which the process changes its state from normal to abnormal is referred to as the *change-point* and the time instant $T$ at which we raise an alarm is denoted as the *stopping time* or the *alarm time*. If an alarm is raised before the change occurs (i.e., $T < k_0$), one has a false alarm. On the other hand, if the alarm is raised after the change has occurred (i.e., $T \geq k_0$), one has a correct detection but with the detection delay $T - k_0 + 1$. Hence, a good sequential change-point detection scheme should be able to obtain a trade-off between the loss associated with the detection delay and that associated with raising a false alarm.

The subject of change-point detection started to emerge from the requirement in quality control which is concerned with the monitoring and evaluation of the quality of products from a continuous production process. Firstly, Shewhart [165] introduced the fundamental concept of a "state of statistical control", in which he proposed a process inspection scheme that takes samples of fixed size at regular time intervals and computes from the samples a suitably chosen statistic, which can be presented graphically in the form of a control chart. Efficient sequential detection procedures were developed later in the 1950-1960's, after the introduction of Sequential Analysis, a branch of statistics ushered by Wald [193]. To improve the sensitivity of the Shewhart's charts, Page [139] and Shiryaev [166] modified Wald's theory of sequential hypothesis testing to develop the CUSUM and the Shiryaev-Roberts charts, respectively, that attain certain optimality properties. This platform has paved the way for the development of sequential change-point detection problem, on both theory and practice [10, 103, 113, 166, 175].

Four different approaches have been considered for solving the change-point detection problem [146, 175], including the Bayesian approach, the generalized Bayesian approach, the minimax approach, and the approach related to multicyclic detection of a distant change in a stationary regime. In the following, we follow the minimax approach under which the change-point $k_0$ is considered as unknown but non-random. Interested readers are referred to [175] for recent results of other approaches.

### 2.4.2 Sequential change-point detection

Let $y_1, y_2, \cdots$ be a sequence of independent random observations generated from a parametric family of distributions $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ depending on the parameter $\theta$. Let $k_0 \geq 1$ be the unknown change-point at which the parameter $\theta$ changes its value from $\theta_0$ to $\theta_1 \neq \theta_0$. In other words, the random variables $y_1, y_2, \cdots, y_{k_0-1}$ have the distribution $P_{\theta_0}$ while the random variables $y_{k_0}, y_{k_0+1}, \cdots$ have the distribution $P_{\theta_1}$. The statistical model for the quickest change detection is described as

$$y_k \sim \begin{cases} P_{\theta_0} & \text{if} \quad k < k_0 \\ P_{\theta_1} & \text{if} \quad k \geq k_0 \end{cases}. \tag{2.64}$$

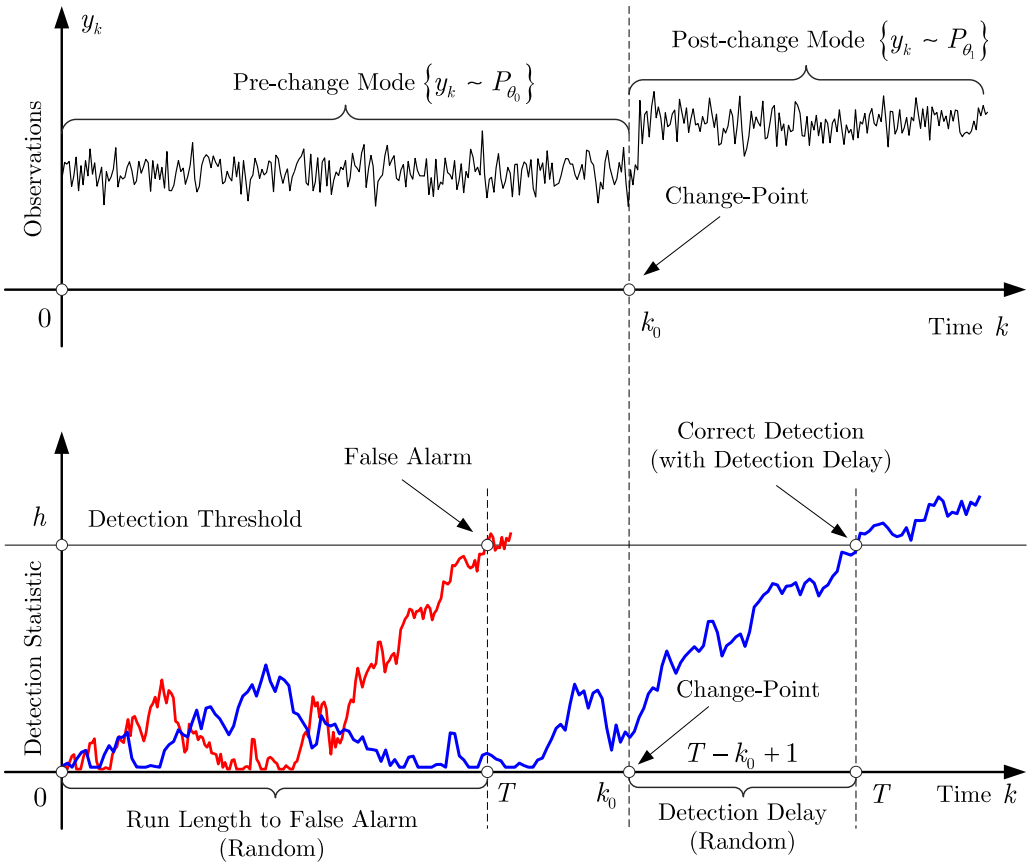



Figure 2.4 – Sequential change-point detection problem.

Let $\mathcal{P}_{k_0}$ denote the probability measure when the observations $y_1, y_2, \cdots, y_{k_0-1} \sim P_{\theta_0}$ and $y_{k_0}, y_{k_0+1}, \cdots \sim P_{\theta_1}$ and $\mathcal{P}_0 \triangleq \mathcal{P}_\infty$ corresponds to $k_0 = \infty$ (i.e., $y_1, y_2, \cdots \sim P_{\theta_0}$). Let $\mathbb{P}_{k_0}$ (res. $\mathbb{P}_0 \triangleq \mathbb{P}_\infty$) and $\mathbb{E}_{k_0}$ (res. $\mathbb{E}_0 \triangleq \mathbb{E}_\infty$) are, respectively, the probability and the expectation w.r.t. the probability measure $\mathcal{P}_{k_0}$ (res. $\mathcal{P}_0 \triangleq \mathcal{P}_\infty$). Suppose that the distributions $P_{\theta_0}$ and $P_{\theta_1}$ receive the densities $f_{\theta_0}$ and $f_{\theta_1}$, respectively.

### Minimax optimality criteria

The objective of an abrupt change detection algorithm is to achieve a trade-off between the risk associated with the detection delay and the risk of raising a false alarm. A large number of optimality criteria have been proposed for interpreting the compromise between these contradictory performance indexes. In general, a criterion of optimality should be favorable of minimizing the average detection delay (ADD) while avoiding frequent false alarms.

Let $T$ be the stopping time of a quickest change detection procedure. The first optimality criterion is due to Lorden [113] who suggested to minimize the following "worst-worst-case" average detection delay (WWADD):

$$\overline{\mathbb{E}}^* [T] = \sup_{k_0 \geq 1} \operatorname{ess\,sup} \mathbb{E}_{k_0} [T - k_0 + 1 \,|T \geq k_0, y_1, y_2, \cdots, y_{k_0-1}] \tag{2.65}$$

among all stopping times $T \in C_\gamma$ in the class

$$C_\gamma = \{T : \mathbb{E}_0[T] \geq \gamma\} \tag{2.66}$$

satisfying the average run length (ARL) to false alarm[6] constraint. The following theorem, whose proof can be found in [113], gives the lower bound for the WWADD $\overline{\mathbb{E}}^*[T]$ defined in (2.65).

**Theorem 2.14.** *(Lorden's Asymptotic Theory [10, 113]). Let $T$ be a stopping time in the class $C_\gamma$ given in (2.66) and $n(\gamma)$ be the lower bound on $\overline{\mathbb{E}}^*[T]$ defined in (2.65). Let also $\rho_{10} = \mathbb{E}_{\theta_1}[\log(f_{\theta_1}(y)/f_{\theta_0}(y))]$ be the Kullback-Leibler distance between $f_{\theta_1}$ and $f_{\theta_0}$, satisfying $0 < \rho_{10} < \infty$. For independent observations $\{y_k\}_{k\geq 1}$, we have*

$$n(\gamma) \sim \frac{\log(\gamma)}{\rho_{10}} \quad as \quad \gamma \to \infty. \tag{2.67}$$

Lorden [113] showed that the Cumulative Sum (CUSUM) procedure, first introduced by Page [139], is first-order asymptotically optimal as $\gamma \to \infty$. By using different arguments, Moustakides [124] and Ritov [157] proved that the CUSUM algorithm is exactly optimal w.r.t. Lorden's criterion (2.65)–(2.66) for any $\gamma > 1$.

Though the optimality of the CUSUM test w.r.t. Lorden's criterion (2.65)–(2.66) is a very strong result, this criterion seems to be too pessimistic since it is, in fact, a double-minimax approach [146]. For this reason, it is more natural to find a procedure that minimizes the following conditional average detection delay (CADD):

$$\tilde{\mathbb{E}}[T] = \mathbb{E}_{k_0}[T - k_0 \,|T \geq k_0], \tag{2.68}$$

for all $k_0 \geq 1$ simultaneously. Since such a uniformly optimal procedure does not exist, Pollak [142] suggested to minimize the following "worst-case" conditional average detection delay (WCADD):

$$\tilde{\mathbb{E}}^*[T] = \sup_{k_0 \geq 1} \mathbb{E}_{k_0}[T - k_0 \,|T \geq k_0], \tag{2.69}$$

among all stopping times $T \in C_\gamma$ satisfying the baseline ARL constraint $\mathbb{E}_0[T] \geq \gamma$, where $\gamma > 1$ is a prescribed value.

Recently, Lai [103, 107] has generalized Lorden's asymptotic theory to non-i.i.d. scenario under the convergence assumption on the conditional probability for the log-likelihood ratio (LLR). Suppose that under $\mathcal{P}_0$, the conditional density function of $y_k$ given $y_1, \cdots, y_{k-1}$ is $f_{\theta_0}(\cdot|y_1, \cdots, y_{k-1})$ for any $k \geq 1$ and that under $\mathcal{P}_{k_0}$, the conditional density function is $f_{\theta_0}(\cdot|y_1, \cdots, y_{k-1})$ for $k < k_0$ and $f_{\theta_1}(\cdot|y_1, \cdots, y_{k-1})$ for $k \geq k_0$. In this non-i.i.d. scenario, the LLR is defined as

$$s_i = \log \frac{f_{\theta_1}(y_i|y_1, \cdots, y_{i-1})}{f_{\theta_0}(y_i|y_1, \cdots, y_{i-1})}. \tag{2.70}$$

To generalize the Lorden's asymptotic theory beyond the i.i.d. setting, Lai [103,107] has imposed the following assumption on the LLR $s_i$ defined in (2.70):

$$\lim_{k\to\infty} \sup_{k_0 \geq 1} \operatorname{ess\,sup} \mathbb{P}_{k_0}\left( \max_{t \leq k} \sum_{i=k_0}^{k_0+t} s_i \geq k\rho_{10}(1+\delta) \middle| y_1, \cdots, y_{k_0-1} \right) = 0, \quad \forall \delta > 0, \tag{2.71}$$

[6]The average run length to false alarm is also denoted as the mean time to false alarm or the mean time between false alarms.

where $\rho_{10} > 0$ is a positive number. For the i.i.d. case, the number $\rho_{10}$ coincides with the Kullback-Leibler distance between $f_{\theta_1}$ and $f_{\theta_0}$.

Under the assumption (2.71), Lai [103,107] has showed that both the WWADD $\overline{\mathbb{E}}^* [T]$ proposed by Lorden [113] and the WCADD $\tilde{\mathbb{E}}^* [T]$ suggested by Pollak [142] are asymptotically lower bounded by

$$\overline{\mathbb{E}}^* [T] \geq \tilde{\mathbb{E}}^* [T] \geq \left(\rho_{10}^{-1} + o(1)\right) \log(\gamma) \quad \text{as} \quad \gamma \to \infty, \tag{2.72}$$

for all stopping times $T \in C_\gamma$ satisfying the baseline ARL constraint $\mathbb{E}_0 [T] \geq \gamma$.

It has been discussed in [103,176] that the baseline ARL constraint $\mathbb{E}_0 [T] \geq \gamma$ implies the asymptotic lower bound $\left(\rho_{10}^{-1} + o(1)\right) \log(\gamma)$ for the CADD $\mathbb{E}_{k_0} [T - k_0 | T \geq k_0]$ for only some unspecified values $k_0$. However, it is the most desirable to obtain the lower bound for $\mathbb{E}_{k_0} [T - k_0 | T \geq k_0]$ uniformly for all $k_0 \geq 1$ subject to the ARL constraint. Since no such detection procedure exists, Lai and Tartakovsky suggested to replace the global false alarm constraint (i.e., the baseline ARL constraint $\mathbb{E}_0 [T] \geq \gamma$) by the worst local (conditional) probability of raising a false alarm within a time window of given length, i.e., $\sup_{l \geq 1} \mathbb{P}_0 (l \leq T < l + m_\alpha) \leq \alpha$ for the non-conditional version [103] and $\sup_{l \geq 1} \mathbb{P}_0 (T < l + m_\alpha | T \geq l) \leq \alpha$ for the conditional version [176], respectively. Moreover, for some practical applications, including intrusion detection in computer networks and a variety of surveillance applications, it is more desirable to control the worst local false alarm rate at a certain value [176]. Let

$$C_\alpha = \left\{ T : \overline{\mathbb{P}}_{\text{fa}} (T; m_\alpha) = \sup_{l \geq 1} \mathbb{P}_0 (l \leq T < l + m_\alpha) \leq \alpha \right\}, \tag{2.73}$$

where

$$\liminf m_\alpha / |\log(\alpha)| > \rho_{10}^{-1} \text{ but } \log(m_\alpha) = o(\log(\alpha)) \text{ when } \alpha \to 0 \tag{2.74}$$

be the class of all stopping times $T$ satisfying the worst-case probability of false alarm within any time window of length $m_\alpha$ upper bounded by a predefined value $\alpha \in (0, 1)$. Lai [103] has given an asymptotic lower bound for $\mathbb{E}_{k_0} \left[(T - k_0)^+\right]$ uniformly over all $k_0 \geq 1$ under the following relaxation of assumption on the convergence of the LLR:

$$\lim_{k \to \infty} \sup_{k_0 \geq 1} \mathbb{P}_{k_0} \left( \max_{t \leq k} \sum_{i=k_0}^{k_0+t} s_i \geq k \rho_{10} (1 + \delta) \right) = 0, \quad \forall \delta > 0. \tag{2.75}$$

The following theorem, whose proof can be found in [103], gives the asymptotic lower bound for $\mathbb{E}_{k_0} \left[(T - k_0)^+\right]$ if the condition (2.75) is satisfied.

**Theorem 2.15.** *(Asymptotic lower bound [103]). Suppose that the conditions* (2.74) *and* (2.75) *hold for some positive number* $\rho_{10}$. *Then as* $\alpha \to 0$

$$\mathbb{E}_{k_0} \left[(T - k_0)^+\right] \geq |\log(\alpha)| \left( \frac{\mathbb{P}_0 (T \geq k_0)}{\rho_{10}} + o(1) \right), \text{ uniformly in } k_0 \geq 1. \tag{2.76}$$

The lower bound (2.76) has been used in [103] to prove the asymptotic optimality of the CUSUM procedure and the window limited (WL) CUSUM procedure. The method is to show that these procedures (i.e., CUSUM and WL CUSUM) with appropriately chosen parameters asymptotically reach the lower bound (2.76) subject to the false alarm constraint (2.73).

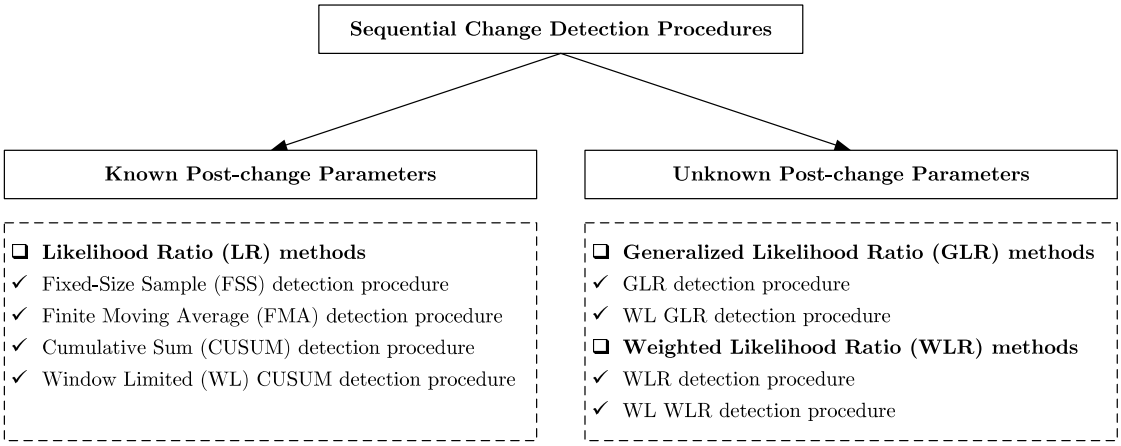



Figure 2.5 – Sequential quickest change detection procedures.

### Detection procedures when the post-change parameter is known

The objective of this sub-subsection is to resume several well-known detection algorithms that may attain the aforementioned optimality criteria. We focus only on the non-Bayesian approach where the change-point $k_0$ is assumed as unknown and non-random.

**Fixed-Size Sample (FSS) procedure.** Let $n \in \mathbb{Z}^+$ be a positive integer. The fixed-size sample (FSS) strategy[7] is, effectively, a repeated hypothesis testing procedure based on the samples of fixed size $L$ observed sequentially. At each time instant $k = nL$, for $n \geq 1$, the FSS algorithm performs a classical hypothesis test $\delta$ between the null hypothesis $\mathcal{H}_0 : \left\{y_{(n-1)L+1}, \cdots, y_{nL} \sim P_{\theta_0}\right\}$ and the alternative hypothesis $\mathcal{H}_1 : \left\{y_{(n-1)L+1}, \cdots, y_{nL} \sim P_{\theta_1}\right\}$. The FSS procedure continues until the decision $d_n$ of the test is favorable of hypothesis $\mathcal{H}_1$ for some $n \geq 1$. Since the solution to the non-sequential hypothesis testing problem is given by the fundamental Neyman-Pearson lemma, the optimal FSS procedure is designed as follows:

$$T_{\text{FSS}} = \inf_{n \geq 1} \{k = nL : d_n = 1\}, \tag{2.77}$$

where the decision $d_n$ of the Neyman-Pearson test is defined as

$$d_n = \begin{cases} 1 & \text{if} \quad S^{nL}_{(n-1)L+1} \geq h \\ 0 & \text{if} \quad S^{nL}_{(n-1)L+1} < h \end{cases}, \quad S^{nL}_{(n-1)L+1} = \sum_{i=(n-1)L+1}^{nL} \log \frac{f_{\theta_1}(y_i)}{f_{\theta_0}(y_i)}, \tag{2.78}$$

where $h$ is a chosen threshold. The demonstration of the FSS detection procedure (2.77)–(2.78) is given in figure (2.6).

Consider now the family of Gaussian distributions

$$y_k \sim \begin{cases} \mathcal{N}(\theta_0, \sigma^2) & \text{if} \quad k < k_0 \\ \mathcal{N}(\theta_1, \sigma^2) & \text{if} \quad k \geq k_0 \end{cases}, \tag{2.79}$$

[7]The fixed-size sample procedure is often denoted as the Shewhart control chart in quality control.

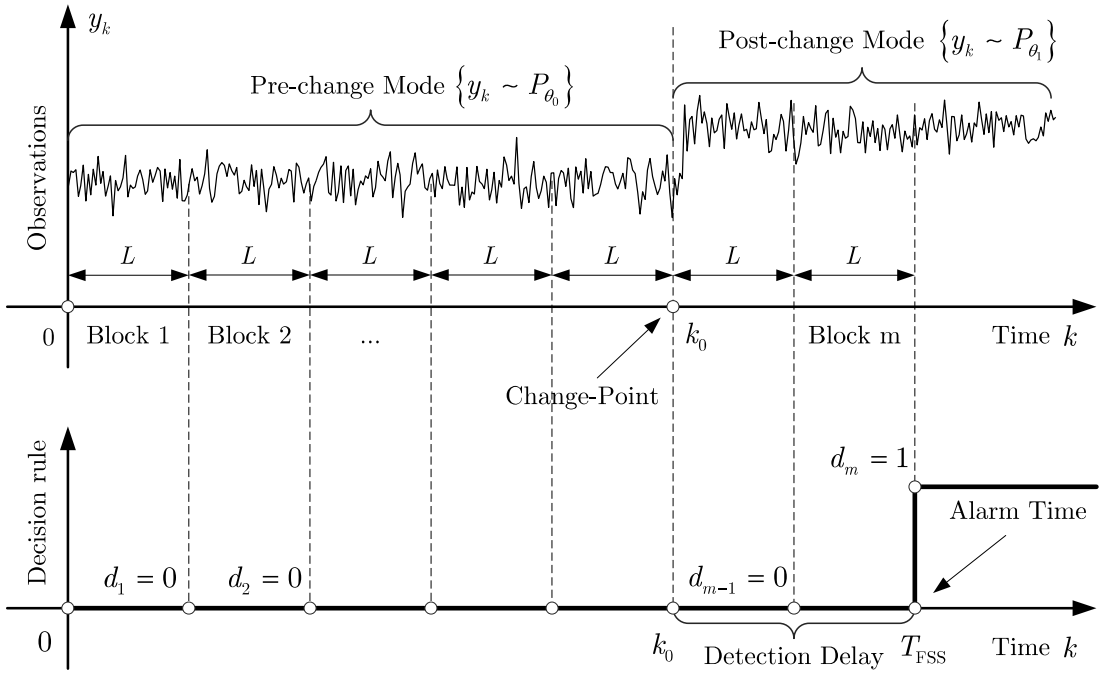



Figure 2.6 – Fixed-size sample (FSS) detection procedure.

where $\theta_0$, $\theta_1$ and $\sigma$ are assumed to be known. In this case, the LLR $S_{(n-1)L+1}^{nL}$ is given as

$$S_{(n-1)L+1}^{nL} = \frac{\theta_1 - \theta_0}{\sigma^2} \sum_{i=(n-1)L+1}^{nL} \left( y_i - \frac{\theta_1 + \theta_0}{2} \right) \tag{2.80}$$

Lorden's criterion of optimality has been studied by Nikiforov [131] in the class of all FSS tests. The main results are given in the following theorem.

**Theorem 2.16.** *(FSS detection procedure [131]). Let us consider the observation model* (2.79)*. Consider the FSS detection procedure* (2.77)–(2.78) *with the LLR computed in* (2.80)*. The optimal FSS algorithm verifies*

$$\overline{\mathbb{E}}^* \left[ T_{\text{FSS}} \right] \simeq 2 \frac{\log \left( \overline{T} \right)}{\rho_{10}}, \quad L \simeq \frac{\log \left( \overline{T} \right)}{\rho_{10}}, \quad h \simeq \log \left( \overline{T} \right), \quad \text{as} \quad \overline{T} \to \infty, \tag{2.81}$$

*where* $\overline{T} = \mathbb{E}_0 \left[ T_{\text{FSS}} \right]$ *is the ARL to false alarm,* $\overline{\mathbb{E}}^* \left[ T_{\text{FSS}} \right]$ *is the WWADD,* $h$ *is the chosen threshold,* $L$ *is the sample size and* $\rho_{10}$ *is the Kullback-Leibler information which is computed in the Gaussian case as* $\rho_{10} = 0.5 \left( \theta_1 - \theta_0 \right) / \sigma^2$*.*

**Finite Moving Average (FMA) procedure.** The Finite Moving Average (FMA) procedure is an algorithm that, for each time instant $k \geq 1$, carries out a test between the null hypothesis $\mathcal{H}_0 : \{y_{k-L+1}, \cdots, y_k \sim P_{\theta_0}\}$ and the alternative hypothesis $\mathcal{H}_1 : \{y_{k-L+1}, \cdots, y_k \sim P_{\theta_1}\}$, based on the block of $L$ observations $y_{k-L+1}, \cdots, y_k$. For the time instant $k+1$, the procedure moves one step by rejecting the last observation $y_{k-L+1}$ and employing the novel one $y_{k+1}$ to form the

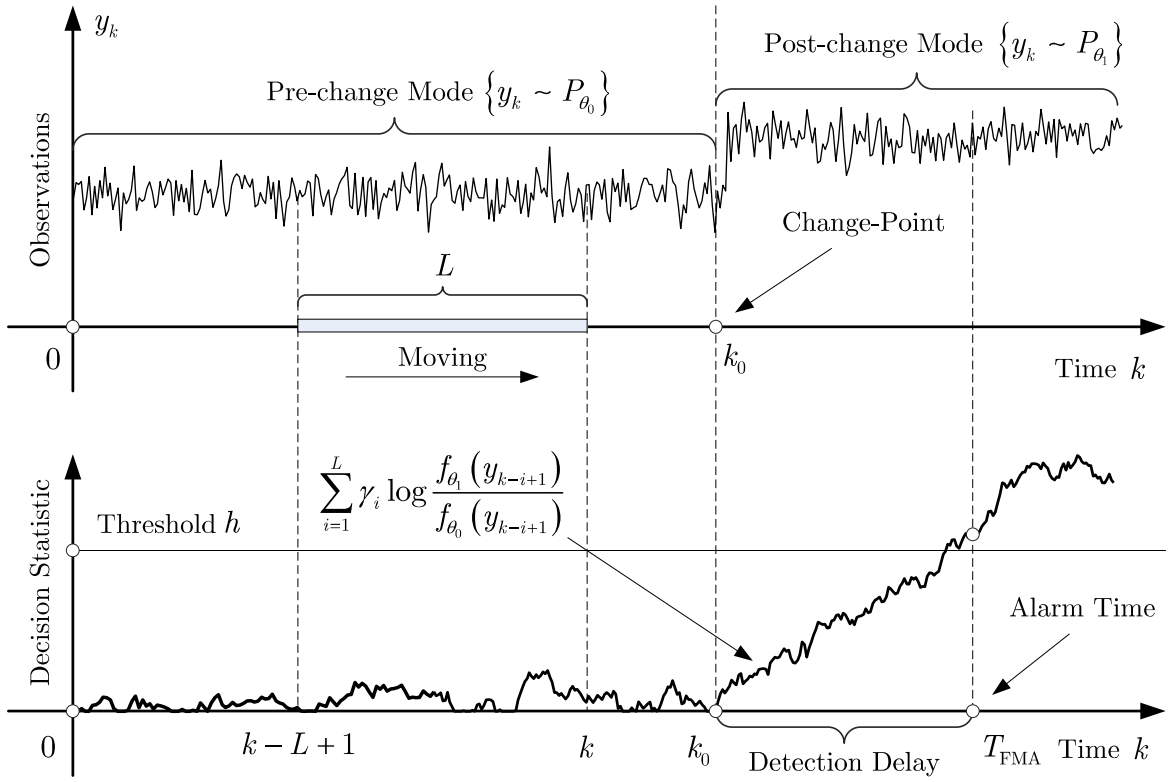



Figure 2.7 – Finite moving average (FMA) detection procedure.

novel block of observations $y_{k-L+2}, \cdots, y_{k+1}$ for constructing the test between $\mathcal{H}_0$ and $\mathcal{H}_1$. The stopping time of the FMA procedure is defined as

$$T_{\text{FMA}} = \inf\left\{k \geq L : \sum_{i=1}^{L} \gamma_i \log \frac{f_{\theta_1}(y_{k-i+1})}{f_{\theta_0}(y_{k-i+1})} \geq h\right\}, \tag{2.82}$$

where $h$ is a chosen threshold and $\gamma_i > 0$, for $i = 1, \cdots, L$, are predefined coefficients. Some results on the FMA test (2.82) were investigated in [106].

**Cumulative Sum (CUSUM) procedure.** By exploiting Wald's theory on sequential analysis [195], Page [139] developed the Cumulative Sum (CUSUM) detection scheme that contains many optimality properties. The idea of the CUSUM procedure is to take into account the variation of the log-likelihood ratio (LLR) $s_i = \log\left[f_{\theta_1}(y_i)/f_{\theta_0}(y_i)\right]$ before and after the change. In fact, the LLR $s_i$ possesses the negative mean before the change (i.e., $\mathbb{E}_{\theta_0}[s_i] < 0$) and it has the positive mean after the change (i.e., $\mathbb{E}_{\theta_1}[s_i] \geq 0$). The are several derivations of the CUSUM procedure [10, 175]. The CUSUM procedure can be described as

$$T_{\text{CS}} = \inf\left\{k \geq 1 : \max_{1 \leq i \leq k} S_i^k \geq h\right\}, \quad S_i^k = \sum_{t=i}^{k} \log \frac{f_{\theta_1}(y_t)}{f_{\theta_0}(y_t)}, \tag{2.83}$$

where $h$ is the chosen threshold. The CUSUM procedure can be also expressed in a recursive manner as

$$T_{\text{CS}} = \inf\{k \geq 1 : g_k \geq h\}, \tag{2.84}$$

where the decision function $g_k = \max_{1 \le i \le k} S_i^k$ is calculated recursively as

$$g_k = \left( g_{k-1} + \log \frac{f_{\theta_1}(y_k)}{f_{\theta_0}(y_k)} \right)^+, \quad g_0 = 0, \tag{2.85}$$

where $(x)^+ = \max(0, x)$.

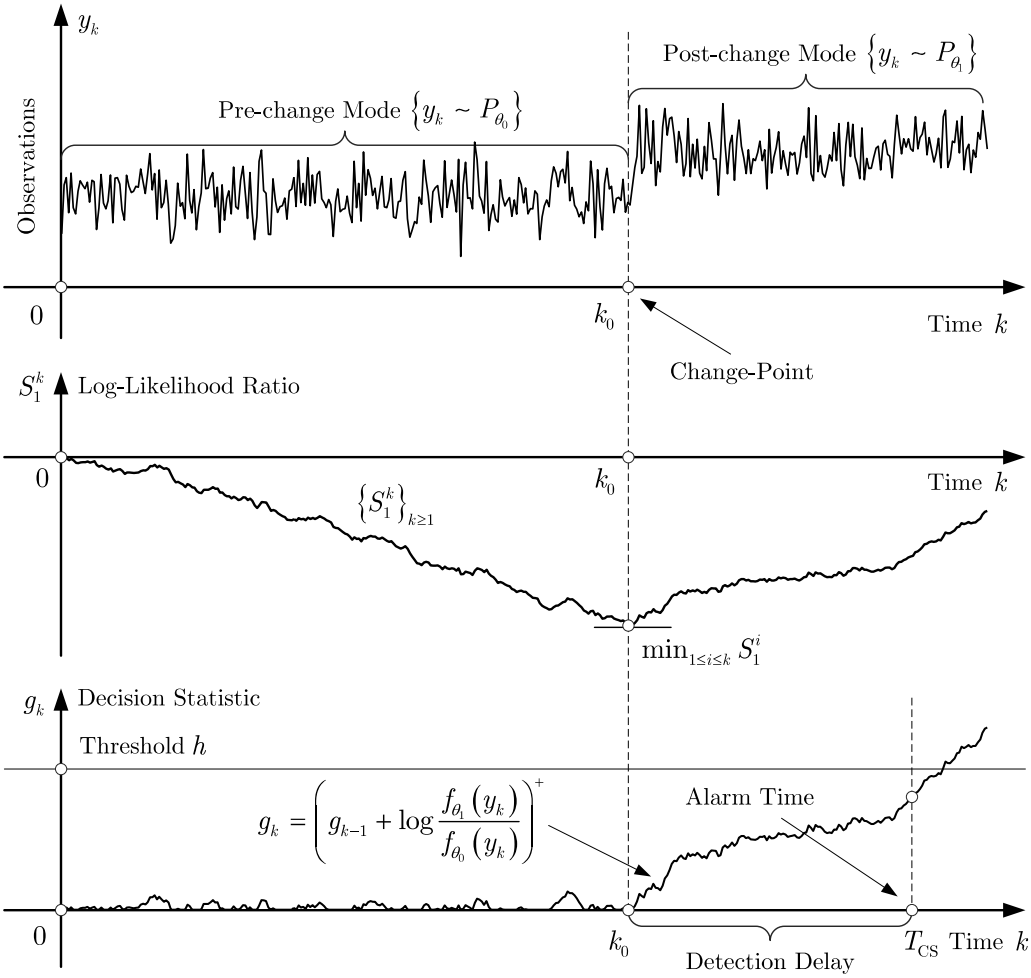



Figure 2.8 – CUSUM detection procedure.

Lorden [113] showed that the CUSUM detection scheme is asymptotically optimal in the sense that it minimizes the WWADD $\overline{\mathbb{E}}^*[T]$ defined in (2.65) among all stopping times $T \in C_\gamma$ satisfying the baseline ARL constraint $\mathbb{E}_0[T] \ge \gamma$. Especially, he showed that if the threshold $h$ is such chosen as $h \sim \log(\gamma)$ and $\mathbb{E}_0[T_{CS}] \sim \gamma$, then

$$\overline{\mathbb{E}}^*[T_{CS}] = \inf_{T \in C_\gamma} \left\{ \overline{\mathbb{E}}^*[T] \right\} \sim \frac{\log(\gamma)}{\rho_{10}}, \quad \text{as} \quad \gamma \to \infty. \tag{2.86}$$

Recently, Lai [103] showed that the CUSUM test with suitably chosen threshold $h$ attains also the asymptotic lower bound for the WCADD $\tilde{\mathbb{E}}^*[T]$ defined in (2.69) among all stopping times $T \in C_\gamma$, i.e.,

$$\tilde{\mathbb{E}}^*[T_{CS}] = \inf_{T \in C_\gamma} \left\{ \tilde{\mathbb{E}}^*[T] \right\} \sim \frac{\log(\gamma)}{\rho_{10}}, \quad \text{as} \quad \gamma \to \infty. \tag{2.87}$$

The exact optimality (i.e., non-asymptotic for any $\gamma \geq 1$) of the CUSUM procedure w.r.t. Lorden's criterion has been studied by Moustakides [124] and Ritov [157], respectively.

**Window Limited CUSUM procedure.** The Window Limited (WL) CUSUM algorithm was first introduced by Willsky and Jones [205] for the detection of abrupt changes in linear systems. The idea of the WL CUSUM procedure is to utilize the last $m_\alpha$ observations for the decision-making. The stopping time $T_{\mathrm{WL}}$ of the WL CUSUM test is defined as

$$T_{\mathrm{WL}} = \inf\left\{k \geq m_\alpha : \max_{k-m_\alpha+1\leq i\leq k} S_i^k \geq h\right\}, \quad S_i^k = \sum_{t=i}^{k} \log \frac{f_{\theta_1}(y_t)}{f_{\theta_0}(y_t)}, \tag{2.88}$$

where $h$ is a chosen threshold. The WL CUSUM algorithm can be utilized also for the non-i.i.d. scenario, where the LLR $S_i^k$ becomes

$$S_i^k = \sum_{t=i}^{k} \log \frac{f_{\theta_1}(y_t|y_1,\cdots,y_{t-1})}{f_{\theta_0}(y_t|y_1,\cdots,y_{t-1})}. \tag{2.89}$$

**Theorem 2.17.** *(Properties of the WL CUSUM detection procedure [103]). Consider the WL CUSUM detection procedure defined in* (2.88)–(2.89)*. If the threshold $h$ is such chosen that $2m_\alpha e^{-h} = \alpha$, where the window size $m_\alpha$ satisfies* (2.74)*, then*

$$\overline{\mathbb{P}}_{\mathrm{fa}}(T_{\mathrm{WL}}; m_\alpha) = \sup_{l\geq m_\alpha} \mathbb{P}_0(l \leq T_{\mathrm{WL}} \leq l + m_\alpha - 1) \leq 2m_\alpha e^{-h}. \tag{2.90}$$

*Moreover, if the constraint* (2.75) *and the following constraint:*

$$\lim_{k\to\infty} \sup_{1\leq k_0\leq t} \operatorname{ess\,sup} \mathbb{P}_{k_0}\left(k^{-1}\sum_{i=t}^{t+k} s_i \leq \rho_{10} - \delta \,\middle|\, y_1, y_2, \cdots, y_{t-1}\right) = 0 \quad \forall \delta > 0, \tag{2.91}$$

*are satisfied, then as $\alpha \to 0$, we have*

$$\mathbb{E}_{k_0}\left[(T_{\mathrm{WL}} - k_0)^+\right] \sim |\log(\alpha)|\left(\frac{\mathbb{P}_0(T_{\mathrm{WL}} \geq k_0)}{\rho_{10}} + o(1)\right) \text{ uniformly in } k_0 \geq 1. \tag{2.92}$$

It follows from the above theorem that if the conditions (2.74), (2.75) and (2.91) are satisfied and the threshold $h$ is chosen as $2m_\alpha e^{-h} = \alpha$, the WL CUSUM procedure is asymptotically optimal in the sense that it minimizes $\mathbb{E}_{k_0}\left[(T_{\mathrm{WL}} - k_0)^+\right]$ uniformly in $k_0 \geq 1$ among all stopping times in the class $C_\alpha$ defined in (2.73), as $\alpha \to 0$.

### Detection procedures when the post-change parameter is unknown

In many practical situations, the pre-change hypothesis is often simple (i.e., $\mathcal{H}_0 = \{\theta = \theta_0\}$) but the post-change hypothesis is composite (i.e., $\mathcal{H}_1 = \{\theta \in \Theta_1\}$, where $\Theta_1 \subseteq \Theta \setminus \{\theta_0\}$). There are two approaches for dealing with such circumstances [103,175]. The first one utilizes a weighting function $G(\theta)$, which is often considered as the *a priori* distribution of the unknown parameter $\theta \in \Theta_1$, for weighting the LR w.r.t. all possible values of the parameter $\theta \in \Theta_1$. The second one involves the generalized LR approach, which replaces the unknown parameter $\theta$ by its maximum likelihood estimate (MLE).

**Weighted Likelihood Ratio (WLR) procedure.** The weighted likelihood ratio (WLR) detection procedure can be defined directly as

$$\tilde{T} = \inf\left\{k \geq 1 : \max_{1\leq i\leq k} \log \tilde{\Lambda}_i^k \geq h\right\}, \tag{2.93}$$

where $h$ is a chosen threshold and $\tilde{\Lambda}_i^k$ is the weighted LR, being calculated by

$$\tilde{\Lambda}_i^k = \int_{\theta\in\Theta_1} \prod_{t=i}^{k} \frac{f_\theta\left(y_t|y_1,\cdots,y_{t-1}\right)}{f_{\theta_0}\left(y_t|y_1,\cdots,y_{t-1}\right)} dG\left(\theta\right), \tag{2.94}$$

where $f_\theta\left(y_t|y_1,\cdots,y_{t-1}\right)$ is the conditional density function of $y_t$ given $y_1,\cdots,y_{t-1}$ and $G\left(\theta\right)$ is the *a priori* distribution of the parameter $\theta$ on $\Theta_1$.

Similar to the case of GLR scheme, let us consider the exponential family of distributions $\mathcal{P} = \{P_\theta|\theta \in \Theta\}$ whose p.d.f. is given in (2.100), where the parameter value $\theta = \theta_0$ before the change and $\theta \neq \theta_0$ after the change. In this case, the weighted LR statistic is computed as

$$\tilde{\Lambda}_i^k = \int_{\Theta_1} \exp\left\{\left(\theta-\theta_0\right) S_i^k - \left(k-i+1\right)\left(d\left(\theta\right) - d\left(\theta_0\right)\right)\right\} dG\left(\theta\right), \tag{2.95}$$

where $S_i^k = \sum_{t=i}^k y_t$. The approximation of the WWADD $\overline{\mathbb{E}}_\theta^*\left[\tilde{T}\right]$ of the WLR detection procedure in the case of exponential family of distributions is shown in Theorem 2.18.

**Theorem 2.18.** *(Properties of the WLR detection procedure [10, 143]). Consider the WLR detection procedure given in* (2.93) *with the weighted LR statistic calculated in* (2.95)*. Suppose that the weighting function $G\left(\theta\right)$ has a positive derivative in the neighborhood of $\theta$ and the threshold $h$ is such chosen that $\overline{T} = \mathbb{E}_0\left[\tilde{T}\right]$. Then, as $\overline{T} \to \infty$, the approximation of the WWADD $\overline{\mathbb{E}}_\theta^*\left[\tilde{T}\right]$ is given as*

$$\overline{\mathbb{E}}_\theta^*\left[\tilde{T}\right] \approx \frac{\log\left(\overline{T}\right) + \frac{1}{2}\log\left[\frac{\log\left(\overline{T}\right)}{\rho\left(\theta,\theta_0\right)}\right]}{\rho\left(\theta,\theta_0\right)} - \frac{1}{2\rho\left(\theta,\theta_0\right)}\left\{\log\left[2\pi\frac{\dot{G}^2\left(\theta\right)}{\ddot{d}\left(\theta\right)}\right] - 1\right\} + o\left(1\right), \tag{2.96}$$

*where the K-L distance is calculated as*

$$\rho\left(\theta,\theta_0\right) = \left(\theta-\theta_0\right)\dot{d}\left(\theta\right) - \left(d\left(\theta\right) - d\left(\theta_0\right)\right). \tag{2.97}$$

It follows from (2.96) that the WWADD $\overline{\mathbb{E}}_\theta^*\left[\tilde{T}\right]$ for the WLR detection rule does not reach the infimum of mean delay $\log\left(\overline{T}\right)/\rho\left(\theta,\theta_0\right)$ for the class of detection procedures satisfying the constraint on the ALR2FA $\overline{T} \geq \gamma$ when $\gamma \to \infty$. The additional term can be considered as the price to be paid for the unknown *a priori* information about the parameter $\theta$.

**Generalized Likelihood Ratio (GLR) procedure.** By replacing the unknown parameter $\theta \in \Theta_1$ with its maximum likelihood estimate, the so-called generalized CUSUM detection procedure is defined as

$$\hat{T} = \inf\left\{k \geq 1 : \max_{1\leq i\leq k} \log \hat{\Lambda}_i^k \geq h\right\}, \tag{2.98}$$

where $h$ is a chosen threshold and $\hat{\Lambda}_i^k$ is the generalized likelihood ratio, being calculated as

$$\hat{\Lambda}_i^k = \sup_{\theta\in\Theta_1} \prod_{t=i}^{k} \frac{f_\theta\left(y_t|y_1,\cdots,y_{t-1}\right)}{f_{\theta_0}\left(y_t|y_1,\cdots,y_{t-1}\right)}, \tag{2.99}$$

where $f_\theta\left(y_t|y_1,\cdots,y_{t-1}\right)$ is the conditional density function of $y_t$ given $y_1,\cdots,y_{t-1}$.

Consider the exponential parametric family of distributions $\mathcal{P} = \{P_\theta|\theta\in\Theta\}$ whose p.d.f. has the form

$$f_\theta\left(y\right) = g\left(y\right)\exp\left\{\theta y - d\left(\theta\right)\right\}, \tag{2.100}$$

where $g : y \mapsto g\left(y\right)$ and $d : \theta \mapsto d\left(\theta\right)$ are two functions from $\mathbb{R}$ to $\mathbb{R}$.

Suppose that the parameter value $\theta = \theta_0$ before the change and $\theta\in\Theta_1 \equiv [\theta_1,\infty)$, where $\theta_1 > \theta_0$, after the change. In this case, the GLR procedure is given by

$$\hat{T} = \inf\left\{k\geq 1 : \max_{1\leq i\leq k}\sup_{\theta\geq\theta_1}\left\{\left(\theta-\theta_0\right)S_i^k - \left(k-i+1\right)\left(d\left(\theta\right)-d\left(\theta_0\right)\right)\right\}\geq h\right\}, \tag{2.101}$$

where $S_i^k = \sum_{t=i}^k y_t$ and $h$ is a chosen threshold. The following theorem, whose proof can be found in [114], gives the upper bound on the WWADD $\overline{\mathbb{E}}_\theta^*\left[\hat{T}\right]$, which depends on the parameter $\theta$, of the GLR procedure (2.101) subject to the ARL constraint $\mathbb{E}_0\left[\hat{T}\right]\geq\gamma$.

**Theorem 2.19.** *(Properties of the GLR detection procedure [10, 67, 114]). Consider the GLR detection procedure given in* (2.101)*. When the threshold $h$ and the error probability $\alpha$ are connected through*

$$e^{-h} = \frac{\alpha}{3\log\left(\alpha^{-1}\right)\left[1+\frac{1}{\rho(\theta_1,\theta_0)}\right]^2}, \tag{2.102}$$

*then the ARL to false alarm $\overline{T}$ satisfies*

$$\overline{T} = \mathbb{E}_0\left[\hat{T}\right]\geq\alpha^{-1}, \tag{2.103}$$

*and the WWADD $\overline{\mathbb{E}}_\theta^*\left[\hat{T}\right]$ is upper bounded by*

$$\overline{\mathbb{E}}_\theta^*\left[\hat{T}\right] \leq \frac{\log\left(\overline{T}\right)+\log\left(\log\left(\overline{T}\right)\right)}{\rho\left(\theta,\theta_0\right)} + \frac{2\log\left(\sqrt{3}\left[1+\frac{1}{\rho(\theta_1,\theta_0)}\right]\right)}{\rho\left(\theta,\theta_0\right)} + \frac{\theta^2}{\rho^2\left(\theta,\theta_0\right)}\frac{\partial^2 d\left(\theta\right)}{\partial\theta^2} + 1, \tag{2.104}$$

*for all $\theta\geq\theta_1$, where $\rho\left(\theta,\theta_0\right) = \mathbb{E}_\theta\left[\log\left(f_\theta\left(y\right)/f_{\theta_0}\left(y\right)\right)\right]$ is the K-L distance between $f_\theta$ and $f_{\theta_0}$.*

Theorem 2.19 allows us to establish the relation between the WWADD $\overline{\mathbb{E}}_\theta^*\left[\hat{T}\right]$ and the ARL2FA $\mathbb{E}_0\left[\hat{T}\right]$. Moreover, the upper bound on the WWADD $\overline{\mathbb{E}}_\theta^*\left[\hat{T}\right]$ of the GLR detection procedure obtained in Theorem 2.19 can be utilized to compare with the upper bound on the WWADD $\overline{\mathbb{E}}_\theta^*\left[\tilde{T}\right]$ of the WLR detection procedure in Theorem 2.18.

**Window Limited Weighted Likelihood Ratio (WL WLR) procedure.** The window-limited weighted likelihood ratio (WL WLR) detection procedure utilizes the mixture LR statistics $\tilde{\Lambda}_i^k$ given in (2.94). The WL WLR procedure is defined as

$$\tilde{T}_{\mathrm{WL}}(h) = \inf\left\{k \geq m_\alpha : \max_{k-m_\alpha+1\leq i\leq k} \log \tilde{\Lambda}_i^k \geq h\right\}, \tag{2.105}$$

where the time window $m_\alpha$ satisfying $m_\alpha/|\log(\alpha)| \to \infty$ but $\log(m_\alpha) = o(\log(\alpha))$ as $\alpha \to 0$. The following theorem, whose proof can be found in [103], proves the asymptotic optimality of the WL WLR detection procedure.

**Theorem 2.20.** *(Properties of the WL WLR detection procedure [103]). Suppose that for every $\delta > 0$, there exist $\Theta_\delta \subset \Theta_1$ and $k(\delta) \geq 1$ such that $\theta_1 \in \Theta_\delta$, $G(\Theta_\delta) > 0$, and*

$$\sup_{k\geq k(\delta)} \sup_{1\leq k_0\leq t_0} \operatorname{ess\,sup} \mathbb{P}_{k_0}^{\theta}\left(\inf_{\theta\in\Theta_\delta} \sum_{i=t_0}^{t_0+k} s_i(\theta) \leq k\left(\rho(\theta_1,\theta_0)-\delta\right) \middle| y_1, y_2, \cdots, y_{t_0-1}\right) \leq \delta, \tag{2.106}$$

*where the LLR $s_i(\theta)$ is calculated as*

$$s_i(\theta) = \log \frac{f_\theta(y_i|y_1,\ldots,y_{i-1})}{f_{\theta_0}(y_i|y_1,\ldots,y_{i-1})},$$

*and $\rho(\theta_1,\theta_0)$ is the K-L distance. If the window size $m_\alpha$ is such chosen that $\alpha = 2m_\alpha e^{-h}$, then $\sup_{l\geq 1}\mathbb{P}_0\left(l \leq \tilde{T}_{WL} < l + m_\alpha\right) \leq \alpha$ as $\alpha \to 0$ and then*

$$\sup_{k_0\geq 1} \operatorname{ess\,sup} \mathbb{E}_{k_0}^{\theta_1}\left[\left(\tilde{T}_{\mathrm{WL}} - k_0 + 1\right)^+ \middle| y_1, \cdots, y_{k_0-1}\right] \leq h\frac{1+o(1)}{\rho(\theta_1,\theta_0)}, \tag{2.107}$$

$$\mathbb{E}_{k_0}^{\theta_1}\left[\left(\tilde{T}_{\mathrm{WL}} - k_0\right)^+\right] \leq h\frac{\mathbb{P}_0\left(\tilde{T}_{\mathrm{WL}} \geq k_0\right)}{\rho(\theta_1,\theta_0)+o(1)}, \quad \textit{uniformly in } k_0 \geq 1. \tag{2.108}$$

It follows from Theorem 2.20 that the WL WLR procedure (2.105) based on the mixture LLRs with appropriately chosen parameters is asymptotically optimal in the sense that it minimizes the ADD $\overline{\mathbb{E}}_{k_0}^{\theta_1}\left[\left(\tilde{T}_{\mathrm{WL}} - k_0\right)^+\right]$ uniformly in $k_0 \geq 1$ over all stopping times satisfying the constraint (2.73) as $\alpha \to 0$.

**Window Limited Generalized Likelihood Ratio (WL GLR) procedure.** The idea of the window limited generalized likelihood ratio (WL GLR) approach is due to Willsky and Jones [205] who proposed to utilize the last $m_\alpha$ observations at each time instant instead of all observed samples. The stopping time of the WL GLR procedure is defined as

$$\hat{T}_{\mathrm{WL}} = \inf\left\{k \geq m_\alpha : \max_{k-m_\alpha+1\leq i\leq k-\tilde{m}_\alpha} \sup_{\theta\in\Theta_1} \log \hat{\Lambda}_i^k \geq h\right\}, \tag{2.109}$$

where the GLR $\hat{\Lambda}_i^k$ is calculated in (2.99) and $\tilde{m}_\alpha < m_\alpha$ is the number of necessary observations for the MLE and $h$ is a chosen threshold. It has been shown in [103, 107] that the WL GLR defined in (2.109), with the parameters are such chosen that $h \sim |\log(\alpha)|$, $\tilde{m}_\alpha \sim o(|\log(\alpha)|)$ and $m_\alpha = \frac{1}{2}\alpha\exp(h)$, attains the lower bound for the ADD $\mathbb{E}_{k_0}\left[(T-k_0)^+\right]$ uniformly for all $k_0 \geq m_\alpha$ among all stopping times $T$ in the class $C_\alpha$ defined in (2.73). Recursive methods for the implementation of the WL GLR detection procedure and numerical examples have been performed in [102, 107].

### 2.4.3 Sequential change-point detection-isolation

In the previous section, we have resumed several optimality criteria and detection procedures for the problem of detecting abrupt changes in a stochastic system. In this section, we continue with the joint detection and isolation problem which was first introduced by Nikiforov [130].

#### Problem statement

In the case of multiple hypotheses, there are several post-change hypotheses $\mathcal{H}_l$, for $1 \leq l \leq K$. As before, let $y_1, y_2, \cdots$ be a sequence of independent random observations generated from a parametric family of distributions $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$, where the parameter space $\Theta = \bigcup_{l=0}^{K} \theta_l$. Similar to the detection problem, the parameter $\theta$ receives its nominal value $\theta = \theta_0$ under normal operation. From an unknown change-point $k_0 \geq 1$, the system shifts to another mode, causing the parameter $\theta$ to change its value from $\theta = \theta_0$ to $\theta = \theta_l$, for $l = 1, \cdots, K$, where $K$ stands for possible change modes. In other words, the random variables $y_1, y_2, \cdots, y_{k_0-1}$ have the distribution $P_{\theta_0}$ while the random variables $y_{k_0}, y_{k_0+1}, \cdots$ have the distributions $P_{\theta_l}$. The statistical model for the change-point detection-isolation is described as

$$y_k \sim \begin{cases} P_{\theta_0} & \text{if} \quad k < k_0 \\ P_{\theta_l} & \text{if} \quad k \geq k_0 \end{cases}, \quad l = 1, \cdots, K. \tag{2.110}$$

Denote by $\mathcal{P}_{k_0}^l$ the probability measure when the observations $y_1, y_2, \cdots, y_{k_0-1} \sim P_{\theta_0}$ and the observations $y_{k_0}, y_{k_0+1}, \cdots \sim P_{\theta_l}$, for $1 \leq l \leq K$. Also, we denote by $\mathcal{P}_0 \triangleq \mathcal{P}_\infty^l \triangleq \mathcal{P}_{k_0}^0$, for all $0 \leq l \leq K$ and all $k_0 \geq 1$, the pre-change probability measure when the observations $y_1, y_2, \cdots \sim P_{\theta_0}$. Let $\mathbb{P}_{k_0}^l$ (res. $\mathbb{P}_0 \triangleq \mathbb{P}_\infty^l \triangleq \mathbb{P}_{k_0}^0$) and $\mathbb{E}_{k_0}^l$ (res. $\mathbb{E}_0 \triangleq \mathbb{E}_\infty^l \triangleq \mathbb{E}_{k_0}^0$) be, respectively, the probability and the expectation with respect to the probability measure $\mathcal{P}_{k_0}^l$ (res. $\mathcal{P}_0 \triangleq \mathcal{P}_\infty^l \triangleq \mathcal{P}_{k_0}^0$).

The change detection and isolation algorithm should calculate a pair $(T, \nu)$ based on the observations $y_1, y_2, \cdots$, where $\nu$, for $1 \leq \nu \leq K$, is the final detection and $T$ is the stopping time at which the change type $\nu$ is declared. It it intuitively obvious that the detection-isolation algorithm should be favorable of small delay for detection-isolation with few false alarm and few false isolation rates.

#### Minimax optimality criteria

In the following, we note several optimality criteria for the quickest change detection-isolation problem by the minimax approach, where the change-point $k_0$ is unknown and non-random.

**Worst-worst-case conditional detection-isolation delay.** For evaluating the false alarm and false isolation rates, suppose that the observations $(y_k)_{k \geq 1}$ are coming from the distribution $P_{\theta_l}$, for $0 \leq l \leq K$. Consider the following sequence of alarm times and final decisions $(T_r, \nu_r)$:

$$T_0 = 0 < T_1 < T_2 < \cdots < T_r < \cdots, \quad \text{and} \quad \nu_1, \nu_2, \cdots, \nu_r, \cdots,$$

where $T_r$ is the alarm time of the detection-isolation algorithm applied to $y_{T_{r-1}+1}, y_{T_{r-1}+2}, \cdots$ and $\nu_r$ is the corresponding final decision. The first false alarm/isolation $T^{\nu=j}$ of the $j$-type is defined as

$$T^{\nu=j} = \inf \{T_r : \nu_r = j\}, \quad 1 \leq j \neq l \leq K,$$

where it is assumed that $\inf\{\emptyset\} = \infty$ and that the system restarts from scratch after each false alarm/isolation.

For measuring the risk associated with the detection-isolation delay, consider the sequence of observations $(y_k)_{k\geq 1}$ which are coming from the observation model (2.110). If the change is detected/isolated correctly after the change-point $k_0$ ($T \geq k_0$ and $\nu = l$), the delay for detection-isolation of the $l$-type change is defined as

$$\tau_l = T - k_0 + 1. \tag{2.111}$$

As discussed in [130], the detection delay $\tau_l = T - k_0 + 1$, for $1 \leq l \leq K$, should be stochastically small and the mean time to false alarm/isolation $T^{\nu=j} = \inf\{T_r : \nu_r = j\}$, for any combination of $j \neq l$, should be stochastically large. By generalizing Lorden's criterion for the detection problem, Nikiforov [130] proposed to minimize the worst-worst-case mean delay for detection-isolation

$$\overline{\tau}^*(\delta) = \sup_{k_0\geq 1, 1\leq l\leq K} \operatorname{ess\,sup} \mathbb{E}^l_{k_0}\left[(T - k_0 + 1) \middle| T \geq k_0, y_1, y_2, \cdots, y_{k_0-1}\right] \tag{2.112}$$

among all procedures $\delta = (T, \nu) \in C_\gamma$ satisfying

$$C_\gamma = \left\{\delta = (T,\nu) : \min_{0\leq j\leq K} \min_{0\leq l\neq j\leq K} \mathbb{E}_l\left[\inf\{T_r : \nu_r = j\}\right] \geq \gamma\right\}, \tag{2.113}$$

where $\mathbb{E}_l[\cdot] \triangleq \mathbb{E}^l_1[\cdot]$, for $1 \leq l \leq K$, and $\gamma$ is the minimum value for the mean time to false alarm/isolation. The asymptotic lower bound $n(\gamma)$ for the worst-worst-case delay (2.112)–(2.113) is obtained in [130] as

$$n(\gamma) = \inf_{(T,\nu)\in C_\gamma} (\overline{\tau}^*) \gtrsim \frac{\log(\gamma)}{\rho^*}, \text{ as } \gamma \to \infty, \tag{2.114}$$

where $\rho^* = \min_{1\leq l\leq K} \min_{0\leq j\neq l\leq K} \rho_{lj}$ and $\rho_{lj} = \mathbb{E}_{\theta_l}\left[\log\left(f_{\theta_l}(y_1)/f_{\theta_j}(y_1)\right)\right]$ is the K-L information between $f_{\theta_l}$ and $f_{\theta_j}$.

**Uniformly constrained conditional probability of false isolation.** The drawback of the criterion (2.112)–(2.113) lies in that the change-point $k_0$ is constrained at the onset time $k_0 = 1$ for evaluating false isolation probabilities. For circumventing this inconvenience, a more tractable criterion has been introduced in [129,132], involving the minimization of the maximum mean delay for the detection-isolation

$$\tilde{\tau}^*(\delta) = \sup_{k_0\geq 1, 1\leq l\leq K} \mathbb{E}^l_{k_0}[T - k_0 + 1 | T \geq k_0] \tag{2.115}$$

among all stopping procedures $\delta = (T, \nu) \in C_{\gamma,\beta}$ satisfying

$$C_{\gamma,\beta} = \left\{\delta = (T,\nu) : \mathbb{E}_0[T] \geq \gamma, \max_{1\leq l\leq K} \max_{1\leq j\neq l\leq K} \sup_{k_0\geq 1} \mathbb{P}^l_{k_0}(\nu = j | T \geq k_0) \leq \beta\right\}. \tag{2.116}$$

An asymptotic lower bound $n(\gamma, \beta)$ for the maximum mean delay (2.115)–(2.116) over all procedures in the class $C_{\gamma\beta}$ is given by [132]:

$$n(\gamma,\beta) \gtrsim \max\left\{\log\frac{\gamma}{\rho^*_{\text{fa}}}, \log\frac{\beta^{-1}}{\rho^*_{\text{fi}}}\right\}, \text{ as } \min\left\{\gamma, \beta^{-1}\right\} \to \infty, \tag{2.117}$$

where $\rho^*_{\text{fa}} = \min_{1\leq j\leq K} \rho_{j0}$ and $\rho^*_{\text{fi}} = \min_{1\leq l\leq K} \min_{1\leq j\neq l\leq K} \rho_{lj}$.

**Uniformly constrained probabilities of false alarm and false isolation within a time window.** In aforementioned criteria, the false alarm constraint $\mathbb{E}_0[T] \geq \gamma$ stipulates a large mean time to false alarm. However, a long expected duration to false alarm does not necessarily imply the small value for the probability of false alarm within any time window of given length [103, 104, 176]. Moreover, for safety-critical applications, it is preferable to warrant that the local probabilities of false alarm and false isolation within a time window of predefined length are upper bounded [176]. For these reasons, Lai [104] suggested to replace the ARL to false alarm and false isolation constraints by the following worst-case probabilities of false alarm and false isolation within a time window:

$$\overline{\mathbb{P}}_{\text{fa}} = \sup_{l \geq 1} \mathbb{P}_0\left(l \leq T < l + m_\alpha\right), \; \overline{\mathbb{P}}_{\text{fi}} = \max_{1 \leq l \leq K} \sup_{k_0 \geq 0} \mathbb{P}_{k_0}^l\left(k_0 \leq T < k_0 + m_\alpha, \nu \neq l\right), \tag{2.118}$$

where $\overline{\mathbb{P}}_{\text{fa}}$ denotes the worst-case probability of false alarm and $\overline{\mathbb{P}}_{\text{fi}}$ stands for the worst-case probability of false isolation and $m_\alpha$ satisfies $\liminf m_\alpha / |\log(\alpha)| > 1/|\rho^*|$ but $\log(m_\alpha) = o(\log(\alpha))$ as $\alpha \to 0$. Let

$$C_{m_\alpha} = \left\{\delta = (T, \nu) : \overline{\mathbb{P}}_{\text{fa}}(\delta) \leq \alpha m_\alpha, \; \overline{\mathbb{P}}_{\text{fi}}(\delta) \leq \alpha m_\alpha\right\}, \tag{2.119}$$

be the class of all detection-isolation procedures satisfying constraints on $\overline{\mathbb{P}}_{\text{fa}}$ and $\overline{\mathbb{P}}_{\text{fi}}$. As $\alpha \to 0$, an asymptotic lower bound for the mean delay for detection-isolation $\mathbb{E}_{k_0}^l\left[(T - k_0 + 1)^+\right]$, for every $1 \leq l \leq K$, in the class $C_{m_\alpha}$ is obtained in [104]:

$$\mathbb{E}_{k_0}^l\left[(T - k_0 + 1)^+\right] \geq \frac{\mathbb{P}_0(T \geq k_0)\,|\log(\alpha)|}{\rho_l + o(1)} \text{ uniformly in } k_0 \geq 1, \tag{2.120}$$

where $\rho_l = \min_{j \neq l} \rho_{lj}$.

### Detection and isolation procedures

Several detection-isolation algorithms which attain different optimality criteria have been proposed in literature. In the following, we consider typical ones, including the generalized CUSUM procedure [130], matrix CUSUM procedure [138], recursive vector CUSUM procedure [129, 132] and the non-recursive window limited vector CUSUM procedure [104].

**Generalized CUSUM procedure.** By utilizing the idea of the class of extended stopping variables, Nikiforov [130] generalized the Page's CUSUM procedure [139] to the problem of joint detection and isolation. The generalized CUSUM procedure $\delta_{\text{GCS}} = (T_{\text{GCS}}, \nu_{\text{GCS}})$ introduced by Nikiforov [130] can be described as

$$T_{\text{GCS}} = \min_{1 \leq l \leq K}\left\{T_{\text{GCS}}^l\right\}, \tag{2.121}$$

$$\nu_{\text{GCS}} = \arg\min_{1 \leq l \leq K}\left\{T_{\text{GCS}}^l\right\}, \tag{2.122}$$

where $T_{\text{GCS}}^l$ is the stopping time responsible for the detection of hypothesis $\mathcal{H}_l$ against other alternative hypotheses $\{\mathcal{H}_j\}_{0 \leq j \neq l \leq K}$ and it is defined as

$$T_{\text{GCS}}^l = \inf\left\{k \geq 1 : \max_{1 \leq i \leq k} \min_{0 \leq j \neq l \leq K} S_i^k(l, j) \geq h\right\}, \; S_i^k(l, j) = \sum_{t=i}^{k} \log \frac{f_{\theta_l}(y_t)}{f_{\theta_j}(y_t)}, \tag{2.123}$$

where $h$ is the chosen threshold and $S_i^k(l, j)$ is the LLR between hypothesis $\mathcal{H}_l$ and hypothesis $\mathcal{H}_j$.

**Theorem 2.21.** *(Asymptotic optimality of generalized CUSUM procedure [130]). Consider the generalized CUSUM procedure* (2.121)–(2.123) *in the class* $C_\gamma$ *defined in* (2.113) *and* $h \sim \log(\gamma)$ *as* $\gamma \to \infty$*, especially* $h = \log(\gamma)$*. Then*

$$\overline{\tau}^* (T_{\mathrm{GCS}}) \leq \max_{1 \leq l \leq K} \mathbb{E}_l [T_{\mathrm{GCS}}] \sim \frac{\log \gamma}{\rho^*} \quad as \quad \gamma \to \infty, \tag{2.124}$$

*where* $\rho^* = \min_{1 \leq l \leq K} \min_{0 \leq j \neq l \leq K} \rho_{lj}$*, where* $0 < \rho_{lj} < \infty$ *for all* $0 \leq l \neq j \leq K$ *is the minimal Kullback-Leibler information between two closet hypotheses.*

**Matrix CUSUM procedure.** As discussed in [138], the main drawback of the generalized CUSUM algorithm (2.121)–(2.123) lies in that it does not permit a recursive form, which makes it computationally prohibitive for on-line applications. For this reason, Oskiper and Poor [138] designed the matrix CUSUM procedure $\delta_{\mathrm{MCS}} = (T_{\mathrm{MCS}}, \nu_{\mathrm{MCS}})$ which can be expressed in a recursive manner. The authors suggested to replace the $\max - \min$ operands in (2.123) by the $\min - \max$ operands, leading to the following extended stopping time $T^l_{\mathrm{MCS}}$ for the matrix CUSUM procedure:

$$T^l_{\mathrm{MCS}} = \inf \left\{ k \geq 1 : \min_{0 \leq j \neq l \leq K} \max_{1 \leq i \leq k} S^k_i (l, j) \geq h \right\}. \tag{2.125}$$

The stopping time and the final decision of the matrix CUSUM algorithm is described as

$$T_{\mathrm{MCS}} = \min_{1 \leq l \leq K} \left\{ T^l_{\mathrm{MCS}} \right\}, \tag{2.126}$$

$$\nu_{\mathrm{MCS}} = \arg \min_{1 \leq l \leq K} \left\{ T^l_{\mathrm{MCS}} \right\}. \tag{2.127}$$

It can be notified from (2.125) that the CUSUM statistic $g_k (l, j) = \max_{1 \leq i \leq k} S^k_i (l, j)$ can be calculated recursively as

$$g_k (l, j) = \left( g_{k-1} (l, j) + s_k (l, j) \right)^+, \; 1 \leq l \leq K, 0 \leq j \neq l \leq K, \tag{2.128}$$

where $(x)^+ = \max(x, 0)$, $s_k (l, j) = \log \left[ f_{\theta_l} (y_k) / f_{\theta_j} (y_k) \right]$ and initial condition $g_0 (l, j) = 0$, for all $1 \leq l \leq K$ and $0 \leq j \neq l \leq K$. The recursive matrix CUSUM procedure can be described as

$$T_{\mathrm{MCS}} = \inf \left\{ k \geq 1 : \max_{1 \leq l \leq K} \min_{0 \leq j \neq l \leq K} g_k (l, j) \geq h \right\}, \tag{2.129}$$

$$\nu_{\mathrm{MCS}} = \arg \max_{1 \leq l \leq K} \left\{ \min_{0 \leq j \neq l \leq K} g_{T_{\mathrm{MCS}}} (l, j) \right\}. \tag{2.130}$$

It has been shown in [138, 175] that the matrix CUSUM procedure (2.129)–(2.130) is fully recursive and its statistical properties asymptotically coincides with the generalized CUSUM algorithm (2.121)–(2.123). Another version of the matrix CUSUM procedure with different thresholds can be found in [175, 181].

**Vector CUSUM procedure.** As discussed in [129, 132, 175, 181], both the generalized CUSUM algorithm and the matrix CUSUM procedure depend heavily on the mutual geometry of the hypotheses. Sometimes, the probability of false isolation increases significantly when the change time $k_0 \to \infty$ due to an uncontrolled growth of some cumulative

sums under the pre-change hypothesis $\mathcal{H}_0$ (see [175, pages 507-508] for detailed explanation). In order to circumvent this difficulty, Nikiforov [129, 132] suggested to replace the statistic $\max_{1\leq i\leq k} S_i^k(l,j)$, which may be stochastically large under $\mathcal{H}_0$ for some $l,j$, by the statistic $\max_{1\leq i\leq k} S_i^k(l,0) - \max_{1\leq i\leq k} S_i^k(j,0)$, which is stochastically small under $\mathcal{H}_0$ for all $1\leq l,j\leq K$, leading to the following recursive vector CUSUM procedure $\delta_{\mathrm{VCS}} = (T_{\mathrm{VCS}}, \nu_{\mathrm{VCS}})$:

$$T_{\mathrm{VCS}} = \inf\left\{k\geq 1 : \max_{1\leq l\leq K}\min_{0\leq j\neq l\leq K}\left[g_k(l,0) - g_k(j,0) - h_{lj}\right]\geq 0\right\}, \tag{2.131}$$

$$\nu_{\mathrm{VCS}} = \arg\max_{1\leq l\leq K}\left\{\min_{0\leq j\neq l\leq K}\left[g_{T_{\mathrm{VCS}}}(l,0) - g_{T_{\mathrm{VCS}}}(j,0) - h_{lj}\right]\right\}. \tag{2.132}$$

where the function $g_k(l,0)$ is defined in a recursive manner as

$$g_k(l,0) = \left(g_{k-1}(l,0) + s_k(l,0)\right)^+, \; 1\leq l\leq K, \tag{2.133}$$

with initial condition $g_0(l,0) = 0 \quad 1\leq l\leq K$ and $g_0(0,0) = 0$. The thresholds $h_{l,j}$ are chosen in the following way

$$h_{l,j} = \begin{cases} h_{\mathrm{fa}} & \text{if} \quad 1\leq l\leq K \quad \text{and} \quad j = 0 \\ h_{\mathrm{fi}} & \text{if} \quad 1\leq l,j\leq K \quad \text{and} \quad j\neq l \end{cases}, \tag{2.134}$$

where $h_{\mathrm{fa}}$ and $h_{\mathrm{fi}}$ stand for the detection and isolation thresholds, respectively.

The statistical properties of the vector CUSUM procedure $\delta_{\mathrm{VCS}} = (T_{\mathrm{VCS}}, \nu_{\mathrm{VCS}})$ have been investigated in [129, 132] with respect to the optimality criterion (2.115)–(2.116).

**Theorem 2.22.** *(Asymptotic optimality of vector CUSUM procedure [129, 132, 175]). Consider the vector CUSUM procedure $\delta_{\mathrm{VCS}} = (T_{\mathrm{VCS}}, \nu_{\mathrm{VCS}})$ given in (2.131)–(2.134). Suppose that $0 < \rho_{lj} < \infty$ for all $0\leq l\neq j\leq K$ and the following regularity condition is fulfilled: the moment-generating function $\varphi(\varsigma) = \mathbb{E}_l\left[e^{\varsigma s_k(l,j)}\right] < \infty$ exists for all real number $\varsigma\in(-\eta,\eta)$, where $\eta>0$, and for all $1\leq l\leq K$ and $0\leq j\neq l\leq K$. Let $h_{lj}$ be given by (2.134) and $h_{\mathrm{fa}} > h_{\mathrm{fi}}$. Let also $\gamma\to\infty$, $\beta\to 0$ and $\log(\gamma)\geq\log\left(\beta^{-1}\right)(1+o(1))$. If the thresholds are selected as $h_{\mathrm{fa}}\sim\log(\gamma)$ as $\gamma\to\infty$ and $h_{\mathrm{fi}}\sim\log\left(\beta^{-1}\right)$ as $\beta\to 0$, then*

$$\mathbb{E}_0\left[T_{\mathrm{VCS}}\right]\geq\gamma, \quad \max_{1\leq l\leq K}\max_{1\leq j\neq l\leq K}\sup_{k_0\geq 1}\mathbb{P}_{k_0}^l\left(\nu_{\mathrm{VCS}} = j | T_{\mathrm{VCS}}\geq k_0\right)\leq\beta\left(1+o(1)\right), \tag{2.135}$$

*and*

$$\tilde{\tau}^*\left(\delta_{\mathrm{VCS}}\right)\leq\max\left(\frac{\log(\gamma)}{\rho_{\mathrm{fa}}^*}, \frac{\log\left(\beta^{-1}\right)}{\rho_{\mathrm{fi}}^*}\right)(1+o(1)). \tag{2.136}$$

It follows from the Theorem 2.22 that the vector CUSUM procedure $\delta_{\mathrm{VCS}} = (T_{\mathrm{VCS}}, \nu_{\mathrm{VCS}})$ is asymptotically optimal in the class $C_{\gamma\beta}$ defined in (2.116).

**Window Limited vector CUSUM procedure.** Pursuing the asymptotic theory for the detection problem [103, 107], Lai [104] has generalized the results of Nikiforov in [130] for the non-i.i.d. case under the convergence condition on the log-likelihood ratio. He proposed also the following window limited vector CUSUM procedure $\delta_{\mathrm{VWL}} = (T_{\mathrm{VWL}}, \nu_{\mathrm{VWL}})$, where the stopping time $T_{\mathrm{VWL}}$ and the final decision $\nu_{\mathrm{VWL}}$ are given by

$$T_{\mathrm{VWL}} = \inf\left\{k\geq 1 : \max_{1\leq l\leq K}\min_{0\leq j\neq l\leq K}\left[\max_{k-m_\alpha+1\leq i\leq k} S_i^k(l,0) - \max_{k-m_\alpha+1\leq i\leq k} S_i^k(j,0)\right]\geq h\right\}, \tag{2.137}$$

$$\nu_{\mathrm{VWL}} = \arg\max_{1\leq l\leq K}\left\{\max_{T_{\mathrm{VWL}}-m_\alpha+1\leq i\leq T_{\mathrm{VWL}}} S_i^{T_{\mathrm{VWL}}}(l,0)\right\}, \tag{2.138}$$

where $h$ is a chosen threshold. The statistical properties of the window limited vector CUSUM procedure have been investigated in [104], whose i.i.d. case is shown in the following theorem.

**Theorem 2.23.** *(Asymptotic optimality of the window limited vector CUSUM procedure [104]). Consider the window limited vector CUSUM procedure* $\delta_{\mathrm{VWL}} = (T_{\mathrm{VWL}}, \nu_{\mathrm{VWL}})$ *given in* (2.137)–(2.138). *Suppose that* $m_\alpha = O(|\log(\alpha)|)$ *and* $h \sim |\log(\alpha)|$ *as* $\alpha \to \infty$. *Especially, if the threshold* $h$ *is such chosen that* $2Ke^{-h} = \alpha$, *then*

$$\overline{\mathbb{P}}_{\mathrm{fa}}(\delta_{\mathrm{VWL}}) = \sup_{l \geq 1} \mathbb{P}_0(l \leq T_{\mathrm{VWL}} < l + m_\alpha) \leq \alpha m_\alpha, \tag{2.139}$$

$$\overline{\mathbb{P}}_{\mathrm{fi}}(\delta_{\mathrm{VWL}}) = \max_{1 \leq l \leq K} \sup_{k_0 \geq 0} \mathbb{P}^l_{k_0}(k_0 \leq T_{\mathrm{VWL}} < k_0 + m_\alpha, \nu \neq l) \leq \frac{K-1}{K} \alpha m_\alpha, \tag{2.140}$$

*and for every* $1 \leq l \leq K$, *as* $\alpha \to 0$

$$\mathbb{E}^l_{k_0}\left[(T_{\mathrm{VWL}} - k_0 + 1)^+\right] \leq \frac{\mathbb{P}_0(T \geq k_0)\,|\log(\alpha)|}{\min_{0 \leq j \neq l \leq K} \rho_{lj} + o(1)} \text{ uniformly in } k_0 \geq 1, \tag{2.141}$$

*thus proving that the window limited vector CUSUM procedure* $\delta_{\mathrm{VWL}} = (T_{\mathrm{VWL}}, \nu_{\mathrm{VWL}})$ *is asymptotically optimal in the class* $C_{m_\alpha}$ *defined in* (2.119) *in the sense that it minimizes the worst-case conditional mean delay for detection-isolation defined in* (2.115).

It follows from (2.120) and (2.141) that the window limited vector CUSUM procedure (2.137)–(2.138) is asymptotically optimal in the sense that it minimizes the average delay for detection-isolation $\mathbb{E}^l_{k_0}\left[(T_{\mathrm{VWL}} - k_0 + 1)^+\right]$, for all $1 \leq l \leq K$, uniformly in $k_0 \geq 1$ over all stopping times $T$ in the class $C_{m_\alpha}$ defined in (2.119).

### 2.4.4 Conclusion

Several criteria and optimal procedures for the sequential change-point detection-isolation problem have been reviewed in this section. In the quickest change detection problem, the criteria of optimality are to minimize the risk associated with the detection delay for a given value on the false alarm rate. For the joint detection-isolation problem, it is proposed to minimize also the risk connected to the delay for detection-isolation subject to the false alarm and false isolation rates.

The abrupt change detection-isolation problem posits that the post-change duration is infinitely long and that the detection probability is unity once the change has occurred. In practice, however, there exist certain situations where the post-change duration is short, including the detection of a "burst" acoustic signature or a "pulse" in radio astronomy signals, the passive underwater surveillance or the on-line monitoring of SCADA systems against cyber-physical attacks. The problem of detecting transient signals will be considered in the following section.

## 2.5 Sequential Detection of Transient Changes

In previous section, we have presented different results on the classical quickest change detection-isolation problem which deals with an abrupt change of infinitely long duration in distribution of a stochastic process. The objective of this section is to introduce recent results on the transient change detection problem. Several transient detectors with respect to different optimality criteria will be discussed.

### 2.5.1 Introduction

The classical quickest change detection-isolation methods are extremely suitable to the on-line surveillance of technological processes against abnormal behaviors of infinitely long duration. The criteria of optimality should be favorable of small mean delay for the detection/isolation subject to acceptable levels on the false alarm and false isolation. The transient change detection problem, on the other hand, is interested in the reliable detection of transient signals. The transient change detection problem can be broadly classified into two types [67, 69]: short-duration signals and safety-critical applications.

#### Transient changes involving short-duration signals

In practice, there exist a large number of applications where the input data contains, in addition to random noises, suddenly arriving signals of short period, including radar and sonar [2], non-destructive testing [164], a "burst" in acoustic signatures [29], a "pulse" in radio astronomy signals [57], the monitoring of water quality in distribution networks [70], or the surveillance of SCADA systems against cyber-physical attacks [80]. In such applications, the transient signals should be detected before their disappearance. The following statistical model is often used for describing short-duration changes in a stochastic system [29, 69, 75, 173, 199]:

$$y_k \sim \begin{cases} P_{\theta_0} & \text{if} \quad k < k_0 \\ P_{\theta_1} & \text{if} \quad k_0 \leq k < k_0 + L \\ P_{\theta_0} & \text{if} \quad k \geq k_0 + L \end{cases}, \tag{2.142}$$

where $P_{\theta_0}$ denotes the distribution of the observations $y_k$ under the pre-change mode and the post-change mode, $P_{\theta_1}$ stands for the distribution of the observations $y_k$ under the transient change mode, $k_0$ is the unknown change-point and $L$ is the transient change duration.

Unlike the traditional quickest change detection problem where the change duration is infinitely long (corresponding to $L \to \infty$), three scenarios may occur in the case of short-duration signals (see figure 2.9) as:

- *False alarm*: The change is detected before its occurrence (i.e., $T < k_0$). The false alarm rate can be evaluated by either the ARL to false alarm [113] or the probability of false alarm within any time window of predefined length [102, 103, 176].
- *Timely (correct) detection*: The change is detected within the transient change period (i.e., $k_0 \leq T \leq k_0 + L - 1$). Generally, the timely (correct) detection rate is measured by the probability of detection, i.e., the probability of detecting the change within the transient window $[k_0, k_0 + L - 1]$.
- *Missed (latent) detection*: The change is detected after its disappearance (i.e., $k_0 + L \leq T < \infty$) or the change is never revealed (i.e., $T \to \infty$). The missed detection rate should be evaluated by the probability of missed detection. The authors in [75] have considered the latent detection (i.e., $k_0 + L \leq T < \infty$) as legitimate detection. In our opinion, the latent detection should be considered as the missed detection since the change has already terminated.

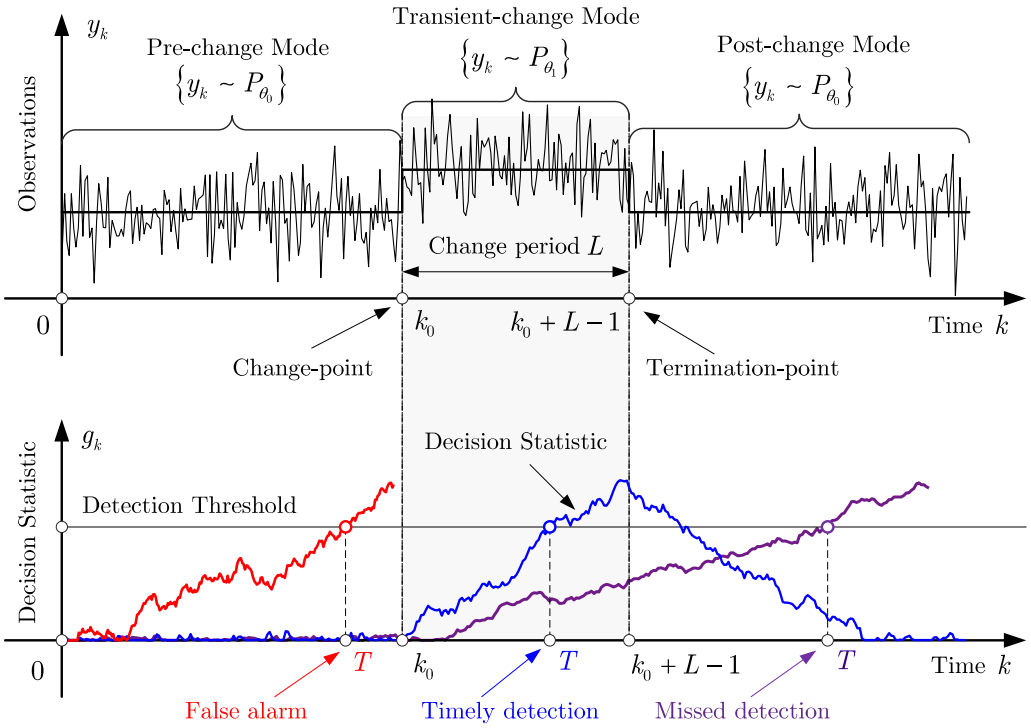



Figure 2.9 – Transient change detection problem for short-duration signals.

In the quickest change detection problem, the change should be detected with the probability of 1 since the change duration is assumed to be infinitely long. Hence, the detection delay is the only quantity of interest for evaluating the detection of the change [75]. The criteria of optimality should be favorable of small detection delay subject to an acceptable level of false alarm. In contrast, the transient change detection problem posits that the change duration is finite and short, leading to the fact that the probability of detection of the change may be smaller than 1. The probability of detection and the probability of missed detection are, therefore, two quantities of interest when dealing with short-duration changes. The criteria of optimality should be favorable of high probability of detection (or small probability of missed detection) subject to an acceptable level of false alarm.

### Transient changes involving safety-critical applications

The second type of transient change detection problem involves safety-critical applications such as the integrity monitoring of GPS systems [9], the quality monitoring of water supply [67,69,70], or the surveillance of SCADA systems against cyber-physical attacks [6,7,141]. For the security of such safety-critical infrastructures, the maximum permitted detection delay is often limited by a prescribed value $L$ even if the changes are of infinitely long duration [9,69,123]. In other words, a predefined hard limit $L$ is imposed on the detection delay. This value $L$ can be calculated from the gravity of the change (i.e., the magnitude of the change) and the permitted consequence of the change. As it can be seen from figure 2.10, following scenarios may occur:

- *False alarm*: Similar to the case of short-duration signals, the false alarm is any declaration that takes place before its occurrence (i.e., $T < k_0$). For safety-critical applications, the

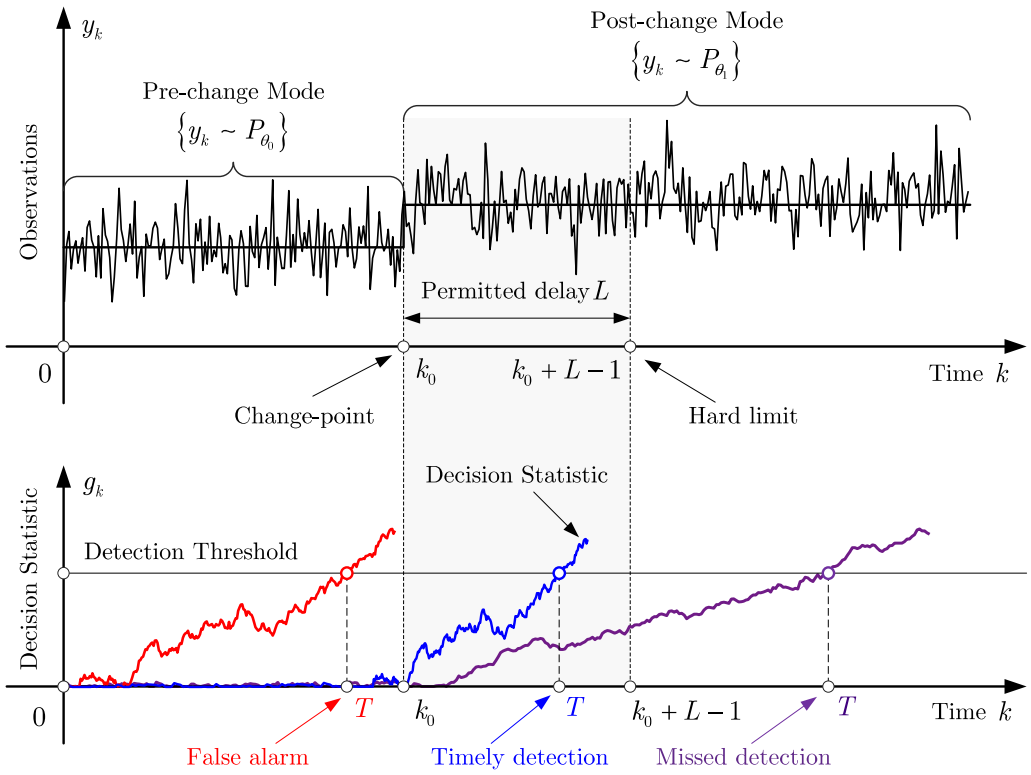



Figure 2.10 – Transient change detection problem for safety-critical applications.

false alarm rate should be measured by the probability of false alarm within any time window of predefined length since this criterion is more stringent than the ARL to false alarm constraint [103, 176].

- *Timely detection*: Since a hard limit $L$ is imposed on the detection delay, the change is said to be correctly detected only if the alarm is raised within the predefined window of size $L$ right after the change (i.e., $k_0 \leq T \leq k_0 + L - 1$). The probability of detection is, therefore, an appropriate performance index for measuring the detection rate.

- *Missed detection*: Any declaration of the change with the detection delay greater than a prescribed value $L$ is considered as missed (i.e., $T \geq k_0 + L$). The missed detection rate is generally evaluated by the probability of missed detection.

As it has been discussed in [9, 69, 123], the drawback of the classical quickest change detection criterion lies in the existence of the right "tail" in the distribution of the detection delay. Roughly speaking, a small average detection delay does not guarantee that the probability of having the detection delay greater than a required time-to-alert $L$ (i.e., the probability of missed detection) is negligible. Moreover, the declaration of the change with detection delay greater than $L$ is undesirable, especially for safety-critical applications, since the latent detection would cause catastrophic damage to the systems. In contrast, in the case of timely detection (i.e., $k_0 \leq T \leq k_0 + L - 1$), the true detection delay $T - k_0 + 1$ is always smaller than or equal to the required time-to-alert $L$. In such cases, the true detection delay has no significance since the impact of the change on the system is negligible. For these reasons, the risk associated with the detection

of the change should be evaluated by either the probability of detection or the probability of missed detection in stead of classical performance indexes involving the mean detection delay.

### Discussion

Following from above analysis, there exists a fundamental difference between two aforementioned types of transient change detection problems. The first type deals with short-duration signals while the second type involves safety-critical applications even if the changes are of infinitely long duration. However, the optimality criteria for both types should favor high probability of detection or small probability of missed detection subject to an acceptable level on the false alarm rate.

In practice, there are several applications comprising of both types of transient changes, including the monitoring of water quality against malevolent activities [67,70] or the surveillance of SCADA systems against cyber-physical attacks [7, 80]. Let us take an example of malicious attacks on SCADA systems. On one hand, the cyber-physical attacks on SCADA systems can be modeled as additive signals of short-duration on both system equations, as it has been discussed in chapter 1. On the other hand, the SCADA systems involve a large number of safety-critical infrastructures such as electric power grids, gas pipelines or water networks. For these reasons, the security of SCADA systems against cyber-physical attacks addressed in this manuscript perfectly fits into the transient change detection framework.

Let us discuss now the attack duration. Let $\overline{L}$ be the true attack duration and $L$ be the required time-to-alert designed by system operators. The putative value (i.e., designed value) $L$ is known *a priori* but the true value $\overline{L}$ is generally unknown. Let us consider three following scenarios. Firstly, two types of transient changes perfectly coincide if the true value is equal to the putative value (i.e., $\overline{L} = L$). Secondly, if the true attack duration is greater than the required time-to-alert (i.e., $\overline{L} > L$), any detection of attack with a delay greater than $L$ is undesirable. Such an alarm should be considered as missed since it may cause catastrophic damage to safety-critical infrastructures. Hence, any detection of attack with detection delay greater than $L$ is considered as missed even if the true detection is greater the required time-to-alert $L$. Thirdly, if the true attack duration is smaller than the putative value (i.e., $\overline{L} < L$), it is desirable to detect the attack before its termination, of course. However, the detection of the change after its termination and before the hard limit $L$ (i.e., $\overline{L} < T \leq L$) is still acceptable. The known hard limit $L$ can be used in place of the unknown true value $\overline{L}$ in the case $\overline{L} < L$. In summary, the attack duration could be assumed to be known (i.e., being equal to the hard limit $L$) and any detection of the change with detection delay greater than the prescribed value $L$ is considered as missed.

Staring from now, we denote by $\mathbb{P}$ the probability measure. Let $\mathcal{P}_0 \triangleq \mathcal{P}_\infty$ (resp. $\mathbb{P}_0 \triangleq \mathbb{P}_\infty$ and $\mathbb{E}_0 \triangleq \mathbb{P}_0$) be the joint distribution (resp. probability measure and mathematical expectation) of the observations $y_1, y_2, \cdots, y_{k_0}, \cdots$ when the observations $y_1, y_2, \cdots$ follow the pre-change mode (i.e., $y_k \sim P_0$ for all $k \geq 1$). Let $\mathcal{P}_{k_0}$ (resp. $\mathbb{P}_{k_0}$ and $\mathbb{E}_{k_0}$) denote the joint distribution (resp. probability measure and mathematical expectation) of the observations $y_1, y_2, \cdots, y_{k_0}, \cdots$ when the observation $y_k$ follows the observation model (2.142). Different criteria of optimality for the transient change detection problem will be investigated in the following sub-section.

### 2.5.2 Criteria of optimality

In contrast to the classical quickest change detection problem, the optimality criteria for the transient change detection problem should be of high probability of detection or small probability of missed detection subject to an acceptable level of false alarm. The false alarm rate can be measured by either the ARL to false alarm or the probability of false alarm within any time window of predefined length. In the literature, various criteria of optimality under both Bayesian and non-Bayesian approaches have been proposed for comparing different transient change detection procedures.

#### Bayesian approach

The Bayesian approach considers the change-point $k_0$ as an unknown and random variable following some known *a priori* distribution $Q$. The *a priori* distribution $Q$ is often chosen as either the geometric distribution $Q(p)$ or the zero-modified geometric distribution $Q(\pi, p)$. The geometric distribution $Q(p)$ has the form: $\mathbb{P}(k_0 = k) = p(1-p)^{k-1}$ for any $k \geq 1$ with the parameter $p \in (0, 1]$. On the other hand, the zero-modified geometric distribution $Q(\pi, p)$ has the form: $\mathbb{P}(k_0 \leq 0) = \pi$ and $\mathbb{P}(k_0 = k) = (1-\pi)\, p(1-p)^{k-1}$ for any $k \geq 1$ with the parameters $\pi \in [0, 1]$ and $p \in (0, 1]$. It can be seen clearly that the geometric distribution $Q(p)$ is a special case of the zero-modified geometric distribution $Q(\pi, p)$ with $\pi = 0$.

The first optimality criterion under the Bayesian setting was found in [18] where the author suggested to maximize the probability of detection $\mathbb{P}(|T - k_0 + 1| \leq L)$. By imposing the *a priori* geometric distribution $Q(p)$ on the change-point $k_0$, the Bayesian optimization problem was solved for the case of independent and identically distributed (i.i.d.) observations under simple hypotheses. The optimal solution to the problem, which was obtained for any $L = 1, 2, \cdots$, turned out to be the simple Shewhart control chart [165]. The probability maximizing idea was utilized also in [145, 161]. For example, Sarnowski and Szajnowski [161] extended the results in [18] to the case of dependent observations generated from Markov processes. In addition, Pollak and Krieger [145] considered the i.i.d. observations but the post-change parameter $\theta$ was assumed to follow some *a priori* known parametric family of distributions $G(\theta)$. The Bayesian problem of maximizing $\mathbb{P}(|T - k_0 + 1| \leq L)$ was solved in [145] for the special case $L = 1$. It is worth noting that, in the aforementioned work [18, 144, 161] under the Bayesian framework, the authors did not attempt to control the false alarm rate in any sense, as discussed in Moustakides [123].

In an attempt to control the false alarm rate, the authors in [150] suggested to maximize the probability of detection $\mathbb{P}(k_0 \leq T \leq k_0 + L - 1)$ subject to the the probability of false alarm $\mathbb{P}(T < k_0) \leq \alpha$, where $\alpha \in (0, 1)$ is the prescribed value. Following the same probability maximizing approach, Moustakides [123] studied multiple optimality properties of the Shewhart control chart [165] with respect to different criteria of optimality, under both Bayesian approach and non-Bayesian approach. Under the Bayesian setting, Moustakides [123] imposed the zero-modified *a priori* distribution $Q(\pi, p)$ on the change-point $k_0$ and suggested two new criteria of optimality.

The first criterion proposed by Moustakides [123] involves the maximization of the following conditional probability of detection:

$$\sup_{T \in C_\alpha} \left\{ \mathbb{P}_{\mathrm{d}}^{\mathrm{M}}(T; L) = \mathbb{P}(k_0 \leq T \leq k_0 + L - 1 | T \geq k_0) \right\}, \tag{2.143}$$

over all stopping times $T \in C_\alpha$ in the class $C_\alpha = \{T : \mathbb{P}(T < k_0) \leq \alpha\}$, where $\alpha \in (0,1)$ is a prescribed value on the false alarm rate and $\mathbb{P}_{\mathrm{d}}^{\mathrm{M}}(T;L)$ denotes the conditional probability of detection which depends on the *a priori* distribution $Q(\pi,p)$. It has been discussed in [123] that one of the main drawbacks of the Bayesian approach is the requirement to properly specify the distribution $Q(\pi,p)$ which depends heavily on the parameters $\pi$ and $p$. For this reason, Moustakides [123] suggested an alternative criterion to (2.143), which is independent from the distribution $Q(\pi,p)$. The second criterion proposed by Moustakides [123] consists of maximizing the following worst-case conditional probability of detection:

$$\sup_{T \in C_\gamma} \left\{ \overline{\mathbb{P}}_{\mathrm{d}}^{\mathrm{M}}(T;L) = \inf_{Q(\pi,p)} \mathbb{P}(k_0 \leq T \leq k_0 + L - 1 \,|\, T \geq k_0) \right\}, \tag{2.144}$$

over all stopping times $T \in C_\gamma$ in the class $C_\gamma = \{T : \mathbb{E}_0[T] \geq \gamma\}$, where $\gamma \geq 1$ is an acceptable level on the ARL to false alarm and $\overline{\mathbb{P}}_{\mathrm{d}}^{\mathrm{M}}(T;L)$ stands for the worst-case conditional probability of detection over all *a priori* distributions $Q(\pi,p)$. The problem was solved in [123] for the case $L = 1$. It was shown in [123] that the optimum detection procedures w.r.t. both aforementioned optimality criteria turned out to be the modified Shewhart control chart [165].

### Non-Bayesian approach

Several criteria of optimality have been proposed under the non-Bayesian framework in which the change-point $k_0$ is assumed to be unknown but non-random. The optimality criteria often involve the maximization of the (worst-case, conditional) probability of detection or the minimization of the (worst-case, conditional) probability of missed detection subject to an acceptable level on the false alarm rate.

By modifying the optimality criteria suggested by Lorden [113] and Pollak [143], initially proposed for the classical quickest change detection problem, Moustakides [123] introduced two new performance indexes for measuring the risk associated with the detection of the change. The first criterion of optimality, obtained by modifying the Lorden's criterion [113], consists in maximizing the following worst-worst-case conditional probability of detection:

$$\sup_{T \in C_\gamma} \left\{ \overline{\mathbb{P}}_{\mathrm{d}}^{\mathrm{M}_1}(T;L) = \inf_{k_0 \geq 1} \operatorname{ess\,inf} \mathbb{P}_{k_0}(k_0 \leq T \leq k_0 + L - 1 \,|\, y_1, y_2, \cdots, y_{k_0-1}, T \geq k_0) \right\} \tag{2.145}$$

among all stopping times $T \in C_\gamma$ in the class $C_\gamma = \{T : \mathbb{E}_0[T] \geq \gamma\}$, where $\gamma \geq 1$ is a prescribed value on the ARL to false alarm and $\overline{\mathbb{P}}_{\mathrm{d}}^{\mathrm{M}_1}(T;L)$ denotes the worst-worst-case conditional probability of detection proposed by Moustakides in [123].

The second criterion of optimality, obtained by modifying the Pollak's criterion [143], involves the maximization of the following worst-case conditional probability of detection:

$$\sup_{T \in C_\gamma} \left\{ \overline{\mathbb{P}}_{\mathrm{d}}^{\mathrm{M}_2}(T;L) = \inf_{k_0 \geq 1} \mathbb{P}_{k_0}(k_0 \leq T \leq k_0 + L - 1 | \, T \geq k_0) \right\} \tag{2.146}$$

among all stopping times $T$ satisfying $\mathbb{E}_0[T] \geq \gamma$, where $\gamma \geq 1$ is a prescribed value on the ARL to false alarm and $\overline{\mathbb{P}}_{\mathrm{d}}^{\mathrm{M}_2}(T;L)$ stands for the worst-case conditional probability of detection proposed by Moustakides in [123].

Previously, the optimality criterion (2.146) was adopted by Pollak and Krieger [145] for the special case $L = 1$ under the semi-Bayesian setting where the change-point $k_0$ was supposed to

be unknown and deterministic but the post-change parameter $\theta$ was assumed to be a random variable following some known *a priori* parametric family of distributions $G(\theta)$. The criterion of optimality involves the maximization of the following worst-case conditional probability of detection (for $L = 1$):

$$\sup_{T \in C_\gamma} \left\{ \overline{\mathbb{P}}_{\mathrm{d}}^{\mathrm{P}}(T) = \inf_{k_0 \geq 1} \mathbb{P}_{k_0}(T = k_0 \,|T \geq k_0) \right\}, \tag{2.147}$$

among all stopping times $T \in C_\gamma$ in the class $C_\gamma = \{T : \mathbb{E}_0[T] \geq \gamma\}$, where $\gamma \geq 1$ is a prescribed value on the ARL to false alarm, and $\overline{\mathbb{P}}_{\mathrm{d}}^{\mathrm{P}}(T)$ stands for the worst-case conditional probability of detection suggested by Pollak and Krieger [145], implicitly depending on the distribution $G(\theta)$. It was shown in [145] that the optimal detection procedure was the generalized Shewhart control chart [165].

**Remark 2.2.** *The criterion* (2.147) *suggested by Pollak and Krieger [145] was a special case of the criterion* (2.146) *proposed by Moustakides [123] for the case* $L = 1$. *Though the optimality criteria* (2.145)–(2.146) *were written for any* $L = 1, 2, \cdots$, *Moustakides was able to solve the problem only for the special case* $L = 1$, *which coincided with the work of Pollak and Krieger [145]. The optimal stopping times for both max-min criteria turned out to be the generalized Shewhart control chart [165].*

Under the non-Bayesian framework, the probability minimizing approach has been also considered in [9, 69]. The optimality criteria involved the minimization of the worst-case (conditional) probability of missed detection subject to an acceptable level on the worst-case probability of false alarm within any time window of predefined length. The first probability minimizing idea was proposed by Bakhache and Nikiforov in [9], where the authors suggested to minimize the following worst-case (non-conditional) probability of missed detection:

$$\inf_{T \in C_\alpha^{\mathrm{B}}} \left\{ \overline{\mathbb{P}}_{\mathrm{md}}^{\mathrm{B}}(T; L) = \sup_{k_0 \geq 1} \mathbb{P}_{k_0}(T - k_0 + 1 > L) \right\} \tag{2.148}$$

among all stopping times $T \in C_\alpha$ satisfying

$$C_\alpha^{\mathrm{B}} = \left\{ T : \overline{\mathbb{P}}_{\mathrm{fa}}^{\mathrm{B}}(T; m_\alpha) = \sup_{l \geq 1} \mathbb{P}_0(l \leq T < l + m_\alpha) \leq \alpha \right\}, \tag{2.149}$$

where $\overline{\mathbb{P}}_{\mathrm{md}}^{\mathrm{B}}(T; L)$ denotes the worst-case probability of missed detection and $\overline{\mathbb{P}}_{\mathrm{fa}}^{\mathrm{B}}(T; m_\alpha)$ stands for the worst-case probability of false alarm within any time window of length $m_\alpha$ and $\alpha \in (0, 1)$ is a prescribed value on the false alarm rate.

It has been discussed in [69] that, for safety-critical applications, the worst-case probability of missed detection $\sup_{k_0 \geq 1} \mathbb{P}_{k_0}(T - k_0 + 1 > L)$ should be replaced by the worst-case conditional probability of missed detection $\sup_{k_0 \geq 1} \mathbb{P}_{k_0}(T - k_0 + 1 > L | T \geq k_0)$. Under the assumption that the change does not occur during the "preheating" period (i.e., $k_0 \geq L$), Guépié *et al* [69] suggested to minimize the following worst-case conditional probability of missed detection:

$$\inf_{T \in C_\alpha^{\mathrm{G}}} \left\{ \overline{\mathbb{P}}_{\mathrm{md}}^{\mathrm{G}}(T; L) = \sup_{k_0 \geq L} \mathbb{P}_{k_0}(T - k_0 + 1 > L | T \geq k_0) \right\} \tag{2.150}$$

among all stopping times $T \in C_\alpha^{\mathrm{G}}$ satisfying

$$C_\alpha^{\mathrm{G}} = \left\{ T : \overline{\mathbb{P}}_{\mathrm{fa}}^{\mathrm{G}}(T; m_\alpha) = \sup_{l \geq L} \mathbb{P}_0(l \leq T < l + m_\alpha) \leq \alpha \right\}, \tag{2.151}$$

where $\overline{\mathbb{P}}_{\mathrm{md}}^{\mathrm{G}}(T; L)$ denotes the worst-case conditional probability of missed detection and $\overline{\mathbb{P}}_{\mathrm{fa}}^{\mathrm{G}}(T; m_\alpha)$ stands for the worst-case probability of false alarm within any time window of length $m_\alpha$ and $\alpha \in (0, 1)$ is a prescribed value on the false alarm rate. It should be noted that the window size $m_\alpha$ and the false alarm rate $\alpha$, independent from each other, are decided by system operators.

### 2.5.3 Detection procedures

The objective of this subsection is to resume several detection procedures which can be used for detecting transient changes in the statistical model (2.142). We focus only on the non-Bayesian setting where the change-point $k_0$ is unknown but non-random. For both academic and practical purposes, we consider two scenarios: known transient change parameters and unknown transient change parameters. It should be noted that the *a priori* information about the change plays an extremely important role in the design of detection procedures. In other words, the more *a priori* information about the parameters we have, the better the detection procedures would be designed.

#### Known transient change parameters

The assumption on the known transient parameters, including the shape of the change, the magnitude of the change and the duration of the change, is mainly applicable for the academic purpose. The only unknown parameter is the change-point $k_0$. Under these assumptions, optimal and/or suboptimal detection procedures w.r.t. certain optimality criteria may be obtained (see, for example, in [9, 67–70, 123, 145]).

It is well-known that, when dealing with an abrupt change of infinitely long duration, the CUSUM procedure proposed by Page [139] is optimal in the sense that it minimizes the worst-worst-case mean detection delay for a given value on the ARL to false alarm. Hence, it is reasonable to consider the Page's CUSUM test in detecting temporary signals (i.e., transient signals or signals of short duration). For example, the CUSUM procedure has been employed for detecting transient signals in radio astronomy [57] or transient changes in hidden Markov models [29].

In addition, Han *et al* in [75] have investigated the statistical performance of the CUSUM procedure applied to the detection of transient signals modeled in (2.142). The stopping time $T_{\mathrm{CS}}$ of the CUSUM procedure in the recursive form can be described as

$$T_{\mathrm{CS}} = \inf_{k \geq 1} \{ Z_k \geq h \}, \quad Z_k = \max\{0, Z_{k-1} + \log[f_{\theta_1}(y_k) / f_{\theta_0}(y_k)]\}, \quad Z_0 = 0, \tag{2.152}$$

where $Z_k$ is the CUSUM statistic. The authors in [75] have considered three scenarios for the probability of detection, including standard detection, initial-point detection and latent detection.

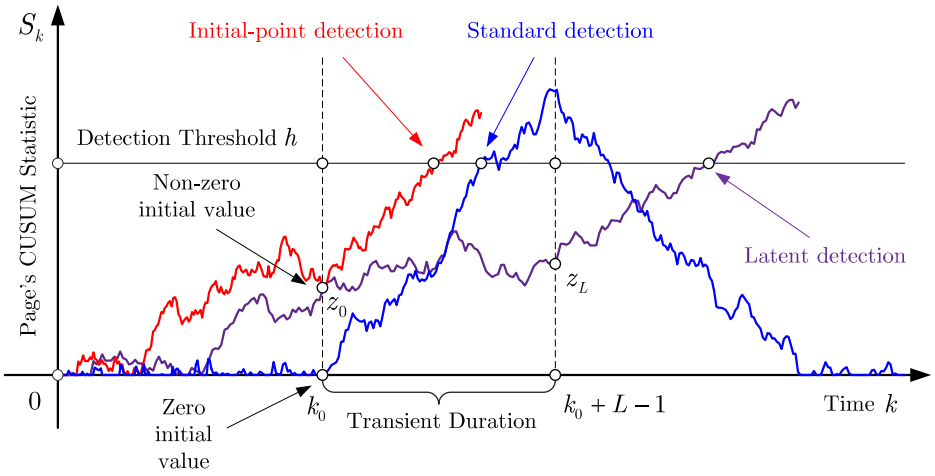



Figure 2.11 – Three scenarios of detection with the Page's CUSUM procedure (from [75]).

- *Standard detection*: The CUSUM statistic $Z_k$ is zero at the change's onset (i.e., $Z_{k_0} = 0$) and the threshold is crossed before the change has disappeared. The probability of detection is described as

$$\mathbb{P}_{\mathrm{d}}^{\mathrm{std}}\left(T_{\mathrm{CS}}; L\right) = \mathbb{P}\left(0 \leq T_{\mathrm{CS}} - k_0 + 1 \leq L \,|Z_{k_0} = 0\right). \tag{2.153}$$

- *Initial-point detection*: The CUSUM statistic $Z_k$ is non-zero when the change starts (i.e., $Z_{k_0} = z_0 \neq 0$) and the threshold is crossed before the change has terminated. The probability of detection is approximated as

$$\mathbb{P}_{\mathrm{d}}^{\mathrm{init}}\left(T_{\mathrm{CS}}; L\right) = \int_{z_0} \mathbb{P}\left(0 \leq T_{\mathrm{CS}} - k_0 + 1 \leq L \,|Z_{k_0} = z_0\right) dF_{ss}\left(z_0\right), \tag{2.154}$$

  where $dF_{ss}\left(z_0\right)$ is the probability distribution of $z_0$ at the change's onset.

- *Latent detection*: The CUSUM statistic $Z_k$ is non-zero at the change's onset and the threshold is crossed after the disappearance of the change. Taking into account the latent detection, the probability of detection is written as

$$\begin{aligned}\mathbb{P}_{\mathrm{d}}^{\mathrm{lat}}\left(T_{\mathrm{CS}}; L\right) = &\int_{z_0} \mathbb{P}\left(0 \leq T_{\mathrm{CS}} - k_0 + 1 \leq L \,|Z_{k_0} = z_0\right) dF_{ss}\left(z_0\right) + \\ &\int_{z_0}\int_{z_L} \mathbb{P}\left(\text{decide } \mathcal{H}_1; h - z_L; -z_L\right) dF\left(z_L \,|T_{\mathrm{CS}} - k_0 + 1 > L, Z_{k_0} = z_0\right) dF_{ss}\left(z_0\right),\end{aligned} \tag{2.155}$$

  where $z_L$ is the value of the CUSUM statistic $Z_k$ at the time instant $k = k_0 + L - 1$, $\mathbb{P}\left(\text{decide } \mathcal{H}_1; h - z_L; -z_L\right)$ denotes the probability of crossing the upper threshold in a standard sequential test with upper and lower thresholds, respectively, $h - z_L$ and $-z_L$, and $dF\left(z_L \,|T_{\mathrm{CS}} - k_0 + 1 > L, Z_{k_0} = z_0\right)$ refers to the probability function of the CUSUM statistic $S_k$ at the end of the transient signal (i.e., $k = k_0 + L - 1$) accounting for both the non-initial value $z_0$ and under the condition that the detection is raised after the termination of the change (i.e., $T_{\mathrm{CS}} \geq k_0 + L$).

It has been discussed in [75] that the latent detection is of legitimate interest. The relationship between three types of detection is, therefore, described as

$$\mathbb{P}_{\mathrm{d}}^{\mathrm{std}}\left(T_{\mathrm{CS}}; L\right) \leq \mathbb{P}_{\mathrm{d}}^{\mathrm{init}}\left(T_{\mathrm{CS}}; L\right) \leq \mathbb{P}_{\mathrm{d}}^{\mathrm{lat}}\left(T_{\mathrm{CS}}; L\right). \tag{2.156}$$

Several methods (three analytical and two numerical) have been proposed for approximating the probability of detection $\mathbb{P}_{\mathrm{d}}^{\mathrm{lat}}\left(T_{\mathrm{CS}}; L\right)$. Three analytical methods include the ternary quantization method, the continuous-time moment matching method and the Brownian motion method. Two numerical methods are the matrix approach and the fast Fourier transform approach. Interested readers are referred to [75] for more details.

In addition, the Window Limited (WL) CUSUM procedure, initially proposed by Willsky and Jones [205], has been shown by Lai [102,103,107] to be an asymptotically optimal detection rule. In order to render the WL CUSUM procedure more flexible, Guépié [67] developed the so-called Variable Threshold Window Limited (VTWL) CUSUM algorithm for the detection of transient signals. Under the assumption that the change does not occur during the "preheating" period (i.e., $k_0 \geq L$), the VTWL CUSUM algorithm can be described as

$$T_{\mathrm{VTWL}} = \inf\left\{k \geq L : \max_{k-L+1 \leq i \leq k}\left(S_i^k - h_{k-i+1}\right) \geq 0\right\}, \tag{2.157}$$

where $S_i^k$ is the log-likelihood ratio (LLR) and the thresholds $h_1, h_2, \cdots, h_L$ are considered as tuning parameters for optimizing the VTWL CUSUM algorithm w.r.t. the transient change detection criterion (2.150)–(2.151).

Consider the following Gaussian independent observation model:

$$y_k \sim \begin{cases} \mathcal{N}\left(0, \sigma^2\right) & \text{if} \quad k < k_0 \\ \mathcal{N}\left(\theta_1, \sigma^2\right) & \text{if} \quad k_0 \leq k < k_0 + L \\ \mathcal{N}\left(0, \sigma^2\right) & \text{if} \quad k \geq k_0 + L \end{cases}, \tag{2.158}$$

where the change-point $k_0$ is unknown but the change duration $L$ is assumed to be known. The parameters of the Gaussian distribution $\theta_1$ and $\sigma$ are completely known. The VTWL CUSUM algorithm is expressed in the Gaussian case as

$$T_{\mathrm{VTWL}} = \inf\left\{k \geq L : \max_{k-L+1 \leq i \leq k}\left(S_i^k - h_{k-i+1}\right) \geq 0\right\}, \quad S_i^k = \sum_{t=i}^{k} \frac{\theta_1}{\sigma^2}\left(y_t - \frac{\theta_1}{2}\right). \tag{2.159}$$

The optimal choice of thresholds $h_1, h_2, \cdots, h_L$ of the VTWL CUSUM procedure (2.159) in the Gaussian model (2.158) w.r.t. the transient change detection criterion (2.150)–(2.151) has been addressed in [67]. It has been shown that the optimized VTWL CUSUM algorithm is equivalent to the following Finite Moving Average (FMA) detection rule:

$$T_{\mathrm{FMA}}\left(\tilde{h}_L\right) = \inf_{k \geq L}\left\{\sum_{t=k-L+1}^{k} y_t \geq \tilde{h}_L\right\}, \tag{2.160}$$

where the threshold $\tilde{h}_L$ is chosen for assuring an acceptable level of false alarm.

In addition, Guépié [67] addressed also the detection of transient signals following some known profiles with constant signs and he obtained similar results in such cases by utilizing the concept of the associated random variables [46]. It has been shown by simulation in [67, 68, 70] that

the FMA detection rule outperforms than the CUSUM algorithm w.r.t. the transient change detection criterion (2.150)–(2.151).

Recently, Moustakides [123] has obtained an exact optimal solution w.r.t. a different transient change detection criterion. Though the criteria of optimality (2.145)–(2.146) suggested in [123] were written for any $L \geq 1$, Moustakides was able to find the optimal solution for the particular case $L = 1$ only. In the case of i.i.d. observations before and after the change with corresponding densities $f_{\theta_0}$ and $f_{\theta_1}$ and for $L = 1$, Moustakides proved that the following simple Shewhart control chart:

$$T_{\mathrm{SH}} = \inf \left\{ k \geq 1 : \log \left[ f_{\theta_1}(y_k) / f_{\theta_0}(y_k) \right] \geq h \right\}, \tag{2.161}$$

where $h$ is a chosen threshold, minimizes both the worst-worst-case conditional probability of detection (2.145) and the worst-case conditional probability of detection (2.146) among all stopping times satisfying the ARL to false alarm constraint. It is worth noting that the Shewhart test (2.161) coincides with the repeated Neyman-Pearson test applied to one observation at each time instant $k \geq 1$. A more general result was obtained previously by Pollak and Krieger [145] under the semi-Bayesian setting where the change-point $k_0$ is unknown and non-random but the transient change parameter $\theta$ follows a known *a priori* distribution $G(\theta)$. The optimal results in [123, 145], which were obtained for the special case $L = 1$, have very limited practical application. It can be applied only to "loud-and-short" transient changes.

### Unknown transient change parameters

In practice, there are a large number of applications involving unknown transient change parameters. In other words, hypotheses on the transient signals are highly composite w.r.t. the duration of the change, the shape of the change and the magnitude of the change. In such circumstances, it is desirable to design detection procedures offering robust performance with minimum information about the transient change parameters. From a literature review, transient detectors for unknown parameters can be classified into four main categories: CUSUM-based algorithms, generalized likelihood ratio (GLR-based) detectors on the basis of some preliminary transformations, min-max detectors, and transient detectors based on power-law statistics in the frequency domain.

The CUSUM-based algorithms have been employed extensively for dealing with the the transient changes of unknown parameters (i.e., location, length, strength and form). For example, the CUSUM procedure has been shown in [200] to perform relatively well regardless of various forms of the transient signals. However, the CUSUM test has been shown [199] to be quite sensitive to the transient length. In other words, the CUSUM procedure designed for long-and-quite transients would perform badly for short-and-loud signals and vice versa [199]. The robustness of the CUSUM procedure with respect to the transient length has been improved by using time-varying thresholds [198, 199, 203, 204]. The stopping time $T_{\mathrm{VTP}}$ of the so-called Variable Threshold Page (VTP) test can be described as

$$T_{\mathrm{VTP}} = \inf_{k \geq 1} \{ Z_k \geq h_k \}, \quad Z_k = \max \left\{ 0, Z_{k-1} + \log \left[ f_{\theta_1}(y_k) / f_{\theta_0}(y_k) \right] \right\}, \quad Z_0 = 0, \tag{2.162}$$

where the thresholds $h_k$ are tuned for assuring an acceptable level of false alarm. The design and implementation of the VTP algorithm for the case of Gaussian shift-in-variance have been elaborated in [198, 199]. In addition, it has been shown by simulation that the VTP test offers competitive performance w.r.t. several transient detectors found in literature.

The integration between the GLR structure and a class of linear transformations has been considered in [59,148] as an alternative solution to CUSUM-based procedures for detecting short-duration signals. This approach has been utilized for comparing different GLR-based transient detectors on the basis of several linear time-frequency analysis techniques, including the short-time Fourier transform [59], the Gabor representation [58] and the wavelet transform [2, 60]. Borrowing from [59, 148], the signal model is described as

$$y = C\theta + e + \xi, \tag{2.163}$$

where $y \in \mathbb{R}^p$ is the vector of observations, $C \in \mathbb{R}^{p \times n}$ is the observation matrix, $\theta \in \mathbb{R}^n$ is the signal descriptor, $e \in \mathbb{R}^p$ stands for the signal mismatch, and $\xi \in \mathbb{R}^p$ denotes the random noises. Let $W \in \mathbb{R}^{m \times p}$ be the matrix of orthonormal rows. The signal model after a linear transform is written as

$$z = Wy = WC\theta + We + W\xi, \tag{2.164}$$

where the signal descriptor $\theta$ is assumed to be zero under $\mathcal{H}_0$ and non-zero under $\mathcal{H}_1$. The transient detector designed for the ideal model (i.e., there is no model mismatch or $e = 0$) raises an alarm if the GLR statistic

$$T_{\mathrm{GLR}}(y) = y^T W^T WC \left(C^T W^T WC\right)^{-1} C^T W^T Wy \tag{2.165}$$

is greater than a threshold $h$ which is normally chosen for assuring an acceptable level of false alarm. Interested readers are referred to [59, 148] for more details.

In addition, the authors in [173] have proposed the *hyperparameter* approach for detecting unknown transient signals, where the unknown parameters are assumed to follow some known *a priori* distribution with unknown parameters. Let $y$ be an observation vector following a distribution $F_\theta(y)$ depending on the parameter $\theta$. The detection problem consists in deciding hypothesis $\mathcal{H}_0 = \{\theta \in \Theta_0\}$ against hypothesis $\mathcal{H}_1 = \{\theta = \Theta_1\}$, where $\Theta_0 \cap \Theta_1 = \emptyset$. Since the hypotheses are composite, the *a priori* distribution $G(\theta)$ is imposed on the parameter $\theta$ where the distribution $G_\theta(z)$ is known but its parameter is unknown. For these reasons, the authors in [173] suggested to jointly estimate the parameter $\theta$ and the parameters of $G_\theta(z)$ via the estimation-maximization (EM) algorithm. The transient detector [173] raises an alarm if the following statistic

$$T_{\mathrm{EM}} = \frac{\max_{\theta \in \Theta_1} \left\{\int_z dF(y|z)\, dG_\theta(z)\right\}}{\max_{\theta \in \Theta_0} \left\{\int_z dF(y|z)\, dG_\theta(z)\right\}}$$

is crossing a threshold. Interested readers are referred to [173] for detailed implementation of the EM algorithm.

In [74], Han *et al* have developed the min-max detector for detecting transient signals. This min-max idea was initiated by Baygun and Hero [12] in the statistical hypothesis testing framework. The criterion of optimality involves the minimization of the maximum probability of missed detection subject to an acceptable level on the probability of false alarm. The minimization is over all tests and the maximization is over all possible alternatives (i.e., change-point and change duration). Let $N$ be the number of observations and $L$ be the minimum value of change duration. It has been shown that, when the number of observations $N$ goes to infinity, the min-max detector raises an alarm once the following statistic

$$T_{\mathrm{MM}} = \sum_{i=1}^{N-L+1} \prod_{j=i}^{i+L-1} \frac{f_{\theta_1}(y_j)}{f_{\theta_0}(y_j)},$$

where $f_{\theta_0}(y_j)$ and $f_{\theta_1}(y_j)$ are the p.d.f. of the observation $y_j$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively, is greater than the threshold $h$. The comparison between the min-max detector and the Page's CUSUM test has been also performed for the case of Gaussian shift-in-mean transient. The min-max test outperforms the CUSUM test for the worst-case scenario while the CUSUM test offers better performance in some others.

The so-called power-law statistics proposed by Nuttall [134, 135], operating on the basis of the magnitude-squared discrete Fourier transform (DFT) bins, have been shown to be simple, effective and reliable detectors when dealing with transient signals with unknown structure, location, length and strength. Especially, when some *a priori* information about the transient signals, i.e., transient length, is available, the "maximum" detector proposed by Nuttall [136] has been shown to perform extremely well compared to other detectors [200]. The drawback of the power-law detector lies in that its data must be pre-normalized and spectrally white, as has been discussed in [202]. In order to circumvent this difficulty, Wang and Willett [201, 202] proposed several novel power-law detectors in both frequency and wavelet domains. These detectors can be considered as all-purpose and plug-in solutions for detecting transient signals since they offer exceptional performance, are easy to implement and require minimal information about transient signals.

### 2.5.4 Conclusion

In this section, we have resumed recent results on the transient change detection problem which consists of two types: short-duration signals and safety-critical applications. For both types of transient changes, the criteria of optimality should be favorable of maximizing the probability of detection or minimizing the probability of missed detection subject to an acceptable level of false alarm. Taking into account the transient change detection problem, various detection procedures have been proposed for detecting transient signals, for both academic and practical purposes.

It has been discussed in [199] that if the information about the transient changes (i.e., structure, length and strength) is available, that information should be exploited. Such essential information may exist in several (though quite limited) practical scenarios. In addition, the assumption on the known transient change parameters enables to establish theoretical results. Optimal and/or suboptimal procedures w.r.t. several transient change detection criteria have been obtained. In the literature, exactly optimal results have been derived in [123, 145] for the special case $L = 1$. Unfortunately, the case $L = 1$ has a very limited practical application. For a more general case $L \geq 1$, suboptimal results have been obtained in [67, 69].

It is of practical interest to design detection rules capable of detecting transient signals regardless of their structure, location, length and strength [199]. However, existing methods for unknown transient change parameters, including the min-max detector, GLR-based detectors on the basis of preliminary transformations and transient detectors based on power-law statistics, are mainly applicable to finite observation intervals, i.e., to *a posteriori* transient change detection. The only exclusions include CUSUM-based detection procedures [29, 57, 75, 198, 199] where infinite observations are processed in the real time.

## 2.6 Conclusion

In this chapter, we have discussed contemporary results on the statistical decision theory, including the classical (non-sequential) hypothesis testing problem, the sequential hypothesis testing problem, the sequential change-point detection and isolation problem and the sequential detection of transient signals. Non-sequential methods utilize a fixed number of observations for designing statistical tests between two (or more) hypotheses. This fixed-size sample approach is particularly suitable to off-line applications but not to on-line monitoring of safety-critical infrastructures. Sequential methods, on the other hand, seem to be more adequate for on-line monitoring applications. The sequential hypothesis testing theory allows us to design optimal (or suboptimal) procedures for deciding between two (or more) hypotheses while reducing the number observations compared to non-sequential detection rules. The sequential hypothesis testing techniques, however, appear inappropriate for the surveillance of safety-critical infrastructures. In such applications, it is assumed that the random observations are firstly generated by a common distribution $P_{\theta_0}$, corresponding to normal behavior of the systems, and then from an unknown change-point $k_0$, these random variables follow another common distribution $P_{\theta_1} \neq P_{\theta_0}$. The sequential change detection-isolation techniques are extremely suitable to the detection and identification of abrupt changes in stochastic systems.

The security of SCADA systems against cyber-physical attacks, involving both short-duration signals and safety-critical infrastructures, has been shown to perfectly fit into the transient change detection framework due to the inevitable effect of random noises. The existing methods working with finite observation intervals are not adequate for the on-line monitoring of SCADA systems since the decision has to be made in real-time. In addition, exactly optimal results obtained in [123, 145] for the case $L = 1$ have a very limited significance. For a more general case of $L \geq 1$, several suboptimal results have been introduced in [67, 69].

In his PhD thesis, Guépié [67] suggested to minimize the worst-case probability of missed detection for a given value on the worst-case probability of false alarm within any time window of predefined length. He designed also sub-optimal detection algorithms w.r.t. the transient change detection criterion. However, Guépié [67] was able to solve the problem for the independent Gaussian variables where transient profiles are of constant sign. The design and the study of the transient change detectors in the previous work depend heavily on the concept of associated random variables, see details and results in [46, 110]. It is questionable whether the results obtained in obtained in [67] remain valid for the dependent observations generated from the discrete-time state space model in the presence of unknown system states (nuisances) and random noises. Moreover, the calculation of the upper bound for the worst-case probability of false alarm depends heavily on the assumption that the transient profiles must be of constant sign. The question arises naturally is whether the results obtained in [67] hold when the sign of the transient profiles is not constant. Finally, Guépié [67] used the simple observation model which may not suitable to such applications as the monitoring of SCADA systems against cyber-physical attacks.

Pursuing the work started in [67], we consider in the following chapter the problem of detecting transient signals on stochastic-dynamical systems. Especially, the discrete-time state space model driven by Gaussian noises is employed to describe SCADA systems. Cyber-physical attacks are modeled as additive signals of short duration on both state evolution and sensor measurement equations. Moreover, the remaining problems after the work of Guépié [67] will be also treated in the next chapter.

# Chapter 3

# Sequential Detection of Transient Signals in Stochastic-dynamical Systems

**Contents**

## 3.1 Introduction

The security of SCADA systems against cyber-physical attacks has been investigated in chapter 1. Several approaches have been considered for protecting, detecting and isolating malicious activities on these large-scale industrial control systems. The majority of safety-critical infrastructures, including electric power grids, gas pipelines or water distribution or irrigation networks, can be described in the discrete-time state space model (see chapter 5 for more details). It has been discussed in chapter 1 that the attack detection and identification problem is closely related to the fault detection and isolation (FDI) problem in automatic control community. The statistical FDI problem is concerned with deciding whether the fault has occurred and then to identify the types of the fault with respect to (w.r.t.) random noises and unknown system states (often regarded as the nuisance parameter). This problem is generally solved by using the analytical redundancy approach, which is comprised of two steps: residual generation and residual evaluation. The residuals are first generated by using some techniques [30, 35, 83] such as the Kalman filter or the parity space to eliminate the negative impact of the nuisance parameter. Next, they are evaluated by utilizing the change detection techniques [10, 175] for circumventing the random noises.

The model-based fault diagnosis methods concentrate mainly on the generation of robust residuals which are decoupled from the model uncertainties (i.e., the disturbances). For example, the unknown input observer (UIO) techniques have been utilized in [4, 6, 7, 140, 141] for detecting and identifying cyber-physical attacks on SCADA systems. However, the negative impact of random noises on the decision-making process has not been considered seriously. On the other hand, the statistical decision theory, which has been excerpted in chapter 2, focuses mainly on the evaluation of random residuals based on relatively simple observation models. Optimal detection-isolation algorithms exist in only limited scenarios with a simple abstraction. The majority of work in this field is, therefore, dedicated to finding asymptotically optimal or sub-optimal detection-isolation algorithms w.r.t. a given criterion of optimality.

It has been discussed in chapter 1 and chapter 2 that the on-line monitoring of SCADA systems against cyber-physical attacks should be formulated as the sequential detection of transient signals in stochastic-dynamical systems. This chapter is organized follows. In section 3.2, we formulate the detection of cyber-physical attacks on SCADA systems as the problem of detecting transient signals in stochastic-dynamical systems. Traditional residual generation methods, including the steady-state Kalman filter approach and the fixed-size parity space approach, are presented in section 3.3. Several sub-optimal detection algorithms w.r.t. the transient change detection criterion for completely known transient change parameters and partially known transient change parameters are considered in section 3.4 and section 3.5, respectively. Finally, some concluding remarks are offered in section 3.6.

## 3.2 Transient Changes in Stochastic-Dynamical Systems

In this section, we formulate the detection of cyber-physical attacks on SCADA systems as the problem of detecting transient changes in stochastic-dynamical systems. The SCADA systems are described as discrete-time state space models driven by Gaussian noises. The cyber-physical attacks are modeled as additive signals of short duration on both state evolution and sensor measurement equations. The criterion of optimality for this problem, which was first introduced

in [67–70], is officially stated. This optimality criterion will be utilized through this chapter for designing sub-optimal detection procedures.

### 3.2.1 System and attack models

The following discrete-time state space model is utilized throughout this manuscript for describing SCADA systems under normal operation[8]:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + v_k \end{cases}; \quad x_1 = \overline{x}_1, \tag{3.1}$$

where $x_k \in \mathbb{R}^n$ is the vector of system states with unknown initial values $\overline{x}_1 \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$ is the vector of control signals, $d_k \in \mathbb{R}^q$ is the vector of disturbances, $y_k \in \mathbb{R}^p$ is the vector of sensor measurements, $w_k \in \mathbb{R}^n$ is the vector of process noises and $v_k \in \mathbb{R}^p$ is the vector of sensor noises; the matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $F \in \mathbb{R}^{n\times q}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, $G \in \mathbb{R}^{p\times q}$ are assumed to be completely known.

The process noises $w_k \sim \mathcal{N}(0, Q)$, where $Q \in \mathbb{R}^{n\times n}$, and the sensor noises $v_k \sim \mathcal{N}(0, R)$, where $R \in \mathbb{R}^{p\times p}$, are assumed to be independent identically distributed (i.i.d.) zero-mean Gaussian random vectors, i.e., $\text{cov}(w_k, w_l) = Q\delta_{kl}$, $\text{cov}(v_k, v_l) = R\delta_{kl}$ and $\text{cov}(w_k, v_l) = 0$, where $\delta_{kl} = 1$ if $k = l$ and $\delta_{kl} = 0$ otherwise. The noise covariance matrices $Q$ and $R$ are assumed to be exactly known and $R$ is positive-definite.

For simplicity, the control signals $u_k$ and the disturbances $d_k$ are assumed to be completely known. The control signals $u_k$ are known since they are the outputs of controllers. In many important applications such as electric power grids, gas pipelines or water distribution and irrigation networks, the disturbances $d_k$ correspond to customers' demands. In such applications, the demands are often estimated by specially-designed software with an acceptable level of error. Generally, these estimation errors are unbiased, so they can be integrated into the process noises $w_k$ and/or sensor noises $v_k$.

The system model under cyber-physical attacks can be described as follows:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + Ka^x_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + Ha^x_k + Ma^y_k + v_k \end{cases}; \quad x_1 = \overline{x}_1, \tag{3.2}$$

where $a^x_k \in \mathbb{R}^r$ is the state attack vector, $a^y_k \in \mathbb{R}^p$ is the sensor attack vector; the attack matrices $K \in \mathbb{R}^{n\times r}$, $H \in \mathbb{R}^{p\times r}$ and $M \in \mathbb{R}^{p\times p}$ are assumed to be known.

**Remark 3.1.** *The vector $a^x_k$ is denoted as the state attack vector since the component $Ka^x_k$ impacts the system dynamics directly. The component $Ha^x_k$ is due to feed-through effects from the state attack vector to the sensor measurements. The vector $a^y_k$ is called the sensor attack vector since the component $Ma^y_k$ impacts the sensor measurements directly. The attack vectors $a^x_k$ and $a^y_k$ are designed by the attacker for realizing his malicious target while the attack matrices $K$, $H$, and $M$ are decided by system operators.*

**Remark 3.2.** *It has been shown that the attack vectors $a^x_k$ and $a^y_k$ could be coordinated to disrupt the systems while remaining stealthy to traditional anomaly detectors [141]. The stealthiness of an attack depends heavily on the model knowledge, the disclosure resources and the disruption*

[8] From this point, the SCADA systems are assumed to start operating at time instant $k = 1$.

*capabilities [187]. Being equipped with perfect model knowledge and necessary resources, powerful attackers could design undetectable attacks by the replay attack strategy [120], the false data injection attack strategy [121], the zero-dynamics attack strategy [186] or the covert attack strategy [169]. To render those stealthy attacks detectable, the security analysis process is required. For example, more secure sensors can be sited in vulnerable points of the systems, making the stealthy attacks detectable (see, for example, [121], [186] or [99]). For these reasons, only detectable attacks are considered in this manuscript.*

**Example 3.1.** *The covert attack strategy* (1.9) *introduced in [169] is based on the coordination of cyber attacks on control signals and sensor measurements only. This attack strategy can be generalized to the cyber-physical attack scenarios as follows:*

- The attack vector $a_k^x$ on control signals can be chosen arbitrarily based on the target and available disruption resources of the attacker.
- The attack vector $a_k^y$ on sensor measurements is calculated by the following equation:

$$\begin{cases} x_{a,k+1} &= Ax_{a,k} + Ka_k^x \\ a_k^y &= -Cx_{a,k} - Ha_k^x \end{cases}; \quad \{x_{a,k}\}_{k\leq k_0} = 0, \tag{3.3}$$

  where $x_{a,k}$ is the vector of "attacked states" reflecting the difference between the system states under normal operation and those under attack.

It should be noted that the difference between the covert attack model (3.3) for cyber-physical attacks and the covert attack model (1.9) for cyber attacks on control signals and sensor measurements lies in the attack vectors $a_k^x$ (i.e., in (3.3)) and $a_k^u$ (i.e., in (1.9)). The covert attack model (3.3) will be utilized throughout this manuscript from this point.

### 3.2.2 Model of transient signals

Let the state attack vector $a_k^x$ and the sensor attack vector $a_k^y$ be grouped into the attack vector $a_k = \left[(a_k^x)^T, (a_k^y)^T\right]^T \in \mathbb{R}^s$, where $s = r + p$. Also, let $B_a = [K, 0] \in \mathbb{R}^{n\times s}$ and $D_a = [H, M] \in \mathbb{R}^{p\times s}$ be attack matrices. The attack components $B_a a_k = Ka_k^x$ and $D_a a_k = Ha_k^x + Ma_k^y$, leading to the following simplified model of SCADA systems under cyber-physical attacks:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + B_a a_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + D_a a_k + v_k \end{cases}; \quad x_1 = \overline{x}_1. \tag{3.4}$$

Let us suppose that the adversary performs his malicious attack during a short period $\tau_a = [k_0, k_0 + L - 1]$, where $k_0$ is the attack instant (unknown) and $L$ is the attack period (assumed to be known). The attack vector $a_k$ is then described by

$$a_k = \begin{cases} 0 & \text{if } k < k_0 \\ \theta_{k-k_0+1} & \text{if } k_0 \leq k < k_0 + L \\ 0 & \text{if } k \geq k_0 + L \end{cases}, \tag{3.5}$$

where $\theta_1, \theta_2, \cdots, \theta_L \in \mathbb{R}^s$ are the attack profiles defined in $\tau_a$. Sometimes, the attack profiles $\theta_1, \theta_2, \cdots, \theta_L$ are denoted as the attack signatures.

**Remark 3.3.** *The a priori information about the attack signatures $\theta_1, \theta_2, \cdots, \theta_L$ is extremely important in designing detection procedures. For the monitoring of safety-critical infrastructures against cyber-physical attacks, this critical information could be obtained via the security analysis process. For example, it is possible to figure out which attack scenarios may occur to the system by investigating the system's vulnerabilities. Since each attack scenario leads to a particular signature (i.e., a specific profile), the "shape" of the attack could calculated from the dynamics of the system. Sometimes, the magnitude of the profile is also available in particular situations. Let us consider a simple SCADA water distribution network described in figure 5.5 in chapter 5, where a pump is utilized for supplying water to a reservoir. It is assumed that the water network is equipped with a constant speed pump which operates in two modes: "on" and "off". It is assumed that the attacker performs his malicious attack for switching the pump "off" while it is functioning (see [213] for a real attack on a water utility). In such an attack scenario, the attack profiles $\theta_1, \theta_2, \cdots, \theta_L$ are completely specified.*

**Remark 3.4.** *In this thesis, we consider two scenarios: the attack profiles are completely known (i.e., in section 3.4) and the attack profiles are partially known (i.e., in section 3.5). The first scenario involves complete information about the attack signatures, including both shape and magnitude. This assumption is important in evaluating the best theoretically achievable performance of detection procedures. In the second scenario, it is assumed that the shape of the attack profiles is known but their magnitude is unknown. It is clear that the second scenario is more practical than the first one. However, theoretical results obtained in such practical cases are often limited.*

### 3.2.3 Criterion of optimality

The detection algorithm consists of calculating the stopping time $T$ at which the attack is declared. Historically, the optimality criteria favor minimizing the risk associated with detection delay (e.g., the worst-worst-case detection delay [113] or the worst-case conditional detection delay [142]) subject to an acceptable level of false alarms, which could be measured by either the ARL to false alarm or the probability of false alarm within any time window of predefined length. It is our opinion that traditional optimality criteria are not adequate for the detection of cyber-physical attacks on SCADA systems due to following reasons.

Firstly, the adversary may prefer to perform his malicious attack within a short period due to limited capabilities (see, for example, [7, 25, 80]). This malevolent action leads to the transient change (i.e., the change of short duration) in sensor measurements. Therefore, it is preferable to detect the change before its disappearance since any detection of the signal after its disappearance makes no sense.

Secondly, in safety-critical applications, the permitted detection delay $L$ is often given by norms or standards. This value $L$ can be calculated from the gravity of the attack (i.e., the magnitude of the attack) and the permitted consequence of the attack. The detection of attack with the delay smaller than $L$ is considered to be negligible (i.e., no matter the detection delay is small or large) since its impact to the system is often small and limited (see [9] for an example about the navigation systems integrity monitoring). Any detection with the delay greater than or equal to the prescribed value $L$ is considered as a missed detection since its impact to the system is negative.

For these reasons, the criterion of optimality for the transient change detection problem, which was first introduced in [67, 69], will be used throughout this thesis. This optimality criterion

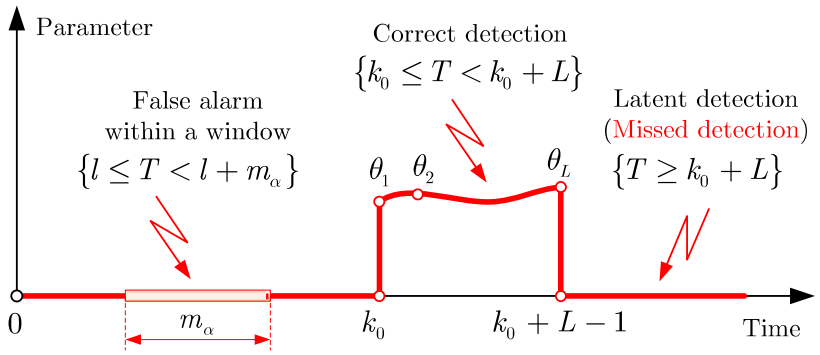



Figure 3.1 – Transient change detection criterion.

involves the minimization of the following worst-case probability of missed detection:

$$\inf_{T \in C_\alpha} \left\{ \overline{\mathbb{P}}_{\text{md}}(T; L) = \sup_{k_0 \ge L} \mathbb{P}_{k_0}(T - k_0 + 1 > L | T \ge k_0) \right\}, \tag{3.6}$$

among all stopping times $T \in C_\alpha$ satisfying

$$C_\alpha = \left\{ T : \overline{\mathbb{P}}_{\text{fa}}(T; m_\alpha) = \sup_{l \ge L} \mathbb{P}_0 \{ l \le T < l + m_\alpha \} \le \alpha \right\}, \tag{3.7}$$

where $\overline{\mathbb{P}}_{\text{md}}$ denotes the worst-case probability of missed detection and $\overline{\mathbb{P}}_{\text{fa}}$ stands for the worst-case probability of false alarm within any time window of length $m_\alpha$ (see figure 3.1).

## 3.3 Residual Generation Methods

In this section, we consider two compelling approaches for generating the residuals, including the steady-state Kalman filter method and the fixed-size parity space method. Specially, we integrate two residual models into the unified statistical model which will be used in designing detection procedures.

### 3.3.1 Steady-state Kalman filter-based residual generation

Let us assume that the steady-state Kalman filter is used for generating the sequence of innovations (i.e., or residuals). In practice, if the system is detectable [85], the Kalman filter converges very fast after several iterations. Consequently, the optimal Kalman gain $K_k$ converges also to its steady-state value $K_\infty$. The steady-state Kalman gain $K_\infty$ is calculated as

$$K_\infty = P_\infty C^T \left( C P_\infty C^T + R \right)^{-1}, \tag{3.8}$$

where $P_\infty$ denotes the steady-state covariance matrix of the state estimation error, which can be found by solving the following discrete-time algebraic Riccati equation:

$$P_\infty = A P_\infty A^T - A P_\infty C^T \left( C P_\infty C^T + R \right)^{-1} C P_\infty A^T + Q. \tag{3.9}$$

The operation of the steady-state Kalman filter is then described as

$$\begin{cases} \hat{x}_{k+1|k} &= A\hat{x}_{k|k-1} + Bu_k + Fd_k + AK_\infty \left(y_k - \hat{y}_{k|k-1}\right) \\ \hat{y}_{k|k-1} &= C\hat{x}_{k|k-1} + Du_k + Gd_k \end{cases}, \quad \hat{x}_{1|0} = \overline{x}_1, \tag{3.10}$$

where $\hat{x}_{k|k-1} \in \mathbb{R}^n$ is state estimate and $\hat{y}_{k|k-1} \in \mathbb{R}^p$ is the output estimate.

Let $r_k = y_k - \hat{y}_{k|k-1} \in \mathbb{R}^p$ be the vector of innovations. It has been shown [10, 116] (see also Appendix A.1) that the innovations $\{r_k\}_{k\geq 1}$ are independent Gaussian vectors with covariance matrix $J \triangleq CP_\infty C^T + R$. Under normal operation, these residual vectors $\{r_k\}_{k\geq 1}$ are independent identically distributed (i.i.d.) zero-mean Gaussian vectors, i.e., $r_k \sim \mathcal{N}(0, J)$. Under abnormal situations (i.e., faults or attacks occurring at an unknown time instant $k_0$), the innovations $\{r_k\}_{k\geq 1}$ are still independent Gaussian vectors but their means change from the baseline value (i.e., $\mathbb{E}_0[r_k] = 0$ for $k < k_0$) to the non-zero profiles (i.e., $\mathbb{E}_{k_0}[r_k] = \psi_{k-k_0+1}$ for $k \geq k_0$), where $\mathbb{E}_0[r_k]$ and $\mathbb{E}_{k_0}[r_k]$ are expectations of the residual vector $r_k$ under normal operation (i.e., $k_0 \to \infty$) and abnormal behavior from time instant $k_0$, respectively, and the change profiles $\psi_1, \psi_2, \cdots \in \mathbb{R}^p$ can be calculated from the system dynamics.

Let $\varrho_1, \varrho_2, \cdots \in \mathbb{R}^p$ be a sequence of i.i.d. random vectors satisfying a zero-mean multivariate Gaussian distribution satisfying $\varrho_k \sim \mathcal{N}(0, J)$. The statistical model of the innovations can be expressed by

$$r_k = \begin{cases} \varrho_k & \text{if} \quad k < k_0 \\ \psi_{k-k_0+1} + \varrho_k & \text{if} \quad k_0 \leq k < k_0 + L \\ \tilde{\psi}_k + \varrho_k & \text{if} \quad k \geq k_0 + L \end{cases}, \tag{3.11}$$

where $\psi_1, \psi_2, \cdots, \psi_L$ are the transient change profiles, being calculated from the attack profiles $\theta_1, \theta_2, \cdots, \theta_L$ by the following equation:

$$\begin{cases} \epsilon_{k+1} &= (A - AK_\infty C)\,\epsilon_k + (B_a - AK_\infty D_a)\,\theta_k \\ \psi_k &= C\epsilon_k + D_a\theta_k \end{cases}; \quad \epsilon_1 = 0, \tag{3.12}$$

and the post-change profiles $\tilde{\psi}_k$ (i.e., for $k \geq k_0 + L$) are of no interest. Interested readers are referred to [10, 107] or Appendix A.1 for more details on the calculation of innovation signatures.

Let $r^k_{k-L+1} = \left[r^T_{k-L+1}, \cdots, r^T_k\right]^T \in \mathbb{R}^{Lp}$ be the concatenated vector of residuals, $\varrho^k_{k-L+1} = \left[\varrho^T_{k-L+1}, \cdots, \varrho^T_k\right]^T \in \mathbb{R}^{Lp}$ be the concatenated vector of random noises, and $\psi^k_{k-L+1}(k_0) \in \mathbb{R}^{Lp}$ be the vector of transient signals depending on the relative position of the change-point $k_0$ within the time window $[k-L+1, k]$ by the following relation:

$$\psi^k_{k-L+1}(k_0) = \begin{cases} [0] & \text{if} \quad k < k_0 \\ \begin{bmatrix} [0] \\ \psi_1 \\ \vdots \\ \psi_{k-k_0+1} \end{bmatrix} & \text{if} \quad k_0 \leq k < k_0 + L \\ \left[\tilde{\psi}^k_{k-L+1}(k_0)\right] & \text{if} \quad k \geq k_0 + L \end{cases}, \tag{3.13}$$

where $[0]$ is the null vector of appropriate dimension and the vector of post-change profiles $\tilde{\psi}^k_{k-L+1}(k_0) \in \mathbb{R}^{Lp}$ is of no interest. Putting together (3.11)–(3.13), the statistical model of the

innovation vector $r_{k-L+1}^{k}$ generated by the steady-state Kalman filter is described as

$$r_{k-L+1}^{k} = \psi_{k-L+1}^{k}(k_0) + \varrho_{k-L+1}^{k}, \tag{3.14}$$

where the random noises $\varrho_{k-L+1}^{k} \sim \mathcal{N}(0, \Sigma_{\varrho})$, where $\Sigma_{\varrho} = \mathrm{diag}(J) \in \mathbb{R}^{Lp \times Lp}$ is a block-diagonal matrix formed of blocks $J$.

**Remark 3.5.** *In his PhD thesis [67], Guépié has addressed the problem of detecting transient changes of constant sign in a sequence of independent Gaussian random variables (i.e., the scalar case). Theoretical results obtained in [67] can be generalized to the vector case without any difficulty if each component constituting the vector of profiles is of constant sign. However, the transient profiles $\psi_1, \psi_2, \cdots, \psi_L$ generated from the steady-state Kalman filter, in general, do not satisfy such a condition. For example, the arguments utilized by Guépié [67] for obtaining sub-optimal detection procedures are inapplicable here.*

### 3.3.2 Fixed-size parity space-based residual generation

In this subsection, we develop the statistical model of the residuals generated by the fixed-size parity space. Suppose that the attack does not occur during the "preheating" period (i.e., $k_0 \geq L$) and that our algorithms operate from time instant $k \geq L$. By utilizing the last $L$ observations, for each time instant $k \geq L$, the observation model is described as

$$\begin{aligned}
\underbrace{\begin{bmatrix} y_{k-L+1} \\ y_{k-L+2} \\ \vdots \\ y_k \end{bmatrix}}_{y_{k-L+1}^{k}} &= \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{L-1} \end{bmatrix}}_{\mathcal{C}} x_{k-L+1} + \underbrace{\begin{bmatrix} D_a & 0 & \cdots & 0 \\ CB_a & D_a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2}B_a & CA^{L-3}B_a & \cdots & D_a \end{bmatrix}}_{\mathcal{M}} \underbrace{\begin{bmatrix} a_{k-L+1} \\ a_{k-L+2} \\ \vdots \\ a_k \end{bmatrix}}_{\theta_{k-L+1}^{k}(k_0)} + \\
&\underbrace{\begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2}B & CA^{L-3}B & \cdots & D \end{bmatrix}}_{\mathcal{D}} \underbrace{\begin{bmatrix} u_{k-L+1} \\ u_{k-L+2} \\ \vdots \\ u_k \end{bmatrix}}_{u_{k-L+1}^{k}} + \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ C & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2} & CA^{L-3} & \cdots & 0 \end{bmatrix}}_{\mathcal{H}} \underbrace{\begin{bmatrix} w_{k-L+1} \\ w_{k-L+2} \\ \vdots \\ w_k \end{bmatrix}}_{w_{k-L+1}^{k}} + \\
&\underbrace{\begin{bmatrix} G & 0 & \cdots & 0 \\ CF & G & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2}F & CA^{L-3}F & \cdots & G \end{bmatrix}}_{\mathcal{G}} \underbrace{\begin{bmatrix} d_{k-L+1} \\ d_{k-L+2} \\ \vdots \\ d_k \end{bmatrix}}_{d_{k-L+1}^{k}} + \underbrace{\begin{bmatrix} v_{k-L+1} \\ v_{k-L+2} \\ \vdots \\ v_k \end{bmatrix}}_{v_{k-L+1}^{k}},
\end{aligned} \tag{3.15}$$

or in a simpler form as

$$y_{k-L+1}^{k} = \mathcal{C}x_{k-L+1} + \mathcal{D}u_{k-L+1}^{k} + \mathcal{G}d_{k-L+1}^{k} + \mathcal{M}\theta_{k-L+1}^{k}(k_0) + \mathcal{H}w_{k-L+1}^{k} + v_{k-L+1}^{k}, \tag{3.16}$$

where $y_{k-L+1}^{k} \in \mathbb{R}^{Lp}$ is the concatenated vector of measurements, $u_{k-L+1}^{k} \in \mathbb{R}^{Lm}$ is the concatenated vector of control signals, $d_{k-L+1}^{k} \in \mathbb{R}^{Lq}$ is the concatenated vector of disturbances, $w_{k-L+1}^{k} \in \mathbb{R}^{Ln}$ is the concatenated vector of process noises, $v_{k-L+1}^{k} \in \mathbb{R}^{Lp}$ is the concatenated

vector of sensor noises, $\theta^k_{k-L+1}(k_0) \in \mathbb{R}^{Ls}$ is the concatenated vector of transient signals; the matrices $\mathcal{C} \in \mathbb{R}^{Lp\times n}$, $\mathcal{D} \in \mathbb{R}^{Lp\times Lm}$, $\mathcal{G} \in \mathbb{R}^{Lp\times Lq}$, $\mathcal{H} \in \mathbb{R}^{Lp\times Ln}$ and $\mathcal{M} \in \mathbb{R}^{Lp\times Ls}$. The process noises $w^k_{k-L+1} \sim \mathcal{N}(0, \mathcal{Q})$ and the sensor noises $v^k_{k-L+1} \sim \mathcal{N}(0, \mathcal{R})$, where $\mathcal{Q} = \mathrm{diag}(Q) \in \mathbb{R}^{Ln\times Ln}$ and $\mathcal{R} = \mathrm{diag}(R) \in \mathbb{R}^{Lp\times Lp}$ are block-diagonal matrices formed of blocks $Q$ and $R$, respectively. Let also $\eta^k_{k-L+1} = \mathcal{H}w^k_{k-L+1} + v^k_{k-L+1}$ be a concatenated vector of random noises, integrating both process noises and sensor noises. It is clear that $\eta^k_{k-L+1} \sim \mathcal{N}(0, \mathcal{S})$, where the covariance matrix $\mathcal{S} = \mathcal{H}\mathcal{Q}\mathcal{H}^T + \mathcal{R} \in \mathbb{R}^{Lp\times Lp}$ is symmetric and positive-definite.

The concatenated vector of attack profiles $\theta^k_{k-L+1}(k_0)$, depending on the relative position of the change-point $k_0$ within the window $[k-L+1, k]$, is described as

$$\theta^k_{k-L+1}(k_0) = \begin{cases} [0] & \text{if} \quad k < k_0 \\ \begin{bmatrix} [0] \\ \theta_1 \\ \vdots \\ \theta_{k-k_0+1} \end{bmatrix} & \text{if} \quad k_0 \le k < k_0 + L \\ \left[\tilde{\theta}^k_{k-L+1}(k_0)\right] & \text{if} \quad k \ge k_0 + L \end{cases}, \tag{3.17}$$

where $[0]$ is a null vector of appropriate dimension and the post-change profiles $\tilde{\theta}^k_{k-L+1}(k_0) \in \mathbb{R}^{Ls}$ are of no interest.

Since the vector of control signals $u_k$ and the vector of disturbances $d_k$ are assumed to be exactly known, they can be eliminated by subtraction from the observation model (3.15)–(3.16), leading to the following statistical model:

$$z^k_{k-L+1} = y^k_{k-L+1} - \left(\mathcal{D}u^k_{k-L+1} + \mathcal{G}d^k_{k-L+1}\right) = \mathcal{C}x_{k-L+1} + \mathcal{M}\theta^k_{k-L+1}(k_0) + \eta^k_{k-L+1}, \tag{3.18}$$

where $z^k_{k-L+1} \in \mathbb{R}^{Lp}$ is the simplified observation vector.

It is worth noting that the nuisance parameter $x_{k-L+1}$ has to be eliminated from (3.18) in order to avoid its negative impact on detection algorithms. The rejection of the nuisance parameter has been discussed in [52] by applying the invariant hypothesis testing theory. Specially, the method used in [52] coincides with the parity space approach in the fault diagnosis community. The main idea is as follows. The simplified observation vector $z^k_{k-L+1}$ is projected onto the orthogonal complement space $R(\mathcal{C})^\perp$ of the column space $R(\mathcal{C})$ of matrix $\mathcal{C}$ (i.e., the left-null space of matrix $\mathcal{C}$), which is assumed to be full column rank (i.e., $\mathrm{rank}(\mathcal{C}) = n$). The residual vector is calculated as $r^k_{k-L+1} = \mathcal{W}z^k_{k-L+1}$, where the rows of the matrix $\mathcal{W} \in \mathbb{R}^{(Lp-n)\times Lp}$ are composed of the eigenvectors of the projection matrix $\mathcal{P}^\perp_{\mathcal{C}} = \mathcal{I} - \mathcal{C}\left(\mathcal{C}^T\mathcal{C}\right)^{-1}\mathcal{C}^T$ corresponding to eigenvalue 1, where $\mathcal{I}$ is the identity matrix of appropriate dimension. The rejection matrix $\mathcal{W}$ satisfies the following conditions: $\mathcal{W}\mathcal{C} = 0$, $\mathcal{W}^T\mathcal{W} = \mathcal{P}^\perp_{\mathcal{C}}$ and $\mathcal{W}\mathcal{W}^T = \mathcal{I}$. Hence, the residual vector $r^k_{k-L+1}$ is independent from the nuisance vector $x_{k-L+1}$. The statistical model of the residuals generated by the fixed-size parity space is expressed by

$$r^k_{k-L+1} = \mathcal{W}z^k_{k-L+1} = \mathcal{W}\mathcal{M}\theta^k_{k-L+1}(k_0) + \mathcal{W}\eta^k_{k-L+1}. \tag{3.19}$$

In order to develop a statistical model similar to (3.14), let us define, respectively, the vector of transient profiles $\varphi^k_{k-L+1}(k_0) = \mathcal{W}\mathcal{M}\theta^k_{k-L+1}(k_0) \in \mathbb{R}^{Lp-n}$ and the vector of random noises $\varsigma^k_{k-L+1} = \mathcal{W}\eta^k_{k-L+1} \in \mathbb{R}^{(Lp-n)\times(Lp-n)}$. The statistical model of residual vector $r^k_{k-L+1}$ in (3.19) is then reduced to

$$r^k_{k-L+1} = \varphi^k_{k-L+1}(k_0) + \varsigma^k_{k-L+1}, \tag{3.20}$$

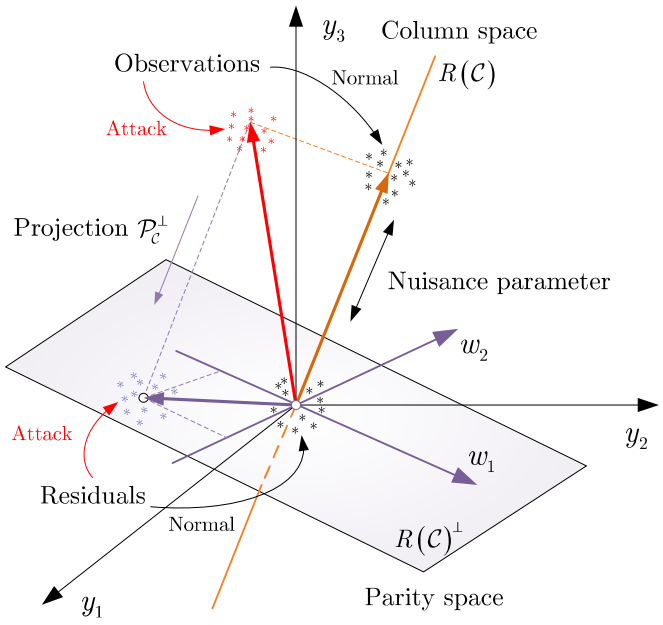



Figure 3.2 – Nuisance parameter rejection by the orthogonal projection of the observations onto the parity space.

where the random noises $\varsigma_{k-L+1}^{k} \sim \mathcal{N}(0, \Sigma_{\varsigma})$, where the covariance matrix $\Sigma_{\varsigma} = \mathcal{W}\mathcal{S}\mathcal{W}^{T} \in \mathbb{R}^{(Lp-n)\times(Lp-n)}$.

### 3.3.3 Relation to sliding window Kalman filter approach

In this subsection, we investigate the relation between the fixed-size parity space approach and the so-called "sliding window Kalman filter" approach [72, 73] for residual generation. In order to eliminate the negative impact of the nuisance parameter, Gustafsson suggested to utilize the least-square estimate $\hat{x}_{k-L+1}$ of system state $x_{k-L+1}$. Under the linear and Gaussian assumptions, the least-square estimate coincides with the maximum likelihood estimate [10], which is written as

$$\hat{x}_{k-L+1} = \left(\mathcal{C}^{T}\mathcal{S}\mathcal{C}\right)^{-1}\mathcal{C}^{T}\mathcal{S}^{-1}z_{k-L+1}^{k}. \tag{3.21}$$

Since the matrix $\mathcal{S}$ is symmetric and positive-definite, it can be decomposed into $\mathcal{S} = \mathcal{V}\mathcal{V}^{T}$ and $\mathcal{S}^{-1} = \mathcal{V}^{-T}\mathcal{V}^{-1}$, then

$$z_{k-L+1}^{k} - \mathcal{C}\hat{x}_{k-L+1} = \left[\mathcal{I} - \mathcal{C}\left(\mathcal{C}^{T}\mathcal{S}\mathcal{C}\right)^{-1}\mathcal{C}^{T}\mathcal{S}^{-1}\right]z_{k-L+1}^{k} = \mathcal{V}\mathcal{P}_{\mathcal{V}^{-1}\mathcal{C}}^{\perp}\mathcal{V}^{-1}z_{k-L+1}^{k}, \tag{3.22}$$

where $\mathcal{P}_{\mathcal{V}^{-1}\mathcal{C}}^{\perp}$ is the projection matrix onto the left-null space $R\left(\mathcal{V}^{-1}\mathcal{C}\right)^{\perp}$ of matrix $\mathcal{V}^{-1}\mathcal{C}$, which is calculated as

$$\mathcal{P}_{\mathcal{V}^{-1}\mathcal{C}}^{\perp} = \mathcal{I} - \left(\mathcal{V}^{-1}\mathcal{C}\right)\left[\left(\mathcal{V}^{-1}\mathcal{C}\right)^{T}\left(\mathcal{V}^{-1}\mathcal{C}\right)\right]^{-1}\left(\mathcal{V}^{-1}\mathcal{C}\right)^{T}. \tag{3.23}$$

Since the projection matrix $\mathcal{P}_{\mathcal{V}^{-1}\mathcal{C}}^{\perp}$ is singular, i.e., rank $\left(\mathcal{P}_{\mathcal{V}^{-1}\mathcal{C}}^{\perp}\right) = Lp-n$, the covariance matrix of $z_{k-L+1}^{k} - \mathcal{C}\hat{x}_{k-L+1}$ is singular, as well. In order to circumvent this difficulty, Gustafsson

[72, 73] suggested to replace the idempotent (but not symmetric) matrix $\mathcal{V}\mathcal{P}^{\perp}_{\mathcal{V}^{-1}\mathcal{C}}\mathcal{V}^{-1}$ by the matrix $\mathcal{W}_{\mathrm{LS}} \in \mathbb{R}^{(Lp-n)\times Lp}$, where its rows form a basis for the row space of matrix $\mathcal{V}\mathcal{P}^{\perp}_{\mathcal{V}^{-1}\mathcal{C}}\mathcal{V}^{-1}$, thus satisfying $\mathcal{W}_{\mathrm{LS}}\mathcal{C} = 0$. It is clear that the rows of $\mathcal{W}_{\mathrm{LS}}$ form also a basis for the left-null space $R(\mathcal{C})^{\perp}$ of matrix $\mathcal{C}$. The statistical model of the residuals generated by the least-square estimation method

$$r^k_{k-L+1} = \mathcal{W}_{\mathrm{LS}}\mathcal{M}\theta^k_{k-L+1}(k_0) + \mathcal{W}_{\mathrm{LS}}\eta^k_{k-L+1} \tag{3.24}$$

coincides with the statistical model (3.19) of the residuals generated fixed-size parity space.

The difference between two residual-generation methods lies in the choice of the rejection matrix $\mathcal{W}$, where the rows of $\mathcal{W}$ form a basis for the left-null space $R(\mathcal{C})^{\perp}$ of matrix $\mathcal{C}$. It has been discussed in [72, 73] that the sliding window Kalman filter method generates the residuals with minimum covariance. However, the residuals with minimum covariance do not guarantee the statistical performance of a detection procedure since a small noise covariance matrix $\mathcal{W}\mathcal{S}\mathcal{W}^T$ is often associated with small value of the change magnitude $\mathcal{W}\mathcal{M}\theta^k_{k-L+1}(k_0)$. A more appropriate performance index for comparing residual-generation methods will be considered in subsection 3.3.5.

### 3.3.4 Unified statistical model of the residuals

In this subsection, we propose a unified statistical model of the residuals generated by either the steady-state Kalman filter approach or the fixed-size parity approach. It follows from (3.14) and (3.20) that both residual-generation methods lead to the following unified statistical model:

$$r^k_{k-L+1} = \phi^k_{k-L+1}(k_0) + \xi^k_{k-L+1}, \tag{3.25}$$

where $r^k_{k-L+1}$ is the vector of residuals, $\phi^k_{k-L+1}(k_0)$ is the vector of transient signals and $\xi^k_{k-L+1} \sim \mathcal{N}(0, \Sigma)$ is the vector of random noises. For the steady-state Kalman filter approach, the transient profiles $\phi^k_{k-L+1}(k_0) = \psi^k_{k-L+1}(k_0)$ and the random noises $\xi^k_{k-L+1} = \varrho^k_{k-L+1}$ (i.e., $\Sigma = \Sigma_\varrho$). On the other hand, we have that the transient profiles $\phi^k_{k-L+1}(k_0) = \varphi^k_{k-L+1}(k_0)$ and the random noises $\xi^k_{k-L+1} = \varsigma^k_{k-L+1}$ (i.e., $\Sigma = \Sigma_\varsigma$) for the fixed-size parity space approach.

Let us add some comments on the transient profiles $\phi^k_{k-L+1}(k_0)$ and the random noises $\xi^k_{k-L+1}$ of the unified statistical model (3.25). Firstly, the vector of transient profiles $\phi^k_{k-L+1}(k_0)$ reflects the impact of the attack to the statistical model of the residuals. Under normal operation (i.e., $k < k_0$), $\phi^k_{k-L+1}(k_0)$ is the null vector. During the attack period (i.e., $k_0 \le k < k_0+L$), the vector $\phi^k_{k-L+1}(k_0)$ depends on the relative position of index $k_0$ within the time window $[k-L+1, k]$. For the post-change period (i.e., $k \ge k_0 + L$), the post-change profiles $\tilde{\phi}^k_{k-L+1}(k_0)$ are of no interest since any detection of attack with the detection delay equal to or greater than $L$ is considered as missed.

Secondly, the random noises $\xi^k_{k-L+1}$, for $k \ge L$, are exchangeable (i.e., $\xi^L_1, \xi^{L+1}_2, \cdots, \xi^k_{k-L+1}, \cdots$ follow the same distribution) and the covariance matrix $\Sigma$ is positive-definite. This property of the random noises $\xi^k_{k-L+1}$ is important in investigating the statistical performance of the detection procedures proposed in the following sections. For the steady-state Kalman filter approach, the vector of random noises $\varrho^k_{k-L+1} = \left[\varrho^T_{k-L+1}, \cdots, \varrho^T_k\right]^T$, where $\{\varrho_k\}_{k\ge1}$ are i.i.d. zero-mean Gaussian random vectors with positive-definite covariance matrix $J$. Hence, it is clear that the random noises $\varrho^L_1, \varrho^{L+1}_2, \cdots, \varrho^k_{k-L+1}, \cdots$ follow the same distribution (i.e., $\varrho^k_{k-L+1} \sim \mathcal{N}(0, \Sigma_\varrho)$, where $\Sigma_\varrho$ is positive-definite). For the fixed-size parity space approach, the vector of random noises is $\varsigma^k_{k-L+1} = \mathcal{W}\left(\mathcal{H}w^k_{k-L+1} + v^k_{k-L+1}\right)$, where $w^k_{k-L+1} = \left[w^T_{k-L+1}, \cdots, w^T_k\right]^T$ and

$v_{k-L+1}^{k} = \left[v_{k-L+1}^{T}, \cdots, v_{k}^{T}\right]^{T}$. Since the process noises $\{w_k\}_{k\geq 1}$ and the sensor noises $\{v_k\}_{k\geq 1}$ are i.i.d. zero-mean Gaussian vectors, the random noises $\varsigma_1^L, \varsigma_2^L, \cdots, \varsigma_{k-L+1}^k, \cdots$ follow the same distribution, i.e., $\varsigma_{k-L+1}^k \sim \mathcal{N}(0, \Sigma_\varsigma)$, where $\Sigma_\varsigma = \mathcal{W}\mathcal{S}\mathcal{W}^T$ is positive-definite.

### 3.3.5 Comparison of residual-generation methods

This section is dedicated to the comparison of residual-generation methods by means of the Kullback-Leibler (K-L) information number (or the K-L distance). It is well-known [10] that the residuals with higher K-L distance should offer better statistical performance than the residuals with lower K-L distance.

Starting from now, let $\mathcal{P}_{k_0}$ (resp. $\mathcal{P}_0 \triangleq \mathcal{P}_\infty$) be the joint distribution of the residuals $r_1^L, r_2^{L+1}, \cdots, r_{k-L+1}^k, \cdots$ when they follow the statistical model (3.25). Let also $\mathbb{E}_{k_0}$ (resp. $\mathbb{E}_0 \triangleq \mathbb{E}_\infty$) denote the corresponding mathematical expectations, and $p_{k_0}$ (resp. $p_0 \triangleq p_\infty$) stand for the probability density function. It is assumed, for the sake of simplicity, that $k = L$ and $k_0 = 1$. Then, the K-L distance between the distribution $\mathcal{P}_0$ and the distribution $\mathcal{P}_1$ is defined as

$$\rho = \int_{-\infty}^{+\infty} p_0\left(r_1^L\right) \log \frac{p_0\left(r_1^L\right)}{p_1\left(r_1^L\right)} dr_1^L, \tag{3.26}$$

where $\rho$ is the K-L distance. Let us stack transient vectors $\psi_1, \psi_2, \cdots, \psi_L$ (resp. $\varphi_1, \varphi_2, \cdots, \varphi_L$) into the concatenated vector $\psi_1^L(1)$ (resp. $\varphi_1^L(1)$), corresponding to the steady-state Kalman filter approach (resp. the fixed-size parity space approach). The K-L distances are calculated for the Gaussian noises [10] as

$$\rho_{\mathrm{KF}} = \frac{1}{2}\left[\psi_1^L(1)\right]^T \left[\Sigma_\varrho^{-1}\right] \left[\psi_1^L(1)\right], \tag{3.27}$$

$$\rho_{\mathrm{PS}} = \frac{1}{2}\left[\varphi_1^L(1)\right]^T \left[\Sigma_\varsigma^{-1}\right] \left[\varphi_1^L(1)\right], \tag{3.28}$$

where $\rho_{\mathrm{KF}}$ and $\rho_{\mathrm{PS}}$ are the K-L distances generated by the steady-state Kalman filter and the fixed-size parity space approaches, respectively.

In the following, we consider the choice of rejection matrix $\mathcal{W}$ by the parity space approach by means of K-L distance. The comparison between the Kalman filter and the parity space will be performed numerically later.

**Lemma 3.1.** *(Choice of rejection matrix). Let $\mathcal{W} \in \mathbb{R}^{(Lp-n)\times n}$ be a matrix such that the rows of $\mathcal{W}$ form a basis (not necessarily orthonormal) for the left-null space $R(\mathcal{C})^\perp$ of matrix $\mathcal{C}$, thus satisfying $\mathcal{W}\mathcal{C} = 0$. The following K-L distance*

$$\rho_{\mathrm{PS}} = \frac{1}{2}\left[\mathcal{M}\theta_1^L(1)\right]^T \left[\mathcal{W}^T\left(\mathcal{W}\mathcal{S}\mathcal{W}^T\right)^{-1}\mathcal{W}\right] \left[\mathcal{M}\theta_1^L(1)\right] \tag{3.29}$$

*does not depend on the choice of the rejection matrix $\mathcal{W}$.*

*Proof.* Since matrix $\mathcal{S}$ is symmetric and positive-definite, it can be decomposed (i.e., by Cholesky factorization) as $\mathcal{S} = \mathcal{V}\mathcal{V}^T$ which satisfies $\mathcal{S}^{-1} = \mathcal{V}^{-T}\mathcal{V}^{-1}$, where the matrix $\mathcal{V}$ is lower-triangular and non-singular. It follows from [117, page 210] that $\mathrm{rank}(\mathcal{W}\mathcal{V}) = \mathrm{rank}(\mathcal{W}) = Lp - n$ and $\mathrm{rank}(\mathcal{V}^{-1}\mathcal{C}) = \mathrm{rank}(\mathcal{C}) = n$ since matrix $\mathcal{V}$ is non-singular. Putting together

with $(\mathcal{W}\mathcal{V})(\mathcal{V}^{-1}\mathcal{C}) = 0$, the columns of matrix $(\mathcal{W}\mathcal{V})^T$ form a basis (not necessarily orthonormal) for the left-null space $R(\mathcal{V}^{-1}\mathcal{C})^{\perp}$ of matrix $\mathcal{V}^{-1}\mathcal{C}$. It follows from [117, pages 429-430] that projection matrix $\mathcal{P}^{\perp}_{\mathcal{V}^{-1}\mathcal{C}}$ from $\mathbb{R}^{Lp}$ onto $R(\mathcal{V}^{-1}\mathcal{C})^{\perp}$ is calculated as $\mathcal{P}^{\perp}_{\mathcal{V}^{-1}\mathcal{C}} = (\mathcal{W}\mathcal{V})^T\left[(\mathcal{W}\mathcal{V})(\mathcal{W}\mathcal{V})^T\right]^{-1}(\mathcal{W}\mathcal{V})$ and that $\mathcal{P}^{\perp}_{\mathcal{V}^{-1}\mathcal{C}}$ does not depend on the choice of $\mathcal{W}\mathcal{V}$, thus being independent from the choice of $\mathcal{W}$. Let $\mathcal{W}_1$ and $\mathcal{W}_2$ be two different choices of $\mathcal{W}$, then

$$\begin{aligned}
(\mathcal{W}_1\mathcal{V})^T\left[(\mathcal{W}_1\mathcal{V})(\mathcal{W}_1\mathcal{V})^T\right]^{-1}(\mathcal{W}_1\mathcal{V}) - (\mathcal{W}_2\mathcal{V})^T\left[(\mathcal{W}_2\mathcal{V})(\mathcal{W}_2\mathcal{V})^T\right]^{-1}(\mathcal{W}_1\mathcal{V}) = 0 \Leftrightarrow \\
\mathcal{V}^T\mathcal{W}_1^T\left[\mathcal{W}_1\mathcal{S}\mathcal{W}_1^T\right]^{-1}\mathcal{W}_1\mathcal{V} - \mathcal{V}^T\mathcal{W}_2^T\left[\mathcal{W}_2\mathcal{S}\mathcal{W}_2^T\right]^{-1}\mathcal{W}_2\mathcal{V} = 0 \Leftrightarrow \\
\mathcal{V}^T\left(\mathcal{W}_1^T\left[\mathcal{W}_1\mathcal{S}\mathcal{W}_1^T\right]^{-1}\mathcal{W}_1 - \mathcal{W}_2^T\left[\mathcal{W}_2\mathcal{S}\mathcal{W}_2^T\right]^{-1}\mathcal{W}_2\right)\mathcal{V} = 0 \Leftrightarrow \\
\mathcal{V}^{-T}\mathcal{V}^T\left(\mathcal{W}_1^T\left[\mathcal{W}_1\mathcal{S}\mathcal{W}_1^T\right]^{-1}\mathcal{W}_1 - \mathcal{W}_2^T\left[\mathcal{W}_2\mathcal{S}\mathcal{W}_2^T\right]^{-1}\mathcal{W}_2\right)\mathcal{V}\mathcal{V}^{-1} = 0 \Leftrightarrow \\
\mathcal{W}_1^T\left[\mathcal{W}_1\mathcal{S}\mathcal{W}_1^T\right]^{-1}\mathcal{W}_1 - \mathcal{W}_2^T\left[\mathcal{W}_2\mathcal{S}\mathcal{W}_2^T\right]^{-1}\mathcal{W}_2 = 0,
\end{aligned}$$

leading to $\mathcal{W}_1^T\left[\mathcal{W}_1\mathcal{S}\mathcal{W}_1^T\right]^{-1}\mathcal{W}_1 = \mathcal{W}_2^T\left[\mathcal{W}_2\mathcal{S}\mathcal{W}_2^T\right]^{-1}\mathcal{W}_2$, thus proving that the K-L distance $\rho_{\text{PS}}$ defined in (3.29) is independent from the choice of rejection matrix $\mathcal{W}$. □

An analogous problem of optimal fault detection has been addressed within the statistical framework in [51]. A linear model with nuisance parameters and a general covariance matrix (not necessarily diagonal) has been considered in the context of the unknown but non-random nuisance parameters. Two different invariant tests have been designed in such a case. The first invariant statistics was based on the knowledge of the observation matrix and the noise covariance matrix and the second one was based on the observation matrix only. It was shown that the two methods are equivalent. The numerical examples are given in chapter 6 for demonstrating theoretical results obtained in this subsection.

### 3.3.6 Discussion

The results of Lemma 3.1 helps in choosing the rejection matrix $\mathcal{W}$ for the fixed-size parity space approach. Under the K-L distance criterion, the rejection matrix $\mathcal{W}$, which satisfies $\mathcal{W}\mathcal{C} = 0$, can be chosen arbitrarily. Though the sliding window Kalman filter method generates the residuals with minimum noise covariance [72, 73], this method is just as efficient as the traditional fixed-size parity space approach. In addition, from the least-square estimation point of view, the authors in [10, pages 230-231] calculated the K-L information number by

$$\rho_{\text{LS}} = \frac{1}{2}\left[\mathcal{M}\theta_1^L(1)\right]^T\left[\mathcal{V}^{-T}\mathcal{P}^{\perp}_{\mathcal{V}^{-1}\mathcal{C}}\mathcal{V}^{-1}\right]\left[\mathcal{M}\theta_1^L(1)\right] = \rho_{\text{PS}}, \tag{3.30}$$

since the projection matrix $\mathcal{P}^{\perp}_{\mathcal{V}^{-1}\mathcal{C}} = (\mathcal{W}\mathcal{V})^T\left[(\mathcal{W}\mathcal{V})(\mathcal{W}\mathcal{V})^T\right]^{-1}(\mathcal{W}\mathcal{V})$.

Let us discuss now the comparison between the the steady-state Kalman filter approach and the fixed-size parity space approach. Firstly, we have not found any analytical expression for comparing the compelling residual-generation methods (i.e., Kalman filter and parity space). However, the comparison between these methods can be performed easily by the numerical calculation of the K-L distances.

Secondly, under perfect conditions (i.e., the model matches the real system, the process noises and the sensor noises are white, the noises covariance matrices are exactly known, the initial condition is Gaussian, and the system is detectable), the steady-state Kalman filter is an optimal estimator. At each time instant $k \geq L$, the steady-state Kalman filter utilizes the information about the *a priori* state estimate $\hat{x}_{k|k-1}$ for estimating the system state $\hat{x}_{k+1|k}$. For this reason, it is intuitive that the steady-state Kalman filter-based detectors will perform better than the fixed-size parity space-based detectors. This point will be shown by numerical examples in chapter 6.

Finally, the Kalman filter is no longer optimal in many practical situations, including modeling errors or unknown noise covariance matrices. The residuals are no longer independent and the proposed statistical model (3.14) is not valid. In such circumstances, the parity space approach may offer better statistical performance than the Kalman filter approach does. This point will be investigated by the simulation results in chapter 6.

## 3.4 Detection Algorithms under Known Transient Change Parameters

This section is organized as follows. The VTWL CUSUM algorithm is designed in subsection 3.4.1. Next, the statistical properties of the VTWL CUSUM algorithm as well as the optimal choice of thresholds are solved in subsection 3.4.2. It is shown that the optimal choice of thresholds leads to the simple Finite Moving Average (FMA) detection rule. In addition, a numerical method is proposed in subsection 3.4.3 for estimating the error probabilities of both VTWL CUSUM and FMA detectors. Finally, the robustness of the proposed FMA test w.r.t. several operational parameters is investigated in subsection **??**.

### 3.4.1 Variable Threshold Window Limited (VTWL) CUSUM algorithm

In this subsection, we adapt the VTWL CUSUM algorithm (2.159), which was first introduced by Guépié [67,69] for the i.i.d. Gaussian observations, to the unified statistical model (3.25). The idea of the VTWL CUSUM algorithm is derived from the off-line point of view of the change detection problem [10]. Based on the statistical model (3.25), it is convenient to introduce, for each time instant $k \geq L$, the following hypotheses about the change-point $k_0$:

$$\mathcal{H}_0 \triangleq \{k_0 > k\} \quad \text{and} \quad \mathcal{H}_j \triangleq \{k_0 = k - L + j\}, \text{ for } j = 1, 2, \cdots, L, \tag{3.31}$$

where $r_{k-L+1}^k \sim \mathcal{N}(0, \Sigma)$ under hypothesis $\mathcal{H}_0$ and $r_{k-L+1}^k \sim \mathcal{N}\left(\phi_{k-L+1}^k (k - L + j), \Sigma\right)$ under hypothesis $\mathcal{H}_j$. The change-point detection problem reduces to the problem of testing the null hypothesis $\mathcal{H}_0$ against $L$ alternative hypotheses $\mathcal{H}_j$, for $1 \leq j \leq L$. The alarm is raised if one of the hypotheses $\mathcal{H}_j$, for $1 \leq j \leq L$, is declared.

The standard statistical method consists in estimating the change-point $k_0$ by the maximum likelihood ratio (MLE) principle. Let $i = k - L + j$, the log-likelihood ratio (LLR) between hypothesis $\mathcal{H}_j$ and hypothesis $\mathcal{H}_0$ is calculated as

$$S_i^k = \log \frac{p_{\phi_{k-L+1}^k(i)}\left(r_{k-L+1}^k\right)}{p_0\left(r_{k-L+1}^k\right)}, \tag{3.32}$$

where $p_{\phi^k_{k-L+1}(i)}\left(r^k_{k-L+1}\right)$ and $p_0\left(r^k_{k-L+1}\right)$ is the probability density function (p.d.f.) of the residual vector $r^k_{k-L+1}$ under hypothesis $\mathcal{H}_j$ and hypothesis $\mathcal{H}_0$, respectively.

By utilizing the MLE principle with small modification (i.e., using variable thresholds), we introduce the following VTWL CUSUM algorithm:

$$T_{\mathrm{VTWL}} = \inf\left\{k \geq L : \max_{k-L+1\leq i\leq k}\left(S^k_i - h_{k-i+1}\right) \geq 0\right\}, \tag{3.33}$$

where $T_{\mathrm{VTWL}}$ is the alarm time of the VTWL CUSUM algorithm, $h_1, h_2, \cdots, h_L$ are chosen thresholds and the LLR $S^k_i$ is calculated in the Gaussian case as

$$S^k_i = \left[\phi^k_{k-L+1}(i)\right]^T\left[\Sigma^{-1}\right]\left[r^k_{k-L+1} - \frac{1}{2}\phi^k_{k-L+1}(i)\right]. \tag{3.34}$$

The VTWL CUSUM algorithm proceeds as follows. For each instant $k \geq L$, the algorithm uses the last $L$ measurements $y_{k-L+1}, \cdots, y_k$ for decision making. For each time index $i$ from $k-L+1$ to $k$, the LLR $S^k_i$ is first calculated by (3.34), depending on either the steady-state Kalman filter or the fixed-size parity space is employed. Next, the LLR $S^k_i$ is compared to each threshold $h_{k-i+1}$ and the alarm time $T_{\mathrm{VTWL}}$ is raised if one of the LLRs is greater than or equal to its corresponding threshold. Especially, the thresholds $h_1, h_2, \cdots, h_L$ are considered as tuning parameters for optimizing the VTWL CUSUM algorithm.

### 3.4.2 Optimization of the VTWL CUSUM algorithm and the FMA test

This subsection is dedicated to investigate the statistical properties of the proposed VTWL CUSUM algorithm (3.33)–(3.34). The properties of the worst-case probability of false alarm $\overline{\mathbb{P}}_{\mathrm{fa}}$ and the worst-case probability of missed detection $\overline{\mathbb{P}}_{\mathrm{md}}$ are given in Theorem 3.1.

**Theorem 3.1.** *Consider the VTWL CUSUM algorithm defined in* (3.33)–(3.34). *Then,*

*1. The worst-case probability of false alarm within any time window of length $m_\alpha$ corresponds to the first time window, i.e.,*

$$\overline{\mathbb{P}}_{\mathrm{fa}}\left(T_{\mathrm{VTWL}}; m_\alpha; h_1, h_2, \cdots, h_L\right) = \mathbb{P}_0\left(L \leq T_{\mathrm{VTWL}} \leq L + m_\alpha - 1\right). \tag{3.35}$$

*2. The worst-case probability of missed detection is upper bounded by*

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{VTWL}}; L; h_1, h_2, \cdots, h_L\right) \leq \tilde{\mathbb{P}}_{\mathrm{md}}\left(T_{VTWL}; h_L\right) \triangleq \Phi\left(\frac{h_L - \mu_{S^L_1}}{\sigma_{S^L_1}}\right), \tag{3.36}$$

*where $\Phi(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}t^2\right\}dt$ is the c.d.f. of the standard normal distribution, $\tilde{\mathbb{P}}_{\mathrm{md}}\left(T_{VTWL}; h_L\right)$ is the proposed upper bound for the worst-case probability of missed detection $\overline{\mathbb{P}}_{\mathrm{md}}$, and the parameters $\mu_{S^L_1}$ and $\sigma_{S^L_1}$ are calculated by*

$$\mu_{S^L_1} = \frac{1}{2}\left[\phi^L_1(1)\right]^T\left[\Sigma^{-1}\right]\left[\phi^L_1(1)\right], \tag{3.37}$$

$$\sigma^2_{S^L_1} = \left[\phi^L_1(1)\right]^T\left[\Sigma^{-1}\right]\left[\phi^L_1(1)\right]. \tag{3.38}$$

*Proof.* The proof is given in Appendix A.2. □

It is worth noting that the simultaneous minimization of both the worst-case probability of missed detection $\overline{\mathbb{P}}_{\mathrm{md}}$ and the worst-case probability of false alarm $\overline{\mathbb{P}}_{\mathrm{fa}}$ is contradictory. Moreover, their analytical expression is not available due to mathematical complexity. For these reasons, we propose minimizing the upper bound $\tilde{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{VTWL}}; h_L\right)$ for the worst-case probability of missed detection subject to an acceptable level of the worst-case probability of false alarm within any time window of length $m_\alpha$. Before considering the optimization problem, let us impose the following assumption of the transient change profiles $\phi_1^L(1)$. This assumption is essential in solving the optimization problem.

**Assumption 3.1.** *It is assumed that the vector of transient change profiles $\phi_1^L(1)$ defined in (3.25) is non-null (i.e., $\psi_1^L(1) \neq 0$ for the steady-state Kalman filter and $\varphi_1^L(1) \neq 0$ for the fixed-size parity space).*

Assumption 3.1 plays an extremely important role in choosing the thresholds of the VTWL CUSUM algorithm. This assumption provides sufficient condition for the following lemma.

**Lemma 3.2.** *Let $\mathcal{S} \in \mathbb{R}^{m_\alpha}$ be a Gaussian random vector consisting of $m_\alpha$ log-likelihood ratios (LLRs) $S_1^L, S_2^{L+1}, \cdots, S_{m_\alpha}^{L+m_\alpha-1}$. If Assumption 3.1 is satisfied, then the covariance matrix $\Sigma_{\mathcal{S}} \in \mathbb{R}^{m_\alpha \times m_\alpha}$ of the random vector $\mathcal{S}$ is positive-definite.*

*Proof.* The proof is given in Appendix A.3. □

By exploiting the results of Lemma 3.2, the optimal choice of thresholds w.r.t. the criterion (3.6)–(3.7) is formulated and solved in Theorem 3.2.

**Theorem 3.2.** *Consider the VTWL CUSUM algorithm defined in (3.33)–(3.34). Then,*

*1. The optimal choice of the thresholds $h_1, h_2, \cdots, h_L$ leads to the following optimization problem:*

$$\begin{cases} \inf_{h_1,h_2,\cdots,h_L} & \tilde{\mathbb{P}}_{\mathrm{md}}\left(T_{VTWL}; h_L\right) \\ \textit{subject to} & \overline{\mathbb{P}}_{\mathrm{fa}}\left(T_{\mathrm{VTWL}}; m_\alpha; h_1, h_2, \cdots, h_L\right) \leq \alpha \end{cases}, \tag{3.39}$$

*where $\alpha \in (0,1)$ is the acceptable level for the worst-case probability of false alarm within any time window of length $m_\alpha$. The optimization problem (3.39) has the unique solution $(h_1^*, h_2^*, \cdots, h_L^*)$ for a given $\alpha \in (0,1)$, where $h_1^*, h_2^*, \cdots, h_{L-1}^* \to \infty$ and $h_L^*$ is calculated from the following equation:*

$$\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\left\{S_{k-L+1}^k < h_L^*\right\}\right) = 1-\alpha. \tag{3.40}$$

*2. The optimized VTWL CUSUM algorithm is equivalent to the following FMA detection rule:*

$$T_{\mathrm{FMA}}\left(\tilde{h}_L\right) = \inf\left\{k \geq L : \left[\phi_1^L(1)\right]^T\left[\Sigma^{-1}\right] r_{k-L+1}^k \geq \tilde{h}_L\right\}, \tag{3.41}$$

*where the threshold $\tilde{h}_L = h_L^* + \mu_{S_1^L}$. Especially, the upper bound for the worst-case probability of missed detection of the FMA test (3.41) is calculated as*

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{FMA}}; \tilde{h}_L\right) \leq \tilde{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{FMA}}; \tilde{h}_L\right) \triangleq \Phi\left(\frac{\tilde{h}_L - 2\mu_{S_1^L}}{\sigma_{S_1^L}}\right). \tag{3.42}$$

*Proof.* The proof of is given in Appendix A.4. □

### 3.4.3 Numerical calculation of error probabilities

In this subsection, we propose a numerical method for estimating the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ and the worst-case probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ for both VTWL CUSUM algorithm and FMA detection rule. This numerical method includes both steady-state Kalman filter approach and fixed-size parity space approach. The results are obtained by utilizing the numerical calculation of the multivariate Gaussian cumulative distribution function (c.d.f.) introduced in [63]. This algorithm has been implemented in Matlab's Statistics Toolbox by the function MVNCDF.

**Proposition 3.1.** *The worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ and the worst-case probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ for the VTWL CUSUM algorithm in* (3.33)–(3.34) *and the FMA detection rule in* (3.41) *are calculated numerically by the following formulas:*

*1. The worst-case probability of false alarm is computed as*

$$\overline{\mathbb{P}}_{\text{fa}}\left(T_{\text{VTWL}}; m_\alpha; h_1, h_2, \cdots, h_L\right) = 1 - \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \bigcap_{i=k-L+1}^{k} \left\{S_i^k < h_{k-i+1}\right\}\right), \tag{3.43}$$

$$\overline{\mathbb{P}}_{\text{fa}}\left(T_{\text{FMA}}; m_\alpha; \tilde{h}_L\right) = 1 - \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \left\{S_{k-L+1}^k < \tilde{h}_L - \mu_{S_1^L}\right\}\right). \tag{3.44}$$

*2. The worst-case probability of missed detection is calculated as*

$$\overline{\mathbb{P}}_{\text{md}}\left(T_{\text{VTWL}}; h_1, h_2, \cdots, h_L\right) = \sup_{k_0 \geq L} \frac{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0+L-1} \bigcap_{i=k-L+1}^{k} \left\{S_i^k < h_{k-i+1}\right\}\right)}{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0-1} \bigcap_{i=k-L+1}^{k} \left\{S_i^k < h_{k-i+1}\right\}\right)}, \tag{3.45}$$

$$\overline{\mathbb{P}}_{\text{md}}\left(T_{\text{FMA}}; \tilde{h}_L\right) = \sup_{k_0 \geq L} \frac{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0+L-1} \left\{S_{k-L+1}^k < \tilde{h}_L - \mu_{S_1^L}\right\}\right)}{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0-1} \left\{S_{k-L+1}^k < \tilde{h}_L - \mu_{S_1^L}\right\}\right)}. \tag{3.46}$$

*Proof.* The proof of equations (3.43)–(3.46) is given in Appendix A.5. □

**Remark 3.6.** *The formulas for numerical calculation of the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ and the worst-case probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ are given in Proposition 3.1. In order to calculate the c.d.f. of a multivariate Gaussian distribution by utilizing the* MVNCDF *function, it is required to formulate and to compute the threshold vector, the mean vector and the covariance matrix. Such calculations are elaborated in Appendix A.5.*

**Remark 3.7.** *The equations* (3.43)–(3.44) *are derived from the results of Theorem 3.1 which shows that the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ within any time window of length $m_\alpha$ corresponds to the first time window $[L; L + m_\alpha - 1]$. The worst-case probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$, on the other hand, involves the "supremum" operation over all change-point $k_0 \geq L$. In other words, the worst-case probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ does not correspond to the*

*first time window* $[L; 2L-1]$. *Fortunately, simulation results show that the probability of missed detection* $\mathbb{P}_{k_0}\left(T \geq k_0 + L | T \geq k_0\right)$ *receives high values for some small values of* $k_0$, *where* $T$ *can be stopping time of the VTWL CUSUM algorithm or the FMA detection rule. For these reasons, we replace the "supremum" operation over all* $k_0 \geq L$ *by "maximum" operation over some* $k_0 \in [L, L+\delta L]$, *where* $\delta L \in \mathbb{N}^+$, *for approximating the worst-case probability of missed detection* $\overline{\mathbb{P}}_{\text{md}}$.

**Remark 3.8.** *The numerical method permits us to estimate the worst-case probability of false alarm* $\overline{\mathbb{P}}_{\text{fa}}$ *and the worst-case probability of missed detection* $\overline{\mathbb{P}}_{\text{md}}$ *instead of the traditional Monte Carlo simulation method. It is worth noting that the proposed method is more efficient than the Monte Carlo simulation regarding the computational time. Moreover, this numerical method can be exploited for investigating the robustness of the FMA test, which will be introduced in subsection* **??**.

### 3.4.4 Sensitivity analysis of FMA test

In this subsection, we perform the sensitivity analysis of the FMA test given in (3.41) in order to evaluate its robustness w.r.t. several operational parameters, including the attack duration $L$, the attack profiles $\theta_1, \theta_2, \cdots, \theta_L$, the process noise covariance matrix $Q$, and the sensor noise covariance matrix $R$. This sensitivity analysis process is important in practical circumstances since the operational parameters are generally not exactly known. In other words, their true values are often associated with their putative values through some levels of deterministic uncertainty.

Let $L$ and $\overline{L}$ be the putative and true values of the attack duration and $\theta_1, \theta_2, \cdots, \theta_L$ and $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$ be the putative and true values of the attack profiles, respectively. Let also $Q$ and $\overline{Q}$ be the putative and true values of the process noise covariance matrix, and $R$ and $\overline{R}$ be the putative and true values of the sensor noise covariance matrix, respectively. It is worth noting that the putative operational parameters (i.e., attack duration $L$, attack profiles $\theta_1, \theta_2, \cdots, \theta_L$, process noise covariance matrix $Q$ and sensor noise covariance matrix $R$) remain unchanged and they are considered as the designed parameters. The variation in true operational parameters (i.e., attack duration $\overline{L}$, attack profiles $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$, process noise covariance matrix $\overline{Q}$ and sensor noise covariance matrix $\overline{R}$) leads to the change in parameters of the unified statistical model (3.25). However, the numerical method introduced in Proposition 3.1 can also be used for investigating the robustness of the FMA test w.r.t. these parameters.

The worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}\left(T_{\text{FMA}}; m_\alpha; \tilde{h}_L\right)$ and the worst-case probability of missed detection $\overline{\mathbb{P}}_{\text{md}}\left(T_{\text{FMA}}; \tilde{h}_L\right)$ can be calculated numerically by (3.44) and (3.46), respectively. The mean vector and the covariance matrix, for both $\overline{\mathbb{P}}_{\text{fa}}\left(T_{\text{FMA}}; m_\alpha; \tilde{h}_L\right)$ and $\overline{\mathbb{P}}_{\text{md}}\left(T_{\text{FMA}}; \tilde{h}_L\right)$, can be formulated in exactly the same manner as in Appendix A.5. However, the mathematical expectations $\mathbb{E}_0\left[S_i^k\right]$ and $\mathbb{E}_{k_0}\left[S_i^k\right]$ and the covariance $\text{cov}\left(S_{i_1}^{k_1}, S_{i_2}^{k_2}\right)$ need to be revised since the true parameters are different from their putative values (i.e., $\overline{L} \neq L$, $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L \neq \theta_1, \theta_2, \cdots, \theta_L$, $\overline{Q} \neq Q$ and $\overline{R} \neq R$).

The mathematical expectations $\mathbb{E}_0\left[S_i^k\right]$ and $\mathbb{E}_{k_0}\left[S_i^k\right]$ depend only on the true attack duration $\overline{L}$ and the true attack profiles $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$ while the covariance $\text{cov}\left(S_{i_1}^{k_1}, S_{i_2}^{k_2}\right)$ depends on the true process noise covariance $\overline{Q}$ and the true sensor noise covariance $\overline{R}$. For the fixed-size parity space approach, the computation of $\text{cov}\left(S_{i_1}^{k_1}, S_{i_2}^{k_2}\right)$ can be generalized from Appendix A.5

without difficulty. On the other hand, for the steady-state Kalman filter, since the true noise covariances are different from their putative values, the Kalman filter is no longer optimal and the innovations are no longer independent. For this reason, it is required to re-calculate the covariance between two innovations (see Appendix A.1). In short, the calculation of $\mathbb{E}_0\left[S_i^k\right]$, $\mathbb{E}_{k_0}\left[S_i^k\right]$ and $\operatorname{cov}\left(S_{i_1}^{k_1}, S_{i_2}^{k_2}\right)$, for both steady-state Kalman filter approach and fixed-size parity space approach, are elaborated in Appendix A.6.

## 3.5 Detection Algorithms under Partially Known Transient Change Parameters

It is of practical interest to consider circumstances where the attack profiles $\theta_1, \theta_2, \cdots, \theta_L$ are completely unknown. For the quickest change detection problem, several scenarios on the *a priori* information about the post-change profiles have been investigated in [10]. In this section, we consider a special case where the change direction is exactly known but the change magnitude is unknown.

This special scenario is motivated by the detection of cyber-physical attacks on SCADA systems. The SCADA systems have been playing an extremely important role in safety-critical infrastructures and the security of SCADA systems against malicious attacks has received increasing concern from both research institutions, industries and governments. Therefore, the security analysis process is required in investigating vulnerable points that could be exploited for performing malevolent activities. Through this analysis process, the attack profiles $\theta_1, \theta_2, \cdots, \theta_L$ are often partially known. For example, if we know exactly which command signals, control signals and/or sensor measurements will be compromised but the power of the attack (i.e., the magnitude of attack signals) is unknown, then the shape of attack profiles $\theta_1, \theta_2, \cdots, \theta_L$ are known but their magnitude is unknown. This section treats such cases.

Assume that the putative values $\theta_1, \theta_2, \cdots, \theta_L$ are known but their true values $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$ are partially known. More precisely, the true attack profiles can be described in terms of putative profiles as $\overline{\theta}_k = \gamma\theta_k$, where the coefficient $\gamma$ is unknown. It can be shown without difficulty that

$$\overline{\phi}_{k-L+1}^k(k_0) = \gamma\phi_{k-L+1}^k(k_0), \tag{3.47}$$

where the transient vector $\overline{\phi}_{k-L+1}^k(k_0) \in \mathbb{R}^{Lp}$ can be calculated from the true attack profiles $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$ in the same manner as $\phi_{k-L+1}^k(k_0)$ in (3.25). Since the attack magnitude $\gamma$ is unknown, the generalized likelihood ratio (GLR) and the weighted likelihood ratio (WLR) approaches are considered for solving the problem.

### 3.5.1 Generalized Likelihood Ratio (GLR) Approach

The generalized likelihood ratio (GLR) approach consists of replacing the unknown parameter $\gamma$ by its maximum likelihood estimate (MLE). The generalized log-likelihood ratio (generalized LLR) $\hat{S}_i^k$ can be computed as

$$\hat{S}_i^k = \sup_{\gamma}\left[\gamma\phi_{k-L+1}^k(i)\right]^T\left[\Sigma^{-1}\right]\left[r_{k-L+1}^k - \frac{1}{2}\gamma\phi_{k-L+1}^k(i)\right]. \tag{3.48}$$

The generalized LLR $\hat{S}_i^k$ can be calculated, after some simple transformations, as follows:

$$\hat{S}_i^k = \left[r_{k-L+1}^k\right]^T \left[\overline{\Sigma}\,(i)\right] \left[r_{k-L+1}^k\right], \tag{3.49}$$

where the matrix $\overline{\Sigma}\,(i)$, which depends on the index $i$, is computed as

$$\overline{\Sigma}\,(i) = \frac{\left[\Sigma^{-1}\right] \left[\phi_{k-L+1}^k\,(i)\right] \left[\phi_{k-L+1}^k\,(i)\right]^T \left[\Sigma^{-1}\right]}{2 \left[\phi_{k-L+1}^k\,(i)\right]^T \left[\Sigma^{-1}\right] \left[\phi_{k-L+1}^k\,(i)\right]}. \tag{3.50}$$

The VTWL GLR detection rule, which utilizes the generalized LLR statistic $\hat{S}_i^k$, is described as

$$\hat{T}_{\mathrm{GLR}} = \inf\left\{k \geq L : \max_{k-L+1 \leq i \leq k} \left(\hat{S}_i^k - h_{k-i+1}\right) \geq 0\right\} \tag{3.51}$$

where the thresholds $h_1, h_2, \cdots, h_L$ are considered as the tuning parameters for optimizing the VTWL GLR algorithm.

### 3.5.2 Weighted Likelihood Ratio (WLR) Approach

The weighted likelihood ratio (WLR) approach assumes that the unknown parameter $\gamma$ follows the *a priori* distribution. The weighted log-likelihood ratio (weighted LLR) $\check{S}_i^k$ is then calculated as

$$\check{S}_i^k = \log \frac{\int \left[p_{\phi_{k-L+1}^k(i)}\left(r_{k-L+1}^k\right)\right] p_\gamma d\gamma}{p_0\left(r_{k-L+1}^k\right)}, \tag{3.52}$$

where $p_\gamma$ is the density distribution function of the unknown parameter $\gamma$.

For the sake of simplicity, let us suppose that the unknown parameter $\gamma$ follows the uniform distribution $\mathcal{U}\,(\gamma_0, \gamma_1)$, where the bounds $0 < \gamma_0 < \gamma_1$ are assumed to be known. The density distribution function $p_\gamma = {}^1\!/_{(\gamma_1 - \gamma_0)}$. After some calculations, we obtain

$$\begin{aligned} \check{S}_i^k \;=\; & \left[r_{k-L+1}^k\right]^T \left[\overline{\Sigma}\,(i)\right] \left[r_{k-L+1}^k\right] + \log\left[\frac{\sqrt{2\pi}}{b\,(i)\,(\gamma_1 - \gamma_0)}\right] + \\ & \log\left[\Phi\left(b\,(i)\,\gamma_1 - \frac{a\,(i)}{b\,(i)}\right) - \Phi\left(b\,(i)\,\gamma_0 - \frac{a\,(i)}{b\,(i)}\right)\right], \end{aligned} \tag{3.53}$$

where the coefficients $a\,(i)$ and $b\,(i)$ are calculated as

$$a\,(i) = \left[\phi_{k-L+1}^k\,(i)\right]^T \left[\Sigma^{-1}\right] \left[r_{k-L+1}^k\right], \tag{3.54}$$

$$b\,(i)^2 = \left[\phi_{k-L+1}^k\,(i)\right]^T \left[\Sigma^{-1}\right] \left[\phi_{k-L+1}^k\,(i)\right]. \tag{3.55}$$

The VTWL WLR detection rule, which utilizes the weighted LLR statistic $\check{S}_i^k$, is described as

$$\check{T}_{\mathrm{WLR}} = \inf\left\{k \geq L : \max_{k-L+1 \leq i \leq k} \left(\check{S}_i^k - h_{k-i+1}\right) \geq 0\right\} \tag{3.56}$$

where the thresholds $h_1, h_2, \cdots, h_L$ are considered as the tuning parameters for optimizing the VTWL WLR algorithm.

### 3.5.3 Statistical properties of VTWL GLR and VTWL WLR

In this subsection, we investigate the statistical properties of the VTWL GLR and VTWL WLR detection rules. Main results are given in Theorem 3.3 and Theorem 3.4.

**Theorem 3.3.** *Consider the VTWL GLR test defined in* (3.51) *and the VTWL WLR test defined in* (3.56)*, respectively. Then,*

*1. The worst-case probability of false alarm within any time window of length $m_\alpha$ is calculated as*

$$\overline{\mathbb{P}}_{\text{fa}}\left(\hat{T}_{\text{GLR}}\right) = \mathbb{P}_0\left(L \leq \hat{T}_{\text{GLR}} \leq L + m_\alpha - 1\right), \tag{3.57}$$

$$\overline{\mathbb{P}}_{\text{fa}}\left(\check{T}_{\text{WLR}}\right) = \mathbb{P}_0\left(L \leq \check{T}_{\text{WLR}} \leq L + m_\alpha - 1\right). \tag{3.58}$$

*2. The worst-case probability of missed detection is upper bounded by*

$$\overline{\mathbb{P}}_{\text{md}}\left(\hat{T}_{\text{GLR}}\right) \leq \tilde{\mathbb{P}}_{\text{md}}\left(\hat{T}_{GLR}; h_L\right) = \mathbb{P}_1\left(\hat{S}_1^L < h_L\right), \tag{3.59}$$

$$\overline{\mathbb{P}}_{\text{md}}\left(\check{T}_{\text{WLR}}\right) \leq \tilde{\mathbb{P}}_{\text{md}}\left(\check{T}_{WLR}; h_L\right) = \mathbb{P}_1\left(\check{S}_1^L < h_L\right), \tag{3.60}$$

*where $\tilde{\mathbb{P}}_{\text{md}}\left(\hat{T}_{GLR}; h_L\right)$ and $\tilde{\mathbb{P}}_{\text{md}}\left(\check{T}_{WLR}; h_L\right)$ are the upper bounds for the worst-case probability of missed detection of the VTWL GLR and VTWL WLR algorithms, respectively.*

*Proof.* Theorem 3.3 can be proved by utilizing the same arguments as Theorem 3.1. □

In the following theorem, we wish to minimize the upper bound $\tilde{\mathbb{P}}_{\text{md}}\left(\hat{T}_{\text{GLR}}; h_L\right)$ (resp. the upper bound $\tilde{\mathbb{P}}_{\text{md}}\left(\check{T}_{\text{WLR}}; h_L\right)$) for the worst-case probability of missed detection $\overline{\mathbb{P}}_{\text{md}}\left(\hat{T}_{\text{GLR}}\right)$ (res. $\overline{\mathbb{P}}_{\text{md}}\left(\check{T}_{\text{WLR}}\right)$) subject to a given value $\alpha \in (0,1)$ on the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}\left(\hat{T}_{\text{GLR}}\right)$ (resp. $\overline{\mathbb{P}}_{\text{fa}}\left(\check{T}_{\text{WLR}}\right)$).

**Theorem 3.4.** *Consider the VTWL GLR test defined in* (3.51) *and the VTWL WLR test defined in* (3.56)*. Then,*

*1. The optimal choice of the thresholds $h_1, h_2, \cdots, h_L$ leads to the following optimization problem:*

$$\inf_{h_1,\cdots,h_L} \tilde{\mathbb{P}}_{\text{md}}\left(\hat{T}_{GLR}; h_L\right) \text{ subject to } \overline{\mathbb{P}}_{\text{fa}}\left(\hat{T}_{\text{GLR}}; m_\alpha; h_1, h_2, \cdots, h_L\right) \leq \alpha, \tag{3.61}$$

$$\inf_{h_1,\cdots,h_L} \tilde{\mathbb{P}}_{\text{md}}\left(\check{T}_{WLR}; h_L\right) \text{ subject to } \overline{\mathbb{P}}_{\text{fa}}\left(\check{T}_{\text{WLR}}; m_\alpha; h_1, h_2, \cdots, h_L\right) \leq \alpha, \tag{3.62}$$

*where $\alpha \in (0,1)$ is the acceptable level on the false alarm rates. Let $\hat{h}_L^*$ and $\check{h}_L^*$ be, respectively, the minimum real numbers satisfying following inequalities:*

$$\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\left\{\hat{S}_{k-L+1}^k < \hat{h}_L^*\right\}\right) \geq 1-\alpha, \tag{3.63}$$

$$\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\left\{\check{S}_{k-L+1}^k < \check{h}_L^*\right\}\right) \geq 1-\alpha. \tag{3.64}$$

*Then, the optimization problem* (3.61) *(resp.* (3.62)*) has the solution* $\hat{h}_1^*, \cdots, \hat{h}_{L-1}^* \to +\infty$ *(resp.* $\check{h}, \cdots, \check{h}_{L-1}^* \to +\infty$*) and* $\hat{h}_L^*$ *(resp.* $\check{h}_L^*$*).*

*2. The optimized VTWL GLR and VTWL WLR algorithms lead to the following FMA detection rules:*

$$\hat{T}_{\text{FMA}} = \inf\left\{k \geq L : \hat{S}_{k-L+1}^k \geq \hat{h}_L^*\right\} \tag{3.65}$$

$$\check{T}_{\text{FMA}} = \inf\left\{k \geq L : \check{S}_{k-L+1}^k \geq \check{h}_L^*\right\} \tag{3.66}$$

*where* $\hat{T}_{\text{FMA}}$ *is the stopping time of the FMA GLR test and* $\check{T}_{FMA}$ *is the stopping time of the FMA WLR test, and the thresholds* $\hat{h}_L^*$ *and* $\check{h}_L^*$ *are chosen for assuring acceptable levels of false alarms.*

*Proof.* The proof is given in the Appendix A.7. □

**Remark 3.9.** *Let us add some comments on the results of Theorem 3.3 and Theorem 3.4. The numerical estimation of the probability of false alarm and the probability of missed detection for the FMA GLR test given in* (3.65) *and the FMA WLR test given in* (3.66) *have not been found due to mathematical complexity. The statistical performance of the FMA GLR test and FMA WLR test will be investigated by Monte Carlo simulation in chapter 6.*

## 3.6 Conclusion

In this chapter, we have considered the sequential detection of transient signals in stochastic-dynamical systems, applied to the detection of cyber-physical attacks on SCADA systems. The SCADA systems are described as a discrete-time linear time-invariant state space model driven by Gaussian noises. The cyber-physical attacks are modeled as additive signals of short duration on both state evolution and sensor measurement equations. The optimality criterion involves the minimization of the worst-case probability of missed detection subject to an acceptable level on the worst-case probability of false alarm within any time window of predefined length.

The traditional two-step approach, including the residual-generation step and the residual-evaluation step, has been considered for solving the problem. For the first step, the residuals are generated by utilizing well-known techniques: the steady-state Kalman filter approach and the fixed-size parity space approach. The unified statistical model of the residuals generated by both aforementioned methods has been developed. Moreover, the Kullback-Leibler (K-L) information number has been considered as the performance index for comparing residual-generation methods. The problem of choosing the parity space has been long discussed in the fault diagnosis community. It has been shown in this chapter that the K-L distance of the residuals generated by the fixed-size parity space is independent from the choice of parity space (i.e., Lemma 3.1).

Based on the unified statistical model (3.25), the VTWL CUSUM algorithm, which was initiated by Guépié in [67, 68, 70] for detecting transient changes in a sequence of independent Gaussian observations, has been adapted to the detection of transient signals on the discrete-time state space model. The idea of utilizing the variable thresholds is to make the algorithm flexible and the thresholds are considered as tuning parameters for optimizing the VTWL CUSUM algorithm. In order to find optimal thresholds w.r.t. the transient change detection criterion, it is required to investigate the properties of the worst-case probability of false alarm and the

worst-case probability of missed detection. It has been shown in Theorem 3.1 that the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ corresponds to the first time window $[L; L + m_\alpha - 1]$ and the upper bound $\tilde{\overline{\mathbb{P}}}_{\text{md}}$ for the worst-case probability of missed detection is proposed instead of its exact value $\overline{\mathbb{P}}_{\text{md}}$.

The optimization problem has been formulated and solved in Theorem 3.2, taking into account the transient change detection criterion. Due to the mathematical complexity, the optimization problem is considered as the optimal choice of thresholds for the VTWL CUSUM algorithm, being favor of minimizing the upper bound $\tilde{\overline{\mathbb{P}}}_{\text{md}}$ for the worst-case probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ for a given value on the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ within any time window of length $m_\alpha$. It has been shown that the optimal choice of thresholds leads to the simple FMA test.

Since their analytical expressions (i.e., $\overline{\mathbb{P}}_{\text{fa}}$ and $\overline{\mathbb{P}}_{\text{md}}$) are not available, we have proposed a numerical method for estimating the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ and the worst-case probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$, for both the VTWL CUSUM algorithm and the FMA detection rule. This numerical method is based on the numerical computation of the c.d.f. of a multivariate Gaussian distribution. Specially, the proposed method has been exploited for investigating the robustness of the FMA test w.r.t. several operational parameters, including the attack duration, the attack profiles, the process and sensor noise covariances. Especially, a recursive algorithm has been proposed for calculating the covariance between two innovations generated by the discrete-time Kalman filter under imperfect conditions (i.e., the true noise covariances are different from their putative values).

The attack profiles are generally unknown in practical situations. In the final section of this chapter, we consider a special scenario where the attack profiles are partially known. More precisely, the shape of the change is assumed to be completely known but the magnitude of the change is unknown. Two standard approaches, including the GLR approach and the WLR approach, have been considered for solving the problem. Similar to the previous cases, the corresponding VTWL GLR and VTWL WLR algorithms have been considered. It has been shown that the optimal choice of thresholds w.r.t. the transient change detection criterion leads also to the FMA GLR test and the FMA WLR test, respectively. However, the numerical method for estimating the worst-case probability of false alarm and the worst-case probability of missed detection for the FMA GLR test and the FMA WLR test has not been found. This point is dedicated to future study.

# Chapter 4

# Sequential Isolation of Transient Signals in Stochastic-dynamical Systems

**Contents**



## 4.1 Introduction

The problem of detecting cyber-physical attacks on SCADA systems has been addressed in chapter 3. The attack detection problem is concerned with making a binary decision of whether a malicious attack has been performed or the system is operating normally. The criterion of optimality involves the minimization of the worst-case probability of missed detection subject to an acceptable level on the worst-case probability of false alarm within any time window of predefined length.

It follows from the security analysis process performed in chapter 1 that there are multiple vulnerable points (i.e., attack types or attack scenarios) which might be exploited for launching malicious attacks on SCADA systems. It is of great interest to determine not only whether the system is under attack (i.e., detection problem) but also the attack types (i.e, isolation problem).

The problem of jointly detecting and identifying cyber-physical attacks (and/or faults) on SCADA systems has been considered in [4, 6, 7, 140, 141] in the deterministic framework (i.e., without random noises). The continuous-time (resp. time-delay continuous-time) state space model has been utilized to describe SCADA systems [140, 141] (resp. SCADA water irrigation networks [4, 6, 7]). The cyber-physical attacks are modeled as additive signals to both state evolution and sensor measurement equations. The detection-isolation schemes have been designed by utilizing the Unknown Input Observer (UIO) techniques. However, the negative impact of random noises has not been considered.

This chapter is dedicated to the joint detection-isolation of transient changes in stochastic-dynamical systems. The organization of this chapter is as follows. Firstly, the problem formulation is given in section 4.2. Secondly, we develop in section 4.3 the unified statistical model of the residuals generated by either the steady-state Kalman filter approach and the fixed-size parity space approach. This unified statistical model is the generalization of the unified statistical model (3.25) developed in chapter 3 to the joint detection-isolation problem. Thirdly, several detection-isolation schemes are introduced in section 4.4 for jointly detecting and identifying transient changes of known profiles. Finally, some concluding remarks are drawn in section 4.5.

## 4.2 Problem Formulation

In this section, we formulate the attack detection-isolation problem as the problem of jointly detecting and isolating transient changes in stochastic-dynamical systems. The model of transient changes in stochastic-dynamical systems is introduced in subsection 4.2.1. A novel criterion of optimality, dedicated to the detection-isolation of suddenly arrived signals of short (and known) duration, is proposed in subsection 4.2.2.

### 4.2.1 System and attack models

Similar to the detection problem, the following discrete-time state space model is employed to describe SCADA systems and cyber-physical attacks:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + B_a a_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + D_a a_k + v_k \end{cases}; \quad x_1 = \overline{x}_1, \tag{4.1}$$

where $x_k \in \mathbb{R}^n$ is the vector of system states with unknown initial values $\overline{x}_1$, $u_k \in \mathbb{R}^m$ is the vector of control signals, $d_k \in \mathbb{R}^q$ is the vector of disturbances, $y_k \in \mathbb{R}^p$ is the vector of sensor measurements, $a_k \in \mathbb{R}^s$ is the vector of attack signals, $w_k \in \mathbb{R}^n$ is the vector of process noises and $v_k \in \mathbb{R}^p$ is the vector of sensor noises; the matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $F \in \mathbb{R}^{n\times q}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, $G \in \mathbb{R}^{p\times q}$, $B_a \in \mathbb{R}^{n\times s}$ and $D_a \in \mathbb{R}^{p\times s}$ are assumed to be completely known.

It is assumed also that the control signals $u_k$ and the disturbances $d_k$ are exactly known. The process noises $w_k$ and the sensor noises $v_k$ are assumed to be i.i.d. zero-mean Gaussian random

vectors, i.e., $\mathrm{cov}\,(w_k, w_l) = Q\delta_{kl}$, $\mathrm{cov}\,(v_k, v_l) = R\delta_{kl}$ and $\mathrm{cov}\,(w_k, v_l) = 0$, where $\delta_{kl} = 1$ if $k = l$ and $\delta_{kl} = 0$ otherwise. The noise covariance matrices $Q$ and $R$ are assumed to be exactly known and the matrix $R$ is positive-definite.

The adversary performs his malicious attack during a short period $\tau_a = [k_0, k_0 + L - 1]$, where $k_0$ is the unknown attack instant and $L$ is the attack duration, assumed to be known. For the detection-isolation problem, there are $K$ distinct (isolated) attack profiles associated with possible attack scenarios. The attack vector $a_k$ is then described as follows:

$$a_k = \begin{cases} 0 & \text{if} \quad k < k_0 \\ \theta_{k-k_0+1}\,(l) & \text{if} \quad k_0 \leq k < k_0 + L\,, \\ 0 & \text{if} \quad k \geq k_0 + L \end{cases} \tag{4.2}$$

where $l$, for $1 \leq l \leq K$, is the attack type and $K$ is the number of transient hypotheses. The attack profiles $\theta_1\,(l)\,, \theta_2\,(l)\,, \cdots, \theta_L\,(l)$ of type $l$, for $1 \leq l \leq K$, are assumed to be completely known.

**Definition 4.1.** *A change detection-isolation algorithm has to compute a pair $(T, \nu)$ based on the observations $y_1, y_2, \cdots$, where $T > 0$ is the stopping time at which the final decision $\nu$, with $1 \leq \nu \leq K$, is decided.*

This chapter is dedicated to designing a detection-isolation procedure $\delta = (T, \nu)$ for jointly detecting and isolating the transient signals modeled in (4.2) in the discrete-time state space model described in (4.1) subject to certain criteria of optimality.

### 4.2.2 Criterion of optimality

There are several criteria for evaluating the performance of a change detection-isolation algorithm. Traditional quickest change detection-isolation criteria involve the minimization of the mean detection-isolation delay under the constraint on the false alarm and/or false isolation rates (see, for example, [104, 128–130, 132]). For safety-critical applications (see, for example, [127]), it is essential to minimize the worst-case probability of missed detection subject to acceptable levels on the risks of false alarm and/or false isolation.

For the transient change detection-isolation problem, there are four scenarios (see also figure 4.1):

- *False alarm*: The change is declared (i.e., detected and isolated) before its occurrence (i.e., $T \leq k_0$). Similar to the quickest change detection-isolation problem, the false alarm rate can be measured by either the ARL to false alarm or the probability of false alarm within any time window of predefined length (see figure 4.1).

- *False isolation*: The change is detected within the transient change window (i.e., $k_0 \leq T < k_0 + L$) but it is incorrectly classified. For example, the procedure $\delta = (T, \nu)$ in figure 4.1 raises the alarm $T \in [k_0, k_0 + L - 1]$ but the final decision $\nu = 3$ while the true change type is $l = 1$ (i.e., $\nu \neq l$). The false isolation rate should be measured by the probability of false isolation within the transient change window.

- *Correct detection and isolation*: The change is detected within the transient change window (i.e., $k_0 \leq T < k_0 + L$) and it is correctly classified. For example, the procedure $\delta = (T, \nu)$

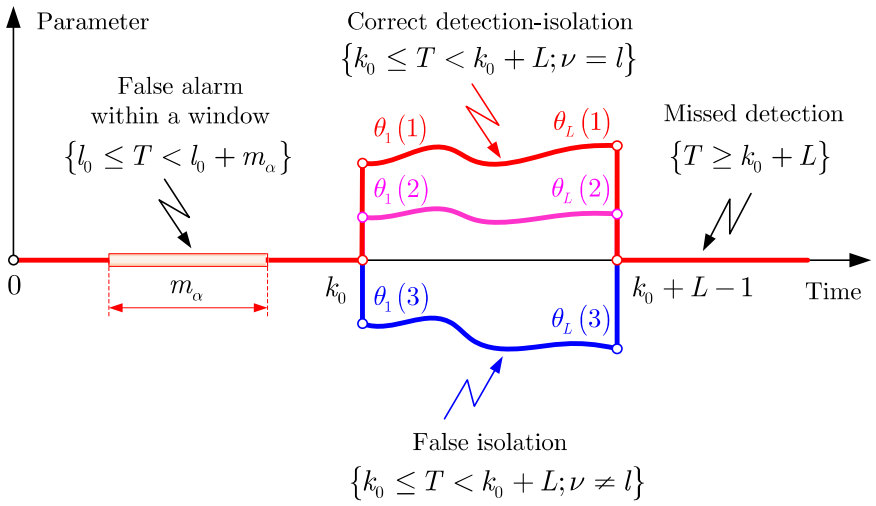



Figure 4.1 – Transient change detection-isolation problem.

in figure 4.1 raises the alarm $T \in [k_0, k_0 + L - 1]$ and the final decision $\nu = 1$ while the true change type is $l = 1$ (i.e., $\nu = l$). The correct detection-isolation rate should also be gauged by the probability of correct detection-isolation within the transient change window.

- *Missed detection*: The change is declared after its disappearance (i.e., $T \ge k_0 + L$). Similar to the detection problem, the missed detection rate should be evaluated by the probability of missed detection, i.e., the probability of detecting and isolating the transient signal after its disappearance.

Following the above analysis, we propose in this manuscript a novel optimality criterion for the transient change detection-isolation problem. The criterion of optimality involves the minimization of the worst-case probability of missed detection subject to acceptable levels on the worst-case probability of false alarm within any time window of predefined length and the worst-case probability of false isolation within the transient change window. The mathematical formulation of such an optimality criterion is given in the following.

Let $\overline{\mathbb{P}}_{\text{md}}(T; L)$ be the worst-case probability of missed detection, $\overline{\mathbb{P}}_{\text{fa}}(T; m_\alpha)$ be the worst-case probability of false alarm within any time window of length $m_\alpha$ and $\overline{\mathbb{P}}_{\text{fi}}(T; L)$ be the worst-case probability of false isolation within the transient change window. Similar to the transient change detection problem, let us assume that the change does not occur before the "preheating" period (i.e., $k_0 \ge L$) and the detection-isolation procedure does not operate in this period (i.e., $k \ge L$). The false alarm and false isolation rates are defined mathematically, respectively, as

$$\overline{\mathbb{P}}_{\text{fa}}(T; m_\alpha) = \sup_{l_0 \ge L} \mathbb{P}_0(l_0 \le T < l_0 + m_\alpha), \tag{4.3}$$

$$\overline{\mathbb{P}}_{\text{fi}}(T; L) = \sup_{k_0 \ge L} \max_{1 \le l \le K} \mathbb{P}^l_{k_0}(k_0 \le T < k_0 + L; \nu \ne l), \tag{4.4}$$

where $\mathbb{P}_0$ denotes the probability under the pre-change mode and $\mathbb{P}^l_{k_0}$ stands for the probability under the change-point $k_0$ and the change-type $l$.

The criterion of optimality involves the minimization of the following worst-case conditional probability of missed detection:

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(T;L\right)=\sup_{k_0\geq L}\max_{1\leq l\leq K}\mathbb{P}_{k_0}^{l}\left(T-k_0+1>L|T\geq k_0\right) \tag{4.5}$$

among all stopping times $T$ in the class $C_\alpha$ satisfying

$$C_\alpha=\left\{T:\overline{\mathbb{P}}_{\mathrm{fa}}\left(T;m_\alpha\right)\leq\alpha;\overline{\mathbb{P}}_{\mathrm{fi}}\left(T;L\right)\leq\alpha\right\}, \tag{4.6}$$

where $\alpha\in(0,1)$ denotes an acceptable level on the false alarm and false isolation rates.

## 4.3 Residual Generation Methods

Both classical residual-generation techniques, steady-state Kalman filter and fixed-size parity space, are utilized for generating the sequence of residuals. The unified statistical model adapted to the transient change detection-isolation problem is also developed.

### 4.3.1 Steady-state Kalman filter approach

In this subsection, we develop the statistical model of residuals generated by the steady-state Kalman filter approach. Similar to the detection problem, let us assume that the steady-state Kalman filter is used for generating the sequence of residuals. The steady-state Kalman gain $K_\infty$ is calculated as

$$K_\infty=P_\infty C^T\left(CP_\infty C^T+R\right)^{-1}, \tag{4.7}$$

where $P_\infty$ denotes the steady-state covariance matrix of the state estimation error, which can be found by solving the following discrete-time algebraic Riccati equation:

$$P_\infty=AP_\infty A^T-AP_\infty C^T\left(CP_\infty C^T+R\right)^{-1}CP_\infty A^T+Q. \tag{4.8}$$

The operation of the steady-state Kalman filter is described by the following equations:

$$\begin{cases}\hat{x}_{k+1|k} & =A\hat{x}_{k|k-1}+Bu_k+Fd_k+AK_\infty\left(y_k-\hat{y}_{k|k-1}\right)\\ \hat{y}_{k|k-1} & =C\hat{x}_{k|k-1}+Du_k+Gd_k\end{cases},\quad \hat{x}_{1|0}=\overline{x}_1, \tag{4.9}$$

where $\hat{x}_{k|k-1}\in\mathbb{R}^n$ is state estimate and $\hat{y}_{k|k-1}\in\mathbb{R}^p$ is the output estimate.

Let $\{\varrho_k\}_{k\geq1}\in\mathbb{R}^p$ be a sequence of independent identically distributed (i.i.d.) zero-mean Gaussian random vectors with covariance matrix $J\triangleq CP_\infty C^T+R$ and $r_k=y_k-\hat{y}_{k|k-1}\in\mathbb{R}^p$ be the innovations (or the residuals). The statistical model of the innovations is described as

$$r_k=\begin{cases}\varrho_k & \text{if}\quad k<k_0\\ \psi_{k-k_0+1}\left(l\right)+\varrho_k & \text{if}\quad k_0\leq k<k_0+L\\ \tilde{\psi}_k\left(l\right)+\varrho_k & \text{if}\quad k\geq k_0+L\end{cases}, \tag{4.10}$$

where transient profiles $\psi_1\left(l\right),\psi_2\left(l\right),\cdots,\psi_L\left(l\right)\in\mathbb{R}^p$ are calculated from the attack profiles $\theta_1\left(l\right),\theta_2\left(l\right),\cdots,\theta_L\left(l\right)$ of type $l$ by the following equation:

$$\begin{cases}\epsilon_{k+1} & =\left(A-AK_\infty C\right)\epsilon_k+\left(B_a-AK_\infty D_a\right)\theta_k\left(l\right)\\ \psi_k\left(l\right) & =C\epsilon_k+D_a\theta_k\left(l\right)\end{cases};\quad \epsilon_1=0, \tag{4.11}$$

and the post-change profiles $\tilde{\psi}_k(l)$ (i.e., for $k \geq k_0 + L$) are of no interest.

Similar to the transient change detection problem, let $r_{k-L+1}^k = \left[r_{k-L+1}^T, \cdots, r_k^T\right]^T \in \mathbb{R}^{Lp}$ be the concatenated vector of innovations, $\varrho_{k-L+1}^k = \left[\varrho_{k-L+1}^T, \cdots, \varrho_k^T\right] \in \mathbb{R}^{Lp}$ be the concatenated vector of random noises, and $\psi_{k-L+1}^k(k_0, l) \in \mathbb{R}^{Lp}$ be the concatenated vector of transient signals, depending on the relative position of the change-point $k_0$ within the window $[k-L+1, k]$ and the change-type $l$ by the following relation:

$$\psi_{k-L+1}^k(k_0, l) = \begin{cases} [0] & \text{if} \quad k < k_0 \\ \begin{bmatrix} [0] \\ \psi_1(l) \\ \vdots \\ \psi_{k-k_0+1}(l) \end{bmatrix} & \text{if} \quad k_0 \leq k < k_0 + L \\ \left[\tilde{\psi}_{k-L+1}^k(k_0, l)\right] & \text{if} \quad k \geq L \end{cases}, \tag{4.12}$$

where the vector of post-change profiles $\tilde{\psi}_{k-L+1}^k(k_0, l) \in \mathbb{R}^{Lp}$ is of no interest. Putting together (4.10)–(4.12), the statistical model of the residual vector $r_{k-L+1}^k$ generated by the steady-state Kalman filter is described as

$$r_{k-L+1}^k = \psi_{k-L+1}^k(k_0, l) + \varrho_{k-L+1}^k, \tag{4.13}$$

where the random noises $\varrho_{k-L+1}^k \sim \mathcal{N}(0, \Sigma_\varrho)$, where $\Sigma_\varrho = \mathrm{diag}(J) \in \mathbb{R}^{Lp \times Lp}$ is a block-diagonal matrix formed of blocks $J$.

### 4.3.2 Fixed-size parity space approach

In this subsection, we develop the statistical model of residuals generated by the fixed-size parity space approach. Similar to the detection problem, the observation model obtained by grouping the last $L$ measurements is described as

$$\begin{aligned} \underbrace{\begin{bmatrix} y_{k-L+1} \\ y_{k-L+2} \\ \vdots \\ y_k \end{bmatrix}}_{y_{k-L+1}^k} &= \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{L-1} \end{bmatrix}}_{\mathcal{C}} x_{k-L+1} + \underbrace{\begin{bmatrix} D_a & 0 & \cdots & 0 \\ CB_a & D_a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2}B_a & CA^{L-3}B_a & \cdots & D_a \end{bmatrix}}_{\mathcal{M}} \underbrace{\begin{bmatrix} a_{k-L+1} \\ a_{k-L+2} \\ \vdots \\ a_k \end{bmatrix}}_{\theta_{k-L+1}^k(k_0,l)} + \\ &\underbrace{\begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2}B & CA^{L-3}B & \cdots & D \end{bmatrix}}_{\mathcal{D}} \underbrace{\begin{bmatrix} u_{k-L+1} \\ u_{k-L+2} \\ \vdots \\ u_k \end{bmatrix}}_{u_{k-L+1}^k} + \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ C & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2} & CA^{L-3} & \cdots & 0 \end{bmatrix}}_{\mathcal{H}} \underbrace{\begin{bmatrix} w_{k-L+1} \\ w_{k-L+2} \\ \vdots \\ w_k \end{bmatrix}}_{w_{k-L+1}^k} + \\ &\underbrace{\begin{bmatrix} G & 0 & \cdots & 0 \\ CF & G & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2}F & CA^{L-3}F & \cdots & G \end{bmatrix}}_{\mathcal{G}} \underbrace{\begin{bmatrix} d_{k-L+1} \\ d_{k-L+2} \\ \vdots \\ d_k \end{bmatrix}}_{d_{k-L+1}^k} + \underbrace{\begin{bmatrix} v_{k-L+1} \\ v_{k-L+2} \\ \vdots \\ v_k \end{bmatrix}}_{v_{k-L+1}^k}, \end{aligned} \tag{4.14}$$

or in a simpler form as

$$y_{k-L+1}^{k} = \mathcal{C}x_{k-L+1} + \mathcal{D}u_{k-L+1}^{k} + \mathcal{G}d_{k-L+1}^{k} + \mathcal{M}\theta_{k-L+1}^{k}(k_0, l) + \mathcal{H}w_{k-L+1}^{k} + v_{k-L+1}^{k}, \quad (4.15)$$

where $y_{k-L+1}^{k} \in \mathbb{R}^{Lp}$ is the concatenated vector of measurements, $u_{k-L+1}^{k} \in \mathbb{R}^{Lm}$ is the concatenated vector of control signals, $d_{k-L+1}^{k} \in \mathbb{R}^{Lq}$ is the concatenated vector of disturbances, $w_{k-L+1}^{k} \in \mathbb{R}^{Ln}$ is the concatenated vector of process noises, $v_{k-L+1}^{k} \in \mathbb{R}^{Lp}$ is the concatenated vector of sensor noises, $\theta_{k-L+1}^{k}(k_0, l) \in \mathbb{R}^{Ls}$ is the concatenated vector of transient signals depending on the change-point $k_0$ and the change-type $l$; the matrices $\mathcal{C} \in \mathbb{R}^{Lp\times n}$, $\mathcal{D} \in \mathbb{R}^{Lp\times Lm}$, $\mathcal{G} \in \mathbb{R}^{Lp\times Lq}$, $\mathcal{H} \in \mathbb{R}^{Lp\times Ln}$ and $\mathcal{M} \in \mathbb{R}^{Lp\times Ls}$. The process noises $w_{k-L+1}^{k} \sim \mathcal{N}(0, \mathcal{Q})$ and the sensor noises $v_{k-L+1}^{k} \sim \mathcal{N}(0, \mathcal{R})$, where $\mathcal{Q} = \operatorname{diag}(Q) \in \mathbb{R}^{Ln\times Ln}$ and $\mathcal{R} = \operatorname{diag}(R) \in \mathbb{R}^{Lp\times Lp}$ are block-diagonal matrices formed of blocks $Q$ and $R$, respectively. Let also $\eta_{k-L+1}^{k} = \mathcal{H}w_{k-L+1}^{k} + v_{k-L+1}^{k}$ be concatenated vector of random noises, integrating both process noises and sensor noises. It is clear that $\eta_{k-L+1}^{k} \sim \mathcal{N}(0, \mathcal{S})$, where the covariance matrix $\mathcal{S} = \mathcal{H}\mathcal{Q}\mathcal{H}^T + \mathcal{R} \in \mathbb{R}^{Lp\times Lp}$ is symmetric and positive-definite.

The concatenated vector of attack profiles $\theta_{k-L+1}^{k}(k_0, l)$, depending on the relative position of the change-point $k_0$ within the window $[k-L+1, k]$ and the change type $l$, is described as

$$\theta_{k-L+1}^{k}(k_0, l) = \begin{cases} [0] & \text{if} \quad k < k_0 \\ \begin{bmatrix} [0] \\ \theta_1(l) \\ \vdots \\ \theta_{k-k_0+1}(l) \end{bmatrix} & \text{if} \quad k_0 \le k < k_0 + L \\ \left[\tilde{\theta}_{k-L+1}^{k}(k_0, l)\right] & \text{if} \quad k \ge k_0 + L \end{cases}, \quad (4.16)$$

where the post-change profiles $\tilde{\theta}_{k-L+1}^{k}(k_0, l) \in \mathbb{R}^{Ls}$ are of no interest.

Since the control signals $u_k$ and the disturbances $d_k$ are assumed to be exactly known, they can be eliminated from the observation model (4.14)–(4.15) by subtraction, leading to the following simplified statistical model:

$$z_{k-L+1}^{k} = y_{k-L+1}^{k} - \left(\mathcal{D}u_{k-L+1}^{k} + \mathcal{G}d_{k-L+1}^{k}\right) = \mathcal{C}x_{k-L+1} + \mathcal{M}\theta_{k-L+1}^{k}(k_0, l) + \eta_{k-L+1}^{k}, \quad (4.17)$$

The rejection of unknown system state vector $x_{k-L+1}$ can be performed in the same manner as it has been done in the detection problem. The simplified observation vector $z_{k-L+1}^{k}$ is projected onto the orthogonal complement space $R(\mathcal{C})^{\perp}$ of the column space $R(\mathcal{C})$ of matrix $\mathcal{C}$ (i.e., the left-null space of matrix $\mathcal{C}$), which is assumed to be full column rank (i.e., $\operatorname{rank}(\mathcal{C}) = n$). The residual vector is calculated as $r_{k-L+1}^{k} = \mathcal{W}z_{k-L+1}^{k}$, where the rows of the matrix $\mathcal{W} \in \mathbb{R}^{(Lp-n)\times Lp}$ are composed of the eigenvectors of the projection matrix $\mathcal{P}_{\mathcal{C}}^{\perp} = \mathcal{I} - \mathcal{C}\left(\mathcal{C}^T\mathcal{C}\right)^{-1}\mathcal{C}^T$ corresponding to eigenvalue 1, where $\mathcal{I}$ is the identity matrix of appropriate dimension. The rejection matrix $\mathcal{W}$ satisfies the following conditions: $\mathcal{W}\mathcal{C} = 0$, $\mathcal{W}^T\mathcal{W} = \mathcal{P}_{\mathcal{C}}^{\perp}$ and $\mathcal{W}\mathcal{W}^T = \mathcal{I}$. The statistical model of the residuals, which are independent from the nuisance parameter $x_{k-L+1}$, can be described as

$$r_{k-L+1}^{k} = \varphi_{k-L+1}^{k}(k_0, l) + \varsigma_{k-L+1}^{k}, \quad (4.18)$$

where the transient profiles $\varphi_{k-L+1}^{k}(k_0, l) = \mathcal{W}\mathcal{M}\theta_{k-L+1}^{k}(k_0, l)$ and the random noises $\varsigma_{k-L+1}^{k} = \mathcal{W}\eta_{k-L+1}^{k}$, thus satisfying $\varsigma_{k-L+1}^{k} \sim \mathcal{N}(0, \Sigma_\varsigma)$, where the covariance matrix $\Sigma_\varsigma = \mathcal{W}\left(\mathcal{H}\mathcal{Q}\mathcal{H}^T + \mathcal{R}\right)\mathcal{W}^T$ is positive-definite.

### 4.3.3 Unified statistical model

In this subsection, we develop the unified statistical model of the residuals generated by either steady-state Kalman filter approach or fixed-size parity space approach. It follows from (4.13) and (4.18) that both residual-generation methods lead to the following unified statistical model:

$$r_{k-L+1}^{k} = \phi_{k-L+1}^{k}\left(k_0, l\right) + \xi_{k-L+1}^{k}, \tag{4.19}$$

where $r_{k-L+1}^{k}$ is the vector of residuals, $\phi_{k-L+1}^{k}\left(k_0, l\right)$ is the vector of transient signals and $\xi_{k-L+1}^{k} \sim \mathcal{N}\left(0, \Sigma\right)$ is the vector of random noises. For the steady-state Kalman filter approach, the transient profiles $\phi_{k-L+1}^{k}\left(k_0, l\right) = \psi_{k-L+1}^{k}\left(k_0, l\right)$ and the random noises $\xi_{k-L+1}^{k} = \varrho_{k-L+1}^{k}$ (i.e., $\Sigma = \Sigma_{\varrho}$). On the other hand, the transient profiles $\phi_{k-L+1}^{k}\left(k_0, l\right) = \varphi_{k-L+1}^{k}\left(k_0, l\right)$ and the random noises $\xi_{k-L+1}^{k} = \varsigma_{k-L+1}^{k}$ (i.e., $\Sigma = \Sigma_{\varsigma}$) for the fixed-size parity space approach.

Similar to the detection problem, the Kullback-Leibler (K-L) distance is employed for comparing two residual-generation methods (i.e., the steady-state Kalman filter ad the fixed-size parity space approaches). Let $\mathcal{P}_{k_0}^{l}$ (resp. $\mathcal{P}_0 \triangleq \mathcal{P}_\infty \triangleq \mathcal{P}_{k_0}^{0}$) be the probability measure when the sequence of residuals $r_1^L, r_2^{L+1}, \cdots, r_{k-L+1}^{k}, \cdots$ follows the unified statistical model (4.19), $\mathbb{E}_{k_0}^{l}$ (resp. $\mathbb{E}_0 \triangleq \mathbb{E}_\infty \triangleq \mathbb{E}_{k_0}^{0}$) denote the corresponding mathematical expectations, and $p_{k_0}^{l}$ (resp. $p_0 \triangleq p_\infty \triangleq p_{k_0}^{0}$) stand for the corresponding probability density function.

Without loss of generality, let us assume that the change-point $k_0 = 1$. The K-L distance $\rho\left(j, l\right)$ between $\mathcal{P}_1^j$ and $\mathcal{P}_1^l$, for $0 \leq j \neq l \leq K$, is defined as

$$\rho\left(j, l\right) = \int_{-\infty}^{+\infty} p_1^j\left(r_1^L\right) \log \frac{p_1^j\left(r_1^L\right)}{p_1^l\left(r_1^L\right)} dr_1^L, \tag{4.20}$$

where the residual vector $r_1^L \sim \mathcal{N}\left(\phi_1^L\left(1, l\right), \Sigma\right)$ under the probability measure $\mathcal{P}_1^l$, for $0 \leq l \leq K$. For the Gaussian case, the K-L distances obtained by the steady-state Kalman filter approach and the fixed-size parity space approach are calculated as

$$\rho_{\mathrm{KF}}\left(j, l\right) = \frac{1}{2}\left[\psi_1^L\left(1, l\right) - \psi_1^L\left(1, j\right)\right]^T \left[\Sigma_{\varrho}^{-1}\right] \left[\psi_1^L\left(1, l\right) - \psi_1^L\left(1, j\right)\right], \tag{4.21}$$

$$\rho_{\mathrm{PS}}\left(j, l\right) = \frac{1}{2}\left[\varphi_1^L\left(1, l\right) - \varphi_1^L\left(1, j\right)\right]^T \left[\Sigma_{\varsigma}^{-1}\right] \left[\varphi_1^L\left(1, l\right) - \varphi_1^L\left(1, j\right)\right], \tag{4.22}$$

where $\rho_{\mathrm{KF}}\left(j, l\right)$ and $\rho_{\mathrm{PS}}\left(j, l\right)$ are the K-L distances between $\mathcal{P}_1^j$ and $\mathcal{P}_1^l$ of the residuals generated by the steady-state Kalman filter and the fixed-size parity space, respectively.

**Remark 4.1.** *By exploiting the results of Lemma 3.1, it can be shown that the K-L distances $\rho_{PS}(j,l)$ between $\mathcal{P}_1^j$ and $\mathcal{P}_1^l$, for $0 \leq j \neq l \leq K$, are independent from the choice of rejection matrix $\mathcal{W}$. The comparison between the steady-state Kalman filter approach and the fixed-size parity space approach for the transient change detection-isolation problem will be performed by simulation in chapter 6.*

## 4.4 Detection-isolation Algorithms

In this section, we design several detection-isolation schemes for jointly detecting and isolating transient changes in the unified statistical model (4.19). By generalizing the traditional CUSUM-based algorithms (i.e., see subsection 2.4.3), we propose several detection-isolation schemes (i.e., generalized, matrix and vector Window Limited (WL) CUSUM algorithms). In addition, the FMA detection-isolation rule will be also introduced.

### 4.4.1 Generalized WL CUSUM algorithm

For the joint detection and isolation problem, Nikiforov [130] and Lai [104] have introduced, respectively, the generalized CUSUM test and the generalized WL CUSUM test (see also subsection 2.4.3). Let us define directly the generalized WL CUSUM algorithm $\delta_{\mathrm{GWL}} = (T_{\mathrm{GWL}}, \nu_{\mathrm{GWL}})$, which utilizes the last $L$ observations at each time instant $k \geq L$, as follows:

$$T_{\mathrm{GWL}} = \inf \left\{ k \geq L : \max_{1 \leq l \leq K} \max_{k-L+1 \leq i \leq k} \min_{0 \leq j \neq l \leq K} \left( S_i^k (l,j) - h \right) \geq 0 \right\}, \tag{4.23}$$

$$\nu_{\mathrm{GWL}} = \arg \max_{1 \leq l \leq K} \max_{T_{\mathrm{GWL}}-L+1 \leq i \leq T_{\mathrm{GWL}}} \min_{0 \leq j \neq l \leq K} S_i^{T_{\mathrm{GWL}}} (l,j), \tag{4.24}$$

where $h$ is the chosen threshold and $S_i^k (l,j)$, for $k-L+1 \leq i \leq k$, $1 \leq l \leq K$ and $0 \leq j \neq l \leq K$, is the log-likelihood ratio (LLR), which is calculated in the Gaussian case as

$$S_i^k (l,j) = \left[ \left( \phi_{k-L+1}^k (i,l) - \phi_{k-L+1}^k (i,j) \right) \right]^T \left[ \Sigma^{-1} \right] \left[ r_{k-L+1}^k - \frac{\phi_{k-L+1}^k (i,l) + \phi_{k-L+1}^k (i,j)}{2} \right]. \tag{4.25}$$

The generalized WL CUSUM algorithm (4.23)–(4.24) proceeds as follows. For each time instant $k \geq L$, the generalized WL CUSUM algorithm uses a block of $L$ last measurements $y_{k-L+1}, \cdots, y_k$ for decision-making. Firstly, the unified statistical model (4.19) is formulated by either the steady-state Kalman filter approach or the fixed-size parity space approach. Secondly, for each time index $i$ from $k-L+1$ to $k$, the LLRs $S_i^k (l,j)$, for all $1 \leq l \leq K$ and $0 \leq j \neq l \leq K$, are calculated. The alarm time $T_{\mathrm{GWL}}$ is raised if there exists such $l$, for $1 \leq l \leq K$, that for some $i \in [k-L+1, k]$, all LLRs $S_i^k (l,j)$, for $0 \leq j \neq l \leq K$, are greater than or equal to the threshold $h$.

### 4.4.2 Matrix WL CUSUM algorithm

The matrix CUSUM algorithm was first introduced in [138] by revising the generalized CUSUM algorithm to obtain the recursive form. Let us define directly the matrix WL CUSUM algorithm $\delta_{\mathrm{MWL}} = (T_{\mathrm{MWL}}, \nu_{\mathrm{MWL}})$, which utilizes the last $L$ observations at each time instant $k \geq L$, as follows:

$$T_{\mathrm{MWL}} = \inf \left\{ k \geq L : \max_{1 \leq l \leq K} \min_{0 \leq j \neq l \leq K} \max_{k-L+1 \leq i \leq k} \left( S_i^k (l,j) - h \right) \geq 0 \right\}, \tag{4.26}$$

$$\nu_{\mathrm{MWL}} = \arg \max_{1 \leq l \leq K} \min_{0 \leq j \neq l \leq K} \max_{T_{\mathrm{MWL}}-L+1 \leq i \leq T_{\mathrm{MWL}}} S_i^{T_{\mathrm{MWL}}} (l,j), \tag{4.27}$$

where $h$ is the chosen threshold and the LLRs $S_i^k (l,j)$, for $k-L+1 \leq i \leq k$, $1 \leq l \leq K$ and $0 \leq j \neq l \leq K$, are calculated in (4.25).

**Remark 4.2.** *The matrix WL CUSUM algorithm* (4.26)–(4.27) *proceeds in the same manner as the generalized WL CUSUM algorithm* (4.23)–(4.24) *except for the replacement of the "max-min" operation in* (4.23)–(4.24) *by the "min-max" operation in* (4.26)–(4.27).

### 4.4.3 Vector WL CUSUM algorithm

The vector WL CUSUM algorithm is obtained by replacing the statistic $\max_{k-L+1 \leq i \leq k} S_i^k (l,j)$ in the matrix WL CUSUM algorithm (4.26)–(4.27) by the following statistic:

$$g_k (l,j) = \max_{k-L+1 \leq i \leq k} S_i^k (l,0) - \max_{k-L+1 \leq i \leq k} S_i^k (j,0). \tag{4.28}$$

The vector WL CUSUM algorithm $\delta_{\mathrm{VWL}} = (T_{\mathrm{VWL}}, \nu_{\mathrm{VWL}})$ is then defined as follows:

$$T_{\mathrm{VWL}} = \inf\left\{k \geq L : \max_{1\leq l\leq K} \min_{0\leq j\neq l\leq K} \left(g_k\left(l,j\right) - h\right) \geq 0\right\}, \tag{4.29}$$

$$\nu_{\mathrm{VWL}} = \arg \max_{1\leq l\leq K} \min_{0\leq j\neq l\leq K} g_{T_{\mathrm{VWL}}}\left(l,j\right), \tag{4.30}$$

where $h$ is the chosen threshold and the LLRs $S_i^k\left(l,j\right)$, for $k-L+1 \leq i \leq k$, $1 \leq l \leq K$ and $0 \leq j \neq l \leq K$, are calculated in (4.25).

### 4.4.4 FMA detection-isolation rule

The FMA detection-isolation rule $\delta_{\mathrm{FMA}}\left(T_{\mathrm{FMA}}; \nu_{\mathrm{FMA}}\right)$, which is the FMA version of the generalized WL CUSUM, the matrix WL CUSUM and the vector WL CUSUM algorithms, can be described as

$$T_{\mathrm{FMA}} = \inf\left\{k \geq L : \max_{1\leq l\leq K} \min_{0\leq j\neq l\leq K} \left(S_{k-L+1}^k\left(l,j\right) - h\right) \geq 0\right\}, \tag{4.31}$$

$$\nu_{\mathrm{FMA}} = \arg \max_{1\leq l\leq K} \min_{0\leq j\neq l\leq K} S_{T_{\mathrm{FMA}}-L+1}^{T_{\mathrm{FMA}}}\left(l,j\right), \tag{4.32}$$

where $h$ is the chosen threshold and the LLRs $S_{k-L+1}^k\left(l,j\right)$, for $1 \leq l \leq K$ and $0 \leq j \neq l \leq K$, are calculated in (4.25).

**Remark 4.3.** *It can be seen that the FMA detection-isolation rule* (4.31)–(4.32) *is the generalization of the FMA detection rule* (3.41) *for the detection problem. The detection-isolation rule* (4.31)–(4.32) *is also the FMA version of the generalized WL CUSUM algorithm* (4.23)–(4.24)*, the matrix WL CUSUM algorithm* (4.26)–(4.27)*, and the vector CUSUM algorithm* (4.29)–(4.30)*. The statistical properties of the FMA detection-isolation rule* (4.31)–(4.32) *will be investigated in the following subsection.*

### 4.4.5 Statistical properties of FMA detection-isolation rule

In this section, we investigate the statistical performance of the FMA detection rule (4.31)–(4.32). Especially, we calculate the upper bound on the worst-case probability of false alarm, the upper bound on the worst-case probability of false isolation and the upper bound on the worst-case probability of missed detection. Main results are given in Theorem 4.1.

**Theorem 4.1.** *Consider the FMA detection rule given in* (4.31)–(4.32)*. Let* $\tilde{\mathbb{P}}_{\mathrm{fa}}$*,* $\tilde{\mathbb{P}}_{\mathrm{fi}}$ *and* $\tilde{\mathbb{P}}_{\mathrm{md}}$ *be, respectively, the upper bounds for* $\overline{\mathbb{P}}_{\mathrm{fa}}\left(\delta_{\mathrm{FMA}}\right)$*,* $\overline{\mathbb{P}}_{\mathrm{fi}}\left(\delta_{\mathrm{FMA}}\right)$ *and* $\overline{\mathbb{P}}_{\mathrm{md}}\left(\delta_{\mathrm{FMA}}\right)$*. Then,*

*1. The worst-case probability of false alarm within any time window of length* $m_\alpha$ *corresponds to the first time window* $[L; L+m_\alpha-1]$*, i.e.,*

$$\overline{\mathbb{P}}_{\mathrm{fa}}\left(\delta_{\mathrm{FMA}}; m_\alpha; h\right) = \mathbb{P}_0\left(L \leq T_{\mathrm{FMA}} \leq L+m_\alpha-1\right), \tag{4.33}$$

*and it is upper bounded by*

$$\overline{\mathbb{P}}_{\mathrm{fa}}\left(\delta_{\mathrm{FMA}}; m_\alpha; h\right) \leq \tilde{\mathbb{P}}_{fa}\left(\delta_{FMA}; m_\alpha; h\right) \triangleq 1 - \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \bigcap_{l=1}^{K} \left\{S_{k-L+1}^k\left(l,0\right) < h\right\}\right). \tag{4.34}$$

*2. The worst-case probability of false isolation within any transient change window corresponds to the first time window* $[L; 2L-1]$*, i.e.,*

$$\overline{\mathbb{P}}_{\text{fi}}\left(\delta_{\text{FMA}}; L; h\right) = \max_{1\leq l\leq K} \mathbb{P}_L^l\left(L \leq T_{\text{FMA}} < 2L; \nu_{\text{FMA}} \neq l\right), \tag{4.35}$$

*and it is upper bounded for the case of threshold* $h \geq 0$ *as*

$$\overline{\mathbb{P}}_{\text{fi}}\left(\delta_{\text{FMA}}; L; h\right) \leq \tilde{\mathbb{P}}_{fi}\left(\delta_{FMA}; L; h\right) \triangleq \max_{1\leq l\leq K}\left[1 - \max_{0\leq \tilde{j}\leq K} \mathbb{P}_L^l\left(\bigcap_{k=L}^{2L-1}\bigcap_{\substack{j=1\\ j\neq \tilde{j}, l}}^{K}\left\{S_{k-L+1}^k\left(j, \tilde{j}\right) < h\right\}\right)\right]. \tag{4.36}$$

*3. The worst-case probability of missed detection is upper bounded by*

$$\overline{\mathbb{P}}_{\text{md}}\left(\delta_{\text{FMA}}; L; h\right) \leq \tilde{\mathbb{P}}_{md}\left(T_{FMA}; L; h\right) \triangleq \max_{1\leq l\leq K} \sum_{\substack{j=0\\ j\neq l}}^{K} \Phi\left(\frac{h - \mu_{S_1^L(l,j)}}{\sigma_{S_1^L(l,j)}}\right), \tag{4.37}$$

*where* $\mu_{S_1^L(l,j)}$ *and* $\sigma_{S_1^L(l,j)}$ *are calculated as*

$$\mu_{S_1^L(l,j)} = \frac{1}{2}\left[\phi_1^L\left(1,l\right) - \phi_1^L\left(1,j\right)\right]^T\left[\Sigma^{-1}\right]\left[\phi_1^L\left(1,l\right) - \phi_1^L\left(1,j\right)\right], \tag{4.38}$$

$$\sigma^2_{S_1^L(l,j)} = \left[\phi_1^L\left(1,l\right) - \phi_1^L\left(1,j\right)\right]^T\left[\Sigma^{-1}\right]\left[\phi_1^L\left(1,l\right) - \phi_1^L\left(1,j\right)\right]. \tag{4.39}$$

*Proof.* The proof of Theorem 4.1 is given in Appendix A.8. □

Let us add some comments on the results of Theorem 4.1. The upper bound $\tilde{\mathbb{P}}_{\text{md}}$ for the worst-case probability of missed detection can be calculated analytically. On the other hand, the upper bound $\tilde{\mathbb{P}}_{\text{fa}}$ for the worst-case probability of false alarm and the upper bound $\tilde{\mathbb{P}}_{\text{fi}}$ for the worst-case probability of false isolation can be estimated numerically by utilizing the same technique as in Proposition 3.1. In addition, the threshold $h$ can be such chosen that the upper bound $\tilde{\mathbb{P}}_{\text{fa}}\left(\delta_{\text{FMA}}; m_\alpha; h\right) \leq \alpha$ and the upper bound $\tilde{\mathbb{P}}_{\text{fi}}\left(\delta_{\text{FMA}}; L; h\right) \leq \alpha$, thus assuring the true worst-case error probabilities $\overline{\mathbb{P}}_{\text{fa}}\left(\delta_{\text{FMA}}\right) \leq \alpha$ and $\overline{\mathbb{P}}_{\text{fi}}\left(\delta_{\text{FMA}}\right) \leq \alpha$.

## 4.5 Conclusion

The attack detection-isolation problem has been formulated as the problem of jointly detecting and identifying transient changes in stochastic-dynamical systems. Similar to the detection problem, the discrete-time state space model driven by Gaussian noises is utilized to describe SCADA systems. The cyber-physical attacks are modeled as additive signals of short duration on both state evolution and sensor measurement equations. In order to eliminate the nuisance parameter, the steady-state Kalman filter and the fixed-size parity space are employed. For the detection-isolation problem, there are multiple attack types (i.e., or attack scenarios) where each attack kind produces a specific attack signature (i.e., or attack profile) after the residual-generation engine. It has been also shown that the utilization of both residual-generation methods leads to the unified statistical model which is then utilized for designing detection-isolation schemes.

In order to compare various detection-isolation algorithms, we have proposed a novel criterion of optimality which aims at minimizing the worst-case probability of missed detection subject

to acceptable levels on the worst-case probability of false alarm and the worst-case probability of false isolation. Several detection-isolation schemes have been adapted to the detection and isolation of transient changes. The FMA detection rule proposed in chapter 3 has been revised for jointly detecting and isolating transient changes in the unified statistical model. The upper bounds on the worst-case probability of false alarm, false isolation and missed detection have been introduced. Though no optimal (or sub-optimal) algorithms have been obtained, we have proposed a simple and efficient detection-isolation test. The comparison between different algorithms will be investigated by the Monte Carlo simulation in chapter 6.

# Part II

# Sequential Monitoring of SCADA Systems against Cyber-physical Attacks

In the first part, we have proposed several algorithms for detecting and isolating transient changes in stochastic-dynamical systems. The target of the second part is to apply the theoretical results to the sequential monitoring of SCADA systems against cyber-physical attacks. This part consists of two chapters. The models of SCADA systems and cyber-physical attacks are developed in chapter 5. Two safety-critical infrastructures, including a simple SCADA gas transmission pipeline and a simple SCADA water distribution network, are described in the discrete-time state space form driven by Gaussian noises. Several types of cyber-physical attacks, including DoS attacks, simple integrity attacks and stealthy integrity attacks are also considered.

The models of SCADA systems and cyber-physical attacks will be utilized in chapter 6 for demonstrating theoretical results obtained in chapter 3 and chapter 4. The negative impact of DoS attacks, simple integrity attacks and stealthy integrity attacks on closed-loop control systems will be demonstrated by performing these malicious attack strategies on the simple SCADA gas pipeline. Theoretical results obtained in chapter 3 (i.e., detection schemes) will be applied for detecting the covert attack strategy on the simple SCADA water distribution network. A more complex water network will be used for showing the performance of detection-isolation schemes proposed in chapter 4.

# Chapter 5

# Models of SCADA Systems and Cyber-physical Attacks

**Contents**



## 5.1 Introduction

The objective of this chapter is to develop the models of SCADA systems and cyber-physical attacks. Generally, the physical layer of almost SCADA systems can be described by a set of partial differential equations (PDEs). The PDEs can be also linearized around the operating points for obtaining the discrete-time state space model driven by Gaussian noises. The cyber-physical attacks are then modeled as additive signals of short duration impacting both system equations.

For the demonstration purpose, we develop in this chapter the models of a simple SCADA gas pipeline and a simple SCADA water distribution network. These geographically dispersed assets are generally controlled and monitored by the SCADA technology, becoming more and more susceptible to malicious attacks. Over the last decades, there has been an increasing

number of cyber incidents involving gas pipelines [152, 156] and water networks [33, 168, 213]. Therefore, the monitoring of these safety-critical infrastructures against malicious attacks plays an extremely important role in ensuring system normal operation and avoiding catastrophic consequences.

A great deal of effort has been devoted to the security of SCADA gas pipelines against cyber-physical attacks. For example, the vulnerabilities and protection measures for gas transmission and distribution systems have been considered in [23, 197]. Numerous techniques have been proposed to detect and isolate gas leaks in transmission and distribution pipelines [15, 125]. However, up to our best knowledge, the monitoring of SCADA gas pipelines against cyber-physical attacks has not been considered seriously.

Water distribution networks, on the other hand, have received much more attention from the research community. The surveillance of these safety-critical infrastructures can be classified into two categories: hydraulic surveillance and quality monitoring. The quality monitoring problem requires the detection and isolation of contaminants injected into water distribution networks. The problem of modeling contaminant dynamics and fault diagnosis in water distribution networks has been considered in [45]. Moreover, Guépié [67] has proposed sequential methods for monitoring the water quality based on the concentration of chlorine level in the water network. The hydraulic surveillance problem consists in developing the hydraulic model of the systems and then utilizing the fault detection-isolation techniques for detecting and identifying any faults occurring to the systems. This approach has been considered in [5–7] for studying the security of water irrigation canals against cyber-physical attacks.

This chapter is organized as follows. In section 5.2, we develop the models of a simple SCADA gas pipeline and several cyber attack scenarios on the gas pipeline. The models of a simple SCADA water network and cyber-physical attacks on the water network are derived in section 5.3. Finally, some concluding remarks are drawn in section 5.4.

## 5.2 Model of SCADA Gas Pipelines

In this section, we develop the model of a simple SCADA gas pipeline and several attack scenarios. The architecture of the gas pipeline is described in subsection 5.2.1. The components of the physical layer and the cyber layer are modeled in subsection 5.2.2 and subsection 5.2.3, respectively. The discrete-time state space model of the gas pipeline is described subsection 5.2.4. Finally, we consider in subsection 5.2.5 several attack scenarios on the gas pipeline.

### 5.2.1 System architecture

The architecture of a typical SCADA gas distribution network consists of two layers: the physical layer and the cyber layer [86, 90, 170]. The physical layer includes physical elements such as gas pipelines, compressors, valves and sensors. The co-operation of these components helps in transporting and distributing gas from production plants to final consumers. The cyber layer is comprised of Programmable Logic Controllers (PLCs), Remote Terminal Units (RTUs), Master Terminal Units (MTUs) and communication devices. In gas pipeline technology, local control algorithms are programmed in PLCs/RTUs and global control algorithms are implemented in MTUs. The communication between the MTUs and the PLCs/RTUs are normally carried out by wireless technology such as radio frequencies or satellites.

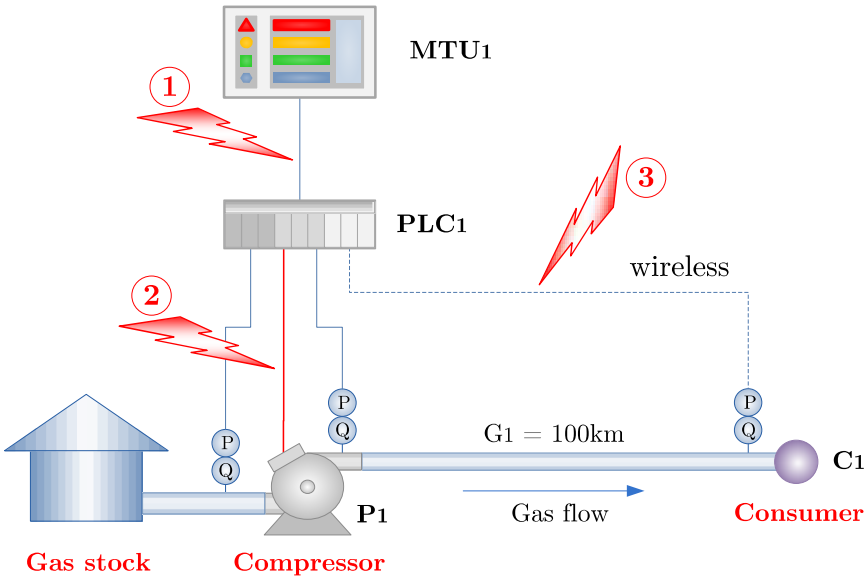



Figure 5.1 – A simple SCADA gas distribution network.

For the sake of simplicity, we study a simple SCADA gas pipeline as shown in figure 5.1. The physical layer consists of a gas pipeline $G_1$, a compressor $P_1$, a customer $C_1$, and pressure and flow rate sensors. The gas flow in the pipeline is controlled and monitored by the cyber layer comprised of the $\text{PLC}_1$ and the $\text{MTU}_1$. The $\text{PLC}_1$ is in charge of controlling the outlet pressure of the pipeline $G_1$ by regulating the speed of the compressor $P_1$ based on the set-point transmitted from the $\text{MTU}_1$.

### 5.2.2 Model of physical layer

In this subsection, we develop the model of each physical element of the SCADA gas pipeline described in figure 5.1. The physical components include the gas pipeline $G_1$, the compressor $P_1$, the customer $C_1$ and the pressure and flow rate sensors.

**Model of gas flow in a pipeline.** Under the isothermal conditions, transient gas flow through a pipeline is governed by the following set of partial differential equations (PDEs) [1, 13, 76, 96]:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u)}{\partial x} = 0, \tag{5.1}$$

$$\frac{\partial (\rho u)}{\partial t} + \frac{\partial (\rho u^2 + p)}{\partial x} = -\frac{\rho u |u|}{2D} f - \rho g \sin \alpha, \tag{5.2}$$

$$p = \rho Z R T, \tag{5.3}$$

where $\rho$ is the gas density, $p$ is the gas pressure, $u$ is the gas axial velocity, $g$ is the gravitational acceleration, $\alpha$ is the pipe inclination, $f$ is the friction factor, $Z$ is the gas compressibility factor, and $D$ is the pipeline diameter.

The variables of interest are the pressure $p(x,t)$ and the mass flow rate $q(x,t)$ at the position $x$ and the time $t$. The inputs to the model are the outlet flow rate $q_{\text{out}}(t) = q(L,t)$ and the inlet

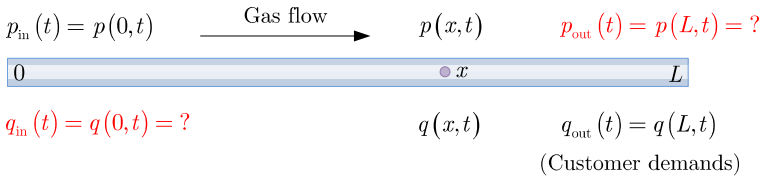



Figure 5.2 – Inputs ($p_{\text{in}}(t)$ and $q_{\text{out}}(t)$) and outputs ($p_{\text{out}}(t)$ and $q_{\text{in}}(t)$) of the gas pipeline model.

pressure $p_{\text{in}}(t) = p(0,t)$. The outlet flow rate $q_{\text{out}}(t)$ corresponds to the customer's demands and the inlet pressure $p_{\text{in}}(t)$ is equivalent to the pressure supplied by the compressor. It is required to calculate the outlet pressure $p_{\text{out}}(t) = p(L,t)$ and the inlet flow rate $q_{\text{in}}(t) = q(0,t)$ (see figure 5.2) from the inlet pressure $p_{\text{in}}(t)$, the outlet flow rate $q_{\text{out}}(t)$, the PDEs (5.1)–(5.3), and the initial conditions.

For an isothermal process [87], the following relation satisfies $p = c^2\rho$ and $q = \rho uA = \rho Q = \rho_n Q_n$, where $c$ is the speed of sound, $q$ is mass flow rate, $Q$ is the volumetric flow rate in the pipeline, $\rho_n$ and $Q_n$ are gas density and volumetric flow rate at standard conditions[9] and $A$ is the pipeline cross section and $A = \pi(D/2)^2$.

Some methods have been proposed for solving the PDEs (5.1)–(5.3), including the numerical methods [76, 87, 96, 137], the transfer function method [13] and the state space method [1]. The numerical methods appear inappropriate for the design of control schemes and monitoring algorithms. For these reasons, the transfer function method and the state space method, which are based on the linearized model of gas flow through a pipeline, are considered. The transfer function model is useful in designing control algorithms while the state space model has an advantage in developing monitoring schemes.

Putting together the PDEs (5.1)–(5.3), we obtain the following simplified PDEs:

$$\frac{1}{ZRT}\frac{\partial p}{\partial t} = -\frac{1}{A}\frac{\partial q}{\partial x}, \tag{5.4}$$

$$\frac{\partial p}{\partial x} + \frac{1}{A}\frac{\partial q}{\partial t} + \frac{ZRT}{A^2}\frac{\partial}{\partial x}\left(\frac{q^2}{p}\right) = -\frac{f}{2D}\frac{ZRT}{A^2}\frac{q\,|q|}{p} - g\sin\alpha\frac{p}{ZRT}. \tag{5.5}$$

For simplicity, let $p_0 = p(x,0)$ and $q_0 = q(x,0)$ be the initial pressure and the initial flow rate at a given position $x$ (i.e., the subscript $x$ is eliminated), respectively. Let also $u_0$ be the initial average gas velocity which is calculated [96] as

$$u_0 = \frac{(q_{\text{in}}(0) + q_{\text{out}}(0))\,ZRT}{(p_{\text{in}}(0) + p_{\text{out}}(0))\,A}, \tag{5.6}$$

where $p_{\text{in}}(0)$, $p_{\text{out}}(0)$, $q_{\text{in}}(0)$ and $q_{\text{out}}(0)$ are the initial values of inlet pressure, outlet pressure, inlet flow rate and outlet flow rate, respectively. Let us assume that $q_{\text{in}}(0) = q_{\text{out}}(0) = \rho_n Q_n(0)$. Then, the initial outlet pressure $p_{\text{out}}(0)$ can be calculated as in [76] by the following equation:

$$p_{\text{out}}(0) = \sqrt{p_{\text{in}}^2(0) - \frac{fL}{D}\left(\frac{2c\rho_n Q_n(0)}{A}\right)^2}. \tag{5.7}$$

[9] Standard conditions are sets of conditions on the temperature and pressure for comparing between different sets of data. The National Institute of Standards and Technology (NIST) uses the temperature of 20°C (293.15 K, 68 °F) and the absolute pressure of 101.325 kPa (14.696 psi, 1 atm).

In order to obtain the transfer function model, the PDEs (5.4)–(5.5) are linearized around the operating points. The variables $p_0$, $q_0$ and $u_0$ are considered as references and some dimensionless variables are defined [13] as follows:

$$x^* = \frac{x}{L}, \quad L^* = \frac{L}{D}, \quad t^* = \frac{tc}{L}, \quad p^* = \frac{p}{p_0}, \quad q^* = \frac{qc}{p_0 A}, \quad u^* = \frac{u_0}{c}. \tag{5.8}$$

The linearized model of transient gas flow through a pipeline is described in terms of dimensionless variables as

$$\frac{\partial \Delta q^*}{\partial x^*} = -\frac{\partial \Delta p^*}{\partial t^*}, \tag{5.9}$$

$$\left(1 - u^{*2}\right) \frac{\partial \Delta p^*}{\partial x^*} = -\frac{\partial \Delta q^*}{\partial t^*} + 2u^* \frac{\partial \Delta p^*}{\partial t^*} - |u^*| f L^* \Delta q^* + \left(\frac{f L^*}{2} u^* \left|u^*\right| - \frac{g \Delta h}{c^2}\right) \Delta p^*. \tag{5.10}$$

Since $u^* \ll 1$ for the practical subsonic transient flow, the component $u^{*2}$ at the left-hand side of (5.10) is omitted. Therefore, the Laplace transform of (5.9)–(5.10) leads to the following linear ordinary differential equations:

$$\frac{\partial \Delta q^*(s)}{\partial x^*} = -s \Delta p^*(s) \tag{5.11}$$

$$\frac{\partial \Delta p^*(s)}{\partial x^*} = -\left(u^* f L^* + s\right) \Delta q^*(s) + \left(\frac{f L^*}{2} u^* \left|u^*\right| - \frac{g \Delta h}{c^2} + 2u^* s\right) \Delta p^*(s) \tag{5.12}$$

By solving the equation (5.11)–(5.12) with the boundary conditions (i.e. the inlet pressure $\Delta p_{\text{in}}$ and the outlet flow rate $\Delta q_{\text{out}}$) and returning to their real values, we obtain that

$$\Delta p_{\text{out}}(s) = F_{p_{\text{out}} p_{\text{in}}} \Delta p_{\text{in}}(s) + F_{p_{\text{out}} q_{\text{out}}} \Delta q_{\text{out}}(s), \tag{5.13}$$

$$\Delta q_{\text{in}}(s) = F_{q_{\text{in}} p_{\text{in}}} \Delta p_{\text{in}}(s) + F_{q_{\text{in}} q_{\text{out}}} \Delta q_{\text{out}}(s), \tag{5.14}$$

where the transfer functions $F_{p_{\text{out}} p_{\text{in}}}(s)$, $F_{p_{\text{out}} q_{\text{out}}}(s)$, $F_{q_{\text{in}} p_{\text{in}}}(s)$ and $F_{q_{\text{in}} q_{\text{out}}}(s)$ are obtained by taking into account only the zero-order and the first-order differential components as

$$F_{p_{\text{out}} p_{\text{in}}}(s) = k_1 \frac{1}{a_1 s + 1} \quad \text{and} \quad F_{q_{\text{in}} p_{\text{in}}}(s) = \frac{c_1 s}{\hat{a}_1 s + 1}, \tag{5.15}$$

$$F_{p_{\text{out}} q_{\text{out}}}(s) = -k_2 \frac{b_1 s + 1}{\hat{a}_1 s + 1} \quad \text{and} \quad F_{q_{\text{in}} q_{\text{out}}}(s) = \frac{1}{d_1 s + 1}, \tag{5.16}$$

where the coefficients in the equations (5.15)–(5.16) can be found in [1, 13].

**Model of a compressor.** In gas distribution systems, compressors are used to increase the gas pressure at the inlet of a pipeline so that it can have enough energy to reach to its outlet. A simple model of a centrifugal compressor, comprised of a motor and a compressing chamber where the pressure is increased, is described in figure 5.3.

The increased pressure $\Delta p$ through the compressor depends on the inlet pressure, the outlet pressure, the inlet mass flow rate, the outlet mass flow rate, and the motor speed by a nonlinear relationship. More details about the characteristics of a centrifugal compressor can be found in [191]. For simplicity, let us assume that the increased pressure $\Delta p$ is controlled by

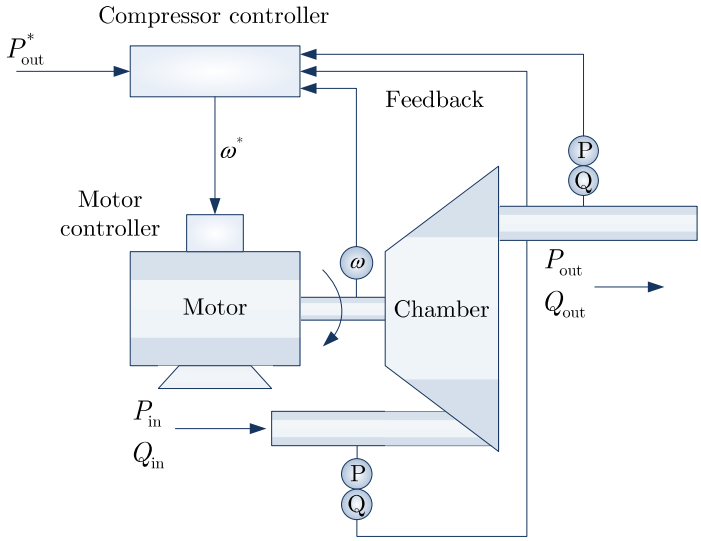



Figure 5.3 – A centrifugal compressor under control.

such a high performance controller that it is related to the control signal $\Delta u$ by the following first-order differential equation:

$$\Delta p(s) = \frac{K_a}{T_a s + 1} \Delta u(s) \tag{5.17}$$

where $K_a$ is the gain factor of the compressor, $T_a$ is the time constant of the compressor, $s$ is the Laplace operand, $\Delta p(s)$ and $\Delta u(s)$ are the Laplace transform of the increased pressure $\Delta p(t)$ and the control signal $\Delta u(t)$, respectively.

**Model of a sensor.** There are several types of sensors utilized in gas distribution systems, such as pressure sensors, flow rate sensors, or temperature sensors. Under the assumption that the gas transmission process is isothermal, we are interested in modeling the pressure sensors and flow rate sensors. Due to of the slow dynamics of the transient gas flow in the network, the model of a pressure sensor and a flow rate sensor can be described as

$$y_p(t) = K_p p(t) + v_p(t), \tag{5.18}$$
$$y_q(t) = K_q q(t) + v_q(t), \tag{5.19}$$

where $p(t)$ and $q(t)$ are the pressure and flow rate at a measured point, $y_p(t)$ and $y_q(t)$ are the measurements of pressure $p(t)$ and flow rate $q(t)$, $K_p$ and $K_q$ are gain coefficients of the sensors and $v_p(t)$ and $v_q(t)$ are sensor noises, respectively. Generally, the sensor noises are assumed to be zero-mean normal variables, i.e., $v_p(t) \sim \mathcal{N}\left(0, \sigma_p^2\right)$ and $v_q(t) \sim \mathcal{N}\left(0, \sigma_q^2\right)$, for all $t \geq 0$.

**Model of a customer.** In a gas distribution network, the customer demands fluctuate significantly during a given period (i.e., one day). This fluctuation is due to the difference in gas consumption of individuals and industries in different hours. In such safety-critical infrastructures as gas pipelines, the variation in customer demands can be estimated by specially-designed software (i.e., using a neural network). For this reason, the customer demand $d(t)$ is assumed to be completely known.

### 5.2.3 Model of cyber layer

In this subsection, we develop the model of the cyber layer which consists of the $\text{MTU}_1$ and the $\text{PLC}_1$. The $\text{PLC}_1$ is responsible for regulating the outlet pressure of the pipeline $G_1$ based on the set-point transmitted from the $\text{MTU}_1$.

**Model of a PLC.** The control algorithm is designed and implemented in the $\text{PLC}_1$ for regulating the outlet pressure at the pipeline $G_1$. Seeking for simplicity, we design a simple control algorithm whose architecture is shown in figure 5.4.

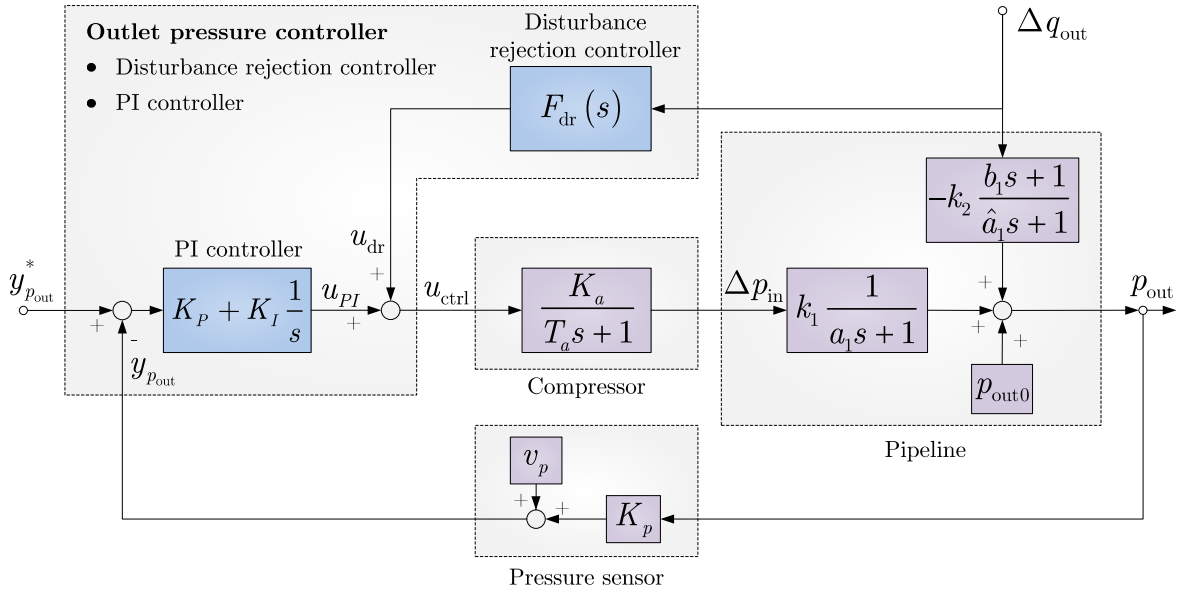



Figure 5.4 – Structure of the outlet pressure controller.

The controller is comprised of two parts. The disturbance rejection controller $F_{\text{dr}}(s)$, which is an open-loop controller, is designed to compensate for the variation in the outlet pressure $\Delta p_{\text{out}}$ due to the change in customer's demand $\Delta q_{\text{out}}$. The Proportional-Integral (PI) controller is designed to regulate the outlet pressure $\Delta p_{\text{out}}$ at a desired value by using closed-loop control techniques. By utilizing simple control design techniques, the disturbance rejection controller and the PI controller can be written as

$$F_{\text{dr}}(s) = \frac{k_2}{k_1 K_a} \frac{(a_1 + b_1)s + 1}{\hat{a}_1 s + 1}, \quad F_{\text{PI}}(s) = K_P + \frac{K_I}{s}, \tag{5.20}$$

where the coefficients $K_P$ and $K_I$ can be tuned by utilizing well-known techniques in automatic control theory.

**Model of a MTU.** For simplicity, let us assume that the $\text{MTU}_1$ takes responsibility for sending command signals $y^*_{p_{\text{out}}}(t)$ to the $\text{PLC}_1$ for regulating the outlet pressure $p_{out}(t)$ of the gas pipeline $G_1$. The command signals are transmitted over long distance from the $\text{MTU}_1$ (i.e., from the control center or from a control sub-station) to the $\text{PLC}_1$ (i.e., in the field), therefore, they are susceptible to several types of cyber attacks.

### 5.2.4 Discrete-time state space model

The physical layer and the cyber layer of the simple SCADA gas pipeline have been modeled in subsection 5.2.2 and subsection 5.2.3, respectively. The target of this subsection is to develop the model of transient gas flow through the network, i.e., from the gas stock to final customers, by combining the physical elements (i.e., pipeline $G_1$, compressor $P_1$ and pressure sensors $S_1$ and $S_2$ and flow rate sensors $S_3$ and $S_4$) and cyber elements (i.e., the $\mathrm{MTU}_1$ and the $\mathrm{PLC}_1$).

The method introduced in [1] is utilized for developing the state space model of transient gas flow through the pipeline. Let $x(t) = [x_1(t), \cdots, x_4(t)]^T$ be state vector which is expressed in terms of the inlet pressure $\Delta p_{\mathrm{in}}(t)$ and the outlet flow rate $\Delta q_{\mathrm{out}}(t)$ as follows:

$$\dot{x}_1(t) = -\frac{1}{a_1}x_1(t) + \frac{k_1}{a_1}\Delta p_{\mathrm{in}}(t), \tag{5.21}$$

$$\dot{x}_2(t) = -\frac{1}{\hat{a}_1}x_2(t) + \frac{1}{\hat{a}_1}\Delta p_{\mathrm{in}}(t), \tag{5.22}$$

$$\dot{x}_3(t) = -\frac{1}{\hat{a}_1}x_3(t) - \frac{k_2}{\hat{a}_1}\Delta q_{\mathrm{out}}(t), \tag{5.23}$$

$$\dot{x}_4(t) = -\frac{1}{d_1}x_4(t) + \frac{1}{d_1}\Delta q_{\mathrm{out}}(t). \tag{5.24}$$

The outputs $\Delta p_{\mathrm{out}}(t)$ and $\Delta q_{\mathrm{in}}(t)$ can be calculated from the inputs $\Delta p_{\mathrm{in}}(t)$ and $\Delta q_{\mathrm{out}}(t)$ and the state variables $x_1(t), \cdots, x_4(t)$ as follows:

$$\Delta p_{\mathrm{out}}(t) = x_2(t) + \left(1 - \frac{b_1}{\hat{a}_1}\right)x_3(t) - \frac{b_1 k_2}{\hat{a}_1}\Delta q_{\mathrm{out}}(t), \tag{5.25}$$

$$\Delta q_{\mathrm{in}}(t) = -\frac{c_1}{\hat{a}_1}x_2(t) + x_4(t) + \frac{c_1}{\hat{a}_1}\Delta p_{\mathrm{in}}(t). \tag{5.26}$$

Let also $y(t) = [y_1(t), \cdots, y_4(t)]^T \in \mathbb{R}^4$ be the measurements of $\Delta p_{\mathrm{in}}(t)$, $\Delta p_{\mathrm{out}}(t)$, $\Delta q_{\mathrm{in}}(t)$ and $\Delta q_{\mathrm{out}}(t)$, respectively. In other words, the measurement equations can be described as

$$y_1(t) = K_p\Delta p_{\mathrm{in}}(t) + v_p(t), \quad y_2(t) = K_p\Delta p_{\mathrm{out}}(t) + v_p(t), \tag{5.27}$$

$$y_3(t) = K_q\Delta q_{\mathrm{in}}(t) + v_q(t), \quad y_4(t) = K_q\Delta q_{\mathrm{out}}(t) + v_q(t). \tag{5.28}$$

For simplicity, let us assume that the time constant of the compressor is much smaller than the gas time constants (i.e., $T_a \ll a_1, \hat{a}_1$), and hence the relationship between the control signals $u(t)$ and the inlet pressure of the pipeline $\Delta p_{\mathrm{in}}(t)$ can be approximated as $\Delta p_{\mathrm{in}}(t) \approx K_a u(t)$. Let also $d(t) = \Delta q_{\mathrm{out}}(t)$ be the variation in customer's demands. The transient gas flow through the pipeline is then described as

$$\begin{cases} \dot{x}(t) &= \tilde{A}x(t) + \tilde{B}u(t) + \tilde{F}d(t) + w(t) \\ y(t) &= \tilde{C}x(t) + \tilde{D}u(t) + \tilde{G}d(t) + v(t) \end{cases}; \quad x(0) = \overline{x}_0, \tag{5.29}$$

where $x(t) \in \mathbb{R}^4$ is the vector of system states, $u(t) \in \mathbb{R}$ is the vector of control signals, $d(t) \in \mathbb{R}$ is vector of disturbances, $y(t) \in \mathbb{R}^4$ is the vector of sensor measurements, $w(t) \in \mathbb{R}^4$ is the vector of process noises accounting for the model uncertainty, $v(t) \in \mathbb{R}^4$ is the vector of sensor noises; the matrices $\tilde{A} \in \mathbb{R}^{4\times4}$, $\tilde{B} \in \mathbb{R}^{4\times1}$, $\tilde{F} \in \mathbb{R}^{4\times1}$, $\tilde{C} \in \mathbb{R}^{4\times4}$, $\tilde{D} \in \mathbb{R}^{4\times1}$ and $\tilde{G} \in \mathbb{R}^{4\times1}$ are calculated

as

$$\tilde{A} = \begin{bmatrix} -\frac{1}{a_1} & 0 & 0 & 0 \\ 0 & -\frac{1}{\hat{a}_1} & 0 & 0 \\ 0 & 0 & -\frac{1}{\hat{a}_1} & 0 \\ 0 & 0 & 0 & -\frac{1}{d_1} \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} K_a \frac{k_1}{a_1} \\ K_a \frac{1}{\hat{a}_1} \\ 0 \\ 0 \end{bmatrix}, \quad \tilde{F} = \begin{bmatrix} 0 \\ 0 \\ -\frac{k_2}{\hat{a}_1} \\ \frac{1}{d_1} \end{bmatrix}, \tag{5.30}$$

$$\tilde{C} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ K_p & 0 & K_p\left(1 - \frac{b_1}{\hat{a}_1}\right) & 0 \\ 0 & -K_q \frac{c_1}{\hat{a}_1} & 0 & K_q \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \tilde{D} = \begin{bmatrix} K_p K_a \\ 0 \\ K_q K_a \frac{c_1}{\hat{a}_1} \\ 0 \end{bmatrix}, \quad \tilde{G} = \begin{bmatrix} 0 \\ -K_p \frac{b_1 k_2}{\hat{a}_1} \\ 0 \\ K_q \end{bmatrix}. \tag{5.31}$$

Since the detection-isolation schemes are designed in the discrete-time domain, it is more convenient to convert the continuous-time state space model (5.29) into the discrete-time state space model. This task can be carried out simply by the exploiting the digital control theory [56]:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + v_k \end{cases}; \quad x_0 = \overline{x}_0, \tag{5.32}$$

where $x_k \in \mathbb{R}^n$ is the vector of system states, $u_k \in \mathbb{R}^m$ is the vector of control signals, $d_k \in \mathbb{R}^q$ is the vector of disturbances (corresponding to the consumption by customers), $w_k \in \mathbb{R}^n$ is the vector of process noises, $y_k \in \mathbb{R}^p$ is the vector of sensor measurements, $v_k \in \mathbb{R}^p$ is the vector of sensor noises; the matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $F \in \mathbb{R}^{n\times q}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, and $G \in \mathbb{R}^{p\times q}$ can be calculated from the corresponding matrices $\tilde{A}$, $\tilde{B}$, $\tilde{F}$, $\tilde{C}$, $\tilde{D}$, $\tilde{G}$ in the continuous-time domain and the sample time $T_S$ (i.e., $n = 4$, $m = 1$, $p = 4$ and $q = 1$). The process noises $w_k$ and the sensor noises $v_k$ are assumed to follow zero-mean normal distribution with known covariance matrices $Q$ and $R$ (i.e., $w_k \sim \mathcal{N}(0, Q)$ and $v_k \sim \mathcal{N}(0, R)$, where $R$ is positive-definite), respectively.

### 5.2.5 Model of cyber-physical attacks

As shown in figure 5.1, there are assumed to exist three possible attack points that can be exploited by adversaries for performing malicious attacks on the SCADA gas pipeline, including the introduction of false command signals sent from $\text{MTU}_1$ to the $\text{PLC}_1$, the modification of control signals sent from the $\text{PLC}_1$ to the compressor $P_1$ and the injection of false data into sensor measurements transmitted from sensors to the $\text{PLC}_1$ and/or the $\text{MTU}_1$.

The system model under cyber attacks on the control signals and the sensor measurements can be described as:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + Ka_k^u + w_k \\ y_k &= Cx_k + Du_k + Gd_k + Ha_k^u + Ma_k^y + v_k \end{cases}; \quad x_0 = \overline{x}_0, \tag{5.33}$$

where $a_k^u \in \mathbb{R}^m$ is the attack vector on the control signals and $a_k^y \in \mathbb{R}^p$ is the attack vector on the sensor measurements. The attack matrices $K \in \mathbb{R}^{n\times m}$, $H \in \mathbb{R}^{p\times m}$ and $M \in \mathbb{R}^{p\times p}$. The attack matrices $K$ and $H$ should satisfy the following condition: $\text{span}(K) \subseteq \text{span}(B)$ and

span $(H) \subseteq$ span $(D)$. The matrices $K$ and $H$ are chosen as $K = B$ and $H = D$. The matrix $M$ is assumed to be diagonal, i.e., $M = \text{diag}(\gamma_j)$, where $\gamma_j = 1$ signifies that sensor $S_j$ is vulnerable and $\gamma_j = 0$ means that sensor $S_j$ is secure. In this numerical example, $n = 4$, $m = 1$, $p = 4$, $q = 1$, $r = 1$ and $s = 5$.

For simplifying the model (5.33), let $a_k = \left[(a_k^x)^T, (a_k^y)^T\right]^T$ be the attack vector, $B_a = [K, 0]$ and $D_a = [H, M]$ be the attack matrices. The system model under attack is rewritten as

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + B_a a_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + D_a a_k + v_k \end{cases}; \quad x_0 = \overline{x}_0, \tag{5.34}$$

where $a_k \in \mathbb{R}^s$, with $s = m + p$, is the attack vector, $B_a \in \mathbb{R}^{n \times s}$ and $D_a \in \mathbb{R}^{p \times s}$ are the attack matrices.

## 5.3 Model of SCADA Water Distribution Networks

In this section, we develop the model of a simple SCADA water distribution network and several attack scenarios. The architecture of a simple SCADA water distribution network is described in subsection 5.3.1. The components of the physical layer and the cyber layer of the water network are modeled in subsection 5.3.2 and subsection 5.3.3, respectively. Some possible attacks scenarios on the water network are shown in subsection 5.3.4.

### 5.3.1 System architecture

Similar to SCADA gas pipelines, the architecture of a SCADA water distribution system is also divided into the physical layer and the cyber layer. The physical layer is comprised of a large number of reservoirs, tanks, junctions, pumps, valves, pipelines, sensors and other hydraulic components which help in transmitting and distributing water from production plants to final customers. The cyber layer, including SCADA control center, communication devices, controllers and anomaly detectors, is in charge of monitoring and supervising the operation of the system based on the data acquired from field devices.

For simplicity, we study a simple SCADA water distribution system as shown in figure 5.5. The physical layer consists of a treatment plant $W_1$, a reservoir $R_1$, a pump $P_1$, 3 junctions $N_2$, $N_3$ and $N_4$, 4 pipelines $G_{01}$, $G_{12}$, $G_{23}$, and $G_{24}$ and 2 consumers $d_1$ and $d_2$. Two pressure sensors $S_1$ and $S_2$ are equipped for measuring pressure heads $h_1$ at the reservoir and $h_2$ at the node $N_2$, respectively. The cyber layer is comprised of the SCADA control center which is responsible for regulating the pressure head $h_1$ at the reservoir and monitoring the operation of the network.

### 5.3.2 Model of physical layer

The model of hydraulic components such as treatment plants, reservoirs, tanks, junctions, pipelines, pumps, valves and customer's demands can be found in [21]. By utilizing the model of each element and exploiting laws of mass and energy conservation, the water flow through a network can be described by a set of non-linear equations. These non-linear equations can be linearized around operating points [151] in order to obtain the linearized state space model [140].

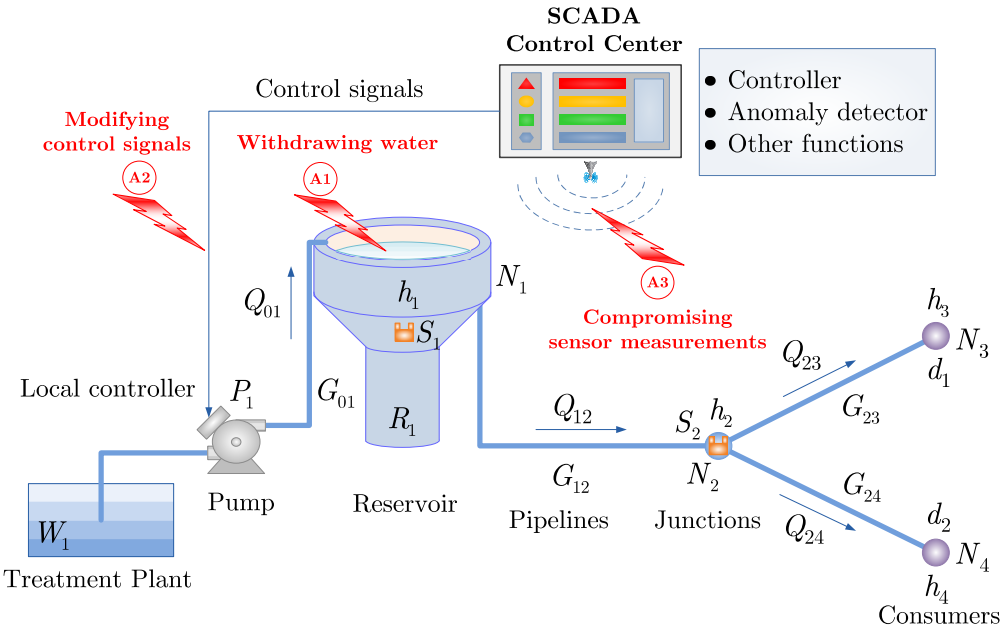



Figure 5.5 – Architecture of a simple SCADA water distribution network.

### System model under normal operation

The linearized model of the water network in figure 5.5 is obtained by exploiting mass and energy balance equations of water flow through the network. The mass balance equations at the reservoir $R_1$ and the junctions $N_2$, $N_3$ and $N_4$ can be written as follows:

$$A_1\dot{h}_1(t) = Q_{01}(t) - Q_{12}(t), \tag{5.35}$$
$$0 = Q_{12}(t) - Q_{23}(t) - Q_{24}(t), \tag{5.36}$$
$$0 = Q_{23}(t) - d_1(t), \tag{5.37}$$
$$0 = Q_{24}(t) - d_2(t), \tag{5.38}$$

where $A_1$ is the cross section of the reservoir, $h_1(t)$ is the pressure head at the reservoir, $Q_{ij}(t)$ is the water flow rate through the pipeline $G_{ij}$, $d_1(t)$ and $d_2(t)$ are the consumption at the junctions $N_3$ and $N_4$, respectively.

For simplicity, let us assume that the pump $P_1$ is regulated by an extremely high performance local controller so that the water flow rate $Q_{01}(t)$ is proportional to the control signal $u(t)$ sent from the SCADA control center. The energy balance equation through the pump $P_1$ is then simplified as

$$0 = u(t) - g_{01}Q_{01}(t), \tag{5.39}$$

where $g_{01}$ is a known coefficient. The energy balance equations through the pipelines $G_{12}$, $G_{23}$ and $G_{24}$ are written as

$$0 = h_1(t) - h_2(t) - g_{12}Q_{12}(t), \tag{5.40}$$
$$0 = h_2(t) - h_3(t) - g_{23}Q_{23}(t), \tag{5.41}$$
$$0 = h_2(t) - h_4(t) - g_{24}Q_{24}(t), \tag{5.42}$$

where $h_2(t)$, $h_3(t)$ and $h_4(t)$ are pressure heads at the junctions $N_2$, $N_3$ and $N_4$, respectively; $g_{12}$, $g_{23}$ and $g_{24}$ are known coefficients obtained linearizing the Hazen-Williams equation for pressure head loss of water flow through the pipelines $G_{12}$, $G_{23}$ and $G_{34}$, respectively.

The mass and energy balance equations can be expressed in the matrix form as

$$\underbrace{\begin{bmatrix} A_1\dot{h}_1(t) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{E\dot{x}(t)} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -g_{01} & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & -g_{12} & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & -g_{23} & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & -g_{24} \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} h_1(t) \\ h_2(t) \\ h_3(t) \\ h_4(t) \\ Q_{01}(t) \\ Q_{12}(t) \\ Q_{23}(t) \\ Q_{24}(t) \end{bmatrix}}_{x(t)} +$$

$$\underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{B} u(t) + \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} d_1(t) \\ d_2(t) \end{bmatrix}}_{d(t)}. \tag{5.43}$$

In practical applications, the customer's demands $d_1(t)$ and $d_2(t)$ can be estimated by some specially-designed software [21] with an acceptable level of accuracy. In addition, the model of the pump under control may not be completely accurate due to some tolerance levels of motors, sensors, etc. These uncertainties are often modeled by so-called the process noises $w(t)$. Taking into account the process noise vector $w(t)$, the state evolution equation can be rewritten as

$$E\dot{x}(t) = Ax(t) + Bu(t) + Fd(t) + w(t); \quad x(0) = \overline{x}_0, \tag{5.44}$$

where $x(t) \in \mathbb{R}^n$ is the vector of system states, $u(t) \in \mathbb{R}^m$ is the vector of control signals, $d(t) \in \mathbb{R}^q$ is the vector of disturbances; the matrices $E \in \mathbb{R}^{n\times n}$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $F \in \mathbb{R}^{n\times q}$, where $n = 8$, $m = 1$, $q = 2$ in this time-continuous state space model.

For simplicity, let us assume that two pressure sensors $S_1$ and $S_2$ are utilized for measuring the pressure head $h_1(t)$ at the reservoir and the pressure head $h_2(t)$ at the junction $N_2$, respectively. The measurement equation is then expressed as

$$\underbrace{\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}}_{y(t)} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} h_1(t) \\ h_2(t) \\ h_3(t) \\ h_4(t) \\ Q_{01}(t) \\ Q_{12}(t) \\ Q_{23}(t) \\ Q_{24}(t) \end{bmatrix}}_{x(t)} + \underbrace{\begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix}}_{v(t)}, \tag{5.45}$$

where $y_1(t)$ and $y_2(t)$ are the measurements of the sensors $S_1$ and $S_2$, respectively; $v_1(t)$ and $v_2(t)$ are sensor noises. The sensor measurement equation can be rewritten in a simpler form as

$$y(t) = Cx(t) + v(t), \tag{5.46}$$

where $y(t) \in \mathbb{R}^p$ is the vector of sensor measurements and $v(t) \in \mathbb{R}^p$ is the vector of sensor noises; the matrix $C \in \mathbb{R}^{p \times n}$ with $p = 2$ in this case.

Putting together (5.44) and (5.46), the linearized model of water flow through the network can be expressed by the following time-continuous state space form:

$$\begin{cases} E\dot{x}(t) &= Ax(t) + Bu(t) + Fd(t) + w(t) \\ y(t) &= Cx(t) + v(t) \end{cases}; \quad x(0) = \overline{x}_0. \tag{5.47}$$

**From singular form to non-singular form**

It is worth noting that the matrix $E$ in (5.47) is singular (i.e., $\det(E) = 0$), therefore, it is necessary to transform the system model (5.47) into a non-singular form. This task can be carried out by exploiting specific results from the index-one singular systems as shown in [140].

Seeking for simplicity but without loss of generality, let us assume that the state space model (5.47) has the following form:

$$\underbrace{\begin{bmatrix} E_{11} & 0 \\ 0 & 0 \end{bmatrix}}_{E} \underbrace{\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix}}_{\dot{x}(t)} = \underbrace{\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}}_{x(t)} + \underbrace{\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}}_{B} u(t) + \underbrace{\begin{bmatrix} F_1 \\ F_2 \end{bmatrix}}_{F} d(t) + \underbrace{\begin{bmatrix} I_1 \\ I_2 \end{bmatrix}}_{I} w(t), \tag{5.48}$$

$$y(t) = \underbrace{\begin{bmatrix} C_1 & C_2 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}}_{x(t)} + v(t), \tag{5.49}$$

where $E_{11} \in \mathbb{R}^{n_1 \times n_1}$ and $A_{22} \in \mathbb{R}^{n_2 \times n_2}$ are non-singular matrices, $I \in \mathbb{R}^{n \times n}$ is the identity matrix, and the system states $x(t) \in \mathbb{R}^n$ are comprised of the dynamic states $x_1(t) \in \mathbb{R}^{n_1}$ and the algebraic states $x_2(t) \in \mathbb{R}^{n_2}$, where $n = n_1 + n_2$. The algebraic states $x_2(t)$ can be calculated from the dynamic states $x_1(t)$ by the following equation:

$$x_2(t) = -A_{22}^{-1}A_{21}x_1(t) - A_{22}^{-1}B_2u(t) - A_{22}^{-1}F_2d(t) - A_{22}^{-1}I_2w(t). \tag{5.50}$$

The elimination of algebraic states $x_2(t)$ leads to a non-singular time-continuous state space model as

$$\begin{aligned} \dot{x}_1(t) &= \underbrace{E_{11}^{-1}\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)}_{\tilde{A}} x_1(t) + \underbrace{E_{11}^{-1}\left(B_1 - A_{12}A_{22}^{-1}B_2\right)}_{\tilde{B}} u(t) + \\ &\quad \underbrace{E_{11}^{-1}\left(F_1 - A_{12}A_{22}^{-1}F_2\right)}_{\tilde{F}} d(t) + \underbrace{E_{11}^{-1}\left(I_1 - A_{12}A_{22}^{-1}I_2\right)}_{\tilde{I}} w(t), \end{aligned} \tag{5.51}$$

$$y(t) = \underbrace{\left(C_1 - C_2A_{22}^{-1}A_{21}\right)}_{\tilde{C}} x_1(t) + \underbrace{\left(-C_2A_{22}^{-1}B_2\right)}_{\tilde{D}} u(t) + \underbrace{\left(-C_2A_{22}^{-1}F_2\right)}_{\tilde{G}} d(t) + v(t), \tag{5.52}$$

and in a simpler form as

$$\begin{cases} \dot{x}_1(t) &= \tilde{A}x_1(t) + \tilde{B}u(t) + \tilde{F}d(t) + \tilde{I}w(t) \\ y(t) &= \tilde{C}x_1(t) + \tilde{D}u(t) + \tilde{G}d(t) + v(t) \end{cases}, \tag{5.53}$$

where the matrices $\tilde{A}$, $\tilde{B}$, $\tilde{F}$, $\tilde{I}$, $\tilde{C}$, $\tilde{D}$ and $\tilde{G}$ can be calculated from the original matrices $E$, $A$, $B$, $F$, $I$ and $C$.

The non-singular continuous-time state space model (5.53) can be transformed into the discrete-time counterpart by the sample time $T_S$ without any difficulty. For notation convenience, we eliminate the "*tilde*" from the matrices, replace the dynamic states $x_1(t)$ by the system states $x_k$ with unknown initial condition $\overline{x}_0$, and employ $n$ as the number of dynamic states in place of $n_1$. As a result, the linearized model of the water distribution network is described in a discrete-time state space form as

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + v_k \end{cases}; \quad x_0 = \overline{x}_0 \tag{5.54}$$

where $x_k \in \mathbb{R}^n$ is the vector of system states corresponding the pressure head $h_1$ at the reservoir, $u_k \in \mathbb{R}^m$ is the vector of control signals transmitted from the control center to the pump $P_1$, $d_k \in \mathbb{R}^q$ is the vector of disturbances (corresponding to customer's demands), $w_k \in \mathbb{R}^n$ is the vector of process noises, $y_k \in \mathbb{R}^p$ is the vector of sensor measurements (sensors $S_1$ and $S_2$), $v_k \in \mathbb{R}^p$ is the vectors of sensor noises; the matrices with appropriate dimension $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $F \in \mathbb{R}^{n\times q}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, and $G \in \mathbb{R}^{p\times q}$, where $n = 1$, $m = 1$, $p = 2$ and $q = 2$ in this numerical example. The process noises $w_k$ and the sensor noises $v_k$ are assumed to follow zero-mean normal distribution with known covariance matrices $Q$ and $R$ (i.e., $w_k \sim \mathcal{N}(0, Q)$ and $v_k \sim \mathcal{N}(0, R)$, where $R$ is positive-definite), respectively.

### 5.3.3 Model of cyber layer

The cyber layer of the water distribution network is responsible for regulating water flow from production plants to final customers, controlling water quality, monitoring abnormal situations occurring to the system, acquiring data for management, or doing other important functions. In this numerical example, we focus on the hydraulic monitoring of water flow through the network. For the sake of simplicity, let us assume that the water flow rate $Q_{01}$ into the reservoir $R_1$ is controlled by a simple algorithm which sends constant control signals (i.e., $u_k = \text{constant}$) from the control center to the local controller for regulating the pump $P_1$ operating at a constant speed. It should be noted that more complicated control algorithms do not alter the principal results since the control signals are completely known to the system operators.

### 5.3.4 Model of cyber-physical attacks

This subsection is dedicated to developing the model of cyber-physical attacks on the water distribution network described in figure 5.5. Firstly, we figure out several possible attack points that can be exploited by the attacker for launching malicious attacks. Secondly, we develop the model of cyber-physical attacks on the water network, from the singular continuous-time state space model to the non-singular discrete-time state space model.

**Possible attack scenarios**

Let us assume that the adversary is able to perform the following malicious activities:

- *Physical attack on the reservoir*: The attacker can launch a physical attack for stealing water from the reservoir (e.g., by utilizing an external pump) with a flow rate $Q_0(t) > 0$, leading to a reduction in water pressure at the reservoir as well as a decrease in the pressure over the network.
- *Cyber attack on control signals*: The malicious agent can also modify the control signals to the pump, forcing the control signals $u(t)$ become $u(t) + a^u(t)$, where $a^u(t)$ is the attack signal added to the control signals $u(t)$.
- *Cyber attack on sensor measurements*: The powerful hacker can compromise sensor measurements (e.g., by breaking into the communication channels between the local devices and the control center), causing the measurements $y_j(t)$ of sensor $S_j$ become $y_j(t)+a^{y_j}(t)$, where $\{a^{y_j}(t)\}_{1\leq j\leq p}$ are the attack signals added to the measurements of sensor $S_j$. The sensor attack vector $a^y(t) \in \mathbb{R}^p$ is then described as $a^y(t) = [a^{y_1}(t), \cdots, a^{y_p}(t)]^T$.

**Model of cyber-physical attacks**

Let us assume that the adversary launches a coordinated attack by withdrawing water from the reservoir (i.e., by the attack vector $a^p(t) = -Q_0(t)$), modifying the control signals (i.e., by the attack vector $a^u(t)$) and compromise sensor measurements (i.e., by the attack vector $a^y(t)$) during the attack period $\tau_a = [k_0, k_0 + L - 1]$, where $k_0$ is an unknown attack instant and $L$ is the attack duration. The state evolution equation under the physical attack on the reservoir and the cyber attack on the control signals can be described as

$$E\dot{x}(t) = Ax(t) + \underbrace{K_1 a^p(t)}_{\text{stealing water}} + Bu(t) + \underbrace{K_2 a^u(t)}_{\text{modifying control signals}} + Fd(t) + w(t), \tag{5.55}$$

where the component $K_1 a^p(t) \in \mathbb{R}^n$ denotes the physical attack for stealing water from the reservoir and $K_2 a^u(t) \in \mathbb{R}^n$ stands for the cyber attack for modifying the control signals. Let $a^x(t) = \begin{bmatrix} a^p(t) \\ a^u(t) \end{bmatrix} \in \mathbb{R}^r$, where $r = 2$, be the state attack vector and $K = [K_1, K_2] \in \mathbb{R}^{n\times r}$ be the attack matrices. The state attack component $Ka^x(t)$ can be described as

$$Ka^x(t) = \underbrace{\underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{K_1} a^p(t) + \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{K_2} a^u(t)}_{\text{stealing water} \quad \text{modifying control signals}} = \underbrace{\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{K} \underbrace{\begin{bmatrix} a^p(t) \\ a^u(t) \end{bmatrix}}_{a^x(t)}}_{\text{state attack component}}. \tag{5.56}$$

The cyber attack on sensors impact their measurements directly and the measurement equation can be expressed as

$$y(t) = Cx(t) + \underbrace{Mu^y(t)}_{\text{compromised sensors}} + v(t), \tag{5.57}$$

where the matrix $M \in \mathbb{R}^{p\times p}$ reflects the attacker's capability to compromise sensor measurements. Seeking for simplicity, it is assumed that $M = \text{diag}(\gamma_j)$, where $\gamma_j = 1$ signifies that the sensor $S_j$ is vulnerable and $\gamma_j = 0$ means that the sensor $S_j$ can not be compromised.

By combining (5.55)–(5.57), the model of the water network under cyber-physical attack can be described in a singular continuous-time state space model as

$$\begin{cases} E\dot{x}(t) &= Ax(t) + Bu(t) + Fd(t) + Ka^x(t) + w(t) \\ y(t) &= Cx(t) + Ma^y(t) + v(t) \end{cases}; \quad x(0) = \overline{x}_0. \tag{5.58}$$

where the matrices $E$, $A$, $B$, $F$, $C$ in (5.58) are the same as those in (5.47).

By utilizing the same techniques as previous subsection, the singular continuous-time state space model (5.58) can be transformed into the following non-singular discrete-time state space model:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + Ka_k^x + w_k \\ y_k &= Cx_k + Du_k + Gd_k + Ha_k^x + Ma_k^y + v_k \end{cases}; \quad x_0 = \overline{x}_0, \tag{5.59}$$

where the system matrices $A$, $B$, $F$, $C$, $D$, $G$, $K$, $H$ and $M$ in the non-singular discrete time state space model (5.59) are different from the system matrices $A$, $B$, $F$, $C$, $K$, $M$ in the singular continuous-time state space model (5.58) due to the transformation from the singular form to the non-singular form and from the continuous-time form to the discrete-time form. We hope that the utilization of such notations does not cause any confusion to readers.

Let $a_k = \left[(a_k^x)^T, (a_k^y)^T\right]^T \in \mathbb{R}^s$, where $s = r + p$, be the attack vector, $B_a = [K, 0] \in \mathbb{R}^{n\times s}$ and $D_a = [H, M] \in \mathbb{R}^{p\times s}$ be the attack matrices. The system model under attack is rewritten as

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + B_a a_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + D_a a_k + v_k \end{cases}; \quad x_0 = \overline{x}_0, \tag{5.60}$$

where $x_k \in \mathbb{R}^n$ is the vector of system states with unknown initial states $\overline{x}_0$, $u_k \in \mathbb{R}^m$ is the vector of control signals, $d_k \in \mathbb{R}^q$ is the vector of disturbances (corresponding to the consumption of customers), $y_k \in \mathbb{R}^p$ is the vector of sensor measurements, $a_k \in \mathbb{R}^{r+p}$ is the vector of attack signals, $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^p$ are the vectors of process noises and sensor noises, respectively; the matrices with appropriate dimension $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $F \in \mathbb{R}^{n\times q}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, $G \in \mathbb{R}^{p\times q}$, $B_a \in \mathbb{R}^{n\times s}$ and $D_a \in \mathbb{R}^{p\times s}$. In this numerical example, the parameters are chosen as $n = 1$, $m = 1$, $p = 2$, $q = 2$, $r = 2$ and $s = r + p = 4$.

## 5.4 Conclusion

The physical layer of the majority of SCADA systems can be described in the discrete-time state space model driven by Gaussian noises. The cyber-physical attacks (i.e., malicious attacks on both physical layer and cyber layer) can be modeled as additive signals of short duration to both state evolution and sensor measurements equations. For the demonstration purpose, we

have developed the models of a simple SCADA gas pipeline and a simple SCADA water network under normal operation as well as under cyber-physical attacks. Especially, we have modeled also several attack strategies found in literature, including DoS attacks, simple integrity attacks and stealthy integrity attacks.

The negative impact of cyber-physical attacks on closed-loop control systems will be demonstrated in the next chapter by launching DoS attacks, simple integrity attacks and stealthy integrity attacks on the command signals, control signals and sensor measurements of the SCADA gas pipelines. In addition, the SCADA water network under the covert attack strategy will be utilized for demonstrating the effectiveness of the detection-isolation schemes proposed in chapter 3 and chapter 4.

# Chapter 6

# Numerical Examples

**Contents**



## 6.1 Introduction

In chapter 3 and chapter 4, we have proposed several sub-optimal algorithms for detecting and isolating transient signals in stochastic-dynamical systems. In order to demonstrate the theoretical findings, we have developed two simulation models, including the model of a simple SCADA gas pipeline and the model of a simple SCADA water distribution network in chapter 5.

The target of this chapter is to apply the proposed algorithms to the detection and identification of several attack scenarios on both the SCADA gas pipeline and the SCADA water network.

This chapter is split into three main sections. In section 6.2, we study the effect of several types of cyber-physical attacks on the SCADA gas pipeline. Specially, we show that DoS attacks and simple integrity attacks (i.e., min-max, scaling and additive attacks) can be detected easily even by traditional anomaly detectors. In contrast, stealthy integrity attacks (i.e., replay attack and covert attack, for example) are much more difficult to detect.

The statistical performance of the proposed detection algorithms are demonstrated in section 6.3. Simulation results are given for comparing between the proposed FMA detection rule and traditional tests, and between the steady-state Kalman filter-based algorithms and the fixed-size parity space-based detection procedures. The comparison between the proposed numerical method and the Monte Carlo simulation method is also carried out. In addition, the robustness of the FMA test with respect to several operational parameters is investigated by both Monte Carlo simulation and numerical method. Furthermore, we examine the statistical performance of several detection schemes when the transient change parameters are partially known.

In section 6.4, we investigate the performance of the proposed detection-isolation schemes. It will be seen that the FMA test is quite effective in detecting and isolating cyber-physical attacks on SCADA systems. Especially, the performance of the FMA test is compared with several traditional tests, including the generalized WL CUSUM test, the matrix WL CUSUM test and the vector WL CUSUM test in different scenarios.

## 6.2 Cyber-Physical Attacks on Gas Pipelines

In this section, we investigate the negative impact of several attack scenarios on the SCADA gas pipeline described in figure 5.1. It has been pointed out in section 5.2 that the adversary can launch his malicious attacks on the command signals, control signals and sensor measurements by either DoS attack strategies, simple integrity attack strategies or stealthy integrity attack strategies.

### 6.2.1 Introduction

A Matlab-Simulink model has been developed for studying the negative impact of cyber-physical attacks on the simple SCADA gas pipeline described in figure 5.1. The simulation model, which is shown in figure 6.1, consists of the physical layer (i.e., the gas stock $GS_1$, the compressor $P_1$, the gas pipeline $G_1$, the pressure sensors $S_1$, $S_2$ and the flow rate sensors $S_3$, $S_4$) and the cyber layer (i.e., the $\text{PLC}_1$ and the $\text{MTU}_1$). The cooperation of the physical layer and the cyber layer helps in transporting and distributing gas from the gas stock $GS_1$ to the customer $C_1$.

The simulation parameters are chosen as follows (i.e., the same as those in [76]). The pipeline parameters include the pipe length $L = 100\,\text{km}$, the pipe diameter $D = 0.6\,\text{m}$, the friction factor $f = 0.03$. The gas parameters include the gas compressibility factor $Z = 0.88$, the gas constant $R = 392\,\text{m}^2/\text{s}^2\text{K}$, the isothermal speed of sound $c = 310\,\text{m/s}$ and the gas density at standard condition $\rho_n = 0.7165\,\text{kg/m}^3$. The environment parameters include the temperature $T = 278\,\text{K}$, the gravity acceleration $g = 9.81\,\text{m/s}^2$ and the pipe inclination $h = 0\,\text{m}$ (i.e., the straight horizontal pipeline). The initial inlet pressure is $p_{\text{in},0} = 50\,\text{bar}$. The volumetric flow

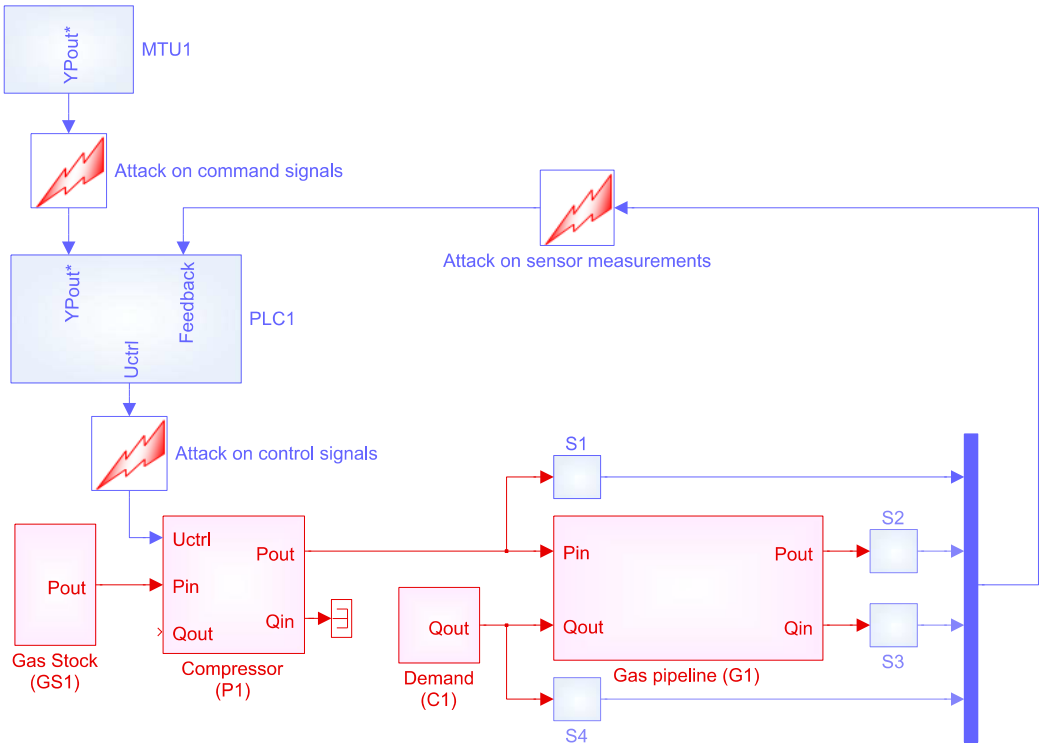


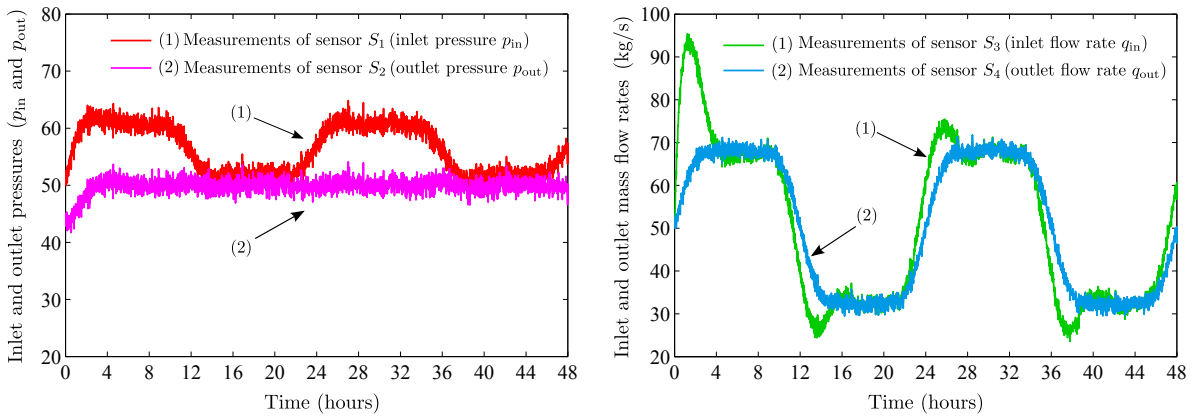

Figure 6.1 – Simulation model of a simple SCADA gas pipeline.

rates at standard conditions are $Q_{n,0} = 70\,\mathrm{m}^3/\mathrm{s}$. The initial mass flow rates are calculated from $Q_{n,0}$ by the following equation: $q_{\mathrm{in},0} = q_{\mathrm{out},0} = \rho_n Q_{n,0} = 50.16\,\mathrm{kg/s}$. The initial outlet pressure is calculated from (5.7) as $p_{\mathrm{out},0} = 43.53\,\mathrm{bar}$. The linearized parameters are obtained from [1,13] as $k_1 = 1.0350$, $k_2 = 2.6178\,10^4$, $a_1 = 3.8213\,10^3$, $b_1 = 1.2751\,10^3$, $c_1 = 0.2993$, $d_1 = 3.8289\,10^3$ and $\hat{a}_1 = 3.8251\,10^3$.

The parameters of the compressor include the gain factor $K_a = 10^5$ and the time constant $T_a = 600\,\mathrm{s}$. The gain factors of the pressure sensors and the flow rate sensors are $K_p = 10^{-5}$ and $K_q = 1$, respectively. The noise variances are $\sigma_p^2 = 1$ for the pressure sensors (i.e., $S_1$ and $S_2$) and $\sigma_q^2 = 2$ for the flow rate sensors (i.e., $S_3$ and $S_4$), respectively. The sample time is chosen as $T_{\mathrm{S}} = 30\,\mathrm{s}$ and the simulation time is $T_{\mathrm{SIM}} = 48\,\mathrm{hours}$.

The parameters of the cyber layer are chosen as follows. The discrete-time PI controller is designed with the proportional gain $K_P = 0.4$ and the integral gain $K_I = 2\,10^{-4}$. The parameters of the disturbance rejection controller $F_{\mathrm{dr}}$ can be transformed from the continuous-time representation (5.20) to the discrete-time representation by either the zero-order hold method, the first-order hold method or the Tustin's method [56]. During normal operation, it is assumed that the command signals transmitted from the $\mathrm{MTU}_1$ to the $\mathrm{PLC}_1$ remain constant at $y^*_{p_{\mathrm{out}}} = 50$ for regulating the outlet pressure at the constant value of $p_{\mathrm{out}} = 50\,\mathrm{bar}$.

The normal behavior of the SCADA gas pipeline is exemplified in figure 6.2. The outlet mass flow rate $q_{\mathrm{out}}$ (i.e., the blue curve), which corresponds to the customers' demands, fluctuates periodically around the nominal value of $50.16\,\mathrm{kg/s}$. In order to regulate the outlet pressure $p_{\mathrm{out}}$ (i.e., the magenta curve) around the value of $50\,\mathrm{bar}$, the $\mathrm{PLC}_1$ performs the control algorithm




(a) Inlet and outlet pressures (i.e., $p_{in}$ and $p_{out}$). (b) Inlet and outlet mass flow rates (i.e., $q_{in}$ and $q_{out}$).

Figure 6.2 – Normal operation of the SCADA gas pipeline.

and sends the control signals to the compressor $P_1$ for regulating the inlet pressure $p_{in}$ (i.e., the red curve). In turn, the change in the inlet pressure leads to the variation in the inlet mass flow rate $q_{in}$ (i.e., the green curve).

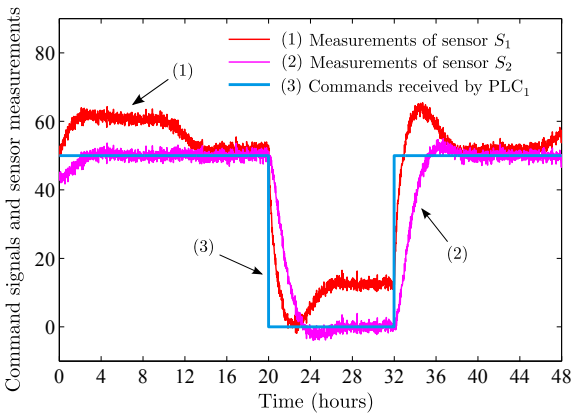
### 6.2.2 DoS attacks

This subsection is dedicated to studying the negative impact of several DoS attack strategies on the SCADA gas pipeline. The DoS attack strategies (1.3), (1.4) and (1.5) are performed on the command signals, on the control signals and on the sensor measurements, respectively. The attack duration is $\tau_a = [20, 32]$ hours.

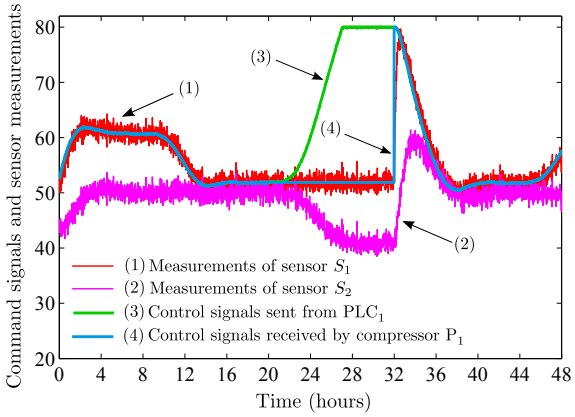
The effect of the DoS attack strategy (1.3) on the command signals transmitted from the $\text{MTU}_1$ to the $\text{PLC}_1$ is described in sub-figure 6.3a. During the attack duration $\tau_a$, the commands received by the $\text{PLC}_1$ are considered as zero since the true signals can not arrive at the $\text{PLC}_1$. In response to this attack, the outlet pressure $p_{out}$ is regulated to keep track of the false commands (i.e., zero). As a consequence, both the inlet pressure $p_{in}$ and the outlet pressure $p_{out}$ are forced to reduce significantly.

In sub-figure 6.3b, we show the effect of the DoS attack strategy (1.4) on the control signals sent from the $\text{PLC}_1$ to the compressor $P_1$. During the attack duration $\tau_a$, the control signals received by the compressor $P_1$ remain the same as the control signals just before the attack (i.e., $u_k = u_{k_0-1}$ for every $k \in [k_0, k_0 + L - 1]$). As a consequence, the inlet pressure $p_{in}$ remains almost constant and the outlet pressure $p_{out}$ reduces slightly in response to the augmentation in customer's demands.

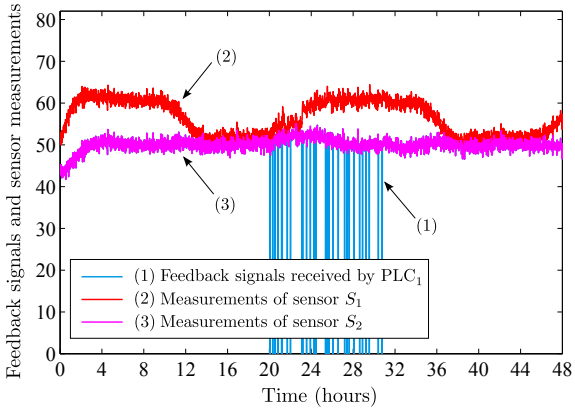
For the DoS attack strategy (1.5), the feedback signals $y_{p_{out}}$ are transmitted successfully to the $\text{PLC}_1$ with the probability $p_1 = \mathbb{P}(\gamma_k = 1)$, where $\gamma_k$ is a random variable following the Bernoulli distribution. We consider two scenarios: $p_1 = 0.95$ (i.e., 95% of the feedback signals are transmitted successfully to the $\text{PLC}_1$) and $p_1 = 0.05$ (i.e., 5% of the feedback signals are transmitted successfully to the $\text{PLC}_1$). The DoS attack strategy (1.5) on the feedback signals with $p_1 = 0.95$ and $p_1 = 0.05$ is described in sub-figure 6.3c and sub-figure 6.3d, respectively. In both cases, the controller is unable to perform its task due to the lack of feedback signals. In the first scenario where the attacker is able to block only 5% of signals, the measurements of



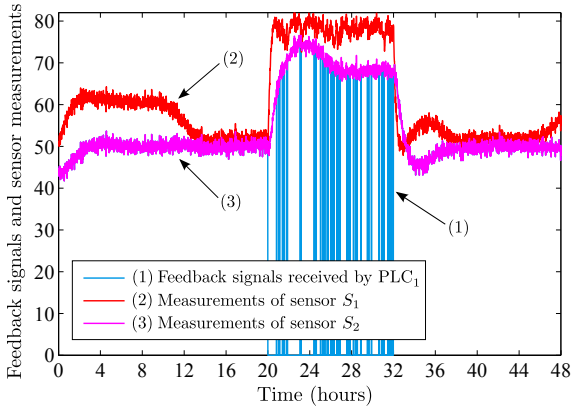

(a) DoS attack strategy (1.3) on the command signals.




(b) DoS attack strategy (1.4) on the control signals.




(c) DoS attack strategy (1.5) on the feedback signals with $p_1 = 0.95$.




(d) DoS attack strategy (1.5) on the feedback signals with $p_1 = 0.05$.

Figure 6.3 – DoS attack strategies on the SCADA gas pipeline.

sensor $S_2$ are deflected slightly from their normal values. In the second scenario, on the other hand, the outlet pressure $p_{\text{out}}$ is out of control since only 5% of feedback signals are transmitted successfully to the PLC$_1$.

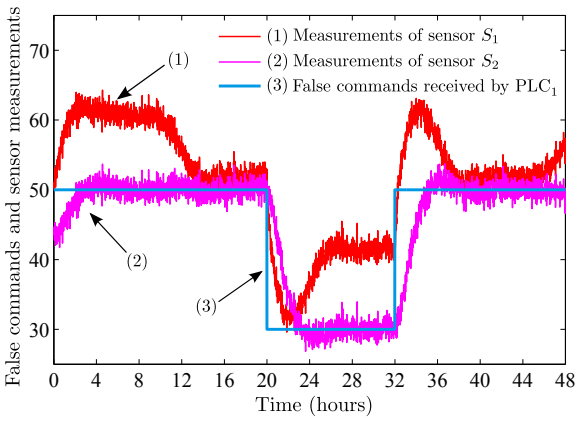
### 6.2.3 Simple integrity attacks

In this subsection, we investigate the negative effect of simple integrity attacks on the SCADA gas pipeline, including the injection of false data into command signals, control signals, and feedback signals. The operational ranges for the command signals, the control signals and the feedback signals are chosen as $\mathcal{Y}^* \triangleq [30, 70]$, $\mathcal{U} \triangleq [30, 80]$ and $\mathcal{Y} \triangleq [30, 70]$, respectively.

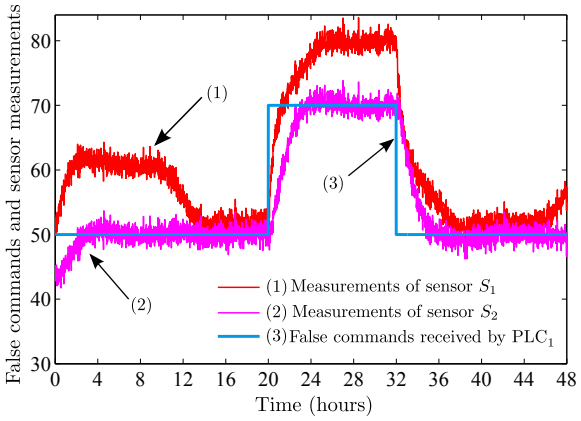
#### Attack on command signals

In figure 6.4, we show the reaction of the SCADA gas pipeline under several simple integrity attack strategies (i.e., min attack, max attack, scaling attack and additive attack) on the command signals sent from the MTU$_1$ to the PLC$_1$. Theoretically, the simple integrity attacks on

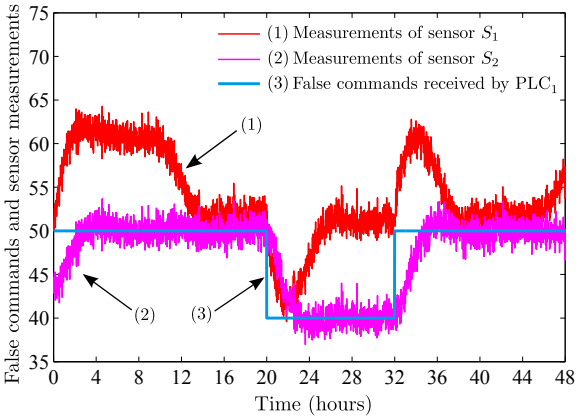
the command signals are equal to the modification of the set-points. For a closed-loop control system, the controller is responsible for regulating the system outputs to keep track of the set-points. Since the SCADA gas pipeline is also a closed-loop control system, the outlet pressure will be controlled for tracking the false commands arrived at the $\text{PLC}_1$. The command signals transmitted from the $\text{MTU}_1$ are always fixed at $y^*_{p_{\text{out}}} = 50$. Since the command signals transmitted from the $\text{MTU}_1$ to the $\text{PLC}_1$ are susceptible to either the min attack, the max attack, the scaling attack or the additive attack, the set-points received by the $\text{PLC}_1$ are different from the original ones (i.e., $\tilde{y}^*_{p_{\text{out}}} \neq y^*_{p_{\text{out}}}$).



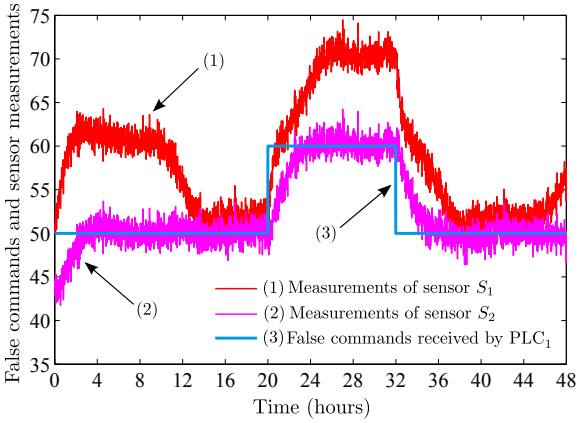

(a) Min attack strategy on command signals.




(b) Max attack strategy on command signals.




(c) Scaling attack strategy on command signals.




(d) Additive attack strategy on command signals.

Figure 6.4 – Simple integrity attacks on command signals transmitted from the $\text{MTU}_1$ to the $\text{PLC}_1$.

Since the set-point transmitted from the $\text{MTU}_1$ is fixed at $y^*_{p_{\text{out}}} = 50$, the $\text{PLC}_1$ receives the false command signals of $\tilde{y}^*_{p_{\text{out}}} = \min\{\mathcal{Y}^*\} = 30$, $\tilde{y}^*_{p_{\text{out}}} = \max\{\mathcal{Y}^*\} = 70$, $\tilde{y}^*_{p_{\text{out}}} = \alpha y^*_{p_{\text{out}}} = 40$ and $\tilde{y}^*_{p_{\text{out}}} = y^*_{p_{\text{out}}} + \delta y^*_{p_{\text{out}}} = 60$ under the min attack strategy, the max attack strategy, the scaling attack strategy with $\alpha = 0.8$ and the additive attack strategy with $\delta y^*_{p_{\text{out}}} = 10$, respectively. The measurements of sensor $S_1$ (i.e., inlet pressure $p_{\text{in}}$) and sensor $S_2$ (i.e., outlet pressure $p_{\text{out}}$) under aforementioned attack strategies are shown in sub-figure 6.4a, sub-figure 6.4b, sub-figure 6.4c and sub-figure 6.4d, respectively.

For simple integrity attack strategies on command signals, the outlet pressure $p_{\text{out}}$ is tracking the

false commands arriving at the $PLC_1$. It is worth noting that each attack strategy results in a specific attack signature (i.e., attack profile) of the outlet pressure $p_{\text{out}}$. If the attack information is known *a priori* (i.e., min-max values, scaling factor or additive value), the attack profile will be available. As discussed in chapter 3 and chapter 4, this information is essential in designing detection-isolation schemes.

**Attack on control signals**

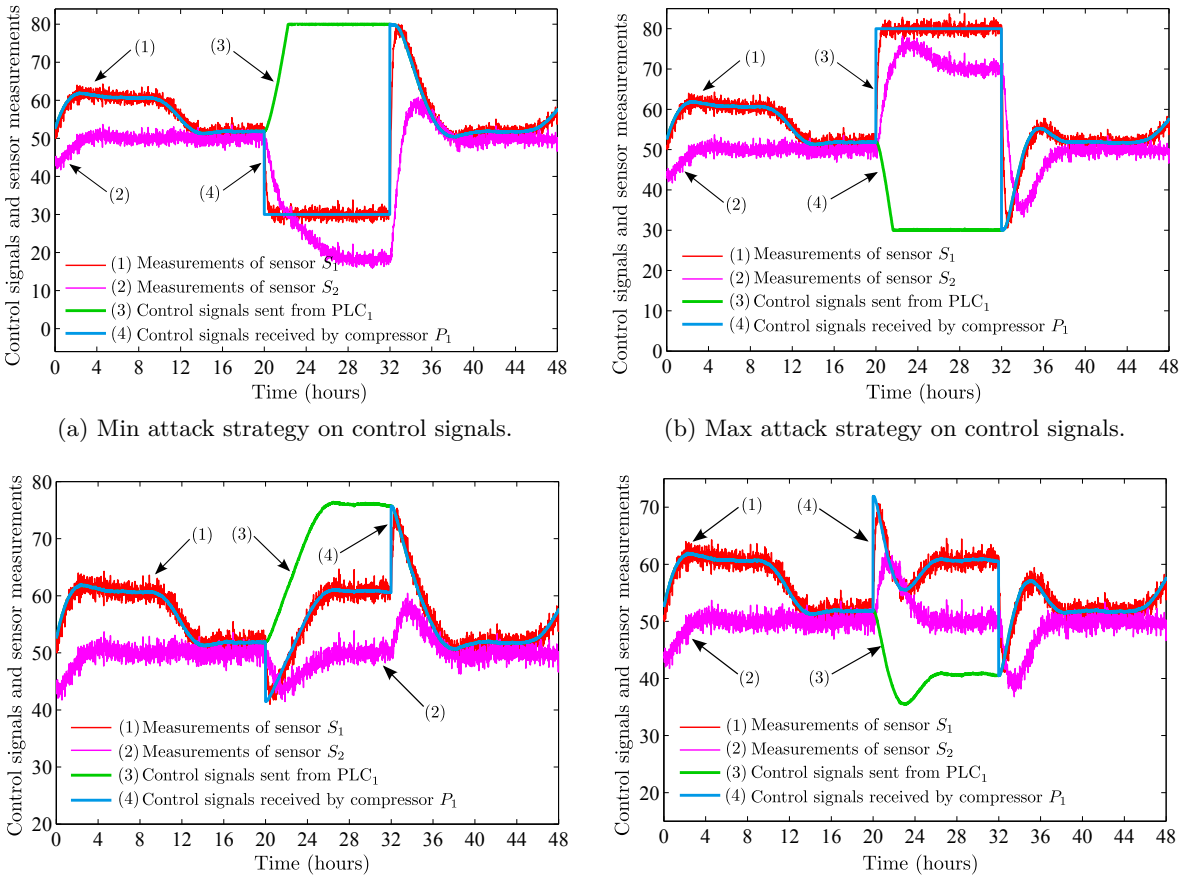
The reaction of closed-loop control systems under several simple integrity attack strategies on control signals is investigated in figure 6.5. The min attack strategy, where the control signals $u_k$ are replaced with the minimum value $u_{\min} = \min\{\mathcal{U}\} = 30$ during the attack duration $\tau_a = [20, 32]$ hours, is shown in sub-figure 6.5a. Under this attack strategy, the controller is unable to perform its task since the control signals received by the compressor $P_1$ (i.e., the blue curve) are fixed at $u_{\min} = 30$. As a consequence, the inlet pressure of the pipeline (i.e., the red curve) is regulated at a fixed value of $p_{\text{in}} = 30\,\text{bar}$, leading to a significant reduction in the outlet pressure (i.e., the magenta curve). Since the outlet pressure is decreasing during the attack period, the controller perceives that it should increase the control signals. Therefore, the true control signals (i.e., the green curve) issued by the $PLC_1$ increases dramatically until it touches the maximum value $u_{\max} = 80$.

The max attack strategy, where the control signals $u_k$ are replaced with the maximum value $u_{\max} = \max\{\mathcal{U}\} = 80$ during the attack period $\tau_a = [20, 32]$ hours, is described in sub-figure 6.5b. Under this attack strategy, the inlet pressure (i.e., the red curve) increases to the value of $p_{\text{in}} = 80\,\text{bar}$ since the control signals received by the compressor $P_1$ (i.e., the blue curve) are fixed at $u_{\max} = 80$. The increase in the inlet pressure $p_{\text{in}}$ leads to the augmentation in the outlet pressure $p_{\text{out}}$. Since the outlet pressure is increasing, the controller performs its task by reducing the control signals (i.e., the green curve). However, those signals can not reach the compressor $P_1$ due to the max attack strategy.

The scaling attack strategy and the additive attack strategy on control signals are shown in sub-figure 6.5c and sub-figure 6.5d, respectively. The reaction of the system under these malicious attacks can be analyzed in the same manner as the min-max attack strategies. The scaling attack and the additive attack on control signals can be considered as the disturbances on the system since the system variables are deflected from their nominal values for a certain amount of time and afterward they return to their normal states thanks to the closed-loop controller.

**Attack on sensor measurements**

In figure 6.6, we investigate the reaction of closed-loop control system under several simple integrity attack strategies (i.e., min attack, max attack, scaling attack and additive attack) on feedback signals (i.e., the measurements of sensor $S_2$). Under the min attack strategy (see sub-figure 6.6a), the measurements $y_{p_{\text{out}}}$ of sensor $S_2$ are replaced with the minimum value of $y_{\min} = \min\{\mathcal{Y}\} = 30$. Since the feedback signals (i.e., $y_{p_{\text{out}}} = 30$) are always smaller than the set-points (i.e., $y^*_{p_{\text{out}}} = 50$), the $PLC_1$ orders the compressor $P_1$ to speed up so as to enhance the inlet pressure up to the maximum value of around $p_{\text{in}} = 80\,\text{bar}$. The augmentation in the inlet pressure leads to the increase in the outlet pressure, thus raising the measurements of sensor $S_2$. Other attack strategies (i.e., max attack, scaling attack and additive attack) can be analyzed in the same manner.




(a) Min attack strategy on control signals.

(b) Max attack strategy on control signals.

(c) Scaling attack strategy on control signals.

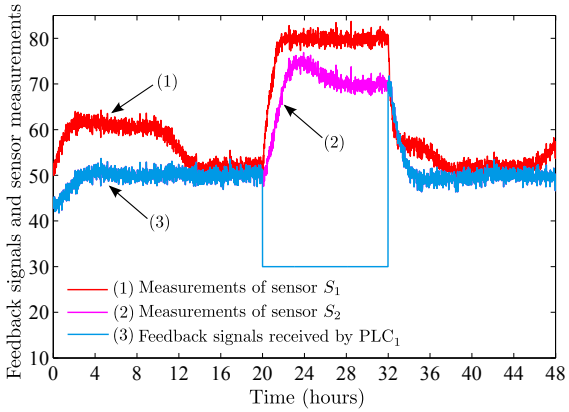
(d) Additive attack strategy on control signals.

Figure 6.5 – Simple integrity attacks on control signals transmitted from the PLC$_1$ to the compressor P$_1$.

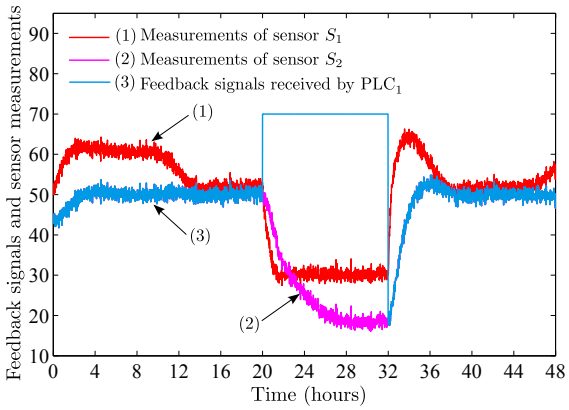
It follows from the simulation results that each simple integrity attack strategy (i.e., min attack, max attack, scaling attack and additive attack) leads to a change in sensor measurements with a specific attack signature. If the information about the attack (i.e., attack strategies and attack parameters) is known *a priori*, efficient detection-isolation schemes can be designed for jointly detecting the attack and identifying attack scenarios. In some situations, powerful attackers are able to perform stealthy attacks for disrupting the system while bypassing traditional anomaly detection schemes. The negative effect of such undetectable attacks on the SCADA gas pipeline and several countermeasures will be investigated in next subsection.

### 6.2.4 Stealthy integrity attacks

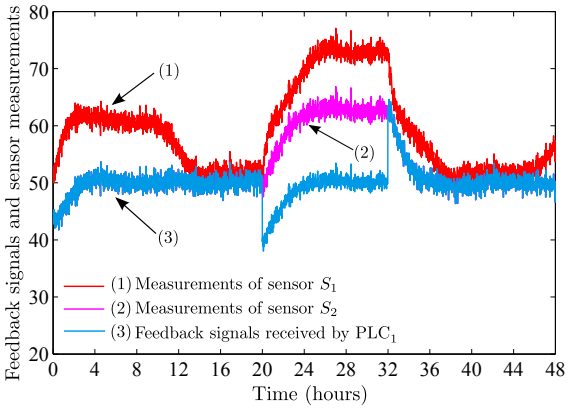
This subsection is dedicated to studying the negative impact of stealthy integrity attack strategies on the SCADA gas pipeline. For the demonstration purpose, we consider only two attack strategies: replay attack and covert attack.



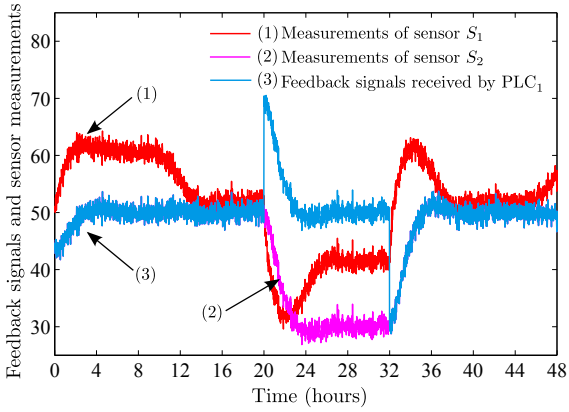

(a) Min attack strategy on feedback signals.




(b) Max attack strategy on feedback signals.




(c) Scaling attack strategy on feedback signals.




(d) Additive attack strategy on feedback signals.

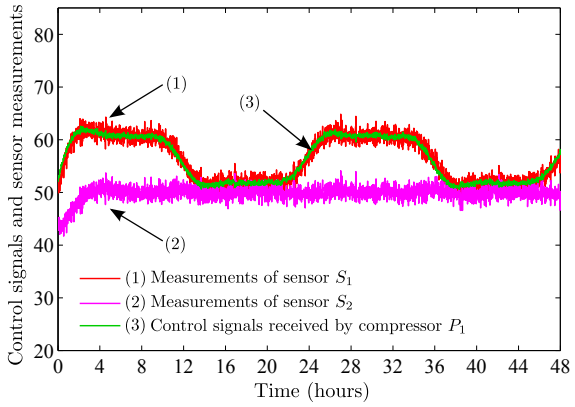
Figure 6.6 – Simple integrity attacks on feedback signals transmitted from sensor $S_2$ to the PLC$_1$.

### Replay attack strategy

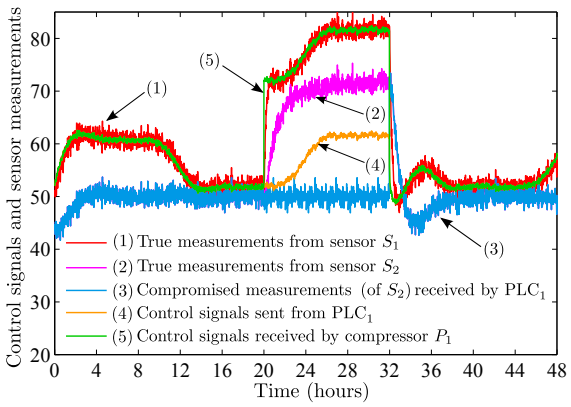
The negative impact of the replay attack on the SCADA gas pipeline is shown in figure 6.7. Under normal operation (see sub-figure 6.7a), the outlet pressure (i.e., the magenta curve) is regulated at the constant value of $p_{\text{out}} = 50\,\text{bar}$. The control signals sent from the PLC$_1$ (i.e., the green curve) are the same as those received by the compressor $P_1$. The inlet pressure $p_{\text{in}}$ (i.e., the red curve) is tracking the control signals arriving at the compressor $P_1$. The measurements of both sensors $S_1$ and $S_2$ are transmitted successfully to the PLC$_1$.

The replay attack strategy on the gas pipeline is performed as follows. During the recording period $\tau_r = [16, 18]$ hours, the feedback signals (i.e., the measurements of sensor $S_2$) are recorded. During the attack period $\tau_a = [20, 30]$ hours, the true measurements are replaced with the previously recorded signals and the control signals are modified by adding a value of $\delta u_k = 20$.

It can be seen from sub-figure 6.7b that, during the attack period, the feedback signals received by the PLC$_1$ (i.e., the blue curve) are almost constant (i.e., around 50) and the control signals issued by the PLC$_1$ (i.e., the orange curve) are almost the same as those from normal operation. Therefore, the relay attack is stealthy to any anomaly detectors which utilize only the command signals, the control signals sent from the PLC$_1$ and the feedback signals received by the PLC$_1$.




(a) Control signals and sensor measurements under normal operation.




(b) Control signals and sensor measurements under replay attack strategy.

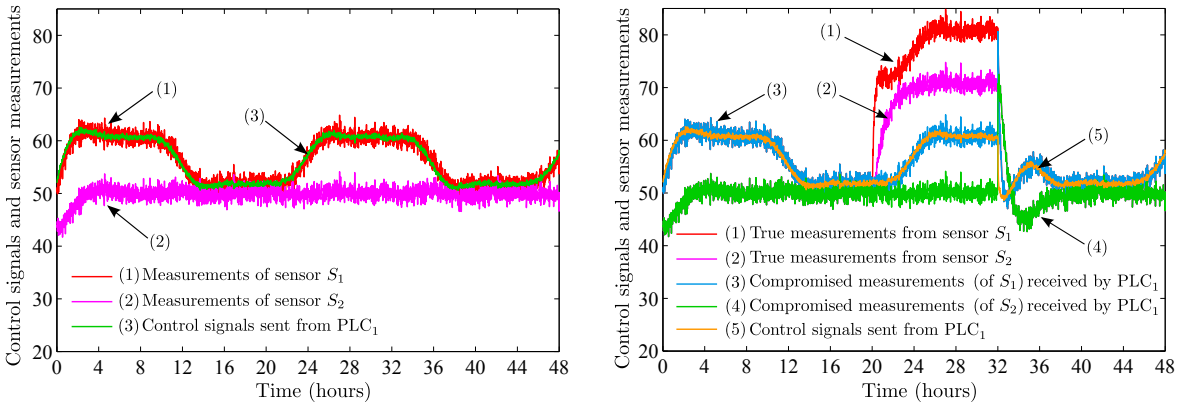
Figure 6.7 – Replay attack strategy on the SCADA gas pipeline. The recording period is $\tau_r = [16, 18]$ hours and the attack period is $\tau_a = [20, 32]$ hours. The attacker increases the control signals by a value of $\delta u_k = 20$ while replaying previously recorded signals during the attack duration.

The negative impact of the replay attack depends mostly on the modification of the control signals. In this numerical example, the control signals are modified by a value of $\delta u_k = 20$, leading to the augmentation in both inlet pressure (i.e., the red curve) and outlet pressure (i.e., the magenta curve) by a value of about 20 bar. It can be noticed that if the measurements of sensor $S_1$ (i.e., the inlet pressure $p_{\text{in}}$) are utilized, the replay attack is no longer stealthy since the information about the attack is contained in these measurements. It should be noted that the measurements of sensor $S_1$ can not be replayed successfully since the inlet pressure $p_{\text{in}}$ depends on the consumer's demands.

### Covert attack strategy

The negative impact of the covert attack on the SCADA gas pipeline is shown in figure 6.8. The normal operation of the system is described in sub-figure 6.8a, where the outlet pressure (i.e., the magenta curve) is regulated at the constant value of 50 bar. The reaction of the system under the covert attack strategy is described in sub-figure 6.8b. During the attack period $\tau_a = [20, 32]$ hours, the control signals are modified by a value of $\delta u_k = 20$. At the same time, the attack signals to the sensor measurements are calculated in such a way that they can compensate for the modification of the control signals (i.e., by the covert attack strategy (1.9)). As a result, the control signals sent from the PLC$_1$ (i.e., the orange curve) and the sensor measurements (i.e., the blue curve and the green curve) received by the PLC$_1$ are the same as those in normal operation. However, the true inlet pressure $p_{\text{in}}$ and outlet pressure $p_{\text{out}}$ (i.e., measured by sensor $S_1$, the red curve, and sensor $S_2$, the magenta curve, respectively) are increased significantly. Therefore, the covert attack strategy has the potential to cause huge damage (i.e., explosion of gas pipeline for example) without being detected by traditional anomaly detectors.

Up to the author's best knowledge, the covert attack strategy can not be revealed by analytical methods. For rendering the covert attack detectable, we propose to implement the sensor protection framework which consists of a sensor protection scheme and a sensor placement strategy.




(a) Control signals and sensor measurements under normal operation.

(b) Control signals, sensor measurements and feedback signals under covert attack.

Figure 6.8 – Covert attack strategy on the SCADA gas pipeline. The attack duration is $\tau_a = [20, 32]$ hours.

The sensor protection scheme consists in implementing some protection countermeasures so that the measurements of several critical sensors can not be modified by the attacker. The sensor placement strategy, on the other hand, deals with the equipment of new secure sensors for transmitting trusted measurements to the control center. In this numerical example, the covert attack becomes detectable of the measurements of sensor(s) $S_1$ and/or $S_2$ are transmitted successfully to the PLC$_1$.

### 6.2.5 Conclusion

In this section, the security of the simple SCADA gas pipeline has been investigated. Several types of cyber-physical attacks found in literature, including DoS attack strategies, simple integrity attack strategies and stealthy attack strategies, have been considered. Theoretical DoS attack strategies (1.3)–(1.4) can be detected easily by our methods since the attack signatures are known. A real DoS attack strategy (1.5), on the other hand, is more difficult to detect since the attack profiles depend heavily on the percentage of the successfully transmitted signals, which is generally unknown to system operators. Simple integrity attack strategies, including the min attack, max attack, scaling attack and additive attack, may cause huge damage to gas pipeline in particular and the closed-loop control systems in general. However, these naive attacks can be detected easily even by traditional anomaly detectors. Stealthy/deception attacks, on the other hand, have been demonstrated to be more difficult to detect. The replay attack strategy is stealthy to several detection schemes in particular scenarios. Well-designed detection schemes can detect the replay attack (see also [120, 122]). The covert attack strategy has been shown to be completely stealthy to traditional anomaly detectors if the attackers are able to compromise all sensors. In order to render stealthy attacks (i.e., replay attack, covert attack, and others) detectable, it is suggested to utilize the hardware redundancy approach for providing the detection-isolation algorithms with trusted measurements which contain information about the attacks.

## 6.3 Detection Algorithms Applied to Simple Water Network

In this section, the detection algorithms proposed in chapter 3 are applied to the detection of cyber-physical attacks on the simple SCADA water distribution network described in chapter 5.

### 6.3.1 Simulation parameters

Let us consider the simple SCADA water distribution network as shown in figure 5.5. Under normal operation, the linearized model of the water network is expressed in the discrete-time state space form (5.54). In this model, $x_k \in \mathbb{R}$ is the pressure head $h_1$ at the reservoir with initial value $\overline{x}_0$, $u_k \in \mathbb{R}$ is the control signals transmitted from the control center to the local controller for regulating the flow rate $Q_{01}$ through the pump, $d_k \in \mathbb{R}^2$ is the disturbances corresponding to the consumption of customers at nodes $N_3$ and $N_4$, $y_k \in \mathbb{R}^2$ is the measurements of sensors $S_1$ and $S_2$. The process noises $w_k \sim \mathcal{N}(0, Q)$ and the sensor noises $v_k \sim \mathcal{N}(0, R)$; the matrices $A \in \mathbb{R}^{1\times1}$, $B \in \mathbb{R}^{1\times1}$, $F \in \mathbb{R}^{1\times2}$, $C \in \mathbb{R}^{2\times1}$, $D \in \mathbb{R}^{2\times1}$, $G \in \mathbb{R}^{2\times2}$, $Q \in \mathbb{R}^{1\times1}$, and $R \in \mathbb{R}^{2\times2}$ (corresponding to $n = 1$, $m = 1$, $p = 2$, $q = 2$).

Under cyber-physical attacks, the system model can be described in (5.59) (resp. in (5.60)), where the attack vectors $a_k^x$ and $a_k^y$ (resp. $a_k$) and the attack matrices $K$, $H$ and $M$ (resp. $B_a$ and $D_a$) are determined by the capabilities of the adversary to disrupt the system. For the purpose of demonstration, let us consider an attack scenario where the attacker performs a coordinated attack by stealing water from the reservoir with a constant flow rate $Q_0$, turning off the pump $P_1$ and compromising the measurements of sensors $S_1$ and $S_2$ during the attack period $\tau_a = [k_0, k_0 + L - 1]$, where $k_0$ is the unknown attack instant and $L$ is the known attack duration. This attack scenario is motivated by a real attack on city water utility where the pump was burned out after being turned on and off, as reported in [213]. Hence, the attack vectors $a_k^x \in \mathbb{R}^2$ and $a_k^y \in \mathbb{R}^2$ (resp. $a_k \in \mathbb{R}^4$) are designed by the adversary and the attack matrices $K \in \mathbb{R}^{1\times2}$, $H \in \mathbb{R}^{2\times2}$ and $M \in \mathbb{R}^{2\times2}$ (resp. $B_a \in \mathbb{R}^{1\times4}$ and $D_a \in \mathbb{R}^{2\times4}$) are decided by system operators (corresponding to $r = 2$ and $s = r + p = 4$).

The linearized parameters are chosen as follows. The sample time $T_S = 100$s and the initial pressure head $\overline{x}_0 = 100$m. The system matrices $A = 1$, $B = 0.5$, $F = \begin{bmatrix} -0.5 & -0.5 \end{bmatrix}$, $C = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $D = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $G = \begin{bmatrix} 0 & 0 \\ -10 & -10 \end{bmatrix}$. The attack matrices $K = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$, $H = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ and $M = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, leading to $B_a = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \end{bmatrix}$ and $D_a = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. The sensor noise covariance matrix $R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and the process noise covariance matrix $Q = 0.02$ and $Q = 0.2$. Without loss of generality, it is assumed that the control signal $u_k = u_0 = 1$ for supplying the reservoir with $Q_{01} = 1\text{m}^3/\text{s}$ and the customer's demands fluctuate around the value $d_{1,k} \approx d_{2,k} \approx 0.5\text{m}^3/\text{s}$.

**Remark 6.1.** *It has been discussed that the covert attack is completely stealthy to traditional anomaly detectors if the attacker is able to compromise all sensors. Therefore, we propose in this numerical example a countermeasure for rendering the covert attack detectable. This method consists of protecting sensor $S_1$ so that its measurements can not be modified by the attacker. This sensor protection scheme is reflected in the matrix $M$, where $M(1,1) = 0$ means that sensor $S_1$ is secure and $M(2,2) = 1$ signifies that sensor $S_2$ is vulnerable.*

The attack parameters are chosen as follows. The stolen flow rate is $Q_0 = 0.2\text{m}^3/\text{s}$. The attack duration is $L = 8$ observations, corresponding to a period of 13.3min. The false alarm rate is measured by the time window of length $m_\alpha = 3L = 24$ observations, being equivalent to a duration of 40min. The attack vector $a_k \in \mathbb{R}^4$ is designed by the covert attack model (3.3), which was first introduced in [169], as follows:

$$a_k = \begin{cases} [0] & \text{if } k < k_0 \\ \begin{bmatrix} -0.2 \\ -1 \\ 0.6\,(k-k_0) \\ 0.6\,(k-k_0) \end{bmatrix} & \text{if } k_0 \le k < k_0 + L, \\ [0] & \text{if } k \ge k_0 + L \end{cases} \tag{6.1}$$

where [0] is the null vector. The attack profiles $\theta_1, \theta_2, \cdots, \theta_L \in \mathbb{R}^4$ can be calculated from the attack vector $a_k$ from (6.1) as $\theta_j = [-0.2,\ -1,\ 0.6\,(j-1)\,,\ 0.6\,(j-1)]^T$, for $1 \le j \le L = 8$.

**Remark 6.2.** *The information about the attack is contained in the attack vector $a_k$ (i.e., the attack profiles $\theta_1, \theta_2, \cdots, \theta_L$). The first element reflects the physical attack to withdraw water from the reservoir with the flow rate $Q_0 = 0.2\text{m}^3/\text{s}$. The second element reflects the cyber attack on the control signals for turning off the pump. The modification of the sensor measurements is reflected by the two last elements.*

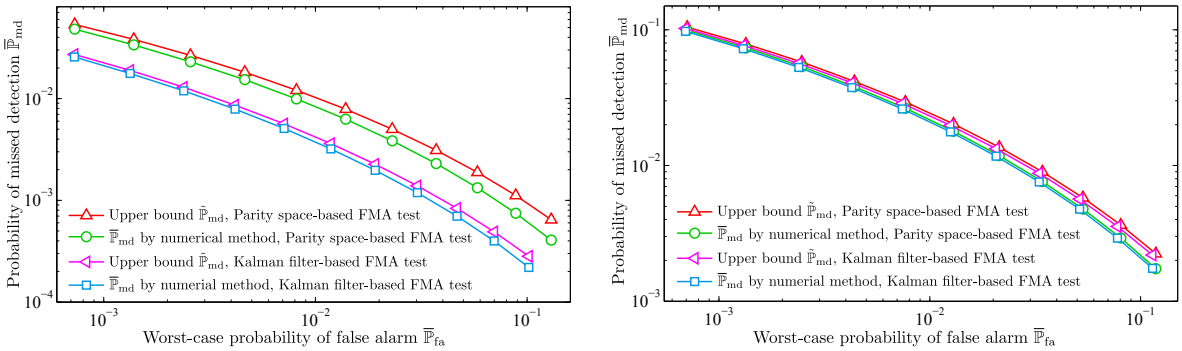
The simulation results are organized as follows. The statistical performance of the FMA tests (i.e., for both the Kalman filter approach and the parity space approach) will be investigated in subsection 6.3.2, by the Monte Carlo simulation as well as the proposed numerical method. In subsection 6.3.3, we study the robustness of the FMA test with respect to (w.r.t.) several operational parameters, including the attack duration, the attack profiles, the process and sensor noise covariance matrices by both numerical method and Monte Carlo simulation. Simulation results for the partially known transient change parameters are given in subsection 6.3.4 for demonstrating the superiority of our proposed detection rules in comparison with traditional detection algorithms.

### 6.3.2 Completely known transient change parameters

This subsection is dedicated to investigating the statistical performance of the proposed detection algorithms under the perfect conditions where system parameters are exactly known. In other words, true parameters are equal to putative parameters. Various simulation results are given and compared for demonstrating theoretical findings.

#### Upper bound on the worst-case probability of missed detection

In figure 6.9, we demonstrate the sharpness of the upper bound $\tilde{\mathbb{P}}_{\text{md}}$ on the worst-case probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ of the FMA detector. The analytical calculation of the upper bound $\tilde{\mathbb{P}}_{\text{md}}$ is compared with the numerical method for $\overline{\mathbb{P}}_{\text{md}}$ with the precision of $10^{-5}$. The change-point is chosen as $k_0 = L + 1 = 9$. We compare the analytical upper bound to the numerical method instead of the Monte Carlo simulation since the Monte Carlo simulation requires a large

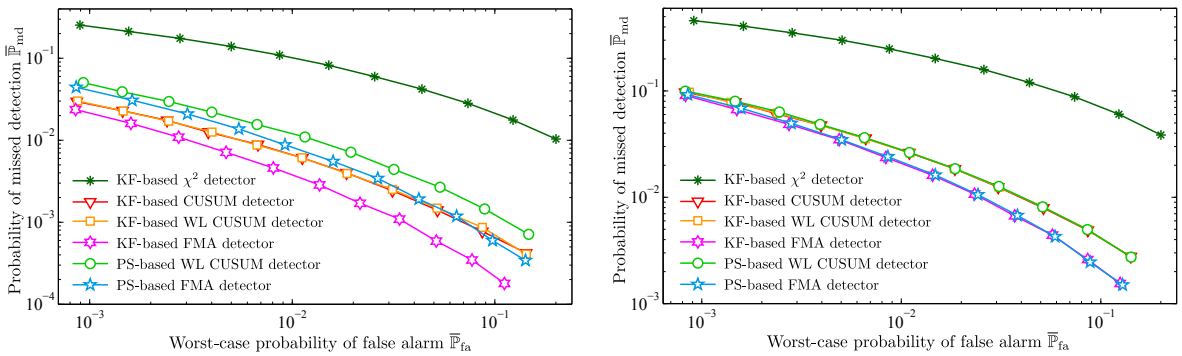



(a) Process noise variance $Q = 0.02$. (b) Process noise variance $Q = 0.2$.

Figure 6.9 – Upper bound $\tilde{\mathbb{P}}_{md}$ for the worst-case probability of missed detection $\overline{\mathbb{P}}_{md}$ of the FMA detector. The simulation has been performed with the process noise variances $Q = 0.02$ and $Q = 0.2$, respectively. The change-point for the numerical method is chosen as $k_0 = L + 1 = 9$.

amount of time for obtaining the precision of $10^{-5}$. The comparison between the numerical method and the Monte Carlo simulation method will be investigated later. It can be seen from the figure 6.9 that the proposed upper bound $\tilde{\mathbb{P}}_{md}$, for both steady-state Kalman filter approach and the fixed-size parity space approach, are quite closed to the numerical values of $\overline{\mathbb{P}}_{md}$.

### Comparison between FMA test and traditional tests




(a) Process noise variance $Q = 0.02$. (b) Process noise variance $Q = 0.2$.

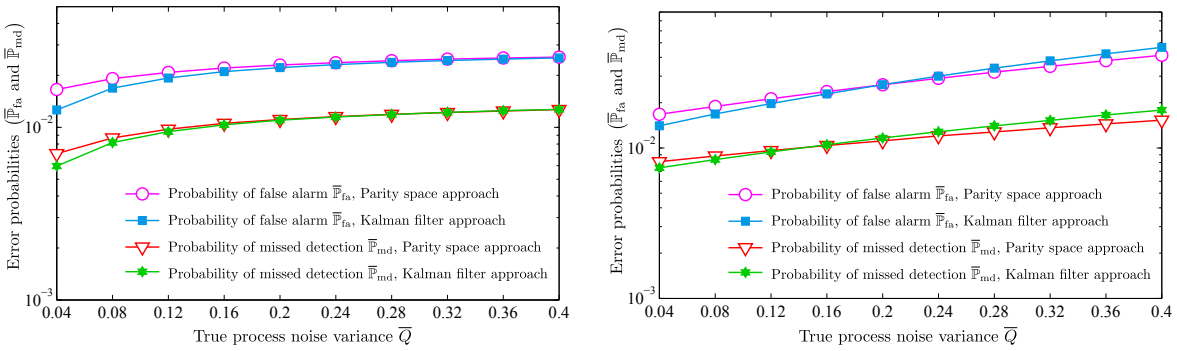
Figure 6.10 – Comparison between the steady-state Kalman filter-based detectors (i.e., KF-based $\chi^2$ detector, KF-based CUSUM detector, KF-based WL CUSUM detector and KF-based FMA detector) and the fixed-size parity space-based detectors (i.e., PS-based WL CUSUM detector and PS-based FMA detector).

The statistical performance of several detection rules by the Monte Carlo simulation of $10^6$ repetitions are shown in figure 6.10. The WL CUSUM test is, in fact, the VTWL CUSUM test with equal thresholds (i.e., $h_1 = h_2 = \cdots = h_L$). The following remarks can be drawn from the simulation results. Firstly, the proposed algorithms (i.e., the CUSUM test, WL CUSUM test, and FMA test) are much better than the traditional non-parametric $\chi^2$ detector. This phenomenon can be explained from the fact that the $\chi^2$ test does not exploit the information

about the transient change profiles while the others utilize this essential information. Secondly, given an acceptable level on the probability of false alarm, the probability of missed detection of the FMA tests is much smaller than that of both CUSUM and WL CUSUM tests. In other words, the proposed FMA tests perform better than the traditional tests, for both the steady-state Kalman filter approach and the fixed-size parity space approach. These simulation results are due to the fact that the optimization of the WL CUSUM algorithm leads to the FMA detection rule. Finally, the statistical performance of the Kalman filter-based algorithms are much better than those of the parity space-based tests when the process noises are small (i.e., process noise variance $Q = 0.02$ in our example). On the other hand, two approaches are comparable in such scenarios that the process noises are large (i.e., process noise variance $Q = 0.2$). The comparison between the Kalman filter approach and the parity space approach is shown in the following sub-subsection.

### Comparison between Kalman filter approach and parity space approach

The comparison between the steady-state Kalman filter approach and the fixed-size parity space approach is shown in figure 6.11 by the Monte Carlo simulation with $10^6$ repetitions.




(a) The putative process noise variance $Q$ is equal to the true process noise variance $\overline{Q}$, both varying from $Q = \overline{Q} = 0.02$ to $Q = \overline{Q} = 0.4$.

(b) The putative process noise variance is fixed at $Q = 0.1$ and the true process noise variance varies from $\overline{Q} = 0.02$ to $\overline{Q} = 0.4$.

Figure 6.11 – Statistical performance comparison between the steady-state Kalman filter-based FMA test and the fixed-size parity space-based FMA test. The worst-case probability of false alarm $\overline{\mathbb{P}}_{fa}$ and the probability of missed detection $\overline{\mathbb{P}}_{md}$ are described as a function of the true process noise variance $\overline{Q}$ which varies from $\overline{Q} = 0.02$ to $\overline{Q} = 0.4$ with the step of $\delta\overline{Q} = 0.02$.

Let us discuss the perfect condition where the process noise variance is exactly known (i.e., $\overline{Q} = Q$). It can be seen clearly from the sub-figure 6.11a that the steady-state Kalman filter approach performs better than the fixed-size parity space approach, especially when the process noises are small. This phenomenon is explained in figure 6.12, where the Kullback-Leibler (K-L) distances of the residuals generated by two approaches are computed and compared. The steady-state Kalman filter generates the residuals with higher K-L distance than the fixed-sized parity space does. The difference becomes significant in such scenarios that the process noises are extremely small. In contrast, when the process noises are large, the difference is negligible. This phenomenon is explained by the approximation of the Bayesian approach (i.e., the steady-state Kalman filter) by the minimax approach (i.e., the fixed-size parity space) produces a significant error only if the process noise is small and, hence, the *a priori* information plays an important

role.

It can be seen from figure 6.12 that the K-L distance of the residuals generated by the general fixed-size parity space approach is equal to the K-L distance of the residuals generated by the least-square estimation method proposed by Gustafsson [72, 73]. This simulation is also consistent with the theoretical results obtained in section (3.3) and previous findings derived in [51], i.e., the statistical performance of a likelihood ratio-based detection procedure on the basis of parity space approach is independent from the choice of the rejection matrix $\mathcal{W}$.

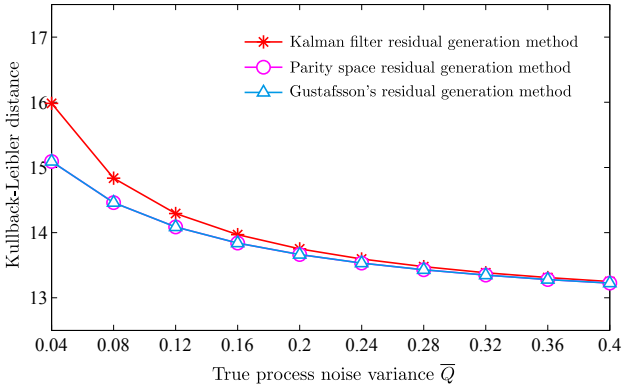


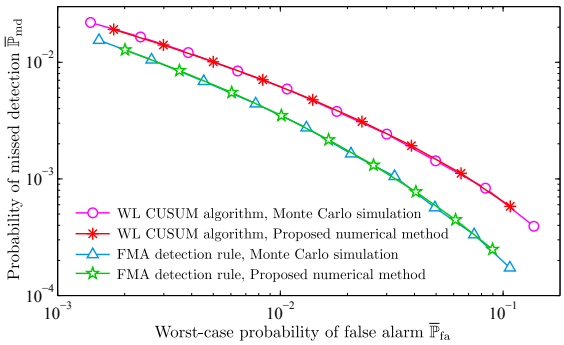

Figure 6.12 – Kullback-Leibler distance of the residuals generated by the steady-state Kalman filter and the fixed-size parity space as a function of true process noise variance $\overline{Q}$.

Let us consider an imperfect scenario where the true process noise variance matrix $\overline{Q}$ is different from the putative value $Q$. The putative process noise variance is chosen as $Q = 0.1$ and the true value varies from $\overline{Q} = 0.02$ to $\overline{Q} = 0.4$ with a step of $\delta\overline{Q} = 0.02$. The statistical performance of the steady-state Kalman filter-based FMA test and the fixed-size parity space-based FMA test is shown and compared in sub-figure 6.11b. It can be seen that the Kalman filter-based FMA test is more sensitive to the true process noises than the parity space-based FMA test. The statistical performance of the Kalman filter-based algorithm reduce significantly with large values of $\overline{Q}$. From some values of $\overline{Q}$ (i.e., when $\overline{Q} \geq 0.2$ in our numerical example), the steady-state Kalman filter-based FMA test performs worse than the fixed-size parity space-based FMA test. This phenomenon can be explained by the fact that the Kalman filter with incorrect process noise information may produce an accumulated state estimation error, especially when the true process noises are larger than their putative value. The statistical performance of Kalman filter-based detection schemes reduces accordingly.

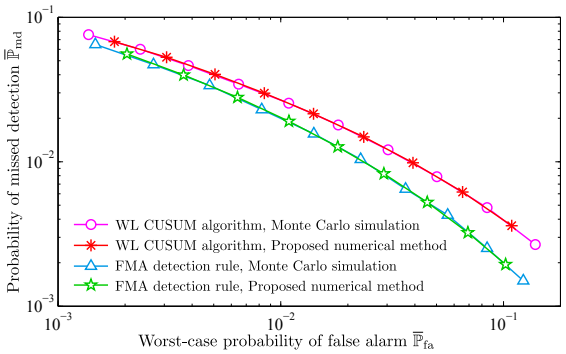
### Numerical calculation of error probabilities

The comparison between the proposed numerical method and the Monte Carlo simulation is given in figure 6.13. The Monte Carlo simulation is executed with $10^6$ repetitions while the numerical method is performed with the precision of $10^{-5}$. This simulation study is executed for both steady-state Kalman filter approach (i.e., sub-figure 6.13a and sub-figure 6.13b) and fixed-size parity space approach (i.e., sub-figure 6.13c and sub-figure 6.13d), and for two values of process noise variance (i.e., $Q = 0.02$ and $Q = 0.2$, respectively).

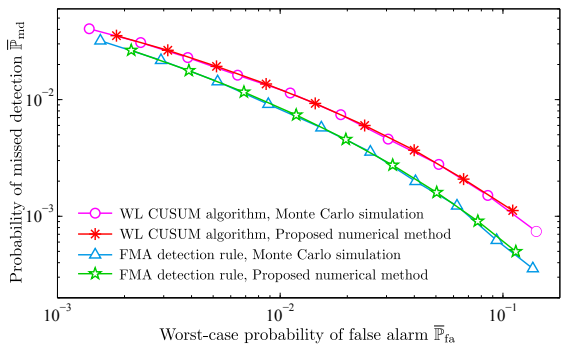
It follows from sub-figures 6.13a, 6.13b, 6.13c and 6.13d that the numerical curves perfectly coincide with the Monte Carlo curves, thus proving the correctness of the proposed numerical



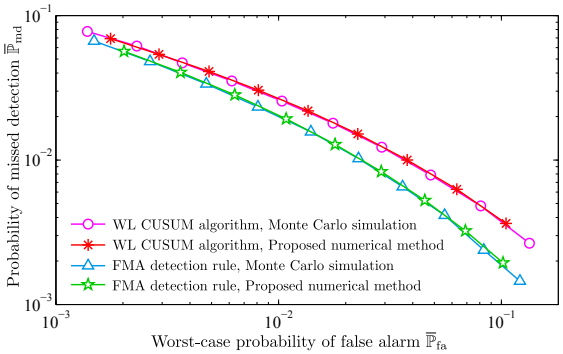

(a) Steady-state Kalman filter approach with process noise variance $Q = 0.02$.




(b) Steady-state Kalman filter approach with process noise variance $Q = 0.2$.




(c) Fixed-size parity space approach with process noise variance $Q = 0.02$.




(d) Fixed-size parity space approach with process noise variance $Q = 0.2$.

Figure 6.13 – Comparison between the numerical method and the Monte Carlo simulation, for both Kalman filter approach and parity space approach.

method. In addition, the numerical method requires smaller amount of time for obtaining the same precision as the Monte Carlo simulation, especial for the FMA test.

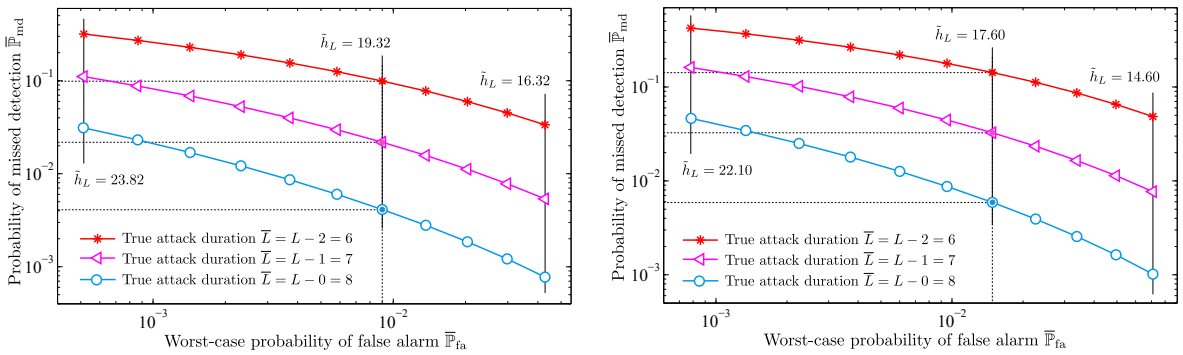
### 6.3.3 Sensitivity analysis of the FMA test

In subsection 3.4.4, we have proposed a numerical method for evaluating the robustness of the FMA test with respect to (w.r.t.) several operational parameters, including the attack duration, the attack profiles, the process noise covariances and the sensor noise covariances. In this subsection, the results in subsection 3.4.4 are applied to investigate the sensitivity of the FMA test w.r.t. these parameters, for both steady-state Kalman filter approach and fixed-size parity space approach. The comparison between the proposed numerical method and the Monte Carlo simulation is also performed.

#### Robustness of the FMA test w.r.t. the attack duration

The sensitivity of the FMA test w.r.t. the attack duration is shown in figure 6.14, for both steady-state Kalman filter approach (i.e., sub-figure 6.14a) and fixed-size parity space approach (i.e., sub-figure 6.14b). In this simulation study, the putative attack duration and the true

attack duration are chosen as $L = 8$ and $\overline{L} = \{6, 7, 8\}$, respectively. The process noise variance is $Q = 0.02$. The probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is described as a function of the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ w.r.t. different values of the true attack duration $\overline{L} = \{6, 7, 8\} \leq L$. Each curve corresponds to one specific value of $\overline{L}$. Some conclusions are drawn as follows.




(a) Steady-state Kalman filter-based FMA test.

(b) Fixed-size parity space-based FMA test.

Figure 6.14 – Sensitivity of the FMA test with respect to the attack duration. The probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is described as a function of the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ for different values of the true attack duration $\overline{L} = \{6, 7, 8\} \leq L = 8$.

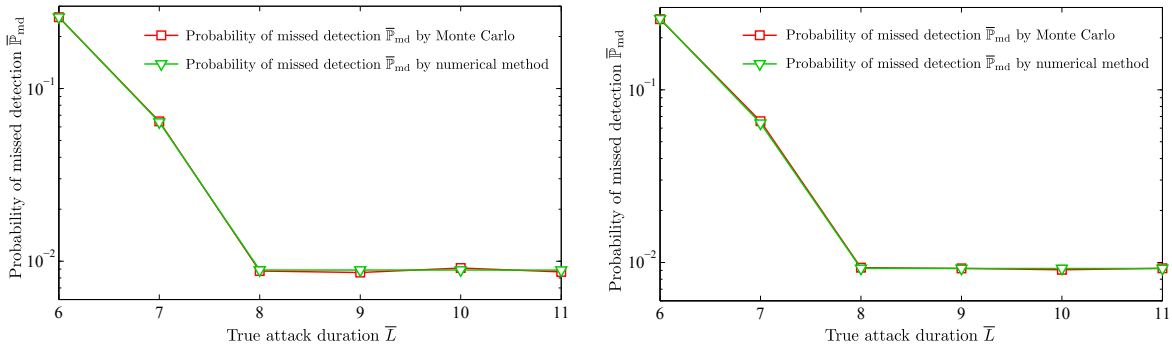
If the true attack duration is greater than the putative value (i.e., $\overline{L} \geq L$), the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ remains unchanged since any detection with the detection delay greater than $L$ is considered as missed. For $\overline{L} = \{6, 7, 8\} \leq L$, the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ depends heavily on the true attack duration $\overline{L}$. The smaller the true attack duration $\overline{L}$, the higher the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$. This phenomenon is explained by the fact that small attack duration $\overline{L}$ causes little changes in the distribution of the observations, thus increasing the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$. On the other hand, the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ is insensitive to the true attack duration $\overline{L}$. This phenomenon can be seen clearly that, for the false alarm case, all the observations are generated from the pre-change distribution.

The interpretation of figure 6.14 is very simple: each value of the probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ corresponds to a certain value of the threshold $\tilde{h}_L$, which is the tuning parameter of the FMA test. By drawing a vertical line, we can estimate the variation of the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ due to a true attack duration smaller than its putative value for a given tuning of the FMA test.

The comparison between the numerical method and the Monte Carlo simulation is also shown in figure 6.15. The precision of the numerical method is chosen as $10^{-5}$ and the Monte Carlo simulation is of $2.10^5$ repetitions. Clearly, the numerical method gives almost the same results as the Monte Carlo simulation does, for both the steady-state Kalman filter-based FMA test (i.e., sub-figure 6.15a) and the fixed-size parity space-based FMA test (i.e., sub-figure 6.15b), thus proving the correctness of the proposed numerical method.

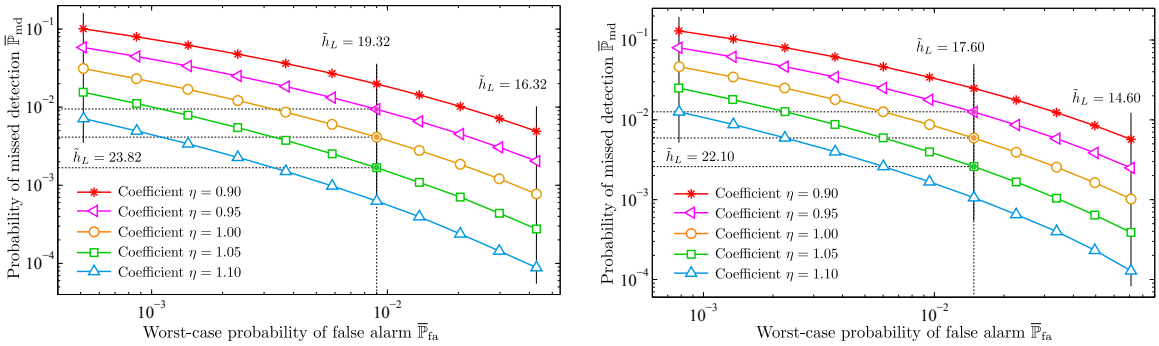
### Robustness of the FMA test w.r.t. the attack profiles

The sensitivity of the FMA test w.r.t. the attack profiles is shown in figure 6.16, for both steady-state Kalman filter approach (i.e., sub-figure 6.16a) and fixed-size parity space approach (i.e.,




(a) Steady-state Kalman filter-based FMA test. (b) Fixed-size parity space-based FMA test.

Figure 6.15 – Comparison between the numerical method and Monte Carlo simulation. The probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is described as a function of the true attack duration $\overline{L} = \{6, 7, 8, 9, 10, 11\}$, for both the Kalman filter approach (left) and the parity space approach (right).

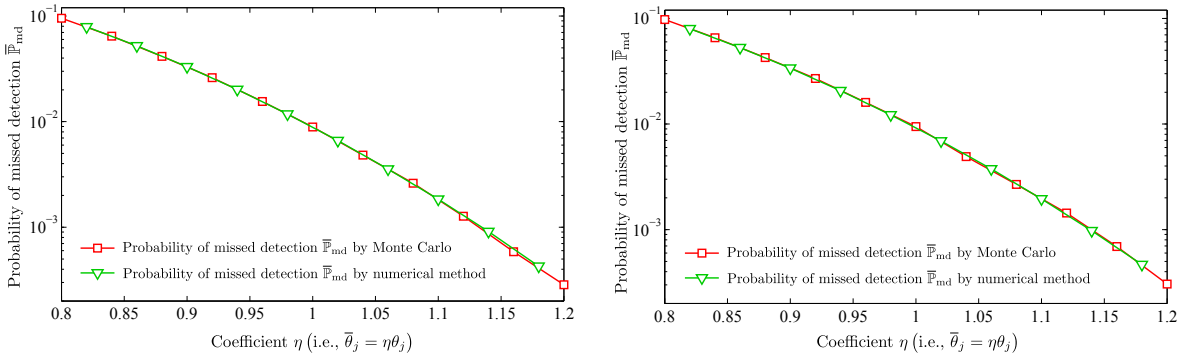



(a) Steady-state Kalman filter-based FMA test. (b) Fixed-size parity space-based FMA test.

Figure 6.16 – Sensitivity of the FMA test with respect to the attack profiles. The probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is described as a function of the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ for different values of $\eta = \{0.90, 0.95, 1.00, 1.05, 1.10\}$. The true attack profiles are related to the putative attack profiles by $\overline{\theta}_j = \eta\theta_j$, for $1 \leq j \leq L$.

sub-figure 6.16b). The true attack profiles $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$ are chosen such as $\overline{\theta}_j = \eta\theta_j$ for $1 \leq j \leq L$, where the coefficient $\eta = \{0.90, 0.95, 1.00, 1.05, 1.10\}$. In other words, the "magnitude" of the change varies from 90% to 110% but the "shape" of the change remains unchanged. Similar to the attack duration case, the probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ is insensitive to the true attack profiles since all the observations are generated from the pre-change distribution. In contrast, the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ depends heavily on the true attack profiles $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$. The smaller the true attack profiles $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$, the higher the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$. This phenomenon can be explained by the fact that small true attack profiles lead to little changes in the distribution of the observations, thus augmenting the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ and vice versa. The variation of the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ due to the difference between the true attack profiles and their putative values w.r.t. the tuning parameter $\tilde{h}_L$ can be determined exactly in the same manner as in the attack duration case.

The comparison between the numerical method and the Monte Carlo simulation is given in

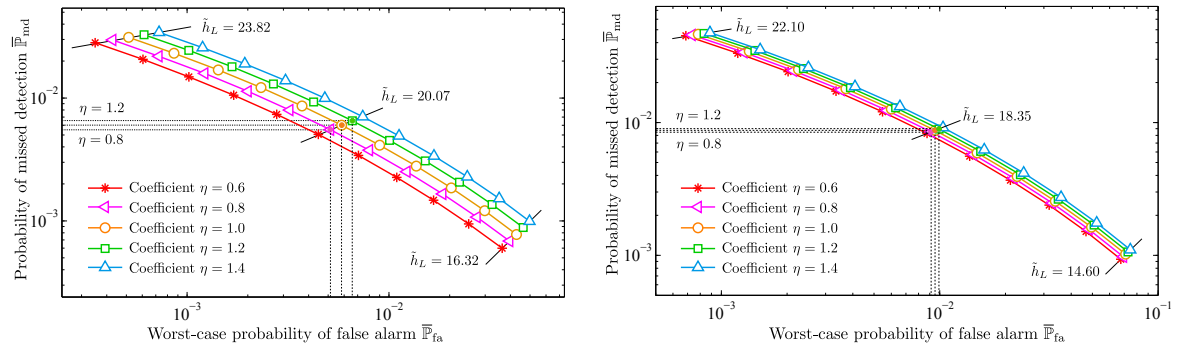



(a) Steady-state Kalman filter-based FMA test. (b) Fixed-size parity space-based FMA test.

Figure 6.17 – Comparison between the numerical method and the Monte Carlo simulation. The probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is described as a function of the coefficient $\eta$, where $\overline{\theta}_j = \eta\theta_j$ for $1 \leq j \leq L$.

figure 6.17. The probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is described as a function of the coefficient $\eta$ which varies from $\eta = 0.8$ to $\eta = 1.2$ with the step of 0.04. Clearly, the numerical curves perfectly coincide with the Monte Carlo curves, for both steady-state Kalman filter approach (i.e., sub-figure 6.17a) and fixed-size parity space approach (i.e., sub-figure 6.17b), thus verifying the precision of the proposed numerical method.

**Robustness of the FMA test w.r.t. the process noises**



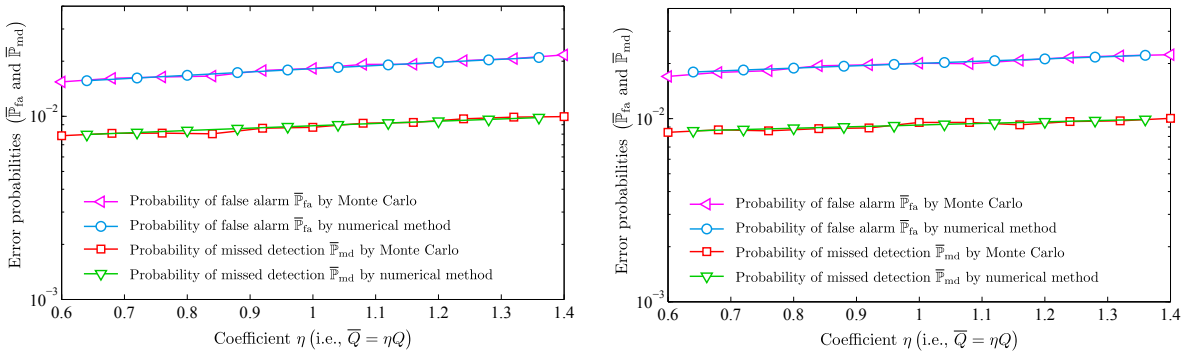

(a) Steady-state Kalman filter-based FMA test. (b) Fixed-size parity space-based FMA test.

Figure 6.18 – Sensibility of the FMA test with respect to the process noises. The probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is described as a function of the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ for different values of $\eta = \{0.6, 0.8, 1.0, 1.2, 1, 4\}$. The true process noise variance is related to its putative value by $\overline{Q} = \eta Q$.

The sensitivity of the FMA test w.r.t. the process noises is described in figure 6.18, for both steady-state Kalman filter approach (i.e., sub-figure 6.18a) and fixed-size parity space approach (i.e., sub-figure 6.18b). In these sub-figures, the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is described as a function of the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ for different values of the true process noise variance $\overline{Q} = \eta Q$, where the putative process noise variance is chosen as $Q = 0.1$

and the coefficient $\eta = \{0.6, 0.8, 1.0, 1.2, 1.4\}$. In this case, the difference $\overline{Q} - Q$ impacts both the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ and the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$. Roughly speaking, the bigger the sensor noises, the higher the error probabilities, i.e, $\overline{\mathbb{P}}_{\text{fa}}$ and $\overline{\mathbb{P}}_{\text{md}}$. For this reason, the interpretation of figure 6.18 w.r.t. the tuning parameter $\tilde{h}_L$ is more complicated.

For simplifying the explanation, three isolines of constant threshold $\tilde{h}_L$ are added to sub-figure 6.18a and sub-figure 6.18b, respectively, for the steady-state Kalman filter approach and fixed-size parity space approach. The tuning parameter $\tilde{h}_L$ is fixed by selecting a point in the curve corresponding to $\eta = 1.0$. The worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ and the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ are determined by drawing, respectively, vertical and horizontal dotted lines from the selected point. The variation of the error probabilities, i.e., $\overline{\mathbb{P}}_{\text{fa}}$ and $\overline{\mathbb{P}}_{\text{md}}$, due to the difference between the true process noise variance $\overline{Q}$ and the its putative value $Q$ can be estimated by utilizing the isoline intersecting the selected point. For example, the isoline of $\tilde{h}_L = 20.07$ in sub-figure 6.18a and the isoline of $\tilde{h}_L = 18.35$ in sub-figure 6.18b are utilized for determining the variation of $\overline{\mathbb{P}}_{\text{fa}}$ and $\overline{\mathbb{P}}_{\text{md}}$ for the steady-state Kalman filter approach and the fixed-size parity space approach, respectively. It can be seen clearly from sub-figure 6.18a and sub-figure 6.18b that the Kalman filter-based FMA test is much more sensitive to the process noises than the parity space-based FMA test. This sensitivity analysis is useful in choosing between the Kalman filter approach and the parity space approach in such situations that the process noises are large.



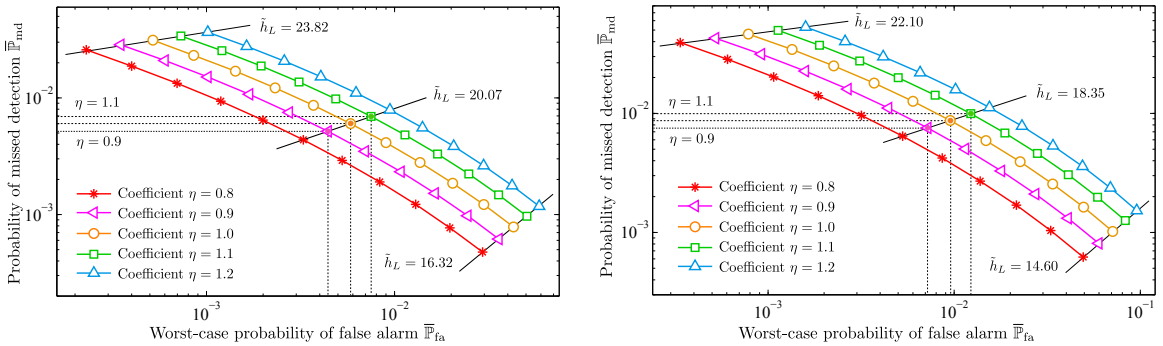

(a) Steady-state Kalman filter-based FMA test.

(b) Fixed-size parity space-based FMA test.

Figure 6.19 – Comparison between the numerical method and the Monte Carlo simulation. The error probabilities ($\overline{\mathbb{P}}_{\text{fa}}$ and $\overline{\mathbb{P}}_{\text{md}}$) are described as a function of the coefficient $\eta$ where $\overline{Q} = \eta Q$.

The comparison between the numerical method and the Monte Carlo simulation is shown in figure 6.19, for both steady-state Kalman filter approach (i.e., sub-figure 6.19a) and fixed-size parity space approach (i.e., sub-figure 6.19b). The error probabilities, i.e., $\overline{\mathbb{P}}_{\text{fa}}$ and $\overline{\mathbb{P}}_{\text{md}}$, are described as a function of the coefficient $\eta$ which varies from $\eta = 0.6$ to $\eta = 1.4$ for the step of 0.04. The true process noise covariance is related to its putative value by $\overline{Q} = \eta Q$, where $Q = 0.1$. It can be seen clearly from the figure that two curves (numerical and Monte Carlo) perfectly coincide, thus proving the correctness of the proposed numerical method. In addition, the coincidence between the numerical curve and the Monte Carlo curve in sub-figure 6.19a validates also the recursive algorithm 2 proposed for calculating the covariance between two residuals generated from the steady-state Kalman filter under imperfect condition, i.e., the true process noise covariance is different from the putative one.

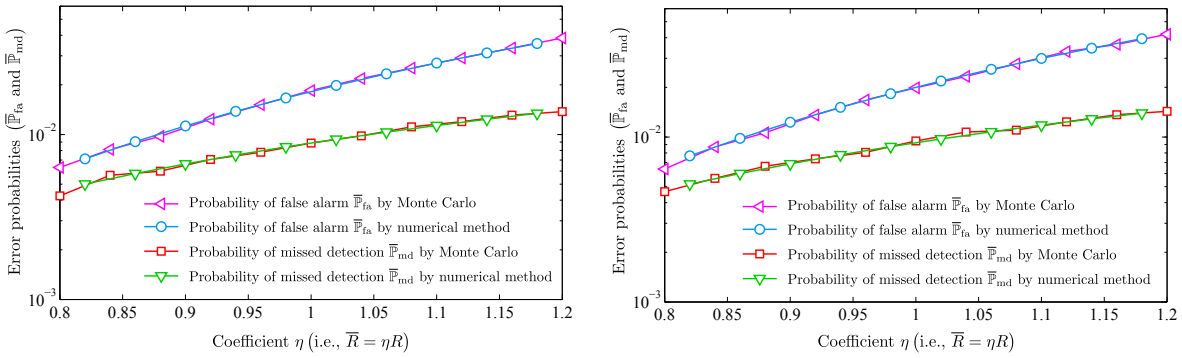
**Robustness of the FMA test w.r.t. sensor noises**




(a) Steady-state Kalman filter-based FMA test. (b) Fixed-size parity space-based FMA test.

Figure 6.20 – Sensitivity of the FMA test with respect to the sensor noises. The probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is described as a function of the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ for different values of $\eta = \{0.8, 0.9, 1.0, 1.1, 1.2\}$. The true sensor noise covariance $\overline{R}$ is related to its putative value by $\overline{R} = \eta R$.

The sensitivity of the FMA test w.r.t. the sensor noises is described in figure 6.20, for both steady-state Kalman filter approach (i.e., sub-figure 6.20a) and fixed-size parity space approach (i.e., sub-figure 6.20b). Similar to the process noise case, the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is described as a function of the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ for different values of true sensor noise covariance $\overline{R} = \eta R$, where the coefficient $\eta = \{0.8, 0.9, 1.0, 1.1, 1.2\}$. Similar to the process noise case, the variation in the true sensor noise covariance matrix $\overline{R}$ leads to a substantial change in both the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ and the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$. The smaller the true sensor noise covariance matrix $\overline{R}$, the better the statistical performance of the FMA test (i.e., the smaller $\overline{\mathbb{P}}_{\text{fa}}$ and $\overline{\mathbb{P}}_{\text{md}}$). The variation of the error probabilities due to the difference between the true sensor noise covariance $\overline{R}$ and its putative value $R$ can be analyzed in exactly the same manner as in the case of process noises. This analysis could help in finding a tradeoff between the performance of the detection algorithms and the price of high-precision sensors.




(a) Steady-state Kalman filter-based FMA test. (b) Fixed-size parity space-based FMA test.

Figure 6.21 – Comparison between the numerical method and the Monte Carlo simulation. The error probabilities ($\overline{\mathbb{P}}_{\text{fa}}$ and $\overline{\mathbb{P}}_{\text{md}}$) are described as a function of the coefficient $\eta$, where $\overline{R} = \eta R$.

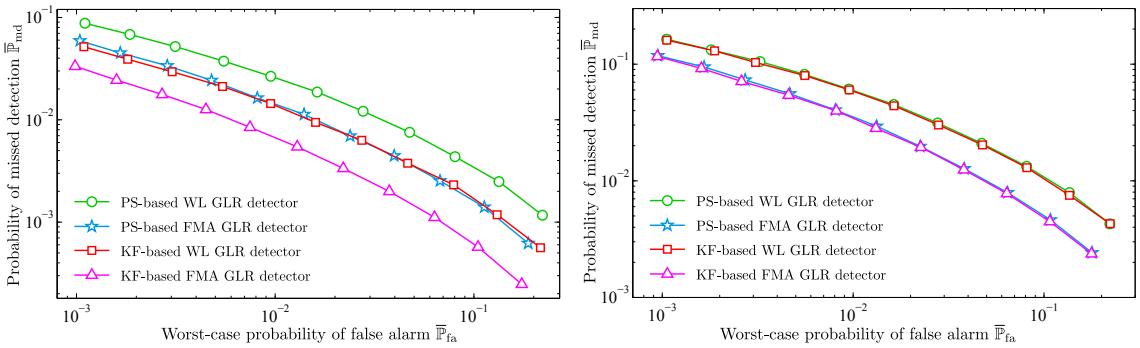
The comparison between the numerical method and the Monte Carlo simulation is shown in figure 6.21, for both steady-state Kalman filter approach (i.e., sub-figure 6.21a) and fixed-size parity space approach (i.e., sub-figure 6.21b). The error probabilities, i.e., $\overline{\mathbb{P}}_{\text{fa}}$ and $\overline{\mathbb{P}}_{\text{md}}$, are described as a function of the coefficient $\eta$ which varies from $\eta = 0.8$ to $\eta = 1.2$ for the step of 0.02. Again, the numerical curves perfectly coincide with the Monte Carlo curves, for both residual-generation methods, thus showing the correctness of our proposed numerical method. Similar to the process noise case, the coincidence between two curves (numerical and Monte Carlo) also validates the recursive algorithm 2 proposed for calculating the covariance between two residuals generated from the steady-state Kalman filter under imperfect conditions, i.e., the true sensor noise covariance is different from the putative one.

### 6.3.4 Partially known transient change parameters

This subsection is dedicated to investigating the statistical performance of the FMA GLR test (3.65) and the FMA WLR test (3.66) by the Monte Carlo simulation. In order to demonstrate the theoretical results obtained in subsection 3.5, we compare the FMA GLR test (resp. FMA WLR test) with the WL GLR test (resp. WL WLR test). It is worth noting that the WL GLR test (resp. WL WLR test) is the special case of the VTWL GLR test (resp. VTWL WLR test) with equal thresholds (i.e., $h_1 = h_2 = \cdots = h_L = h$). The comparison between the FMA GLR test and the FMA WLR test is also performed.

#### Comparison between the FMA GLR test and the WL GLR test

The performance comparison between the FMA GLR test (3.65) and the WL GLR test (3.51) is shown in figure 6.22, for both steady-state Kalman filter approach and fixed-size parity space approach. The simulation parameters remain unchanged. Two values of process noise variance are considered: $Q = 0.02$ (i.e., sub-figure 6.22a) and $Q = 0.2$ (i.e., sub-figure 6.22b). In each sub-figure, the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is described as a function of the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$.




(a) Process noise variance $Q = 0.02$.

(b) Process noise variance $Q = 0.2$.

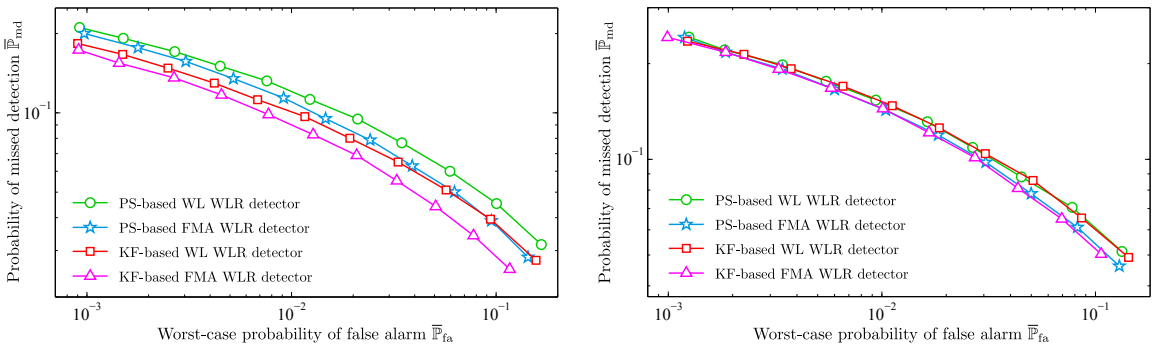
Figure 6.22 – Comparison between the FMA GLR test and the WL GLR test. The probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is described as a function of the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$.

It can be seen clearly that, for a given value on the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$, the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ of the FMA GLR test is smaller than that of the WL

GLR test, for both steady-state Kalman filter approach and fixed-size parity space approach. In other words, the FMA GLR test performs much better than the WL GLR test w.r.t. the transient detection criterion. Moreover, similar to the completely known transient parameters, the steady-state Kalman filter approach gives better statistical performance than the fixed-size parity space approach, especially for small values of process noises (see the difference between sub-figure 6.22a for $Q = 0.02$ and sub-figure 6.22b for $Q = 0.2$).

### Comparison between the FMA WLR test and the WL WLR test

The performance comparison between the FMA WLR test and the WL WLR test is described in figure 6.22. The simulation parameters remain unchanged in comparison to the GLR approach. The *a priori* distribution of the parameter $\gamma$ is chosen as $\gamma \sim \mathcal{U}(\gamma_0, \gamma_1)$, where $\gamma_0 = 0.5$ and $\gamma_1 = 1.5$. The simulation is performed by the following manner. For each Monte Carlo run, the parameter $\gamma$ is generated from the uniform distribution $\mathcal{U}(\gamma_0, \gamma_1)$. The true attack profiles are then calculated from their putative values as $\overline{\theta}_j = \gamma\theta_j$, for $1 \leq j \leq L$. Finally, the WLR-based detection rules are executed obtaining false alarm and missed detection rates.




(a) Process noise variance $Q = 0.02$.

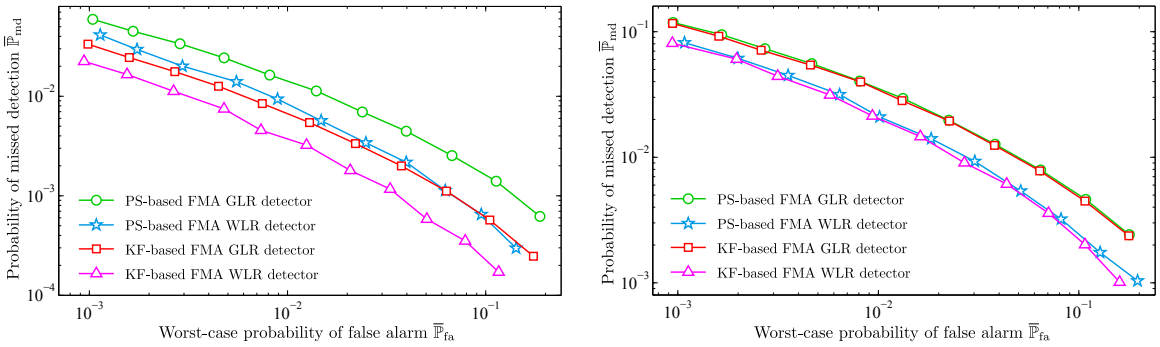
(b) Process noise variance $Q = 0.2$.

Figure 6.23 – Comparison between the FMA WLR test and the WL WLR test. The probability of missed detection $\overline{\mathbb{P}}_{md}$ is described as a function of the worst-case probability of false alarm $\overline{\mathbb{P}}_{fa}$. The *a priori* distribution of the parameter $\gamma$ is chosen as $\gamma \sim \mathcal{U}(0.5, 1.5)$.

It can be concluded from figure 6.22 that the FMA WLR detectors perform much better than the WL WLR detectors, for both steady-state Kalman filter approach and fixed-size parity space approach. As usual, the steady-state Kalman filter approach offers better statistical performance than the fixed-size parity space approach, especially for small process noises. This phenomenon can be seen from sub-figure 6.23a (i.e, for $Q = 0.02$) and sub-figure 6.23b (i.e., for $Q = 0.2$).

### Comparison between the FMA GLR test and the FMA WLR test

It is worth noting that the detection rates are strongly dependent on the parameters $\gamma_0$ and $\gamma_1$ since the true attack profiles $\overline{\theta}_j = \gamma\theta_j$, for $1 \leq j \leq L$. For fixed putative profiles $\theta_1, \theta_2, \cdots, \theta_L$, the higher the parameters $\gamma_0$ and $\gamma_1$, the better the statistical performance of the WLR-based detectors. In order to compare the GLR-based approach to the WLR approach, the parameter $\gamma$ is fixed at $\gamma = 1$, i.e., the true attack profiles are equal to the putative attack profiles. The




(a) Process noise variance $Q = 0.02$.

(b) Process noise variance $Q = 0.2$.

Figure 6.24 – Comparison between the FMA GLR test and the FMA WLR test for $Q = 0.02$ and $Q = 0.2$. The parameter $\gamma$ is fixed at value $\gamma = 1$ for the WLR-based detectors.

FMA GLR detector (3.65) and the FMA WLR detector (3.66) are tested with $2.10^5$ Monte Carlo repetitions and the false alarm and missed detection rates are computed.

The performance comparison between the FMA GLR detectors and the FMA WLR detectors is shown in figure 6.24, for both steady-state Kalman filter approach and fixed-size parity space approach. The WLR-based detectors have been shown to perform much better than the GLR-based detectors for both values of process noise variance $Q = 0.02$ and $Q = 0.2$. This phenomenon can be explained from the fact the WLR approach exploits the *a priori* information about change magnitude while the GLR approach does not utilize this essential information.

## 6.4 Detection-Isolation Algorithms Applied to Complex Water Networks

In this section, the detection-isolation schemes proposed in chapter 4 are applied to the joint detection and isolation of cyber-physical attacks on a more complex water distribution network.

### 6.4.1 Simulation parameters

Consider a more complex SCADA water network as shown in figure 6.25. The water network is comprised of two treatment plants $W_1$ and $W_2$, two reservoirs $R_1$ and $R_2$, a tank $T_3$, two pumps $P_1$ and $P_2$, two consumers $d_1$ and $d_2$, and several nodes and pipelines. Four pressure sensors $S_1$, $S_2$, $S_3$ and $S_4$ are equipped for measuring pressure heads $h_1$ at the reservoir, $h_2$ at the reservoir $R_2$, $h_3$ at the tank $T_3$ and $h_4$ at the node $N_4$, respectively.

The linearized model of the water network can be described in the discrete-time state space model (4.1), where $x_k = [h_1, h_2, h_3]^T \in \mathbb{R}^3$ is vector of system states; $u_k \in \mathbb{R}^2$ are the control signals sent to local controllers for regulating the flow rates $Q_{01}$ and $Q_{02}$ through the pump $P_1$ and $P_2$, respectively; $d_k \in \mathbb{R}^2$ represent the consumption by customers; $y_k \in \mathbb{R}^4$ are the measurements of four sensors $S_1$, $S_2$, $S_3$ and $S_4$; the process noises $w_k \sim \mathcal{N}(0, Q)$ and the sensor noises $v_k \sim \mathcal{N}(0, R)$; the matrices $A \in \mathbb{R}^{3\times3}$, $B \in \mathbb{R}^{3\times2}$, $F \in \mathbb{R}^{3\times2}$, $C \in \mathbb{R}^{4\times3}$, $D \in \mathbb{R}^{4\times2}$, $G \in \mathbb{R}^{4\times2}$, $Q \in \mathbb{R}^{3\times3}$, and $R \in \mathbb{R}^{4\times4}$ (corresponding to $n = 3$, $m = 2$, $p = 4$, $q = 2$).

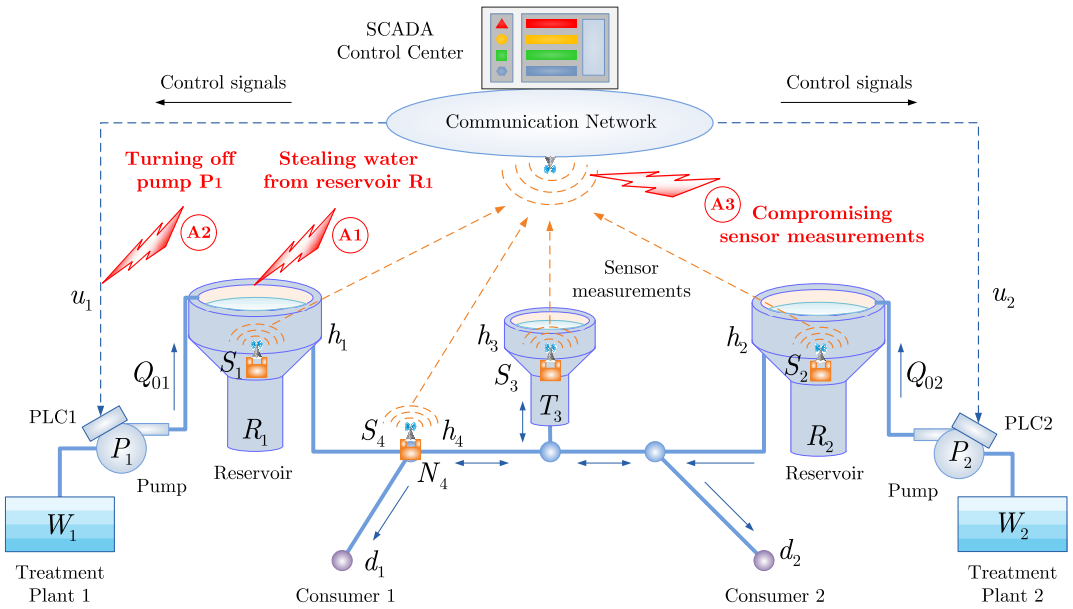
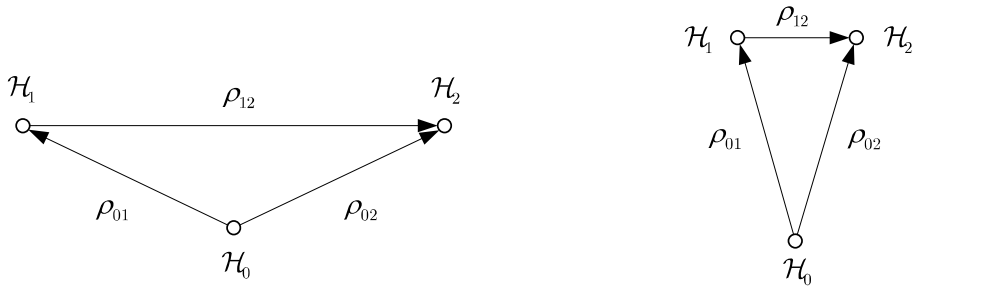



Figure 6.25 – A complex SCADA water distribution network.

Suppose that the attacker has the capabilities to withdraw water from the reservoirs $R_1$ and/or $R_2$, to modify the control signals of the pumps $P_1$ and/or $P_2$ and to compromise the measurements of sensors $S_3$ and $S_4$. It is assumed that trusted measurements are transmitted successfully to the detection-isolation schemes. There may be several attack scenarios that can be launched to the system. For the sake of simplicity, let us assume that the attacker can perform only one of two hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$. The problem is then to determine whether the system is under attack (i.e., between $\mathcal{H}_0$ and $\mathcal{H}_1$, $\mathcal{H}_2$) and then to identify the attack type (i.e., between $\mathcal{H}_1$ or $\mathcal{H}_2$).




(a) Scenario 1: $\rho_{12} > \max\{\rho_{01}, \rho_{02}\}$.

(b) Scenario 2: $\rho_{12} < \min\{\rho_{01}, \rho_{02}\}$.

Figure 6.26 – Two scenarios in the change detection-isolation problem.

For the joint detection-isolation problem, it is essential to consider two scenarios (see also figure 6.26):

- *Scenario 1*: The K-L distance between two alternative hypotheses is higher than K-L distance between an alternative hypothesis and the null hypothesis (i.e., sub-figure 6.26a). In other words, we have $\rho_{12} > \max\{\rho_{10}, \rho_{20}\}$. In this case, the alternative hypotheses $\mathcal{H}_1$

and $\mathcal{H}_2$ are "scattered", resulting in small probability of false isolation. In this scenario, we consider the following hypotheses:

  - Hypothesis $\mathcal{H}_0$: There is no attack on the system.
  - Hypothesis $\mathcal{H}_1$: The attacker performs a coordinated attack by stealing water from the reservoir $R_1$ with a constant flow rate $\delta Q_{01}$, turning off the pump $P_1$ and compromising the measurements of sensors $S_3$ and $S_4$ by the covert attack strategy.
  - Hypothesis $\mathcal{H}_2$: The attacker performs a coordinated attack by stealing water from the reservoir $R_2$ with a constant flow rate $\delta Q_{02}$, turning off the pump $P_2$ and compromising the measurements of sensors $S_3$ and $S_4$ by the covert attack strategy.

- *Scenario 2*: The K-L distance between two alternative hypotheses is smaller than K-L distance between an alternative hypothesis and the null hypothesis (i.e., sub-figure 6.26b). In other words, we have $\rho_{12} < \min\{\rho_{10}, \rho_{20}\}$. In this case, the alternative hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$ are quite "closed", resulting in high probability of false isolation. In this scenario, we consider the following hypotheses:

  - Hypothesis $\mathcal{H}_0$: There is no attack on the system.
  - Hypothesis $\mathcal{H}_1$: The attacker performs a coordinated attack by stealing water from the reservoir $R_1$ with a constant flow rate $\delta Q_{01}$, turning off the pump $P_1$ and compromising the measurements of sensors $S_3$ and $S_4$ by the covert attack strategy.
  - Hypothesis $\mathcal{H}_2$: The attacker aims at turning off the pumps $P_1$ and $P_2$ and compromising the measurements of sensors $S_3$ and $S_4$ by the covert attack strategy.

The system matrices are chosen as:

$$A = \begin{bmatrix} 0.9951 & 0.0009 & 0.0040 \\ 0.0012 & 0.9922 & 0.0066 \\ 0.0162 & 0.0198 & 0.9964 \end{bmatrix}, \quad B = \begin{bmatrix} 0.6250 & 0 \\ 0 & 0.8333 \\ 0 & 0 \end{bmatrix}, \quad F = \begin{bmatrix} -0.2293 & -0.0540 \\ -0.0959 & -0.3657 \\ -1.2950 & -1.1871 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.3669 & 0.1151 & 0.5180 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -29.3525 & -6.9065 \end{bmatrix},$$

and the initial state $\overline{x}_1 = [100, 80, 30]^T$. The noise covariance matrices are

$$Q = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The attack matrices $B_a$ and $D_a$ are

$$B_a = \begin{bmatrix} 0.6250 & 0 & 0.6250 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.8333 & 0 & 0.8333 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, D_a = \begin{bmatrix} 0 & 0 & 0 & 0 & \gamma_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \gamma_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma_4 \end{bmatrix},$$

where $\gamma_1 = \gamma_2 = 0$ (i.e., sensors $S_1$ and $S_2$ are secure) and $\gamma_3 = \gamma_4 = 1$ (i.e., sensors $S_3$ and $S_4$ are vulnerable). The simulation parameters: the attack duration $L = 8$ observations, the

attack instant $k_0 = 9$, the false alarm time window $m_\alpha = 24$ observations. The attack profiles $\theta_1, \theta_2, \cdots, \theta_L$ are composed of the state attack vector $a_k^x$ and the sensor attack vector $a_k^y$. The former depends on the attack scenario and the latter is calculated by the covert attack strategy. It is assumed that the stolen flow rates $\delta Q_{01} = \delta Q_{02} = 0.5\,\mathrm{m}^3/\mathrm{s}$.

- *Scenario 1*: The state attack vector $a_k^x$ is chosen by

$$a_k^x = \left\{ \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\mathcal{H}_0}, \underbrace{\begin{bmatrix} -0.5 \\ 0 \\ -1 \\ 0 \end{bmatrix}}_{\mathcal{H}_1}, \underbrace{\begin{bmatrix} 0 \\ -0.5 \\ 0 \\ -1 \end{bmatrix}}_{\mathcal{H}_2} \right\}, \quad \forall k_0 \leq k < k_0 + L,$$

  and the sensor attack vector $a_k^y$ is calculated by the covert attack strategy (1.9). The K-L distances of the residuals generated by the steady-state Kalman filter approach and the fixed-size parity space approach are computed as follows:

  - Kalman filter approach: $\rho_{01} = 13.9316$, $\rho_{02} = 17.6794$ and $\rho_{12} = 27.5977$.
  - Parity space approach: $\rho_{01} = 12.8467$, $\rho_{02} = 15.3698$ and $\rho_{12} = 24.0241$.

- *Scenario 2*: The state attack vector $a_k^x$ is designed as follows:

$$a_k^x = \left\{ \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\mathcal{H}_0}, \underbrace{\begin{bmatrix} -0.5 \\ 0 \\ -1 \\ 0 \end{bmatrix}}_{\mathcal{H}_1}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ -1 \\ -1 \end{bmatrix}}_{\mathcal{H}_2} \right\}, \quad \forall k_0 \leq k < k_0 + L,$$

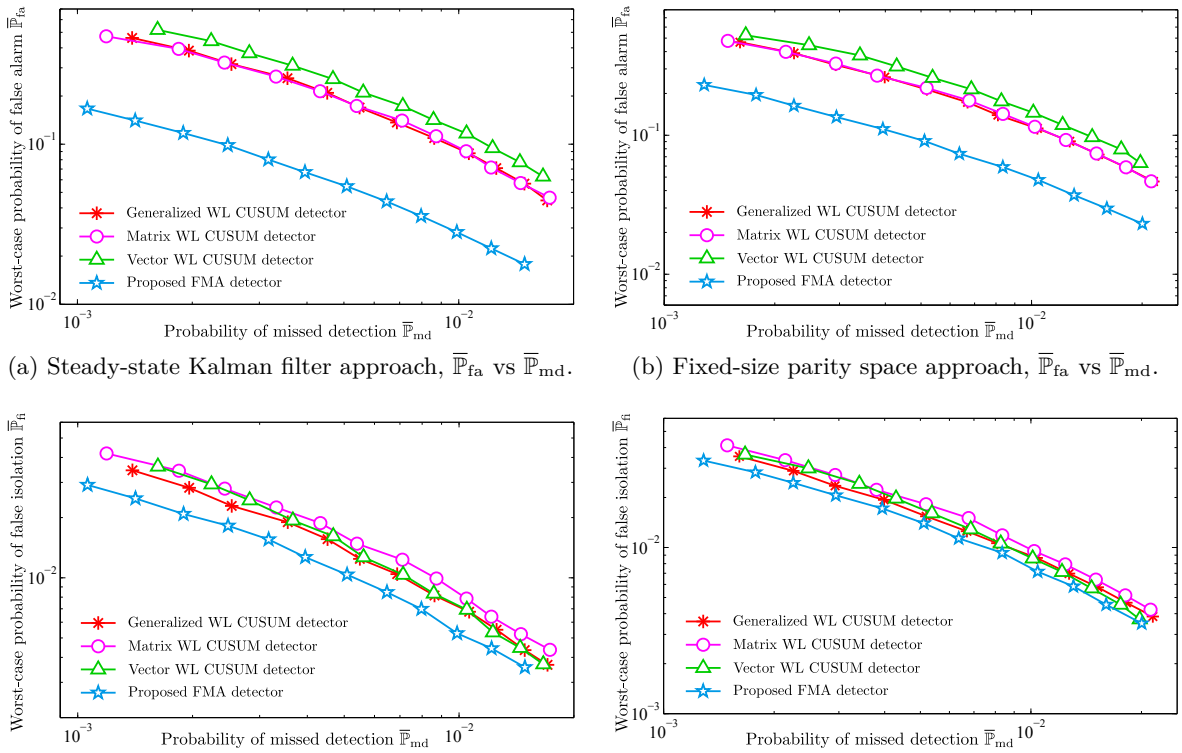
  and the sensor attack vector $a_k^y$ is calculated by the covert attack strategy (3.3). The K-L distances of the residuals generated by the steady-state Kalman filter approach and the fixed-size parity space approach are computed as follows:

  - Kalman filter approach: $\rho_{01} = 13.9316$, $\rho_{02} = 15.8330$ and $\rho_{12} = 8.5136$.
  - Parity space approach: $\rho_{01} = 12.8467$, $\rho_{02} = 14.4040$ and $\rho_{12} = 7.3268$.

### 6.4.2 Comparison between FMA test and WL CUSUM-based tests

This subsection is dedicated to investigating the statistical performance of several detection-isolation schemes, including the generalized WL CUSUM test, the matrix WL CUSUM test, the vector WL CUSUM test and the proposed FMA detection rule. The simulation results are obtained by $2.10^5$ Monte Carlo repetitions. Two aforementioned scenarios are considered: $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$ and $\rho_{12} \leq \min\{\rho_{01}, \rho_{02}\}$.

Figure 6.27 shows the comparison between the FMA detection-isolation rule and the classical WL CUSUM-based algorithms for the scenario 1 where $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$. The results are obtained by the $2.10^5$ Monte Carlo simulation. In sub-figure 6.27a and sub-figure 6.27b, the worst-case probability of false alarm $\overline{\mathbb{P}}_{\mathrm{fa}}$ is described as a function of the probability of missed detection $\overline{\mathbb{P}}_{\mathrm{md}}$ for the steady-state Kalman filter approach and the fixed-size parity approach, respectively. The worst-case probability of false isolation $\overline{\mathbb{P}}_{\mathrm{fi}}$ is drawn as a function of the




(a) Steady-state Kalman filter approach, $\overline{\mathbb{P}}_{\text{fa}}$ vs $\overline{\mathbb{P}}_{\text{md}}$.

(b) Fixed-size parity space approach, $\overline{\mathbb{P}}_{\text{fa}}$ vs $\overline{\mathbb{P}}_{\text{md}}$.

(c) Steady-state Kalman filter approach, $\overline{\mathbb{P}}_{\text{fi}}$ vs $\overline{\mathbb{P}}_{\text{md}}$.

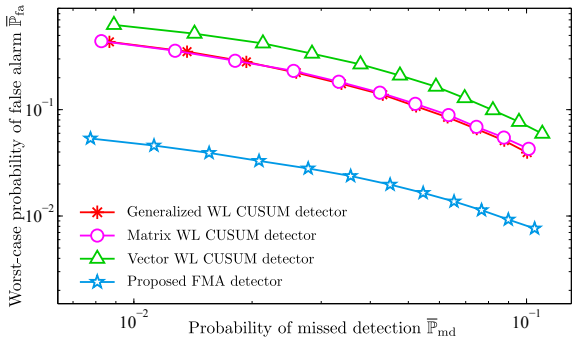
(d) Fixed-size parity space approach, $\overline{\mathbb{P}}_{\text{fi}}$ vs $\overline{\mathbb{P}}_{\text{md}}$.

Figure 6.27 – Comparison between the proposed FMA detection rule and the WL CUSUM-based schemes for the scenario 1, i.e., $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$. The worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ and the worst-case probability of false isolation $\overline{\mathbb{P}}_{\text{fi}}$ are described as a function of the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$. The change-point $k_0$ is chosen as $k_0 = L + 1 = 9$.

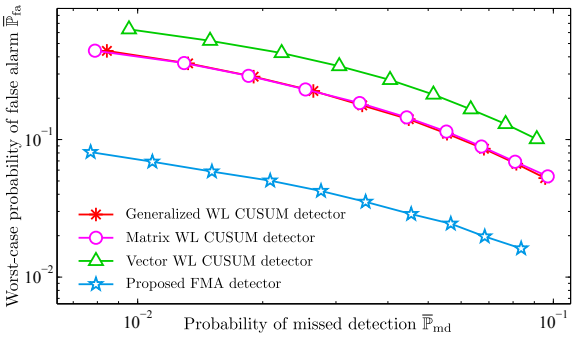
probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$, respectively, in sub-figure 6.27c for the steady-state Kalman filter approach and in sub-figure 6.27d for the fixed-size parity approach.

It can be noticed from those figures that for a given value on the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$, the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ and the worst-case probability of false isolation $\overline{\mathbb{P}}_{\text{fi}}$ of the FMA detection-isolation rule are smaller than those of the classical WL CUSUM-based procedures. In other words, the proposed FMA test performs better than classical tests. In addition, the worst-case probability of false isolation $\overline{\mathbb{P}}_{\text{fi}}$ is much smaller than the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ since $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$.

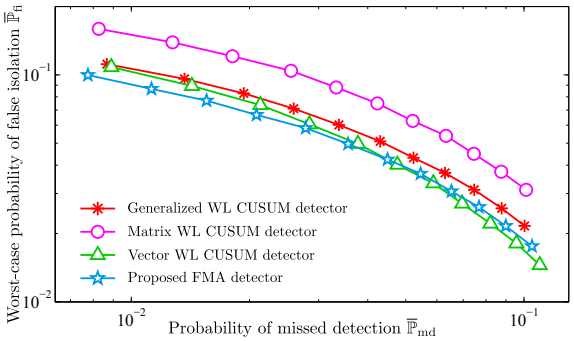
In figure 6.28, the FMA detection-isolation rule is compared with classical WL CUSUM-based algorithms for scenario 2 where $\rho_{12} \leq \min\{\rho_{01}, \rho_{02}\}$. In this case, the worst-case probability of false isolation $\overline{\mathbb{P}}_{\text{fi}}$ is much higher than the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$. In addition, the proposed FMA test performs better than the traditional tests, for both residual-generation methods.



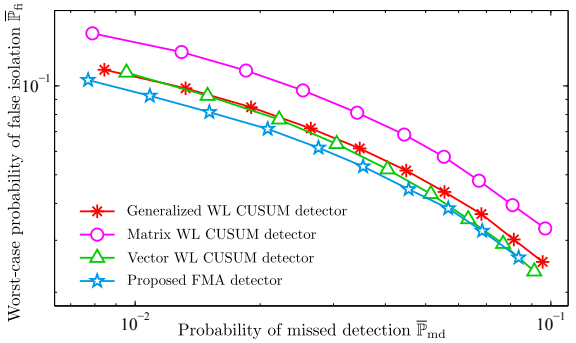

(a) Steady-state Kalman filter approach, $\overline{\mathbb{P}}_{\text{fa}}$ vs $\overline{\mathbb{P}}_{\text{md}}$.




(b) Fixed-size parity space approach, $\overline{\mathbb{P}}_{\text{fa}}$ vs $\overline{\mathbb{P}}_{\text{md}}$.




(c) Steady-state Kalman filter approach, $\overline{\mathbb{P}}_{\text{fi}}$ vs $\overline{\mathbb{P}}_{\text{md}}$.



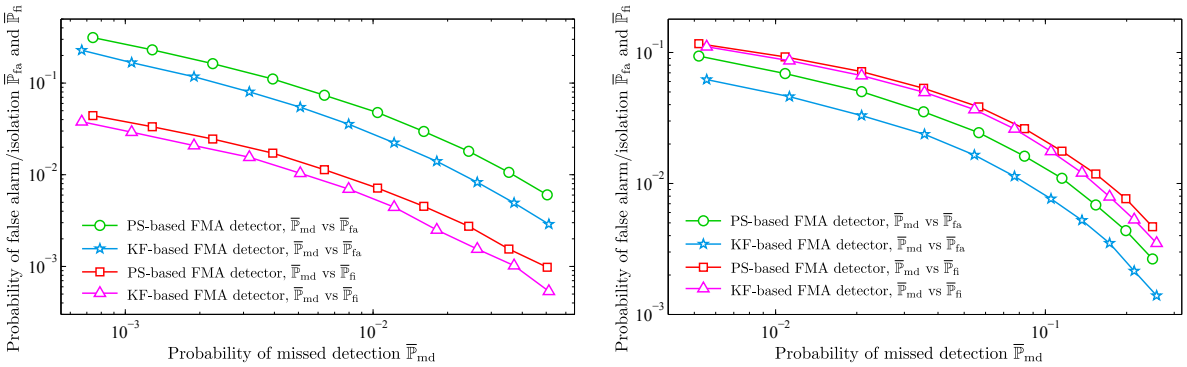

(d) Fixed-size parity space approach, $\overline{\mathbb{P}}_{\text{fi}}$ vs $\overline{\mathbb{P}}_{\text{md}}$.

Figure 6.28 – Comparison between the proposed FMA detection rule and the WL CUSUM-based schemes for the scenario 2, i.e., $\rho_{12} \leq \min\{\rho_{01}, \rho_{02}\}$, by $2.10^5$ Monte Carlo simulation. The worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ and the worst-case probability of false isolation $\overline{\mathbb{P}}_{\text{fi}}$ are described as a function of the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$. The change-point $k_0$ is chosen as $k_0 = L + 1 = 9$.

### 6.4.3 Comparison between steady-state Kalman filter and fixed-size parity space

The Monte Carlo simulation technique is utilized for comparing two residual-generation methods, i.e., the steady-state Kalman filter approach with the fixed-size parity space approach. The simulation results are obtained by $2.10^5$ Monte Carlo repetitions and the change-point $k_0 = L + 1 = 9$.

The simulation results are described in figure 6.29 for both scenarios (i.e., $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$ and $\rho_{12} \leq \min\{\rho_{01}, \rho_{02}\}$). In sub-figure 6.29a, the worst-case probability of false isolation $\overline{\mathbb{P}}_{\text{fi}}$ is smaller than the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ since the K-L distance between two alternative hypotheses is higher than both K-L distances between the null hypothesis and either alternative hypothesis (i.e., $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$). In contrast, sub-figure 6.29b shows that the worst-case probability of false isolation $\overline{\mathbb{P}}_{\text{fi}}$ is higher than the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ since $\rho_{12} \leq \min\{\rho_{01}, \rho_{02}\}$. It follows from both sub-figures that the FMA detector based on the steady-state Kalman filter approach performs much better than the FMA detector based on the fixed-size parity space approach, for both scenarios. This phenomenon can be explained from the fact that the Kalman filter approach generates residuals with higher K-L distances than the parity space approach does (see also subsection 6.4.1).




(a) Scenario 1: $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$.

(b) Scenario 2: $\rho_{12} \leq \min\{\rho_{01}, \rho_{02}\}$.

Figure 6.29 – Comparison between the steady-state Kalman filter approach and the fixed-size parity space approach when using in the proposed FMA detector. The worst-case probability of false alarm $\overline{\mathbb{P}}_{\mathrm{fa}}$ and the worst-case probability of false isolation $\overline{\mathbb{P}}_{\mathrm{fi}}$ are drawn as a function of the probability of missed detection $\overline{\mathbb{P}}_{\mathrm{md}}$. The change-point is chosen as $k_0 = L + 1 = 9$. Both scenarios are considered: $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$ and $\rho_{12} \leq \min\{\rho_{01}, \rho_{02}\}$.

### 6.4.4 Evaluation of upper bounds for error probabilities of FMA detection rule

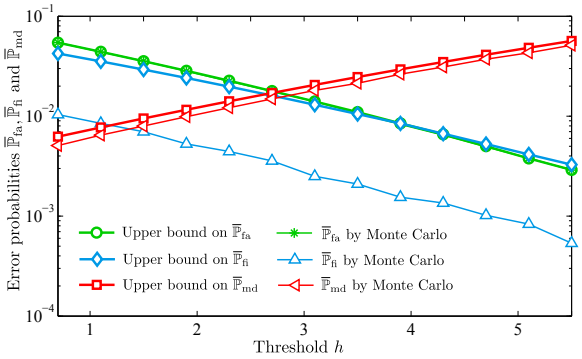
This subsection is dedicated to evaluating the sharpness of the proposed upper bounds for the error probabilities of the FMA detection rule, including the upper bound for the worst-case probability of false alarm, the upper bound for the worst-case probability of false isolation and the upper bound for the probability of missed detection.

The comparison between the proposed upper bounds and the $2.10^5$ Monte Carlo simulation for the error probabilities is shown in figure 6.30. It can be seen that the upper bound for the worst-case probability of false alarm $\overline{\mathbb{P}}_{\mathrm{fa}}$ is extremely tight, for both residual-generation methods, especially for the case of $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$. In contrast, the upper bound for the worst-case probability of false isolation $\overline{\mathbb{P}}_{\mathrm{fi}}$ is not very sharp at all. Finally, the upper bound for the probability of missed detection $\overline{\mathbb{P}}_{\mathrm{md}}$ seems to be acceptable.

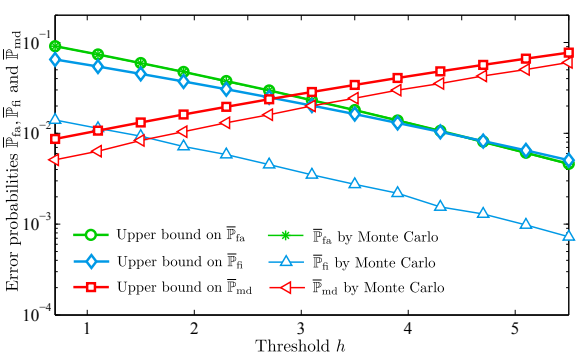
## 6.5 Conclusion

Several sub-optimal algorithms have been proposed in chapter 3 and chapter 4 for detecting and identifying transient changes in stochastic-dynamical systems. The models of two SCADA systems, including the simple SCADA gas pipeline and the simple SCADA water distribution network, have been developed in chapter 5. In this chapter, we have applied the theoretical results obtained in chapter 3 and chapter 4 to the detection and isolation of several types of cyber-physical attacks on both the SCADA gas pipeline and the SCADA water network, whose models have been developed in chapter 5.

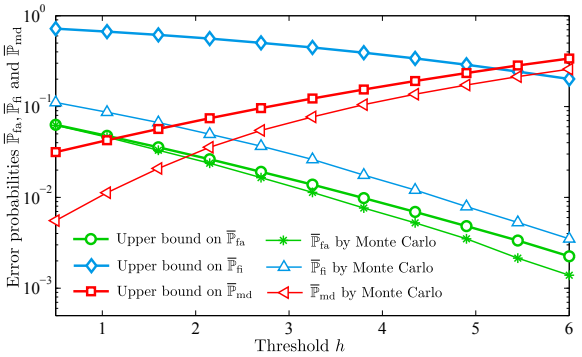
In the first place, we have studied the reaction of the SCADA gas pipeline under several types of cyber-physical attacks (i.e., DoS attacks, simple integrity attacks and stealthy integrity attacks) on the command signals, the control signals and the feedback signals, respectively. The simulation results show that each attack scenario leads to a specific attack signature (i.e., an attack



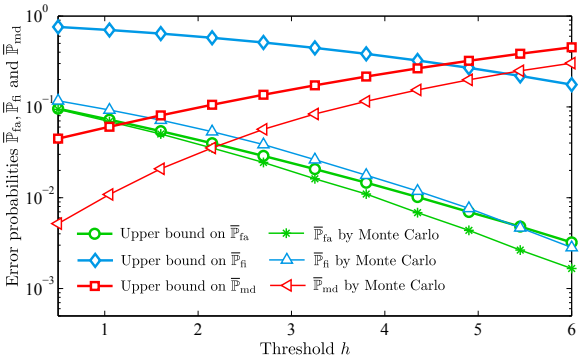

(a) Steady-state Kalman filter approach, scenario 1: $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$.




(b) Fixed-size parity space approach, scenario 1: $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$.




(c) Steady-state Kalman filter approach, scenario 2: $\rho_{12} \leq \min\{\rho_{01}, \rho_{02}\}$.




(d) Fixed-size parity space approach, scenario 2: $\rho_{12} \leq \min\{\rho_{01}, \rho_{02}\}$.

Figure 6.30 – Evaluation of the sharpness of the upper bounds for the error probabilities of the FMA detection rule. The error probabilities $\overline{\mathbb{P}}_{\text{fa}}$, $\overline{\mathbb{P}}_{\text{fi}}$ and $\mathbb{P}_{\text{md}}$ are drawn as a function of the threshold $h$. The change-point is chosen as $k_0 = L + 1 = 9$. Both steady-state Kalman filter and fixed-size parity space approaches associated with two scenarios $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$ and $\rho_{12} \leq \min\{\rho_{01}, \rho_{02}\}$ are considered.

profile) which is essential in designing detection-isolation schemes. In certain circumstances, if the information about the attack (i.e., the attack scenario and attack parameters) is known *a priori*, the attack profile may be available. This essential information helps in improving the statistical performance of the detection-isolation schemes.

Secondly, the statistical performance of the FMA test proposed in chapter 3 has been investigated thoughtfully. The comparison between the proposed FMA detection rule with traditional detection algorithms (i.e., $\chi^2$ test, CUSUM test and WL CUSUM test) has been performed by both Monte Carlo simulation and numerical method. It has been shown that the proposed FMA test performs much better than traditional test w.r.t. the transient change detection criterion. Moreover, the comparison between two residual-generation methods, i.e., the steady-state Kalman filter and the fixed-size parity space, has been also carried out by both Monte Carlo simulation and numerical method. The simulation results have pointed out that Kalman filter-based FMA test outperforms the parity space-based FMA test when the noise covariance matrices are exactly known. On the other hand, when the process noise covariance matrix is unknown, the Kalman filter-based FMA test may perform worse than the parity-space based FMA test.

Furthermore, the robustness of the FMA test w.r.t. several operational parameters has been tested with both Monte Carlo simulation and numerical method. The operational parameters include the attack duration, the attack profiles, the process and sensor noise covariance matrices. It can be noticed that the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ is insensitive to the attack duration and the attack profiles. The probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ depends heavily on these parameters. The probability of missed detection increases significantly when the true attack duration is smaller than the putative value. When the true attack duration is higher than the putative value, however, the probability of missed detection remains unchanged since any detection with the detection delay greater than $L$ is considered as missed. Also, the probability of missed detection $\overline{\mathbb{P}}_{\text{md}}$ is inversely proportional to the true attack profiles since higher attack profiles lead to higher K-L distances, thus reducing the probability of missed detection. It is intuitively obvious that both probability of false alarm and probability of missed detection are inversely proportional to the true values of noise covariances. In other words, the augmentation in the true values of process and sensor noise covariances leads to higher error probabilities.

In practice, the post-change parameters are rarely known. For this reason, we have been considered a more practical scenario where the post-change profiles are partially known. More precisely, the shape of the attack signature is assumed to be known but the magnitude (i.e., or the power) of the attack is unknown. Two standard approaches, the generalized likelihood ratio approach and the weighted likelihood ratio approach, have been considered. It has been shown by Monte Carlo simulation that the FMA GLR test (resp. the FMA WLR) performs much better than the window limited (WL) GLR test (resp. WL WLR test) w.r.t. the transient change detection criterion. The simulation results have also pointed out that the FMA WLR test performs better than the FMA GLR test under the same condition. This phenomenon may be explained by the fact that the WLR approach utilizes the *a priori* information about the power of the attack.

Finally, the comparison between several detection-isolation schemes (i.e., FMA test, generalized WL CUSUM test, matrix WL CUSUM test and vector WL CUSUM test) has been performed by Monte Carlo simulation. It has been shown that the FMA test performs better than the others w.r.t. the transient change detection-isolation criterion. More precisely, for a given value on the probability of missed detection, the probability of false isolation and the probability of false alarm of the proposed FMA test are smaller than those of traditional tests. In addition, the sharpness of the proposed upper bounds on the worst-case probability of false alarm, false isolation and missed detection has also been investigated. It can be seen that the bounds for the false alarm and missed detection rates are quite closed while the bound for the false isolation rate needs to be improved.

# General Conclusion

## Conclusions

This PhD thesis has addressed the problem of detecting and isolating cyber-physical attacks on Supervisory Control And Data Acquisition (SCADA) systems by statistical methods. The state-of-the-art of the problem treated in this thesis has been introduced in chapter 1 and chapter 2. The security of SCADA systems against cyber-physical attacks has been examined in chapter 1. In this chapter, we investigated a large number of cyber incidents involving safety-critical infrastructures as well as the vulnerabilities of SCADA systems. It has been shown that these susceptible points can be exploited by adversaries for performing malicious attacks on SCADA systems. The design of several attack strategies, including DoS attack, simple integrity attack and stealthy integrity attack, has been also presented. Methods proposed for improving the security of SCADA systems could be broadly classified into three classes: information security approach, secure control theory approach and fault detection and isolation (FDI) approach. The information security approach is concerned with protection methods such as authentication, access control or data integrity. The secure control approach, on the other hand, focuses mainly on investigating the vulnerabilities of networked control systems, designing different stealthy/deception attack strategies and proposing countermeasures against these malicious attacks. In contrast, the FDI approach deals with the detection and isolation of abnormal behaviors in stochastic-dynamical systems, thus being suitable to the on-line monitoring of large-scale industrial control systems against cyber-physical attacks. Generally, the statistical FDI problem has been solved by the classical two-step approach: residual generation and residual evaluation. The fault diagnosis community has concentrated mainly on the generation of robust residuals regardless of unknown disturbances and modeling errors. However, process noises and sensor noises are inevitable in almost all technological processes and measurement systems. Hence, the decision-making must take into consideration the negative impact of such random noises. Fortunately, the statistical decision theory, which has been summarized in chapter 2, is equipped with methodologies for dealing with random noises in stochastic systems. The statistical decision theory can be broadly classified into four categories: the non-sequential hypothesis testing, sequential hypothesis testing, sequential detection and isolation of abrupt changes and sequential detection and isolation of transient changes. The sequential detection and isolation of transient changes has been shown to be the most suitable approach to the on-line monitoring of SCADA systems against cyber-physical attacks.

The attack detection and isolation problem has been formulated as the sequential detection and isolation of transient signals on stochastic-dynamical systems. The SCADA systems are described as the discrete-time state space model driven by random noises with unknown system states. The cyber-physical attacks are modeled as additive signals of short duration on both state

evolution and sensor measurement equations. The criteria of optimality for the classical quickest change detection-isolation problem appear inadequate for the security of SCADA systems against cyber-physical attacks. For the transient change detection-isolation problem, the optimality criterion should be in favor of minimizing the worst-case probability of missed detection subject to acceptable levels on the rates of false alarm and false isolation. Sub-optimal detection and isolation algorithms with respect to the transient change detection and isolation criteria have been designed in chapter 3 and chapter 4, respectively. The main contributions of the thesis are as follows:

- *For the detection problem.* Firstly, the detection of cyber-physical attacks has been formulated as the sequential detection of transient signals in stochastic-dynamical systems. The transient change detection criterion, minimizing the worst-case probability of missed detection subject to an acceptable level on the worst-case probability of false alarm within any time window of predefined length, has been utilized throughout this thesis. Secondly, the unified statistical model of the residuals generated from both steady-state Kalman filter and the fixed-size parity space has been developed. This unified statistical model has been utilized for designing the Variable Threshold Window Limited (VTWL) CUSUM algorithm. Thirdly, the optimal choice of thresholds of the VTWL CUSUM algorithm with respect to (w.r.t.) the transient change detection criterion has been solved and it has been shown that the optimized VTWL CUSUM algorithm is equivalent to the simple Finite Moving Average (FMA) detection rule. Fourthly, a numerical method, which is much more efficient than the classical Monte Carlo simulation, has been proposed for estimating the probability of false alarm and the probability of missed detection. Fifthly, the proposed numerical method has been exploited for investigating the robustness of the FMA test w.r.t. several operational parameters, including the attack duration, the attack profiles, the covariance matrices of process noises and sensor noises. Finally, we have considered also a more practical scenario where the attack profiles are partially known, i.e., the "*shape*" of change is known but the "*magnitude*" of the change is unknown. Both the generalized likelihood ratio (GLR) approach and the weighted likelihood ratio (WLR) approach have been considered. It has been shown that the optimal choice of thresholds in such cases turned out to be also the FMA version.

- *For the isolation problem.* The isolation problem is much more difficult than the detection counterpart. Few theoretical results have been obtained. Firstly, the unified statistical model of residuals generated by both aforementioned residual-generation methods has been adapted to the detection and isolation of transient changes in discrete-time state space model. There have been multiple change types (i.e., multiple transient change hypotheses). Secondly, a novel criterion of optimality for the transient change detection and isolation has been introduced. The criterion involves the minimization of the worst-case probability of missed detection subject to acceptable levels on the worst-case probability of false alarm within any time window of given length and on the worst-case probability of false isolation during the transient change window regardless of the change-point. Finally, several quickest change detection-isolation algorithms have been considered for detecting the transient changes, including the generalized WL CUSUM test, the matrix WL CUSUM test and the vector CUSUM test. The FMA version for the detection-isolation problem has been proposed. The upper bounds on the error probabilities of the FMA test have been also obtained.

In order to demonstrate the statistical performance of the proposed algorithms, we have de-

veloped in chapter 5 two simulation models, i.e., a simple SCADA gas pipeline and a simple SCADA water distribution network. The physical layer of almost SCADA systems, including the gas pipeline and the water network considered in this manuscript, can be described in the discrete-time state space model by linearizing the partial differential equations around the operating point. The cyber-physical attacks on boh physical layer (i.e., attacks on physical processes directly) and cyber layer (i.e., attacks on command signals, control signals, sensor measurements) have been modeled as additive signals of short duration on both state evolution and sensor measurement equations.

The theoretical results obtained in chapter 3 and chapter 4 have been applied to the detection and isolation of cyber-physical attacks on the SCADA gas pipeline and the SCADA water in chapter 5. The numerical examples have been shown in chapter 6. The following conclusions can be drawn from the simulation results.

- Firstly, the negative impact of cyber-physical attacks has been investigated by performing several scenarios on the SCADA gas pipeline. Simple attack strategies such as DoS attacks and simple integrity attacks (min attack, max attack, scaling attack or additive attack) can be detected easily by classical anomaly detectors. On the other hand, stealthy integrity attacks such as replay attack or covert attack are much more difficult to detect. For this reason, it is required to implement some *a priori* countermeasures for rendering these deception attacks detectable before applying any detection schemes. Particular methods for revealing several types of undetectable attacks have been considered in literature. This manuscript have not focused on revealing stealthy attacks but on proposing algorithms for detecting and isolating any detectable and identifiable attacks. For this reason, we have proposed a simple sensor protection scheme based on hardware redundancy for rendering stealthy attacks detectable and identifiable.

- Secondly, the statistical performance of the several detection algorithms has been investigated and compared by performing the covert attack on the simple SCADA water network. It has been noticed that the proposed FMA detection rule performs much better than classical algorithms, including the non-parametric $\chi^2$ detector, CUSUM detector, WL CUSUM detector, for both residual-generation methods. The simulation results based on both numerical method and Monte Carlo method have shown that the steady-state Kalman filter approach offers better statistical performance than the fixed-size parity space approach when system parameters are completely known. However, the sensitivity analysis of the FMA test has also proved that the former is much more sensitive than the process noises than the latter. In such scenarios that the true value of process noise covariance is larger than its putative value, the Kalman filter-based FMA test may perform worse than the parity space-based FMA test. Finally, the simulation results about the partially known transient parameters have indicated that the FMA version of both GLR and WLR approaches offers better statistical performance than the window limited counterpart.

- Thirdly, preliminary results on the isolation problem have been demonstrated by performing different attack scenarios on a more complex water network. Simulation results have shown that the proposed FMA test, in general, performs better than classical detection-isolation algorithm, including the generalized WL CUSUM test, the matrix WL CUSUM test and the vector WL CUSUM test w.r.t. the transient change detection-isolation criterion. The proposed upper bounds for the error probabilities of the FMA detection-isolation rule have been also compared with the true error probabilities by Monte Carlo simulation.

It has been shown that the upper bound for the worst-case probability of false alarm is extremely tight, the upper bound for the worst-case probability of missed detection is acceptable but the upper bound for the worst-case probability of false isolation is not really sharp.

## Perspectives

Before finishing this manuscript, we would like to suggest several points for future research.

### For the sequential detection of transient signals

The following points should be taken into consideration for sequential detection of transient signals in stochastic systems in general and in stochastic-dynamical systems in particular:

- *Design of optimal or asymptotically optimal detection rules.* Only sub-optimal algorithms have been designed in this manuscript. It is proposed to minimize the upper bound on the worst-case probability of missed detection in the class of all repeated one-sided truncated sequential tests (i.e., the class of VTWL CUSUM tests) satisfying an acceptable level on the worst-case probability of false alarm within any time window of predefined length. Future work should concentrate on the design of asymptotically optimal (i.e., when the probability of false alarm tends to zero) or exactly optimal tests w.r.t. the transient change detection criterion (3.6)–(3.7). As has been suggested in [67], the preliminary task should focus on calculating the lower bound for the worst-case probability of missed detection in the class $C_\alpha$ defined in (3.7). This lower bound is then compared to the probability of missed detection of the FMA test in order to verify whether the FMA test is (asymptotically) optimal or not. This comparison may suggest some ideas about how to design the (asymptotically) optimal test.

- *Detection of transient signals with variable profiles.* In this manuscript, we consider only the case of fixed transient change profiles $\theta_1, \theta_2, \cdots, \theta_L$, i.e., they are independent from the change-point $k_0$. For some applications, however, the transient change profiles may be varying according to the change-point $k_0$. The future work should also consider this aspect. In our opinion, the detection of variable transient change profiles can be generalized on the basis of this work without much difficulty.

- *Detection of transient signals with completely unknown parameters.* In this manuscript, we consider only two scenarios where the transient profiles are exactly known and the transient profiles are partially known (i.e., the shape of the changes is known but the magnitude of the changes is unknown). The completely unknown transient change parameters, including the change-point, the transient length and the transient change profiles, should be considered in the future.

### For the sequential isolation of transient signals

The following problems remain unsolved when dealing with the joint detection-isolation of transient changes:

- *Calculation of upper bounds for the error probabilities.* In this manuscript, we have tried to propose the upper bounds for the error probabilities. The upper bound for the worst-case probability of missed detection is given in an analytical formula. On the other hands, the upper bounds for the worst-case probability of false alarm and false isolation have been calculated by the numerical method. In addition, the upper bound for the worst-case probability of false isolation is not quite sharp. For these reasons, it is suggested to find "better" and "analytical" bounds for the error probabilities.
- *Distinguish between the false alarm and false isolation rates.* For some situations, it is interesting to differentiate between the false alarm rate and false isolation rate by utilizing different thresholds. This problem has been considered in [129, 132, 175] for the sequential quickest change detection-isolation problem. In the literature, the complete decoupling between the false alarm rate and false isolation rate has not been achieved. Some ideas have been suggested in [175] where the authors proposed to utilize the two step approach: (1) detection and (2) isolation.
- *Design of sub-optimal or asymptotically optimal tests.* Asymptotically optimal detection-isolation rules have been proposed in the quickest change detection-isolation framework. Up to our best knowledge, the problem of jointly detecting and isolating transient signals has not been considered. This problem would be an interesting direction for future research.

**For the security of SCADA systems against cyber-physical attacks**

The security of SCADA systems against cyber-physical attacks can be improved by investigating the following points:

- *Surveillance of SCADA systems.* This manuscript has focused mainly on the detection and isolation of cyber-physical attacks on physical processes, control signals and sensor measurements. Future work should focus on cyber attacks on the supervisory control layer, on the command signals or even on the control algorithms. For example, the on-line monitoring of network traffic [182] may be useful in detecting DoS attacks on computer networks.
- *Revelation of stealthy attacks.* This manuscript has suggested a simple method for revealing several types of stealthy attacks, including the replay attack, the zero-dynamics attack or the covert attack. The proposed method is based on hardware redundancy approach, consisting in protecting some "important" sensors or equipping more secure sensors in such a way that essential information about the attacks is transmitted into monitoring schemes. It is interesting to consider, in the near future, the problem of how many and which sensors should protected and/or equipped. The trade-off between the performance of the algorithms and the equipment costs should be also treated. Moreover, other methods for revealing particular types of stealthy attacks are also welcome.

**For the modeling problem**

In this manuscript, we have modeled SCADA systems as the discrete-time state space model driven by Gaussian noises by linearizing the partial differential equations around the operating points. For practical purpose, following points should be considered:

- *Discrete-time time-variant state space model.* The discrete-time time-variant state space model should be considered in place of the time-invariant counterpart treated in this thesis. The discrete-time Kalman filter approach may be used for generating the sequence of residuals. In contrast, it is questionable whether the parity space approach is applicable or not.
- *Modeling errors.* It is of practical interest to take into consideration the modeling errors in diagnosis schemes. Various techniques for eliminating the modeling errors have been proposed in the fault diagnosis community. For this reason, the integration of advanced residual generation techniques in the FDI community into the statistical framework should be a good research direction.
- *Non-linear systems.* The FDI techniques for non-linear systems should be also considered in future for the detection and isolation of cyber-physical attacks on SCADA systems.

**For long-term perspectives**

In the far future, the following approaches may be useful:

- *On-line monitoring of complex systems.* Generally, practical SCADA systems contain up to thousands or even millions of state variables. The surveillance of such large-scale industrial control systems encounters many problems, especially for the centralized data processing algorithms. Therefore, the decentralized or distributed mechanisms should be considered in future work, see for example in [141, 178, 179].
- *Non-parametric approach.* The parametric model of SCADA systems may be difficult to achieve in many practical situations. The imprecision of system models may lead to an extreme degradation of the statistical performance of detection-isolation schemes. The non-parametric approach, on the other hand, does not require the system and attack models. The machine learning and kernel methods are some examples of the non-parametric approach. These techniques are based on the analysis of the relationship of observed data under the normal operation of the systems. The detection problem can be solved by applying mono-class classification techniques while multi-class classification methods can be employed for the isolation problem.
- *Semi-parametric approach.* The parametric approach utilized in this manuscript depends heavily on the model of SCADA systems and cyber-physical attacks. Sometimes, these models are difficult to obtain. In addition, mathematical models can not describe all real-world phenomena. The non-parametric approach, on the other hand, does not understand the operation of SCADA systems, i.e., the interaction between the physical processes and the cyber layer. The semi-parametric approach is, therefore, the natural integration of the parametric approach and non-parametric approach. Generally, the semi-parametric model consists of two parts: parametric one and non-parametric one. The parametric statistic contains such phenomena that can be described mathematically while the non-parametric statistic consists of the information about non-modeled phenomena.

# Appendix A

# Proofs of Lemmas, Theorems and Propositions

**Contents**

## A.1 Discrete-time Kalman filter

In this section, we introduce some precious properties of the discrete-time Kalman filter. Two essential results include the calculation of innovation signatures and the computation of the covariance matrix between two innovations when noise covariances are not exactly known.

### A.1.1 System model and assumptions

Suppose that the system operates from time instant $k \geq 1$ with initial state $x_1 \sim \mathcal{N}\left(\overline{x}_1, P_{1|0}\right)$, where the mean value $\overline{x}_1$ and the covariance matrix $P_{1|0}$ are assumed to be known. The discrete-time state space model (3.4) can be rewritten as follows:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + B_a a_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + D_a a_k + v_k \end{cases}, \quad x_1 \sim \mathcal{N}\left(\overline{x}_1, P_{1|0}\right), \tag{A.1}$$

where $x_k \in \mathbb{R}^n$ is the vector of system states, $u_k \in \mathbb{R}^m$ is the vector of control signals, $d_k \in \mathbb{R}^q$ is the vector of disturbances, $y_k \in \mathbb{R}^p$ is the vector of sensor measurements, $a_k \in \mathbb{R}^s$ is the vector of attack signals, $w_k \in \mathbb{R}^n$ is the vector of process noises and $v_k \in \mathbb{R}^p$ is the vector of sensor noises; the matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $F \in \mathbb{R}^{n\times q}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, $G \in \mathbb{R}^{p\times q}$, $B_a \in \mathbb{R}^{n\times s}$ and $D_a \in \mathbb{R}^{p\times s}$ are assumed to be exactly known. The control signals $u_k$ and the disturbances $d_k$ are also assumed to be known. The process noises $w_k$ and sensor noises $v_k$ are assumed to be independent identically distributed (i.i.d.) zero-mean Gaussian random vectors, i.e., $w_k \sim \mathcal{N}(0, Q)$ and $v_k \sim \mathcal{N}(0, R)$. It is well known that under aforementioned conditions, the Kalman filter is an optimal estimator in the sense that it minimizes the mean-square of the *a posteriori* state estimation error.

### A.1.2 Discrete-time Kalman filter implementation

Let $\mathbb{E}_0[(\cdot)]$ denote the expectation of a random vector $(\cdot)$ under normal operation (i.e., the attack vector $a_k = 0$). The discrete-time Kalman filter designed for the discrete-time state space model (A.1) under normal operation is implemented by following steps:

1. Initialization step:

$$\hat{x}_{1|0} = \mathbb{E}_0[x_1] = \overline{x}_1, \tag{A.2}$$

$$P_{1|0} = \operatorname{cov}\left(x_1 - \hat{x}_{1|0}\right) = \mathbb{E}_0\left[\left(x_1 - \hat{x}_{1|0}\right)\left(x_1 - \hat{x}_{1|0}\right)^T\right], \tag{A.3}$$

   where $\hat{x}_{1|0}$ is the initial state estimate and $P_{1|0}$ is the initial covariance of state estimation error $x_1 - \hat{x}_{1|0}$.

2. Measurement update step:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k\left(y_k - C\hat{x}_{k|k-1} - Du_k - Gd_k\right), \tag{A.4}$$

$$P_{k|k} = P_{k|k-1} - K_k C P_{k|k-1}, \tag{A.5}$$

   where the optimal Kalman gain $K_k$ is calculated by

$$K_k = P_{k|k-1}C^T\left(CP_{k|k-1}C^T + R\right)^{-1}. \tag{A.6}$$

3. Time update step:

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + Bu_k + Fd_k, \tag{A.7}$$
$$P_{k+1|k} = AP_{k|k}A^T + Q. \tag{A.8}$$

The discrete-time Kalman filter can be also described shortly as follows:

$$\begin{cases} \hat{x}_{k+1|k} & = A\hat{x}_{k|k-1} + Bu_k + Fd_k + AK_k\left(y_k - \hat{y}_{k|k-1}\right) \\ \hat{y}_{k|k-1} & = C\hat{x}_{k|k-1} + Du_k + Gd_k \end{cases}; \quad \hat{x}_{1|0} = \overline{x}_1, \tag{A.9}$$

where the optimal Kalman gain is calculated as

$$K_k = P_{k|k-1}C^T\left(CP_{k|k-1}C^T + R\right)^{-1}, \tag{A.10}$$
$$P_{k+1|k} = AP_{k|k-1}A^T - AP_{k|k-1}C^T\left(CP_{k|k-1}C^T + R\right)^{-1}CP_{k|k-1}A^T + Q, \tag{A.11}$$

with the initial covariance matrix $P_{1|0}$.

### A.1.3 Calculation of innovation signatures

Let $e_k = x_k - \hat{x}_{k|k-1}$ be the state estimation error and $r_k = y_k - \hat{y}_{k|k-1}$ be the measurement estimation error (i.e., the residuals or the innovations). The measurement estimation error is calculated as

$$\begin{aligned} r_k &= y_k - \hat{y}_{k|k-1} \\ &= \underbrace{(Cx_k + Du_k + Gd_k + D_a a_k + v_k)}_{y_k} - \underbrace{\left(C\hat{x}_{k|k-1} + Du_k + Gd_k\right)}_{\hat{y}_{k|k-1}} \\ &= C\underbrace{\left(x_k - \hat{x}_{k|k-1}\right)}_{e_k} + D_a a_k + v_k \\ &= Ce_k + D_a a_k + v_k. \end{aligned}$$

Similarly, the state estimation error is described as

$$\begin{aligned} e_{k+1} &= x_{k+1} - \hat{x}_{k+1|k} \\ &= \underbrace{(Ax_k + Bu_k + Fd_k + B_a a_k + w_k)}_{x_{k+1}} - \underbrace{\left(A\hat{x}_{k|k-1} + Bu_k + Fd_k + AK_k r_k\right)}_{\hat{x}_{k+1|k}} \\ &= A\underbrace{\left(x_k - \hat{x}_{k|k-1}\right)}_{e_k} + B_a a_k + w_k - AK_k r_k \\ &= Ae_k - AK_k\left(Ce_k + D_a a_k + v_k\right) + B_a a_k + w_k \\ &= \left(A - AK_k C\right)e_k + \left(B_a - AK_k D_a\right)a_k + w_k - AK_k v_k. \end{aligned}$$

Finally, the innovation model is described as

$$\begin{cases} e_{k+1} & = \left(A - AK_k C\right)e_k + \left(B_a - AK_k D_a\right)a_k + w_k - AK_k v_k \\ r_k & = Ce_k + D_a a_k + v_k \end{cases}; \quad e_1 = 0. \tag{A.12}$$

In the following, we calculate the innovation signatures (i.e., the profiles of the innovations). The innovation signatures $\psi_1, \psi_2, \cdots, \psi_L$ of the transient changes are defined as the expectation of the innovation vectors during the change period $\tau_a = [k_0, k_0 + L - 1]$. Without loss of generality, let us suppose that the change-point $k_0 = 1$ since the innovation signatures are independent from position of the attack duration. It follows from (3.5) that the attack profiles $\{\theta_k\}_{1 \leq k \leq L}$ are equal to the attack vectors $\{a_k\}_{1 \leq k \leq L}$. The innovation signatures $\psi_1, \psi_2, \cdots \psi_L$ can be calculated from the attack profiles $\theta_1, \theta_2, \cdots, \theta_L$ as follows:

$$\psi_k = C\epsilon_k + D_a\theta_k, \tag{A.13}$$

where the dynamic profiles $\epsilon_k$ are computed by

$$\epsilon_{k+1} = (A - AK_kC)\,\epsilon_k + (B_a - AK_kD_a)\,\theta_k; \quad \epsilon_1 = 0. \tag{A.14}$$

### A.1.4 Calculation of innovation covariance matrices

It has been shown in literature [10, 103, 107, 116] that under perfect conditions (i.e., the system model perfectly matches the real system, the process noises and sensor noises are white, the noise covariances are exactly known, and the initial conditions are Gaussian), the innovations generated by the Kalman filter are independent random vectors with covariance matrix $CP_{k|k-1}C^T + R$. Especially, the innovations $r_k \sim \mathcal{N}\left(0, CP_{k|k-1}C^T + R\right)$ under normal operation and $r_k \sim \mathcal{N}\left(\psi_{k-k_0+1}, CP_{k|k-1}C^T + R\right)$ under the abnormal behavior.

In practical situations, however, the noise covariance matrices are generally unknown. Though there are several methods for estimating noise covariances, they are often associated with some levels of deterministic or stochastic uncertainty. For this reason, it is necessary to investigate the property of innovations generated from the discrete-time Kalman filter when noise covariances are not exactly known.

In this subsection, we calculate the covariance between two innovations $\mathrm{cov}\,(r_{k+l}, r_k) = \mathbb{E}_0\left[r_{k+l}r_k^T\right]$, for any $l \geq 0$, when the true values of process and sensor noise covariances (i.e., $\overline{Q}$ and $\overline{R}$) are different from their putative values (i.e., $Q$ and $R$), respectively. In this case, the value $P_{k|k-1}$ given in (A.11) no longer reflects the true covariance of state estimation error (i.e., $P_{k|k-1} \neq \mathrm{cov}\left(x_k - \hat{x}_{k|k-1}\right)$). Let $\overline{P}_{k|k-1}$ be the true covariance of state estimate error, then it can be calculated recursively as

$$\overline{P}_{k+1|k} = \mathrm{cov}\left(x_{k+1} - \hat{x}_{k+1|k}\right) = \mathrm{cov}\,(e_{k+1}) = \mathbb{E}_0\left[e_{k+1}e_{k+1}^T\right], \tag{A.15}$$

where the state estimation error evolves by the first equation in (A.12) with $a_k = 0$. Then, we

have

$$\begin{aligned}
\overline{P}_{k+1|k} &= \mathbb{E}_0\left[\underbrace{\left\{(A-AK_kC)\,e_k + w_k - AK_kv_k\right\}}_{e_{k+1}}\underbrace{\left\{(A-AK_kC)\,e_k + w_k - AK_kv_k\right\}^T}_{e_{k+1}^T}\right] \\
&= (A-AK_kC)\,\underbrace{\mathbb{E}_0\left[e_ke_k^T\right]}_{\overline{P}_{k|k-1}}(A-AK_kC)^T + (A-AK_kC)\,\underbrace{\mathbb{E}_0\left[e_k\,(w_k-AK_kv_k)^T\right]}_{0} + \\
&\underbrace{\mathbb{E}_0\left[(w_k-AK_kv_k)\,e_k^T\right]}_{0}(A-AK_kC)^T + \mathbb{E}_0\left[(w_k-AK_kv_k)\,(w_k-AK_kv_k)^T\right] \\
&= (A-AK_kC)\,\overline{P}_{k|k-1}\,(A-AK_kC)^T + \overline{Q} + (AK_k)\,\overline{R}\,(AK_k)^T\,.
\end{aligned}$$

We calculate in the following the covariance $\operatorname{cov}(r_{k+l}, r_k) = \mathbb{E}_0\left[r_{k+l}r_k^T\right]$ between two innovation vectors $r_{k+l}$ and $r_k$, for any $k \geq 1$ and $l \geq 0$. For $l = 0$, it is clear that

$$\begin{aligned}
\mathbb{E}_0\left[r_kr_k^T\right] &= \mathbb{E}_0\left[(Ce_k+v_k)\,(Ce_k+v_k)^T\right] \\
&= C\underbrace{\mathbb{E}_0\left[e_ke_k^T\right]}_{\overline{P}_{k|k-1}}C^T + C\underbrace{\mathbb{E}_0\left[e_kv_k^T\right]}_{0} + \underbrace{\mathbb{E}_0\left[v_ke_k^T\right]}_{0}C^T + \underbrace{\mathbb{E}_0\left[v_kv_k^T\right]}_{\overline{R}} \\
&= C\overline{P}_{k|k-1}C^T + \overline{R},
\end{aligned}$$

where $\overline{P}_{k|k-1}$ is the covariance matrix of the state estimation error which can be calculated recursively by (A.16) with the initial value $\overline{P}_{1|0} = P_{1|0}$.

For $l > 0$, we have

$$\begin{aligned}
\mathbb{E}_0\left[r_{k+l}r_k^T\right] &= \mathbb{E}_0\left[(Ce_{k+l}+v_{k+l})\,(Ce_k+v_k)^T\right] \\
&= C\mathbb{E}_0\left[e_{k+l}e_k^T\right]C^T + C\mathbb{E}_0\left[e_{k+l}v_k^T\right] + \underbrace{\mathbb{E}_0\left[v_{k+l}e_k^T\right]}_{0}C^T + \underbrace{\mathbb{E}_0\left[v_{k+l}v_k^T\right]}_{0} \\
&= C\mathbb{E}_0\left[e_{k+l}\,(Ce_k+v_k)^T\right] \\
&= C\mathbb{E}_0\left[e_{k+l}r_k^T\right],
\end{aligned}$$

where the covariance matrix $\mathbb{E}_0\left[e_{k+l}r_k^T\right]$ is calculated as follows:

- For $l = 1$, we have

$$\begin{aligned}
\mathbb{E}_0\left[e_{k+1}r_k^T\right] &= \mathbb{E}_0\left[\underbrace{\left\{(A - AK_kC)\,e_k + w_k - AK_kv_k\right\}}_{e_{k+1}} r_k^T\right] \\
&= (A - AK_kC)\,\mathbb{E}_0\left[e_kr_k^T\right] + \underbrace{\mathbb{E}_0\left[w_kr_k^T\right]}_{0} - AK_k\underbrace{\mathbb{E}_0\left[v_kr_k^T\right]}_{\overline{R}} \\
&= (A - AK_kC)\underbrace{\mathbb{E}_0\left[e_ke_k^T\right]}_{\overline{P}_{k|k-1}}C^T - AK_k\overline{R} \\
&= A\overline{P}_{k|k-1}C^T - AK_kC\overline{P}_{k|k-1}C^T - AK_k\overline{R} \\
&= A\overline{P}_{k|k-1}C^T - AK_k\left(C\overline{P}_{k|k-1}C^T + \overline{R}\right) \\
&= A\overline{P}_{k|k-1}C^T - AP_{k|k-1}C^T\left(CP_{k|k-1}C^T + R\right)^{-1}\left(C\overline{P}_{k|k-1}C^T + \overline{R}\right),
\end{aligned}$$

  since the optimal Kalman gain is calculated by (A.10).

- For $l > 1$, we have

$$\begin{aligned}
\mathbb{E}_0\left[e_{k+l}r_k^T\right] &= \mathbb{E}_0\left[\underbrace{\left\{(A - AK_{k+l-1}C)\,e_{k+l-1} + w_{k+l-1} - AK_{k+l-1}v_{k+l-1}\right\}}_{e_{k+l-1}} r_k^T\right] \\
&= (A - AK_{k+l-1}C)\,\mathbb{E}_0\left[e_{k+l-1}r_k^T\right] + \underbrace{\mathbb{E}_0\left[w_{k+l-1}r_k^T\right]}_{0} - AK_{k+l-1}\underbrace{\mathbb{E}_0\left[v_{k+l-1}r_k^T\right]}_{0} \\
&= (A - AK_{k+l-1}C)\,\mathbb{E}_0\left[e_{k+l-1}r_k^T\right]. \qquad \text{(A.16)}
\end{aligned}$$

It follows from above analysis that the covariance matrix $\mathrm{cov}\,(r_{k+l}, r_k) = \mathbb{E}_0\left[r_{k+l}r_k^T\right]$ between two innovations $r_{k+l}$ and $r_k$, for $k \geq 1$ and $l \geq 0$, when the true noise covariance matrices are different from their putative values can be synthesized into the following algorithm.

## A.2 Proof of Theorem 3.1

The proof of Theorem 3.1 is inspired by [69] and [67, pages 51-54] for the independent Gaussian random observations. In this proof, we generalize the results in [67, 69] to the unified statistical model (3.25). The proof is divided into two parts. In the first part, we investigate the worst-case probability of false alarm $\overline{\mathbb{P}}_{\mathrm{fa}}$ given in (3.35). In the second part, we introduce the upper bound $\tilde{\mathbb{P}}_{\mathrm{md}}$ for the worst-case probability of missed detection $\overline{\mathbb{P}}_{\mathrm{md}}$ given in (3.36).

### A.2.1 Proof of part 1

For the probability of false alarm, let us assume the pre-change mode (i.e., $k_0 \to \infty$). Under the pre-change probability measure $\mathcal{P}_0$, it follows from the unified statistical model (3.25) that

**Algorithm 1** Recursive calculation of covariance matrix $\mathrm{cov}\left(r_{k+l}, r_k\right)$ when true noise covariances (i.e., $\overline{Q}$ and $\overline{R}$) are different from putative noise covariances (i.e., $Q$ and $R$).

1. Initialization of the covariance matrix of state estimation error $P_{1|0} = \mathrm{cov}\left(x_1 - \hat{x}_{1|0}\right)$ and $\overline{P}_{1|0} = P_{1|0}$.

2. Calculation of the putative values of the optimal Kalman gain $K_k$ and the covariance matrix of state estimation error $P_{k|k-1}$:

$$K_k = P_{k|k-1}C^T\left(CP_{k|k-1}C^T + R\right)^{-1}, \tag{A.17}$$

$$P_{k+1|k} = AP_{k|k-1}A^T - AP_{k|k-1}C^T\left(CP_{k|k-1}C^T + R\right)^{-1}CP_{k|k-1}A^T + Q, \tag{A.18}$$

3. Calculation of the covariance matrix of true state estimation error $\overline{P}_{k+1|k}$:

$$\overline{P}_{k+1|k} = \left(A - AK_kC\right)\overline{P}_{k|k-1}\left(A - AK_kC\right)^T + \overline{Q} + \left(AK_k\right)\overline{R}\left(AK_k\right)^T. \tag{A.19}$$

4. Calculation of the covariance matrix $\mathrm{cov}\left(r_{k+l}, r_k\right) = \mathbb{E}_0\left[r_{k+l}r_k^T\right]$:

   - If $l = 0$, then
   $$\mathbb{E}_0\left[r_k r_k^T\right] = C\overline{P}_{k|k-1}C^T + \overline{R}. \tag{A.20}$$
   - Else if $l > 0$, then
   $$\mathbb{E}_0\left[r_{k+l}r_k^T\right] = C\mathbb{E}_0\left[e_{k+l}r_k^T\right], \tag{A.21}$$
   where the matrix $\mathbb{E}_0\left[e_{k+l}r_k^T\right]$ is computed recursively as
     - If $l = 1$, then
     $$\mathbb{E}_0\left[e_{k+1}r_k^T\right] = A\overline{P}_{k|k-1}C^T - AK_k\left(C\overline{P}_{k|k-1}C^T + \overline{R}\right). \tag{A.22}$$
     - Else if $l > 1$, then
     $$\mathbb{E}_0\left[e_{k+l}r_k^T\right] = \left(A - AK_kC\right)\mathbb{E}_0\left[e_{k+l-1}r_k^T\right]. \tag{A.23}$$

$r_{k-L+1}^k = \xi_{k-L+1}^k$, for both the steady-state Kalman filter approach and the fixed-size parity space approach, since the vector of transient signals $\phi_{k-L+1}^k\left(k_0\right)$ is null. Moreover, it has been discussed in subsection 3.3.4 that the random noises $\left\{\xi_{k-L+1}^k\right\}_{k\geq L}$ follow the same distribution (i.e., $\xi_{k-L+1}^k \sim \mathcal{N}\left(0, \Sigma\right)$) and that the vector of transient profiles $\phi_{k-L+1}^k\left(i\right)$ depends only on the relative position of index $i$ within the window $[k-L+1, k]$. Putting together with (3.34), we obtain that the random variables $\left(S_{k-L+1}^k, \cdots, S_k^k\right)$ follow the same distribution as the random variables $\left(S_{k-L+1+j}^{k+j}, \cdots, S_{k+j}^{k+j}\right)$, for all $j \geq 1$.

Let $u_l = \mathbb{P}_0\left(T_{\mathrm{VTWL}} = l\right)$, we show in the following that $u_{l+1} \leq u_l$ for all $l \geq L$. For $l = L$, we

have

$$\begin{aligned}
u_{L+1} &= \mathbb{P}_0\left(T_{\mathrm{VTWL}} = L+1\right) \\
&= \mathbb{P}_0\left(\left\{\max_{1\le i\le L}\left(S_i^L - h_{L-i+1}\right) < 0\right\}\bigcap\left\{\max_{2\le i\le L+1}\left(S_i^{L+1} - h_{L-i+2}\right) \ge 0\right\}\right) \\
&\le \mathbb{P}_0\left(\max_{2\le i\le L+1}\left(S_i^{L+1} - h_{L-i+2}\right) \ge 0\right) \\
&\le \mathbb{P}_0\left(\max_{1\le i\le L}\left(S_i^{L} - h_{L-i+1}\right) \ge 0\right) = u_L,
\end{aligned} \tag{A.24}$$

where the last inequality comes from the above analysis that $\left(S_1^L, \cdots, S_L^L\right)$ and $\left(S_2^{L+1}, \cdots, S_{L+1}^{L+1}\right)$ follow the same distribution, leading to $u_{L+1} \le u_L$. By the same argument, we obtain for the case $l > L$ that

$$\begin{aligned}
u_{l+1} &= \mathbb{P}_0\left(T_{\mathrm{VTWL}} = l+1\right) \\
&= \mathbb{P}_0\left(\bigcap_{k=L}^{l}\left\{\max_{k-L+1\le i\le k}\left(S_i^k - h_{k-i+1}\right) < 0\right\}\bigcap\left\{\max_{l-L+2\le i\le l+1}\left(S_i^{l+1} - h_{l-i+2}\right) \ge 0\right\}\right) \\
&\le \mathbb{P}_0\left(\bigcap_{k=L+1}^{l}\left\{\max_{k-L+1\le i\le k}\left(S_i^k - h_{k-i+1}\right) < 0\right\}\bigcap\left\{\max_{l-L+2\le i\le l+1}\left(S_i^{l+1} - h_{l-i+2}\right) \ge 0\right\}\right) \\
&\le \mathbb{P}_0\left(\bigcap_{k=L}^{l-1}\left\{\max_{k-L+1\le i\le k}\left(S_i^k - h_{k-i+1}\right) < 0\right\}\bigcap\left\{\max_{l-L+1\le i\le l}\left(S_i^{l} - h_{l-i+1}\right) \ge 0\right\}\right) \\
&\le \mathbb{P}_0\left(T_{\mathrm{VTWL}} = l\right) = u_l,
\end{aligned} \tag{A.25}$$

leading to $u_{l+1} \le u_l$. Let $U_l = \mathbb{P}_0\left(l \le T_{VTWL} \le l + m_\alpha - 1\right)$, then

$$U_l - U_{l+1} = \left(\sum_{k=l}^{l+m_\alpha-1} u_k\right) - \left(\sum_{k=l+1}^{l+m_\alpha} u_k\right) = u_l - u_{l+m_\alpha} \ge 0. \tag{A.26}$$

Hence, $\{U_l\}_{l\ge L}$ is a non-increasing sequence, leading to

$$\overline{\mathbb{P}}_{\mathrm{fa}}\left(T_{\mathrm{VTWL}}; m_\alpha; h_1, h_2, \cdots, h_L\right) = U_L = \mathbb{P}_0\left(L \le T_{\mathrm{VTWL}} \le L + m_\alpha - 1\right). \tag{A.27}$$

The proof of part 1 is completed. □.

### A.2.2 Proof of part 2

The worst-case probability of missed detection of the VTWL CUSUM test (3.33)–(3.34) is described as

$$\begin{aligned}
\overline{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{VTWL}}; L; h_1, h_2, \cdots, h_L\right) &= \sup_{k_0\ge L}\mathbb{P}_{k_0}\left(T_{\mathrm{VTWL}} \ge k_0 + L | T_{\mathrm{VTWL}} \ge k_0\right) \\
&= \sup_{k_0\ge L}\frac{\mathbb{P}_{k_0}\left(T_{\mathrm{VTWL}} \ge k_0 + L\right)}{\mathbb{P}_{k_0}\left(T_{\mathrm{VTWL}} \ge k_0\right)},
\end{aligned} \tag{A.28}$$

where it is assumed that $\mathbb{P}_L\left(T_{\mathrm{VTWL}} \geq L\right) = 1$ (i.e., corresponding to the change-point $k_0 = L$). The worst-case probability of missed detection is expressed by

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{VTWL}}; L; h_1, h_2, \cdots, h_L\right) = \sup_{k_0 \geq L} \frac{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0+L-1} \left\{\max_{k-L+1 \leq i \leq k} \left(S_i^k - h_{k-i+1}\right) < 0\right\}\right)}{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0-1} \left\{\max_{k-L+1 \leq i \leq k} \left(S_i^k - h_{k-i+1}\right) < 0\right\}\right)}, \quad \text{(A.29)}$$

where the LLR $S_i^k$, for $k - L + 1 \leq i \leq k$, is rewritten as

$$S_i^k = \left[\phi_{k-L+1}^k\left(i\right)\right]^T \left[\Sigma^{-1}\right] \left[\xi_{k-L+1}^k\right] + \mathbb{E}_{k_0}\left[S_i^k\right], \quad \text{(A.30)}$$

where $\mathbb{E}_{k_0}\left[S_i^k\right]$ is the mathematical expectation of the LLR $S_i^k$ under the probability measure $\mathcal{P}_{k_0}$, which is calculated as

$$\mathbb{E}_{k_0}\left[S_i^k\right] = \left[\phi_{k-L+1}^k\left(i\right)\right]^T \left[\Sigma^{-1}\right] \left[\phi_{k-L+1}^k\left(k_0\right) - \frac{1}{2}\phi_{k-L+1}^k\left(i\right)\right] \quad \text{(A.31)}$$

Let us define three events $A_1$, $A_2$ and $A_3$ as follows:

$$\begin{aligned}
A_1 &= \bigcap_{k=L}^{k_0-1} \left\{\max_{k-L+1 \leq i \leq k} \left(S_i^k - h_{k-i+1}\right) < 0\right\}, \\
A_2 &= \bigcap_{k=k_0}^{k_0+L-2} \left\{\max_{k-L+1 \leq i \leq k} \left(S_i^k - h_{k-i+1}\right) < 0\right\}, \\
A_3 &= \left\{\max_{k_0 \leq i \leq k_0+L-1} \left(S_i^{k_0+L-1} - h_{k_0+L-i}\right) < 0\right\}.
\end{aligned}$$

It follows from (A.30) that the event $A_1$ depends on the random vectors $\xi_1^L, \cdots, \xi_{k_0-L}^{k_0-1}$, the event $A_2$ depends on the random vectors $\xi_{k_0-L+1}^{k_0}, \cdots, \xi_{k_0-1}^{k_0+L-2}$ and the event $A_3$ depends on only the random vector $\xi_{k_0}^{k_0+L-1}$. Moreover, there is no common element between $\xi_1^L, \cdots, \xi_{k_0-L}^{k_0-1}$ and $\xi_{k_0}^{k_0+L-1}$. Hence, the events $A_1$ and $A_3$ are independent, leading to

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{VTWL}}; L; h_1, h_2, \cdots, h_L\right) = \sup_{k_0 \geq L} \frac{\mathbb{P}_{k_0}\left(A_1 \cap A_2 \cap A_3\right)}{\mathbb{P}_{k_0}\left(A_1\right)} \leq \sup_{k_0 \geq L} \frac{\mathbb{P}_{k_0}\left(A_1 \cap A_3\right)}{\mathbb{P}_{k_0}\left(A_1\right)} \leq \sup_{k_0 \geq L} \mathbb{P}_{k_0}\left(A_3\right). \quad \text{(A.32)}$$

By replacing the event $A_3$ with its definition, we obtain that

$$\begin{aligned}
\overline{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{VTWL}}; L; h_1, h_2, \cdots, h_L\right) &\leq \sup_{k_0 \geq L} \mathbb{P}_{k_0}\left(\bigcap_{i=k_0}^{k_0+L-1} \left\{S_i^{k_0+L-1} < h_{k_0+L-i}\right\}\right) \\
&\leq \sup_{k_0 \geq L} \mathbb{P}_{k_0}\left(S_{k_0}^{k_0+L-1} < h_L\right) = \mathbb{P}_1\left(S_1^L < h_L\right). \quad \text{(A.33)}
\end{aligned}$$

Let $\tilde{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{VTWL}}; h_L\right) \triangleq \mathbb{P}_1\left(S_1^L < h_L\right)$ be the upper bound for the worst-case probability of missed detection $\overline{\mathbb{P}}_{\mathrm{md}}$, then

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{VTWL}}; L; h_1, h_2, \cdots, h_L\right) \leq \tilde{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{VTWL}}; h_L\right) \triangleq \Phi\left(\frac{h_L - \mu_{S_1^L}}{\sigma_{S_1^L}}\right), \quad \text{(A.34)}$$

where the parameters $\mu_{S_1^L}$ and $\sigma_{S_1^L}$ are calculated as

$$\mu_{S_1^L} = \frac{1}{2}\left[\phi_1^L(1)\right]^T\left[\Sigma^{-1}\right]\left[\phi_1^L(1)\right], \tag{A.35}$$

$$\sigma^2_{S_1^L} = \left[\phi_1^L(1)\right]^T\left[\Sigma^{-1}\right]\left[\phi_1^L(1)\right]. \tag{A.36}$$

The proof of Theorem 3.1 is completed. □.

## A.3 Proof of Lemma 3.2

Let us suppose that Assumption 3.2 is satisfied. We prove in this section that the covariance matrix $\Sigma_{\mathcal{S}}$ of the Gaussian random vector $\mathcal{S} = \left[S_1^L, \cdots, S_{m_\alpha}^{L+m_\alpha-1}\right]^T \in \mathbb{R}^{m_\alpha}$ is positive-definite. It follows from (3.34) that the LLR $S_i^k$, for $k-L+1 \leq i \leq k$, is rewritten as

$$S_i^k = \left[\phi_{k-L+1}^k(i)\right]^T\left[\Sigma^{-1}\right]\left[\xi_{k-L+1}^k\right] + \mathbb{E}_0\left[S_i^k\right], \tag{A.37}$$

where $\mathbb{E}_0\left[S_i^k\right]$ is the mathematical expectation of the LLR $S_i^k$ under the pre-change probability measure $\mathcal{P}_0$ and it is calculated as

$$\mathbb{E}_0\left[S_i^k\right] = -\frac{1}{2}\left[\phi_{k-L+1}^k(i)\right]^T\left[\Sigma^{-1}\right]\left[\phi_{k-L+1}^k(i)\right] \tag{A.38}$$

### A.3.1 Steady-state Kalman filter approach

For the steady-state Kalman filter approach, the transient profiles $\phi_1^L(1) = \psi_1^L(1)$ and the random noises $\xi_{k-L+1}^k = \varrho_{k-L+1}^k$ with the covariance matrix $\Sigma = \Sigma_\varrho$ is symmetric and positive-definite. The LLR $S_i^k$ can be rewritten for the steady-state Kalman filter approach as

$$S_i^k = \left[\psi_{k-L+1}^k(i)\right]^T\left[\Sigma_\varrho^{-1}\right]\left[\varrho_{k-L+1}^k\right] + \mathbb{E}_0\left[S_i^k\right], \tag{A.39}$$

where the vector of transient profiles $\psi_{k-L+1}^k(i)$ is given by (3.13) and the vector of random noises $\varrho_{k-L+1}^k = \left[\varrho_{k-L+1}^T, \cdots, \varrho_k^T\right]^T$, where $\varrho_{k-L+1}, \cdots, \varrho_k \in \mathbb{R}^p$ are i.i.d. zero-mean Gaussian random vectors. Let the coefficient vector $\lambda_1^L \in \mathbb{R}^{Lp}$ be defined as

$$\lambda_1^L = \left[\lambda_1^T, \cdots, \lambda_L^T\right]^T = \left[\Sigma_\varrho^{-1}\right]\left[\psi_1^L(1)\right], \tag{A.40}$$

where the elements $\lambda_1, \cdots, \lambda_L \in \mathbb{R}^p$ are known. The LLR $S_{k-L+1}^k$ is then described as

$$S_{k-L+1}^k = \left[\lambda_1^L\right]^T\left[\varrho_{k-L+1}^k\right] + \mathbb{E}_0\left[S_{k-L+1}^k\right] = \left[\lambda_1^T, \cdots, \lambda_L^T\right]\begin{bmatrix}\varrho_{k-L+1}\\ \vdots \\ \varrho_k\end{bmatrix} + \mathbb{E}_0\left[S_{k-L+1}^k\right]. \tag{A.41}$$

Then, the LLRs $S_1^L, S_2^{L+1}, \cdots, S_{m_\alpha}^{L+m_\alpha-1}$ can be rewritten as

$$\begin{aligned}
S_1^L &= \lambda_1^T\varrho_1 + \lambda_2^T\varrho_2 + \cdots + \lambda_L^T\varrho_L + \mathbb{E}_0\left[S_1^L\right],\\
S_2^{L+1} &= \lambda_1^T\varrho_2 + \lambda_2^T\varrho_3 + \cdots + \lambda_L^T\varrho_{L+1} + \mathbb{E}_0\left[S_2^{L+1}\right],\\
&\vdots \quad \vdots \quad \vdots\\
S_{m_\alpha}^{L+m_\alpha-1} &= \lambda_1^T\varrho_{m_\alpha} + \lambda_2^T\varrho_{m_\alpha+1} + \cdots + \lambda_L^T\varrho_{L+m_\alpha-1} + \mathbb{E}_0\left[S_{m_\alpha}^{L+m_\alpha-1}\right].
\end{aligned}$$

By rewriting the above equations in matrix form, we obtain

$$\underbrace{\begin{bmatrix} S_1^L \\ S_2^{L+1} \\ \vdots \\ S_{m_\alpha}^{L+m_\alpha-1} \end{bmatrix}}_{\mathcal{S}\in\mathbb{R}^{m_\alpha}} = \underbrace{\begin{bmatrix} \lambda_1^T & \lambda_2^T & \cdots & \lambda_L^T & 0 & \cdots & 0 \\ 0 & \lambda_1^T & \cdots & \lambda_{L-1}^T & \lambda_L^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & \lambda_L^T \end{bmatrix}}_{\mathcal{T}_\varrho\in\mathbb{R}^{m_\alpha\times(L+m_\alpha-1)p}} \underbrace{\begin{bmatrix} \varrho_1 \\ \varrho_2 \\ \vdots \\ \varrho_L \\ \varrho_{L+1} \\ \vdots \\ \varrho_{L+m_\alpha-1} \end{bmatrix}}_{\varrho_1^{L+m_\alpha-1}\in\mathbb{R}^{(L+m_\alpha-1)p}} + \underbrace{\begin{bmatrix} \mathbb{E}_0\left[S_1^L\right] \\ \mathbb{E}_0\left[S_2^{L+1}\right] \\ \vdots \\ \mathbb{E}_0\left[S_{m_\alpha}^{L+m_\alpha-1}\right] \end{bmatrix}}_{\mu_\mathcal{S}\in\mathbb{R}^{m_\alpha}}. \tag{A.42}$$

It is worth noting that $\varrho_1^{L+m_\alpha-1} \sim \mathcal{N}\left(0, \tilde{\Sigma}_\varrho\right)$, where $\tilde{\Sigma}_\varrho \in \mathbb{R}^{(L+m_\alpha-1)p\times(L+m_\alpha-1)p}$ is a positive-definite matrix since $\varrho_1, \varrho_2, \cdots, \varrho_{L+m_\alpha-1}$ are i.i.d. zero-mean Gaussian random vectors (see subsection 3.3.1). The covariance matrix $\Sigma_\mathcal{S}$ is then calculated as

$$\Sigma_\mathcal{S} = \mathcal{T}_\varrho \tilde{\Sigma}_\varrho \mathcal{T}_\varrho^T. \tag{A.43}$$

Let the coefficient vector $\lambda_j \in \mathbb{R}^p$ be described as $\lambda_j = \left[\lambda_j^1, \lambda_j^2, \cdots, \lambda_j^p\right]^T$, where the element $\lambda_j^i \in \mathbb{R}$, for $1 \leq i \leq p$ and $1 \leq j \leq L$. If Assumption 3.1 is satisfied, the transient profiles $\psi_1^L(1)$ is non-null. It follows from (A.40) that the coefficient vector $\lambda_1^L \neq 0$. For this reason, there exists at least one element $\lambda_j^i \neq 0$. Let $T \in \mathbb{R}^{m_\alpha\times m_\alpha}$ be a square matrix formulated by $m_\alpha$ columns containing the non-zero element $\lambda_j^i \neq 0$ extracted from the matrix $\mathcal{T}_\varrho$. Then, the matrix $T$ is described as

$$T = \begin{bmatrix} \lambda_j^i & - & \cdots & - \\ 0 & \lambda_j^i & \cdots & - \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_j^i \end{bmatrix}, \tag{A.44}$$

where the notation "$-$" stands for any real numbers. The matrix $T$ is an upper triangular one with non-zero elements in the diagonal (i.e., $\lambda_j^i \neq 0$), then $\text{rank}(T) = m_\alpha$. Since the columns of $T$ are contained in matrix $\mathcal{T}_\varrho$ and matrix $\mathcal{T}_\varrho$ has $m_\alpha$ rows, we have $\text{rank}(\mathcal{T}_\varrho) = m_\alpha$. In other words, the matrix $\mathcal{T}_\varrho$ is full row rank if Assumption 3.1 is satisfied. As it follows from [91, page 47] that, if matrix $\mathcal{T}_\varrho$ is full row rank and matrix $\tilde{\Sigma}_\varrho$ is non-singular, then the covariance matrix $\Sigma_\mathcal{S} = \mathcal{T}_\varrho \tilde{\Sigma}_\varrho \mathcal{T}_\varrho^T$ is positive-definite. □.

### A.3.2 Fixed-size parity space approach

For the fixed-size parity space approach, the transient profiles $\phi_1^L(1) = \varphi_1^L(1)$ and the random noises $\xi_{k-L+1}^k = \varsigma_{k-L+1}^k$ (see subsection 3.3.2). The LLR $S_i^k$ is described as

$$S_i^k = \left[\varphi_{k-L+1}^k(i)\right]^T \left[\Sigma_\varsigma^{-1}\right] \left[\varsigma_{k-L+1}^k\right] + \mathbb{E}_0\left[S_i^k\right], \tag{A.45}$$

where $\varphi_{k-L+1}^k(i) = \mathcal{W}\mathcal{M}\theta_{k-L+1}^k(i)$ and $\varsigma_{k-L+1}^k = \mathcal{W}\left(\mathcal{H}w_{k-L+1}^k + v_{k-L+1}^k\right)$. Hence, the LLR $S_i^k$ can be rewritten as

$$S_i^k = \left[\varphi_{k-L+1}^k(i)\right]^T \left[\Sigma_\varsigma^{-1}\right] \left[\mathcal{W}\mathcal{H}w_{k-L+1}^k + \mathcal{W}v_{k-L+1}^k\right] + \mathbb{E}_0\left[S_i^k\right]. \tag{A.46}$$

Let us define coefficient vectors $\beta_1^L \in \mathbb{R}^{Ln}$ and $\gamma_1^L \in \mathbb{R}^{Lp}$ be defined as follows:

$$\beta_1^L = \left[\beta_1^T, \cdots, \beta_L^T\right]^T = \left[\left(\varphi_1^L(1)\right)^T \Sigma_\varsigma^{-1} \mathcal{W}\mathcal{H}\right]^T = \mathcal{H}^T \mathcal{W}^T \Sigma_\varsigma^{-1} \varphi_1^L(1), \tag{A.47}$$

$$\gamma_1^L = \left[\gamma_1^T, \cdots, \gamma_L^T\right]^T = \left[\left(\varphi_1^L(1)\right)^T \Sigma_\varsigma^{-1} \mathcal{W}\right]^T = \mathcal{W}^T \Sigma_\varsigma^{-1} \varphi_1^L(1), \tag{A.48}$$

where $\beta_1, \cdots, \beta_L \in \mathbb{R}^n$ and $\gamma_1, \cdots, \gamma_L \in \mathbb{R}^p$ are known. The LLR $S_{k-L+1}^k$ can be described in terms of $w_{k-L+1}^k$ and $v_{k-L+1}^k$ as follows:

$$S_{k-L+1}^k = \left[\beta_1^L\right]^T w_{k-L+1}^k + \left[\gamma_1^L\right]^T v_{k-L+1}^k + \mathbb{E}_0\left[S_{k-L+1}^k\right]. \tag{A.49}$$

Similar to the Kalman filter approach, the Gaussian random vector $\mathcal{S} \in \mathbb{R}^{m_\alpha}$ formed by the LLRs $S_1^L, S_2^L, \cdots, S_{m_\alpha}^{L+m_\alpha-1}$ can be described as

$$\mathcal{S} = \mathcal{T}_w w_1^{L+m_\alpha-1} + \mathcal{T}_v v_1^{L+m_\alpha-1} + \mu_\mathcal{S}, \tag{A.50}$$

where $\mu_\mathcal{S} \in \mathbb{R}^{m_\alpha}$ is non-random mean vector, the random vectors $w_1^{L+m_\alpha-1} = \left[w_1^T, \cdots, w_{L+m_\alpha-1}^T\right]^T$ and $v_1^{L+m_\alpha-1} = \left[v_1^T, \cdots, v_{L+m_\alpha-1}^T\right]^T$, the matrices $\mathcal{T}_w \in \mathbb{R}^{m_\alpha \times (L+m_\alpha-1)n}$ and $\mathcal{T}_v \in \mathbb{R}^{m_\alpha \times (L+m_\alpha-1)p}$ are described, respectively, as

$$\mathcal{T}_w = \begin{bmatrix} \beta_1^T & \beta_2^T & \cdots & \beta_L^T & 0 & \cdots & 0 \\ 0 & \beta_1^T & \cdots & \beta_{L-1}^T & \beta_L^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & \beta_L^T \end{bmatrix}, \tag{A.51}$$

$$\mathcal{T}_v = \begin{bmatrix} \gamma_1^T & \gamma_2^T & \cdots & \gamma_L^T & 0 & \cdots & 0 \\ 0 & \gamma_1^T & \cdots & \gamma_{L-1}^T & \gamma_L^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & \gamma_L^T \end{bmatrix}. \tag{A.52}$$

If Assumption 3.1 is satisfied, i.e., $\varphi_1^L(1) \neq 0$, then $\gamma_1^L = \mathcal{W}^T \Sigma_\varsigma^{-1} \varphi_1^L(1) \neq 0$ since matrix $\Sigma_\varsigma$ is non-singular and matrix $\mathcal{W}^T$ is full column rank. The coefficient vector $\beta_1^L$ may be null or non-null. Similar to the steady-state Kalman filter approach, it can be shown that if $\gamma_1^L \neq 0$, the matrix $\mathcal{T}_v$ is full row rank ($\mathrm{rank}(\mathcal{T}_v) = m_\alpha$). Let $\tilde{\mathcal{Q}}$ and $\tilde{\mathcal{R}}$ be the covariance matrices of random noises $w_1^{L+m_\alpha-1}$ and $v_1^{L+m_\alpha-1}$, then $\tilde{\mathcal{Q}}$ is positive-semidefinite and $\tilde{\mathcal{R}}$ is positive-definite. Hence, the covariance matrix

$$\Sigma_\mathcal{S} = \underbrace{\mathcal{T}_w \tilde{\mathcal{Q}} \mathcal{T}_w^T}_{\text{positive-semidefinite}} + \underbrace{\mathcal{T}_v \tilde{\mathcal{R}} \mathcal{T}_v^T}_{\text{positive-definite}} \tag{A.53}$$

is positive-definite. The proof of Lemma 3.2 is completed. □.

## A.4 Proof of Theorem 3.2

The proof of Theorem 3.2 consists of two parts. The optimization problem is formulated and solved in the first part. It is shown in the second part that the optimized VTWL CUSUM test is equivalent to the FMA test.

### A.4.1 Proof of part 1

Since we wish to minimize the upper bound $\tilde{\mathbb{P}}_{\rm md}\left(T_{\rm VTWL}; h_L\right)$ subject to an acceptable level $\alpha \in (0,1)$ on the worst-case probability of false alarm $\overline{\mathbb{P}}_{\rm fa}$, the optimization problem can be defined as

$$\begin{cases} \inf_{h_1,h_2,\cdots,h_L} & \tilde{\mathbb{P}}_{\rm md}\left(T_{\rm VTWL}; h_L\right) \\ \text{subject to} & \overline{\mathbb{P}}_{\rm fa}\left(T_{\rm VTWL}; m_\alpha; h_1, \cdots, h_L\right) \leq \alpha \end{cases}, \tag{A.54}$$

where the worst-case probability of false alarm is calculated from (3.35) as

$$\begin{aligned} \overline{\mathbb{P}}_{\rm fa}\left(T_{\rm VTWL}; m_\alpha; h_1, \cdots, h_L\right) &= \mathbb{P}_0\left(L \leq T_{\rm VTWL} \leq L + m_\alpha - 1\right) \\ &= 1 - \mathbb{P}_0\left(T_{\rm VTWL} \geq L + m_\alpha\right) \\ &= 1 - \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \left\{ \max_{k-L+1 \leq i \leq k} \left(S_i^k - h_{k-i+1}\right) < 0 \right\}\right) \\ &= 1 - \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \bigcap_{i=k-L+1}^{k} \left\{ S_i^k < h_{k-i+1} \right\}\right). \end{aligned} \tag{A.55}$$

Let us define a function $F_0\left(h_1, h_2, \cdots, h_L\right)$ depending on the thresholds $h_1, h_2, \cdots, h_L$ as

$$F_0\left(h_1, h_2, \cdots, h_L\right) \triangleq \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \bigcap_{i=k-L+1}^{k} \left\{ S_i^k < h_{k-i+1} \right\}\right). \tag{A.56}$$

The optimization problem (A.54) is equivalent to

$$\begin{cases} \inf_{h_1,h_2,\cdots,h_L} & \tilde{\mathbb{P}}_{\rm md}\left(T_{\rm VTWL}; h_L\right) \\ \text{subject to} & F_0\left(h_1, h_2, \cdots, h_L\right) \geq 1 - \alpha \end{cases}, \tag{A.57}$$

where the objective function $\tilde{\mathbb{P}}_{\rm md}\left(T_{\rm VTWL}; h_L\right) = \Phi\left(\frac{h_L - \mu_{S_1^L}}{\sigma_{S_1^L}}\right)$ is monotonically non-decreasing w.r.t. the threshold $h_L$. Let us investigate now the property of the function $F_0\left(h_1, h_2, \cdots, h_L\right)$. Let $\{\delta h_j\}_{1 \leq j \leq L}$ be positive real numbers, then

$$\begin{aligned} F_0\left(h_1, \cdots, h_j + \delta h_j, \cdots, h_L\right) &= \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \left[ \bigcap_{\substack{i=k-L+1 \\ i \neq k-j+1}}^{k} \left\{ S_i^k < h_{k-i+1} \right\} \text{ and } \left\{ S_{k-j+1}^k < h_j + \delta h_j \right\} \right]\right) \\ &= \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \left[ \left\{ \bigcap_{\substack{i=k-L+1 \\ i \neq k-j+1}}^{k} \left\{ S_i^k < h_{k-i+1} \right\} \text{ and } \left\{ S_{k-j+1}^k < h_j \right\} \right\} \bigcup \right.\right. \\ &\qquad \left.\left. \left\{ \bigcap_{\substack{i=k-L+1 \\ i \neq k-j+1}}^{k} \left\{ S_i^k < h_{k-i+1} \right\} \text{ and } \left\{ h_j \leq S_{k-j+1}^k < h_j + \delta h_j \right\} \right\} \right]\right) \\ &\geq \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \left[ \bigcap_{\substack{i=k-L+1 \\ i \neq k-j+1}}^{k} \left\{ S_i^k < h_{k-i+1} \right\} \text{ and } \left\{ S_{k-j+1}^k < h_j \right\} \right]\right) \\ &\geq \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \left[ \bigcap_{i=k-L+1}^{k} \left\{ S_i^k < h_{k-i+1} \right\} \right]\right) \triangleq F_0\left(h_1, \cdots, h_j, \cdots, h_L\right). \end{aligned} \tag{A.58}$$

It follows from (A.58) that the function $F_0(h_1, h_2, \cdots, h_L)$ is monotonically non-decreasing w.r.t. each threshold $h_1, h_2, \cdots, h_L$. By utilizing this property of $F_0(\cdot)$, we prove in the following that the thresholds $h_1^*, h_2^*, \cdots, h_{L-1}^* \to +\infty$ and the threshold $h_L^*$ satisfying

$$F_0(+\infty, \cdots, +\infty, h_L^*) = \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\left\{S_{k-L+1}^k < h_L^*\right\}\right) = 1-\alpha \tag{A.59}$$

are the solution to the optimization problem (A.57). The proof consists of two following steps:

- It follows from Lemma 3.2 that the covariance matrix $\Sigma_{\mathcal{S}}$ of the Gaussian random variables $S_1^L, S_2^{L+1}, \cdots, S_{m_\alpha}^{L+m_\alpha-1}$ is positive-definite, for both steady-state Kalman filter approach and fixed-size parity space approach. The function $F_0(+\infty, \cdots, +\infty, h_L^*)$ is monotonically non-decreasing w.r.t. the threshold $h_L^*$. Its co-domain is $[0, 1]$. Hence, the equation (A.59) has a unique solution $h_L^*$ for a given value $\alpha \in (0, 1)$.
- Let us suppose that a set of thresholds $h_1, \cdots, h_{L-1}, h_L$ satisfying the constraint

  $$F_0(h_1, \cdots, h_{L-1}, h_L) \geq 1-\alpha, \tag{A.60}$$

  defines any alternative solution of the optimization problem (A.57). The goal is to show that $\tilde{\mathbb{P}}_{\mathrm{md}}(T_{\mathrm{VTWL}}; h_L) \geq \tilde{\mathbb{P}}_{\mathrm{md}}(T_{\mathrm{VTWL}}; h_L^*)$. It follows from the monotonically non-decreasing property of the function $F_0(\cdot)$ that

  $$1-\alpha = F_0(+\infty, \cdots, +\infty, h_L^*) \geq F_0(h_1, \cdots, h_{L-1}, h_L^*). \tag{A.61}$$

  Putting together (A.60) and (A.61), we obtain that

  $$F_0(h_1, \cdots, h_{L-1}, h_L) \geq F_0(h_1, \cdots, h_{L-1}, h_L^*), \tag{A.62}$$

  resulting in $h_L \geq h_L^*$ since the function $F_0(\cdot)$ is monotonically non-decreasing w.r.t. each threshold. It follows from (3.36) that the objective function $h_L \mapsto \tilde{\mathbb{P}}_{\mathrm{md}}(T_{\mathrm{VTWL}}; h_L) \triangleq \Phi\left(\frac{h_L - \mu_{S_1^L}}{\sigma_{S_1^L}}\right)$ is monotonically non-decreasing w.r.t. the threshold $h_L$, leading to $\tilde{\mathbb{P}}_{\mathrm{md}}(T_{\mathrm{VTWL}}; h_L) \geq \tilde{\mathbb{P}}_{\mathrm{md}}(T_{\mathrm{VTWL}}; h_L^*)$. $\square$.

### A.4.2 Proof of part 2

The VTWL CUSUM algorithm with optimal thresholds $h_1^*, h_2^*, \cdots, h_L^*$ can be described as

$$\begin{aligned} T_{\mathrm{VTWL}}^* &= \inf\left\{k \geq L : \max_{k-L+1 \leq i \leq k}\left(S_i^k - h_{k-i+1}^*\right) \geq 0\right\} \\ &= \inf\left\{k \geq L : S_{k-L+1}^k \geq h_L^*\right\}, \end{aligned} \tag{A.63}$$

since the optimal thresholds $h_1^*, h_2^*, \cdots, h_{L-1}^* \to +\infty$. In addition, the LLR $S_{k-L+1}^k$ can be re-written as

$$S_{k-L+1}^k = \left[\phi_1^L(1)\right]^T\left[\Sigma^{-1}\right]\left[r_{k-L+1}^k - \frac{1}{2}\phi_1^L(1)\right]. \tag{A.64}$$

Hence, the optimized VTWL CUSUM algorithm is equivalent to the following simple FMA detection rule

$$T_{\mathrm{FMA}}\left(\tilde{h}_L\right) = \inf\left\{k \geq L : \left[\phi_1^L(1)\right]^T\left[\Sigma^{-1}\right] r_{k-L+1}^k \geq \tilde{h}_L\right\}, \tag{A.65}$$

where the threshold $\tilde{h}_L = h_L^* + \mu_{S_1^L}$. The upper bound for the worst-case probability of missed detection for the FMA detection rule as a function of the threshold $\tilde{h}_L$ is calculated as

$$\tilde{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{FMA}}; \tilde{h}_L\right) = \tilde{\mathbb{P}}_{\mathrm{md}}\left(T^*_{\mathrm{VTWL}}; h_L^*\right) = \tilde{\mathbb{P}}_{\mathrm{md}}\left(T^*_{\mathrm{VTWL}}; \tilde{h}_L - \mu_{S_1^L}\right) = \Phi\left(\frac{\tilde{h}_L - 2\mu_{S_1^L}}{\sigma_{S_1^L}}\right). \quad \text{(A.66)}$$

The proof of Theorem 3.2 is completed. □.

## A.5 Proof of Proposition 3.1

The proof of Proposition 3.1 consists of two parts. In the first part, we formulate the threshold vector, the mean vector and the covariance matrix for calculating the worst-case probability of false alarm $\overline{\mathbb{P}}_{\mathrm{fa}}$ and the worst-case probability of missed detection $\overline{\mathbb{P}}_{\mathrm{md}}$ for both the VTWL CUSUM algorithm and the FMA detection rule. The detailed calculation of the elements in vectors and matrices is given in the second part.

### A.5.1 Formulas for calculating error probabilities

The formulas for the numerical calculation of the worst-case probability of false alarm and the worst-case probability of missed detection for both VTWL CUSUM and FMA procedures are given in this part.

#### Worst-case probability of false alarm for VTWL CUSUM algorithm

It follows from (A.55) in the proof of Theorem 3.2 that the worst-case probability of false alarm of the VTWL CUSUM algorithm can be rewritten as

$$\overline{\mathbb{P}}_{\mathrm{fa}}\left(T_{\mathrm{VTWL}}; m_\alpha; h_1, h_2, \cdots, h_L\right) = 1 - \mathbb{P}_0\left(\underbrace{\bigcap_{k=L}^{L+m_\alpha-1} \bigcap_{i=k-L+1}^{k} \left\{S_i^k < h_{k-i+1}\right\}}_{E_1}\right), \quad \text{(A.67)}$$

where the event $E_1$ can be re-written as follows:

$$E_1 = \begin{pmatrix} \left\{S_1^L < h_L\right\} & \bigcap & \left\{S_2^L < h_{L-1}\right\} & \bigcap & \cdots & \bigcap & \left\{S_L^L < h_1\right\} & \bigcap \\ \left\{S_2^{L+1} < h_L\right\} & \bigcap & \left\{S_3^{L+1} < h_{L-1}\right\} & \bigcap & \cdots & \bigcap & \left\{S_{L+1}^{L+1} < h_1\right\} & \bigcap \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \left\{S_{m_\alpha}^{L+m_\alpha-1} < h_L\right\} & \bigcap & \left\{S_{m_\alpha+1}^{L+m_\alpha-1} < h_{L-1}\right\} & \bigcap & \cdots & \bigcap & \left\{S_{L+m_\alpha-1}^{L+m_\alpha-1} < h_1\right\} & \end{pmatrix}.$$

It is worth noting that the event $E_1$ is comprised of $m_\alpha$ rows and $L$ columns. By organizing the event $E_1$ in column-by-column manner, the multivariate Gaussian random variable $\mathcal{S}_1 \in \mathbb{R}^{m_\alpha L}$ with the mean vector $\mu_{\mathcal{S}_1} \in \mathbb{R}^{m_\alpha L}$ and the covariance matrix $\Sigma_{\mathcal{S}_1} \in \mathbb{R}^{m_\alpha L \times m_\alpha L}$ and the

corresponding threshold vector $h_{\mathcal{S}_1} \in \mathbb{R}^{m_\alpha L}$ are described as follows:

$$\mathcal{S}_1 = \begin{bmatrix} S_1^L \\ S_2^{L+1} \\ \vdots \\ S_{L+m_\alpha-1}^{L+m_\alpha-1} \end{bmatrix}; \quad h_{\mathcal{S}_1} = \begin{bmatrix} h_L \\ h_L \\ \vdots \\ h_1 \end{bmatrix}; \quad \mu_{\mathcal{S}_1} = \begin{bmatrix} \mathbb{E}_0\left[S_1^L\right] \\ \mathbb{E}_0\left[S_2^{L+1}\right] \\ \vdots \\ \mathbb{E}_0\left[S_{L+m_\alpha-1}^{L+m_\alpha-1}\right] \end{bmatrix},$$

$$\Sigma_{\mathcal{S}_1} = \begin{bmatrix} \operatorname{cov}\left(S_1^L, S_1^L\right) & \operatorname{cov}\left(S_1^L, S_2^{L+1}\right) & \cdots & \operatorname{cov}\left(S_1^L, S_{L+m_\alpha-1}^{L+m_\alpha-1}\right) \\ \operatorname{cov}\left(S_2^{L+1}, S_1^L\right) & \operatorname{cov}\left(S_2^{L+1}, S_2^{L+1}\right) & \cdots & \operatorname{cov}\left(S_2^{L+1}, S_{L+m_\alpha-1}^{L+m_\alpha-1}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}\left(S_{L+m_\alpha-1}^{L+m_\alpha-1}, S_1^L\right) & \operatorname{cov}\left(S_{L+m_\alpha-1}^{L+m_\alpha-1}, S_2^{L+1}\right) & \cdots & \operatorname{cov}\left(S_{L+m_\alpha-1}^{L+m_\alpha-1}, S_{L+m_\alpha-1}^{L+m_\alpha-1}\right) \end{bmatrix},$$

where the elements of the threshold vector $h_{\mathcal{S}_1}$, the mean vector $\mu_{\mathcal{S}_1}$ and the covariance matrix $\Sigma_{\mathcal{S}_1}$ can be obtained by calculating the expectation $\mathbb{E}_0\left[S_i^k\right]$ of the LLR $S_i^k$ under probability measure $\mathcal{P}_0$ and the covariance $\operatorname{cov}\left(S_{i_1}^{k_1}, S_{i_2}^{k_2}\right)$ between two LLRs $S_{i_1}^{k_1}$ and $S_{i_2}^{k_2}$. This calculation is performed in the second part of this proof. Finally, the formula for the numerical calculation of the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ of the VTWL CUSUM algorithm is given by

$$\overline{\mathbb{P}}_{\text{fa}}\left(T_{\text{VTWL}}; m_\alpha; h_1, h_2, \cdots, h_L\right) = \mathbb{P}\left(\bigcap_{j=1}^{m_\alpha L}\left\{\mathcal{S}_1(j) < h_{\mathcal{S}_1}(j)\right\}\right). \tag{A.68}$$

**Worst-case probability of false alarm for FMA detection rule**

Similar to the VTWL CUSUM algorithm, the worst-case probability of false alarm of the FMA detection rule is given by

$$\overline{\mathbb{P}}_{\text{fa}}\left(T_{\text{FMA}}; m_\alpha; \tilde{h}_L\right) = 1 - \mathbb{P}_0\left(\underbrace{\bigcap_{k=L}^{L+m_\alpha-1}\left\{S_{k-L+1}^k < \tilde{h}_L - \mu_{S_1^L}\right\}}_{E_2}\right), \tag{A.69}$$

where the event $E_2$ is defined by $m_\alpha$ Gaussian random variables $S_1^L, S_2^{L+1}, \cdots, S_{m_\alpha}^{L+m_\alpha-1}$. Let $\mathcal{S}_2 \in \mathbb{R}^{m_\alpha}$ be a multivariate Gaussian random vector with the mean vector $\mu_{\mathcal{S}_2} \in \mathbb{R}^{m_\alpha}$ and the covariance matrix $\Sigma_{\mathcal{S}_2} \in \mathbb{R}^{m_\alpha \times m_\alpha}$ and $h_{\mathcal{S}_2} \in \mathbb{R}^{m_\alpha}$ be the corresponding threshold vector. It is worth noting that the random vector $\mathcal{S}_2$ defined here coincides with the random vector $\mathcal{S}$ defined in Lemma 3.2. However, we prefer using the notation $\mathcal{S}_2$ for distinguishing from the random

vectors $\mathcal{S}_1$, $\mathcal{S}_3$ and $\mathcal{S}_4$ employed in this proof. Then, we get

$$\mathcal{S}_2 = \begin{bmatrix} S_1^L \\ S_2^{L+1} \\ \vdots \\ S_{m_\alpha}^{L+m_\alpha-1} \end{bmatrix}; \quad h_{\mathcal{S}_2} = \begin{bmatrix} \tilde{h}_L - \mu_{S_1^L} \\ \tilde{h}_L - \mu_{S_1^L} \\ \vdots \\ \tilde{h}_L - \mu_{S_1^L} \end{bmatrix}; \quad \mu_{\mathcal{S}_2} = \begin{bmatrix} \mathbb{E}_0\left[S_1^L\right] \\ \mathbb{E}_0\left[S_2^{L+1}\right] \\ \vdots \\ \mathbb{E}_0\left[S_{m_\alpha}^{L+m_\alpha-1}\right] \end{bmatrix},$$

$$\Sigma_{\mathcal{S}_2} = \begin{bmatrix} \operatorname{cov}\left(S_1^L, S_1^L\right) & \operatorname{cov}\left(S_1^L, S_2^{L+1}\right) & \cdots & \operatorname{cov}\left(S_1^L, S_{m_\alpha}^{L+m_\alpha-1}\right) \\ \operatorname{cov}\left(S_2^{L+1}, S_1^L\right) & \operatorname{cov}\left(S_2^{L+1}, S_2^{L+1}\right) & \cdots & \operatorname{cov}\left(S_2^{L+1}, S_{m_\alpha}^{L+m_\alpha-1}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}\left(S_{m_\alpha}^{L+m_\alpha-1}, S_1^L\right) & \operatorname{cov}\left(S_{m_\alpha}^{L+m_\alpha-1}, S_2^{L+1}\right) & \cdots & \operatorname{cov}\left(S_{m_\alpha}^{L+m_\alpha-1}, S_{m_\alpha}^{L+m_\alpha-1}\right) \end{bmatrix},$$

where the elements of the threshold vector $h_{\mathcal{S}_2}$, the mean vector $\mu_{\mathcal{S}_2}$ and the covariance matrix $\Sigma_{\mathcal{S}_2}$ are elaborated in the second part of this proof. Finally, the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}$ of the FMA test is calculated numerically as

$$\overline{\mathbb{P}}_{\text{fa}}\left(T_{\text{FMA}}; m_\alpha; \tilde{h}_L\right) = \mathbb{P}\left(\bigcap_{j=1}^{m_\alpha}\left\{\mathcal{S}_2\left(j\right) < h_{\mathcal{S}_2}\left(j\right)\right\}\right). \tag{A.70}$$

**Worst-case probability of missed detection for VTWL CUSUM algorithm**

The worst-case probability of missed detection of the VTWL CUSUM algorithm can be described as

$$\begin{aligned}
\overline{\mathbb{P}}_{\text{md}}\left(T_{\text{VTWL}}; h_1, h_2, \cdots, h_L\right) &= \sup_{k_0 \geq L} \frac{\mathbb{P}_{k_0}\left(T_{\text{VTWL}}\left(h_1, h_2, \cdots, h_L\right) \geq k_0 + L\right)}{\mathbb{P}_{k_0}\left(T_{\text{VTWL}}\left(h_1, h_2, \cdots, h_L\right) \geq k_0\right)} \\
&= \sup_{k_0 \geq L} \frac{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0+L-1}\left\{\max_{k-L+1 \leq i \leq k}\left(S_i^k - h_{k-i+1}\right) < 0\right\}\right)}{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0-1}\left\{\max_{k-L+1 \leq i \leq k}\left(S_i^k - h_{k-i+1}\right) < 0\right\}\right)} \\
&= \sup_{k_0 \geq L} \frac{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0+L-1}\bigcap_{i=k-L+1}^{k}\left\{S_i^k < h_{k-i+1}\right\}\right)}{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0+L-1}\bigcap_{i=k-L+1}^{k}\left\{S_i^k < h_{k-i+1}\right\}\right)}.
\end{aligned} \tag{A.71}$$

Let us define the following function $F_{k_0}\left(a; b; h_1, h_2, \cdots, h_L\right)$ with $b \geq a \geq L$, for $a, b \in \mathbb{N}^+$ as follows:

$$\begin{aligned}
F_{k_0}\left(a; b; h_1, h_2, \cdots, h_L\right) &= \mathbb{P}_{k_0}\left(\bigcap_{k=a}^{b}\left\{\max_{k-L+1 \leq i \leq k}\left(S_i^k - h_{k-i+1}\right) < 0\right\}\right) \\
&= \mathbb{P}_{k_0}\left(\underbrace{\bigcap_{k=a}^{b}\bigcap_{i=k-L+1}^{k}\left\{S_i^k < h_{k-i+1}\right\}}_{E_3}\right),
\end{aligned} \tag{A.72}$$

where the event $E_3$, which is comprised of $b-a+1$ rows and $L$ columns, can be re-written as follows:

$$E_3 = \begin{pmatrix} \left\{S^a_{a-L+1} < h_L\right\} & \bigcap & \left\{S^a_{a-L+2} < h_{L-1}\right\} & \bigcap & \cdots & \bigcap & \{S^a_a < h_1\} & \bigcap \\ \left\{S^{a+1}_{a-L+2} < h_L\right\} & \bigcap & \left\{S^{a+1}_{a-L+3} < h_{L-1}\right\} & \bigcap & \cdots & \bigcap & \left\{S^{a+1}_{a+1} < h_1\right\} & \bigcap \\ \vdots & & \vdots & & \ddots & \vdots & \vdots & \vdots \\ \left\{S^b_{b-L+1} < h_L\right\} & \bigcap & \left\{S^b_{b-L+2} < h_{L-1}\right\} & \bigcap & \cdots & \bigcap & \left\{S^b_b < h_1\right\} & \end{pmatrix}.$$

The multivariate Gaussian random variable $\mathcal{S}_3 \in \mathbb{R}^{(b-a+1)L}$ with the mean vector $\mu_{\mathcal{S}_3} \in \mathbb{R}^{(b-a+1)L}$ and the covariance matrix $\Sigma_{\mathcal{S}_3} \in \mathbb{R}^{(b-a+1)L\times(b-a+1)L}$ and the corresponding threshold vector $h_{\mathcal{S}_3}$ can be described as

$$\mathcal{S}_3 = \begin{bmatrix} S^a_{a-L+1} \\ S^{a+1}_{a-L+2} \\ \vdots \\ S^b_b \end{bmatrix}; \quad h_{\mathcal{S}_3} = \begin{bmatrix} h_L \\ h_L \\ \vdots \\ h_1 \end{bmatrix}; \quad \mu_{\mathcal{S}_3} = \begin{bmatrix} \mathbb{E}_{k_0}\left[S^a_{a-L+1}\right] \\ \mathbb{E}_{k_0}\left[S^{a+1}_{a-L+2}\right] \\ \vdots \\ \mathbb{E}_{k_0}\left[S^b_b\right] \end{bmatrix},$$

$$\Sigma_{\mathcal{S}_3} = \begin{bmatrix} \operatorname{cov}\left(S^a_{a-L+1}, S^a_{a-L+1}\right) & \operatorname{cov}\left(S^a_{a-L+1}, S^{a+1}_{a-L+2}\right) & \cdots & \operatorname{cov}\left(S^a_{a-L+1}, S^b_b\right) \\ \operatorname{cov}\left(S^{a+1}_{a-L+2}, S^a_{a-L+1}\right) & \operatorname{cov}\left(S^{a+1}_{a-L+2}, S^{a+1}_{a-L+2}\right) & \cdots & \operatorname{cov}\left(S^{a+2}_{a-L+2}, S^b_b\right) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}\left(S^b_b, S^a_{a-L+1}\right) & \operatorname{cov}\left(S^b_b, S^{a+1}_{a-L+2}\right) & \cdots & \operatorname{cov}\left(S^b_b, S^b_b\right) \end{bmatrix}.$$

The function $F_{k_0}(a; b; h_1, h_2, \cdots, h_L)$ can be evaluated numerically as

$$F_{k_0}(a; b; h_1, h_2, \cdots, h_L) = \mathbb{P}_{k_0}\left(\bigcap_{j=1}^{(b-a+1)L} \left\{\mathcal{S}_3(j) < h_{\mathcal{S}_3}(j)\right\}\right). \tag{A.73}$$

The worst-case probability of missed detection for the VTWL CUSUM algorithm is then calculated by utilizing the function $F_{k_0}(\cdot)$ as follows:

$$\overline{\mathbb{P}}_{\mathrm{md}}(T_{\mathrm{VTWL}}; h_1, h_2, \cdots, h_L) = \sup_{k_0 \geq L} \frac{F_{k_0}(L; k_0+L-1; h_1, h_2, \cdots, h_L)}{F_{k_0}(L; k_0-1; h_1, h_2, \cdots, h_L)}, \tag{A.74}$$

where $F_{k_0}(L; k_0-1; h_1, h_2, \cdots, h_L) \triangleq 1$ for $k_0 = L$.

**Worst-case probability of missed detection for FMA detection rule**

Similar to the VTWL CUSUM algorithm, the worst-case probability of missed detection for the FMA detection rule is derived as

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{FMA}}; \tilde{h}_L\right) = \sup_{k_0 \geq L} \frac{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0+L-1} \left\{S^k_{k-L+1} < \tilde{h}_L - \mu_{S^L_1}\right\}\right)}{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0-1} \left\{S^k_{k-L+1} < \tilde{h}_L - \mu_{S^L_1}\right\}\right)}, \tag{A.75}$$

since the threshold $\tilde{h}_L$ of the FMA test is related to the optimal threshold of the VTWL CUSUM test by $\tilde{h}_L = h_L^* + \mu_{S_1^L}$. Let us define also the function $\tilde{F}_{k_0}\left(a; b; \tilde{h}_L - \mu_{S_1^L}\right)$ can be re-written as

$$\tilde{F}_{k_0}\left(a; b; \tilde{h}_L - \mu_{S_1^L}\right) = \mathbb{P}_{k_0}\left(\bigcap_{k=a}^{b}\left\{S_{k-L+1}^k < \tilde{h}_L - \mu_{S_1^L}\right\}\right). \tag{A.76}$$

The multivariate Gaussian random variable $\mathcal{S}_4 \in \mathbb{R}^{(b-a+1)}$, the mean vector $\mu_{\mathcal{S}_4} \in \mathbb{R}^{(b-a+1)}$, the covariance matrix $\Sigma_{\mathcal{S}_4} \in \mathbb{R}^{(b-a+1)\times(b-a+1)}$ and the threshold vector $h_{\mathcal{S}_4} \in \mathbb{R}^{(b-a+1)}$ are defined as

$$\mathcal{S}_4 = \begin{bmatrix} S_{a-L+1}^a \\ S_{a-L+2}^{a+1} \\ \vdots \\ S_{b-L+1}^b \end{bmatrix}; \quad \mu_{\mathcal{S}_4} = \begin{bmatrix} \mathbb{E}_{k_0}\left[S_{a-L+1}^a\right] \\ \mathbb{E}_{k_0}\left[S_{a-L+2}^{a+1}\right] \\ \vdots \\ \mathbb{E}_{k_0}\left[S_{b-L+1}^b\right] \end{bmatrix}; \quad h_{\mathcal{S}_4} = \begin{bmatrix} \tilde{h}_L - \mu_{S_1^L} \\ \tilde{h}_L - \mu_{S_1^L} \\ \vdots \\ \tilde{h}_L - \mu_{S_1^L} \end{bmatrix},$$

$$\Sigma_{\mathcal{S}_4} = \begin{bmatrix} \operatorname{cov}\left(S_{a-L+1}^a, S_{a-L+1}^a\right) & \operatorname{cov}\left(S_{a-L+1}^a, S_{a-L+2}^{a+1}\right) & \cdots & \operatorname{cov}\left(S_{a-L+1}^a, S_{b-L+1}^b\right) \\ \operatorname{cov}\left(S_{a-L+2}^{a+1}, S_{a-L+1}^a\right) & \operatorname{cov}\left(S_{a-L+2}^{a+1}, S_{a-L+2}^{a+1}\right) & \cdots & \operatorname{cov}\left(S_{a-L+2}^{a+1}, S_{b-L+1}^b\right) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}\left(S_{b-L+1}^b, S_{a-L+1}^a\right) & \operatorname{cov}\left(S_{b-L+1}^b, S_{a-L+2}^{a+1}\right) & \cdots & \operatorname{cov}\left(S_{b-L+1}^b, S_{b-L+1}^b\right) \end{bmatrix}.$$

The function $\tilde{F}_{k_0}\left(a; b; \tilde{h}_L\right)$ is calculated numerically as

$$\tilde{F}_{k_0}\left(a; b; \tilde{h}_L - \mu_{S_1^L}\right) = \mathbb{P}_{k_0}\left(\bigcap_{j=1}^{(b-a+1)}\left\{\mathcal{S}_4(j) < h_{\mathcal{S}_4}(j)\right\}\right). \tag{A.77}$$

Finally, the worst-case probability of missed detection for the FMA detection rule is calculated numerically as

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{FMA}}; \tilde{h}_L\right) = \sup_{k_0 \geq L} \frac{\tilde{F}_{k_0}\left(L; k_0 + L - 1; \tilde{h}_L - \mu_{S_1^L}\right)}{\tilde{F}_{k_0}\left(L; k_0 - 1; \tilde{h}_L - \mu_{S_1^L}\right)}, \tag{A.78}$$

where $\tilde{F}_{k_0}\left(L; k_0 - 1; \tilde{h}_L - \mu_{S_1^L}\right) \triangleq 1$ for $k_0 = L$.

### A.5.2 Calculation of expectations and covariances

In this part, we calculate the mathematical expectations $\mathbb{E}_0\left[S_i^k\right]$ and $\mathbb{E}_{k_0}\left[S_i^k\right]$ of the LLR $S_i^k$ under the pre-change probability measure $\mathcal{P}_0$ and the probability measure $\mathcal{P}_{k_0}$, respectively. We compute also the covariance $\operatorname{cov}\left(S_{i_1}^{k_1}, S_{i_2}^{k_2}\right)$ between two LLRs $S_{i_1}^{k_1}$ and $S_{i_2}^{k_2}$, for $k_1 - L + 1 \leq i_1 \leq k_1$ and $k_2 - L + 1 \leq i_2 \leq k_2$.

#### Calculation of mathematical expectations

In this subsection, we calculate the mathematical expectations $\mathbb{E}_0\left[S_i^k\right]$ and $\mathbb{E}_{k_0}\left[S_i^k\right]$ for both steady-state Kalman filter approach and the fixed-size parity space approach. By replacing the

residual vector $r_{k-L+1}^{k}$ in the unified statistical model (3.25) into the LLR $S_i^k$ defined in (3.34), we get

$$S_i^k = \left[\phi_{k-L+1}^k(i)\right]^T \left[\Sigma^{-1}\right] \left[\xi_{k-L+1}^k + \phi_{k-L+1}^k(k_0) - \frac{1}{2}\phi_{k-L+1}^k(i)\right]. \tag{A.79}$$

Under the pre-change probability measure $\mathcal{P}_0$, the transient profiles $\phi_{k-L+1}^k(k_0) = 0$, leading to

$$\mathbb{E}_0\left[S_i^k\right] = -\frac{1}{2}\left[\phi_{k-L+1}^k(i)\right]^T \left[\Sigma^{-1}\right] \left[\phi_{k-L+1}^k(i)\right]. \tag{A.80}$$

Under the probability measure $\mathcal{P}_{k_0}$, the mathematical expectation of the LLR $S_i^k$ is calculated by

$$\mathbb{E}_{k_0}\left[S_i^k\right] = \left[\phi_{k-L+1}^k(i)\right]^T \left[\Sigma^{-1}\right] \left[\phi_{k-L+1}^k(k_0) - \frac{1}{2}\phi_{k-L+1}^k(i)\right]. \tag{A.81}$$

**Calculation of covariance**

We calculate in the following the covariance between two Gaussian random vectors $S_{i_1}^{k_1}$ and $S_{i_2}^{k_2}$, for both the steady-state Kalman filter approach and the fixed-size parity space approach. It follows from (A.79) that the LLR $S_i^k$ can be described as

$$S_i^k = \left[\phi_{k-L+1}^k(i)\right]^T \left[\Sigma^{-1}\right] \left[\xi_{k-L+1}^k\right] + \mathbb{E}_{k_0}\left[S_i^k\right]. \tag{A.82}$$

**Steady-state Kalman filter approach** By this approach, the LLRs $S_{i_1}^{k_1}$ and $S_{i_2}^{k_2}$ can be described as

$$S_{i_1}^{k_1} = \left[\psi_{k_1-L+1}^{k_1}(i_1)\right]^T \left[\Sigma_\varrho^{-1}\right] \left[\varrho_{k_1-L+1}^{k_1}\right] + \mathbb{E}_{k_0}\left[S_{i_1}^{k_1}\right], \tag{A.83}$$

$$S_{i_2}^{k_2} = \left[\psi_{k_2-L+1}^{k_2}(i_2)\right]^T \left[\Sigma_\varrho^{-1}\right] \left[\varrho_{k_2-L+1}^{k_2}\right] + \mathbb{E}_{k_0}\left[S_{i_2}^{k_2}\right]. \tag{A.84}$$

Hence, the covariance between these random variables is calculated as

$$\operatorname{cov}\left(S_{i_1}^{k_1}, S_{i_2}^{k_2}\right) = \left[\psi_{k_1-L+1}^{k_1}(i_1)\right]^T \left[\Sigma_\varrho^{-1}\right] \left[\mathcal{S}_\varrho\right] \left[\Sigma_\varrho^{-1}\right] \left[\psi_{k_2-L+1}^{k_2}(i_2)\right], \tag{A.85}$$

where $\mathcal{S}_\varrho \in \mathbb{R}^{Lp\times Lp}$ is the covariance matrix between the random vectors $\varrho_{k_1-L+1}^{k_1}$ and $\varrho_{k_2-L+1}^{k_2}$, which is calculated as

$$\mathcal{S}_\varrho = \mathbb{E}_0\left[\begin{pmatrix} \varrho_{k_1-L+1} \\ \vdots \\ \varrho_{k_1} \end{pmatrix} \begin{pmatrix} \varrho_{k_2-L+1}^T & \cdots & \varrho_{k_2}^T \end{pmatrix}\right], \tag{A.86}$$

where $\mathbb{E}_0\left[\varrho_{t_1}\varrho_{t_2}^T\right] = CP_\infty C^T + R$ if $t_1 = t_2$ and $\mathbb{E}_0\left[\varrho_{t_1}\varrho_{t_2}^T\right] = 0$ otherwise.

**Fixed-size parity space approach** By this approach, the transient profiles $\phi_{k-L+1}^k(i) = \varphi_{k-L+1}^k(i)$ and the random noises $\varsigma_{k-L+1}^k = \mathcal{W}\left(\mathcal{H}w_{k-L+1}^k + v_{k-L+1}^k\right)$. The LLRs $S_{i_1}^{k_1}$ and $S_{i_2}^{k_2}$ can be decomposed as

$$S_{i_1}^{k_1} = \left[\varphi_{k_1-L+1}^{k_1}(i_1)\right]^T \left[\Sigma_\varsigma^{-1}\right] \left[\varsigma_{k_1-L+1}^{k_1}\right] + \mathbb{E}_{k_0}\left[S_{i_1}^{k_1}\right], \tag{A.87}$$

$$S_{i_2}^{k_2} = \left[\varphi_{k_2-L+1}^{k_2}(i_2)\right]^T \left[\Sigma_\varsigma^{-1}\right] \left[\varsigma_{k_2-L+1}^{k_2}\right] + \mathbb{E}_{k_0}\left[S_{i_2}^{k_2}\right]. \tag{A.88}$$

Hence, the covariance between these random variables is calculated as

$$\operatorname{cov}\left(S_{i_1}^{k_1}, S_{i_2}^{k_2}\right) = \left[\varphi_{k_1-L+1}^{k_1}(i_1)\right]^T \left[\Sigma_\varsigma^{-1}\right] [\mathcal{S}_\varsigma] \left[\Sigma_\varsigma^{-1}\right] \left[\varphi_{k_2-L+1}^{k_2}(i_2)\right], \tag{A.89}$$

where the covariance matrix $\mathcal{S}_\varsigma \in \mathbb{R}^{Lp\times Lp}$ between the random vectors $\varsigma_{k_1-L+1}^{k_1}$ and $\varsigma_{k_2-L+1}^{k_1}$ is calculated as

$$\mathcal{S}_\varsigma = \mathbb{E}_0\left[\left(\mathcal{W}\mathcal{H}w_{k_1-L+1}^{k_1} + \mathcal{W}v_{k_1-L+1}^{k_1}\right)\left(\mathcal{W}\mathcal{H}w_{k_2-L+1}^{k_2} + \mathcal{W}v_{k_2-L+1}^{k_2}\right)^T\right] = \mathcal{W}\left(\mathcal{H}\mathcal{S}_w\mathcal{H}^T + \mathcal{S}_v\right)\mathcal{W}^T, \tag{A.90}$$

where $\mathcal{S}_w \in \mathbb{R}^{Ln\times Ln}$ is the covariance matrix between two random vectors $w_{k_1-L+1}^{k_1}$ and $w_{k_2-L+1}^{k_2}$, and $\mathcal{S}_v \in \mathbb{R}^{Lp\times Lp}$ is the covariance matrix between two random vectors $v_{k_1-L+1}^{k_1}$ and $v_{k_2-L+1}^{k_2}$, which are calculated as follows:

$$\mathcal{S}_w = \mathbb{E}_0\left[\begin{pmatrix} w_{k_1-L+1} \\ \vdots \\ w_{k_1} \end{pmatrix}\begin{pmatrix} w_{k_2-L+1}^T & \cdots & w_{k_2}^T \end{pmatrix}\right], \tag{A.91}$$

$$\mathcal{S}_v = \mathbb{E}_0\left[\begin{pmatrix} v_{k_1-L+1} \\ \vdots \\ v_{k_1} \end{pmatrix}\begin{pmatrix} v_{k_2-L+1}^T & \cdots & v_{k_2}^T \end{pmatrix}\right], \tag{A.92}$$

where $\mathbb{E}_0\left[w_{t_1}w_{t_2}^T\right] = Q$ and $\mathbb{E}_0\left[v_{t_1}v_{t_2}^T\right] = R$ if $t_1 = t_2$ and $\mathbb{E}_0\left[w_{t_1}w_{t_2}^T\right] = 0$ and $\mathbb{E}_0\left[v_{t_1}v_{t_2}^T\right] = 0$ otherwise.

**Remark A.1.** *Let us discuss now the positive definiteness of the covariance matrices $\Sigma_{\mathcal{S}_1}$, $\Sigma_{\mathcal{S}_2}$, $\Sigma_{\mathcal{S}_3}$ and $\Sigma_{\mathcal{S}_4}$. First of all, the random vector $\mathcal{S}_2 \equiv \mathcal{S}$, where the last vector is defined in Lemma 3.2. It follows from Lemma 3.2 that the covariance matrices $\Sigma_{\mathcal{S}_2} \equiv \Sigma_{\mathcal{S}}$, which correspond to the FMA test* (3.41)*, are positive-definite if Assumption 3.1 is satisfied. Second, the covariance matrix $\Sigma_{\mathcal{S}_4}$ is also positive-definite if Assumption 3.1 is satisfied. The proof of this fact is completely analogous to that of Lemma 3.2 in Appendix A.3. Finally, the covariance matrices $\Sigma_{\mathcal{S}_1}$ and $\Sigma_{\mathcal{S}_3}$, which correspond to the VTWL CUSUM test* (3.33)*, are positive-definite in some scenarios. Nevertheless, there are also scenarios where the determinants of $\Sigma_{\mathcal{S}_1}$ and $\Sigma_{\mathcal{S}_3}$ are close to zero, especially with large value of $L$ and $m_\alpha$. Therefore, it is necessary to verify the positive definiteness before executing the numerical computation. The following heuristic solution is proposed in such cases: to use the matrix $\Sigma_{\mathcal{S}_1} + \delta\mathcal{I}$ (resp. $\Sigma_{\mathcal{S}_3} + \delta\mathcal{I}$) instead of covariance matrix $\Sigma_{\mathcal{S}_1}$(resp. $\Sigma_{\mathcal{S}_3}$), where $\mathcal{I}$ is the identity matrix of appropriate size and $\delta > 0$ is a small quantity.*

The proof of Proposition 3.1 is completed. □.

## A.6 Sensibility analysis of FMA test

In this section, we re-calculate the mathematical expectations $\mathbb{E}_0\left[S_i^k\right]$ and $\mathbb{E}_{k_0}\left[S_i^k\right]$ and the covariance $\operatorname{cov}\left(S_{i_1}^{k_1}, S_{i_2}^{k_2}\right)$ when the true values of operational parameters are different from their putative values (i.e., $\overline{L} \neq L$, $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L \neq \theta_1, \theta_2, \cdots, \theta_L$, $\overline{Q} \neq Q$ and $\overline{R} \neq R$).

### A.6.1 Calculation of true mathematical expectations

Let $\overline{\phi}^k_{k-L+1}(k_0)$ be the vector of true transient profiles formulated in the same manner as the putative transient profiles $\phi^k_{k-L+1}(k_0)$ in (3.25), with the putative parameters $L$ and $\theta_1, \theta_2, \cdots, \theta_L$ replaced by the true parameters $\overline{L}$, $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$, respectively. It is worth noting that the vector $\overline{\phi}^k_{k-L+1}(k_0)$ depends also on either the steady-state Kalman filter or the fixed-size parity space is employed. The mathematical expectations $\mathbb{E}_0\left[S^k_i\right]$ and $\mathbb{E}_{k_0}\left[S^k_i\right]$ can be re-calculated, respectively, as

$$\mathbb{E}_0\left[S^k_i\right] = -\frac{1}{2}\left[\phi^k_{k-L+1}(i)\right]^T\left[\Sigma^{-1}\right]\left[\phi^k_{k-L+1}(i)\right], \tag{A.93}$$

$$\mathbb{E}_{k_0}\left[S^k_i\right] = \left[\phi^k_{k-L+1}(i)\right]^T\left[\Sigma^{-1}\right]\left[\overline{\phi}^k_{k-L+1}(k_0) - \frac{1}{2}\phi^k_{k-L+1}(i)\right]. \tag{A.94}$$

It can be noticed that under pre-change mode (i.e., $k_0 \to \infty$), the mathematical expectation $\mathbb{E}_0\left[S^k_i\right]$ given in (A.93) remained unchanged in comparison with $\mathbb{E}_0\left[S^k_i\right]$ calculated in (A.80) when the true parameters are the same as the putative parameters. The quantity $\mathbb{E}_{k_0}\left[S^k_i\right]$, on the other hand, depends on the true transient profiles $\overline{\phi}^k_{k-L+1}(k_0)$ which are calculated from true parameters $\overline{L}$ and $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$.

### A.6.2 Calculation of true covariance

The covariance $\mathrm{cov}\left(S^{k_1}_{i_1}, S^{k_2}_{i_2}\right)$ between two LLRs $S^{k_1}_{i_1}$ and $S^{k_2}_{i_2}$ can be calculated in the same manner as in Appendix A.5. More precisely, the covariance $\mathrm{cov}\left(S^{k_1}_{i_1}, S^{k_2}_{i_2}\right)$ can be computed by (A.85) for the steady-state Kalman filter approach and by (A.89) for the fixed-size parity space approach. Since the true values of random noises are different from their putative values, the covariance matrices $\mathcal{S}_\varrho$ defined in (A.86) and $\mathcal{S}_\varsigma$ defined in (A.90) must be recalculated.

The covariance matrix $\mathcal{S}_\varsigma$ can be calculated by (A.90) in terms of $\mathcal{S}_w$ in (A.91) and $\mathcal{S}_v$ in (A.92), respectively. The elements of $\mathcal{S}_w$ and $\mathcal{S}_v$ are revised by $\mathbb{E}\left[w_{t_1} w^T_{t_2}\right] = \overline{Q}$ and $\mathbb{E}\left[v_{t_1} v^T_{t_2}\right] = \overline{R}$ if $t_1 = t_2$ and $\mathbb{E}\left[w_{t_1} w^T_{t_2}\right] = 0$ and $\mathbb{E}\left[v_{t_1} v^T_{t_2}\right] = 0$ otherwise.

The covariance matrix $\mathcal{S}_\varrho$ can be calculated by (A.86), where its elements $\mathbb{E}\left[\varrho_{t_1}\varrho^T_{t_2}\right] = \mathbb{E}_0\left[r_{t_1} r^T_{t_2}\right]$ need to be re-computed. In such situations that $\overline{Q} \neq Q$ and/or $\overline{R} \neq R$, the Kalman filter is no longer optimal and the residuals are no longer independent. Hence, it is required to calculate $\mathbb{E}_0\left[r_{t_1} r^T_{t_2}\right]$, for $k_1 - L + 1 \leq t_1 \leq k_1$ and $k_2 - L + 1 \leq t_2 \leq k_2$.

The calculation of $\mathbb{E}_0\left[r_k r^T_{k+l}\right]$, for $l \geq 0$, is given Algorithm 2. The idea behind the Algorithm 2 is described in Appendix A.1.

## A.7 Proof of Theorem 3.4

In the following, the optimization problem is formulated and solved for the VTWL GLR test defined in (3.51). However, similar results can be obtained for the VTWL WLR test defined in (3.56). The proof consists of two parts. The optimization problem is formulated and solved in the first part. It is shown in the second part that the optimal choice of thresholds leads to FMA GLR detection rule.

**Algorithm 2** Calculation of the covariance $\mathrm{cov}\left(r_{k+l}, r_k\right)$ between two innovations $r_{k+l}$ and $r_k$ generated by the steady-state Kalman filter when the true noise covariances are different from their putative values (i.e., $\overline{Q} \neq Q$ and/or $\overline{R} \neq R$).

1. Initialization: $\overline{P}_{1|0} = P_\infty$ and $K = AK_\infty$, where $K_\infty$ and $P_\infty$ are given in (3.8)–(3.9).
2. Calculation of the real covariance $\overline{P}_{k+1|k}$:
$$\overline{P}_{k+1|k} = (A - KC)\,\overline{P}_{k|k-1}\,(A - KC)^T + \overline{Q} + K\overline{R}K^T.$$
3. If $(l = 0)$ then
$$\mathbb{E}_0\left[r_k r_k^T\right] = C\overline{P}_{k|k-1}C^T + \overline{R}, \tag{A.95}$$
4. Else if $(l \geq 1)$ then
$$\mathbb{E}_0\left[r_{k+l} r_k^T\right] = C\mathbb{E}_0\left[e_{k+l} r_k^T\right], \tag{A.96}$$
where the matrix $\mathbb{E}_0\left[e_{k+l} r_k^T\right]$ is computed recursively as
$$\mathbb{E}_0\left[e_{k+l} r_k^T\right] = (A - KC)\,\mathbb{E}_0\left[e_{k+l-1} r_k^T\right], \tag{A.97}$$
with initial value (i.e., $l = 1$)
$$\mathbb{E}_0\left[e_{k+1} r_k^T\right] = A\overline{P}_{k|k-1}C^T - K\left(C\overline{P}_{k|k-1}C^T + \overline{R}\right). \tag{A.98}$$

### A.7.1 Proof of part 1

Since we wish to minimize the upper bound $\hat{\mathbb{P}}^*_{\mathrm{md}}\left(h_L\right)$ on the worst-case probability of missed detection $\overline{\mathbb{P}}_{\mathrm{md}}\left(\hat{T}_{\mathrm{GLR}}\right)$ subject to an acceptable level $\alpha \in (0,1)$ on the worst-case probability of false alarm $\overline{\mathbb{P}}_{\mathrm{fa}}\left(\hat{T}_{\mathrm{GLR}}\right)$, the optimization problem can be defined as

$$\begin{cases} \inf_{h_1,h_2,\cdots,h_L} & \hat{\mathbb{P}}^*_{\mathrm{md}}\left(h_L\right) \\ \text{subject to} & \overline{\mathbb{P}}_{\mathrm{fa}}\left(\hat{T}_{\mathrm{GLR}}; m_\alpha; h_1, h_2, \cdots, h_L\right) \leq \alpha \end{cases}, \tag{A.99}$$

where the worst-case probability of false alarm $\overline{\mathbb{P}}_{\mathrm{fa}}\left(\hat{T}_{\mathrm{GLR}}; m_\alpha; h_1, h_2, \cdots, h_L\right)$ is calculated by

$$\overline{\mathbb{P}}_{\mathrm{fa}}\left(\hat{T}_{\mathrm{GLR}}; m_\alpha; h_1, h_2, \cdots, h_L\right) = 1 - \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \bigcap_{i=k-L+1}^{k} \left\{\hat{S}_i^k < h_{k-i+1}\right\}\right). \tag{A.100}$$

Seeking for simplifying the proof, let us define two functions $\hat{F}_0\left(h_1, h_2, \cdots, h_L\right)$ and $\hat{G}_0\left(h_L\right)$ as follows:

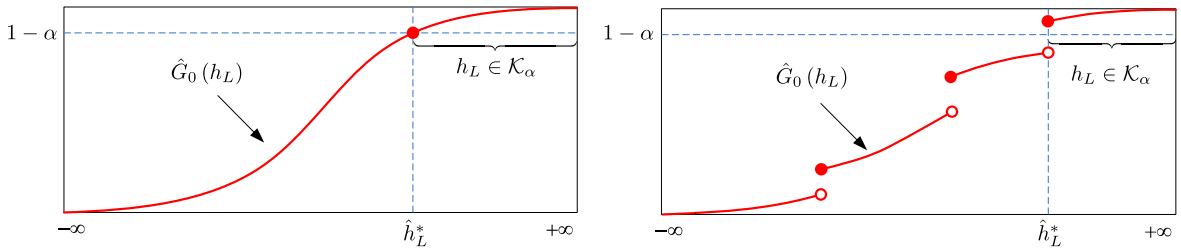
$$\hat{F}_0\left(h_1, h_2, \cdots, h_L\right) \triangleq \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \bigcap_{i=k-L+1}^{k} \left\{\hat{S}_i^k < h_{k-i+1}\right\}\right), \tag{A.101}$$

$$\hat{G}_0\left(h_L\right) \triangleq \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \left\{\hat{S}_{k-L+1}^k < h_L\right\}\right), \tag{A.102}$$

where the function $\hat{F}_0(h_1, h_2, \cdots, h_L)$ is monotonically non-decreasing w.r.t. each threshold and the function $\hat{G}_0(h_L)$ is also monotonically w.r.t. threshold $h_L$. It is clear that $\hat{F}_0(+\infty, \cdots, +\infty, h_L) = \hat{G}_0(h_L)$. The optimization problem (A.99) is reduced to

$$\begin{cases} \inf_{h_1,h_2,\cdots,h_L} & \hat{\mathbb{P}}^*_{\text{md}}(h_L) \\ \text{subject to} & \hat{F}_0(h_1, h_2, \cdots, h_L) \geq 1-\alpha \end{cases}. \tag{A.103}$$

Let $\mathcal{K}_\alpha$ be the set of real numbers satisfying $\hat{G}_0(h_L) \geq 1-\alpha$ for a given value $\alpha \in (0,1)$. It follows from the property of the probability measure that $\hat{G}_0(h_L)$ is a right-continuous function and that $\lim_{h_L \to -\infty} \hat{G}_0(h_L) = 0$ and $\lim_{h_L \to +\infty} \hat{G}_0(h_L) = 1$. For these reasons, the set $\mathcal{K}_\alpha$ is non-null. Let $\hat{h}^*_L = \min\{h_L : h_L \in \mathcal{K}_\alpha\}$ be the minimum value of $h_L$ in the set $\mathcal{K}_\alpha$.




(a) Function $\hat{G}_0(h_L)$ is continuous.

(b) Function $\hat{G}_0(h_L)$ is right-continuous.

Figure A.1 – Function $\hat{G}_0(h_L)$ and optimal solution $\hat{h}^*_L$ in two scenarios.

See figure A.1 for the demonstration in two scenarios: (a) the function $h_L \mapsto \hat{G}_0(h_L)$ is continuous from $\mathbb{R}$ to $\mathbb{R}$ and (b) the function $h_L \mapsto \hat{G}_0(h_L)$ is right-continuous from $\mathbb{R}$ to $\mathbb{R}$. It is clear that if the function $h_L \mapsto \hat{G}_0(h_L)$ is continuous, then the threshold $\hat{h}^*_L$ is the solution of the equation

$$\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \left\{\hat{S}^k_{k-L+1} < h^*_L\right\}\right) = 1-\alpha. \tag{A.104}$$

In the following, we show that the thresholds $h^*_1, \cdots, h^*_{L-1} \to +\infty$ and $\hat{h}^*_L$ are the solution to the optimization problem (A.103). Let us suppose that a set of thresholds $h_1, h_2, \cdots, h_L$ satisfying the constraint

$$\hat{F}_0(h_1, h_2, \cdots, h_L) \geq 1-\alpha, \tag{A.105}$$

defines any alternative solution of the optimization problem (A.103). The goal is to show that $\hat{\mathbb{P}}^*_{md}(h_L) \geq \hat{\mathbb{P}}^*_{md}\left(\hat{h}^*_L\right)$. It is worth noting that the function $\hat{F}_0(h_1, h_2, \cdots, h_L)$ is monotonically non-decreasing w.r.t. each threshold $h_1, h_2, \cdots, h_L$. Hence, we get

$$\hat{F}_0(+\infty, \cdots, +\infty, h_L) \geq \hat{F}_0(h_1, \cdots, h_{L-1}, h_L). \tag{A.106}$$

Putting together (A.105) and (A.106), we obtain that

$$\hat{F}_0(+\infty, \cdots, +\infty, h_L) = \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1} \left\{\hat{S}^k_{k-L+1} < h_L\right\}\right) \geq 1-\alpha, \tag{A.107}$$

leading to $h_L \geq \hat{h}^*_L$ since $\hat{h}^*_L$ is the minimum value in the class $\mathcal{K}_\alpha$. Moreover, the objective function $h_L \mapsto \hat{\mathbb{P}}^*_{\text{md}}(h_L)$ is monotonically non-decreasing. Therefore, $\hat{\mathbb{P}}^*_{\text{md}}(h_L) \geq \hat{\mathbb{P}}^*_{\text{md}}\left(\hat{h}^*_L\right)$, thus proving that $h^*_1, \cdots, h^*_{L-1} \to +\infty$ and $h^*_L = \min\{h_L : h_L \in \mathcal{K}_\alpha\}$ are the optimal thresholds which minimize the upper bound $\hat{\mathbb{P}}^*_{\text{md}}(h_L)$ on the worst-case probability of missed detection $\overline{\mathbb{P}}_{\text{md}}\left(\hat{T}_{\text{GLR}}; h_1, h_2, \cdots, h_L\right)$ of the VTWL GLR algorithm defined in (3.51).

### A.7.2 Proof of part 2

The VTWL GLR algorithm with optimal thresholds $\hat{h}_1^*, \hat{h}_2^*, \cdots, \hat{h}_L^*$ can be described as

$$\begin{aligned}\hat{T}_{\mathrm{GLR}}^* &= \inf\left\{k \geq L : \max_{k-L+1\leq i\leq k}\left(\hat{S}_i^k - \hat{h}_{k-i+1}^*\right) \geq 0\right\} \\ &= \inf\left\{k \geq L : \bigcup_{i=k-L+1}^{k}\left\{\hat{S}_i^k \geq \hat{h}_{k-i+1}^*\right\}\right\} \\ &= \inf\left\{k \geq L : \hat{S}_{k-L+1}^k \geq \hat{h}_L^*\right\}, \end{aligned} \tag{A.108}$$

since the optimal thresholds $\hat{h}_1^*, \cdots, \hat{h}_{L-1}^* \to +\infty$. As a result, the optimized VTWL GLR test $\hat{T}_{\mathrm{GLR}}^*$ is equivalent to the following simple FMA GLR test:

$$\hat{T}_{\mathrm{FMA}} = \inf\left\{k \geq L : \hat{S}_{k-L+1}^k \geq \hat{h}_L^*\right\}, \tag{A.109}$$

where the threshold $\hat{h}_L^*$ is chosen for satisfying some levels of false alarms. The proof of Theorem 3.4 is completed. □.

## A.8 Proof of Theorem 4.1

The proof of Theorem 4.1 consists of three parts. Firstly, it is shown that the worst-case probability of false alarm $\overline{\mathbb{P}}_{\mathrm{fa}}(\delta_{\mathrm{FMA}})$ of the FMA detection-isolation rule (4.31)–(4.32) corresponds to the first time window $[L; L+m_\alpha-1]$. In addition, the upper bound $\tilde{\mathbb{P}}_{\mathrm{fa}}(\delta_{\mathrm{FMA}})$ for the worst-case probability of false alarm $\overline{\mathbb{P}}_{\mathrm{fa}}(\delta_{\mathrm{FMA}})$ is obtained. Secondly, it is proved that the worst-case probability of false isolation $\overline{\mathbb{P}}_{\mathrm{fi}}(\delta_{\mathrm{FMA}})$ corresponds to the first time window $[L; 2L-1]$ and its upper bound $\tilde{\mathbb{P}}_{\mathrm{fi}}(\delta_{\mathrm{FMA}})$ is derived. Finally, the upper bound $\tilde{\mathbb{P}}_{\mathrm{md}}(\delta_{\mathrm{FMA}})$ for the worst-case probability of missed detection $\overline{\mathbb{P}}_{\mathrm{md}}(\delta_{\mathrm{FMA}})$ is calculated analytically.

### A.8.1 Proof of part 1

In this subsection, we show that the worst-case probability of false alarm $\overline{\mathbb{P}}_{\mathrm{fa}}(T_{\mathrm{FMA}})$ of the FMA detection-isolation rule (4.31)–(4.32) corresponds to the first time window $[L; L+m_\alpha-1]$. The FMA algorithm $\delta_{\mathrm{FMA}} = (T_{\mathrm{FMA}}, \nu_{\mathrm{FMA}})$ can be rewritten as follows:

$$T_{\mathrm{FMA}} = \inf\left\{k \geq L : \max_{1\leq l\leq K}\min_{0\leq j\neq l\leq K}\left(S_{k-L+1}^k(l,j) - h\right) \geq 0\right\}, \tag{A.110}$$

$$\nu_{\mathrm{FMA}} = \arg\max_{1\leq l\leq K}\min_{0\leq j\neq l\leq K} S_{T_{\mathrm{FMA}}-L+1}^{T_{\mathrm{FMA}}}(l,j). \tag{A.111}$$

Let $U_{l_0} = \mathbb{P}_0(l_0 \leq T_{\mathrm{FMA}} < l_0 + m_\alpha)$, for $l_0 \geq L$, be the probability of false alarm within the time window $[l_0, l_0+m_\alpha-1]$. Our purpose is to show that $\{U_{l_0}\}_{l_0\geq L}$ is a non-increasing sequence w.r.t. the window position $l_0$. Let also $u_{l_0} = \mathbb{P}_0(T_{\mathrm{FMA}} = l_0)$ be the probability of false alarm at time instant $l_0$. We will show in the following that $\{u_{l_0}\}_{l_0\geq L}$ is a non-increasing sequence w.r.t. time instant $l_0$, i.e., $u_{l_0+1} \leq u_{l_0}$ for all $l_0 \geq L$, in considering two scenarios: $l_0 = L$ and $l_0 > L$.

For $l_0 = L$, we have

$$\begin{aligned} u_{L+1} &= \mathbb{P}_0\left(T_{\mathrm{FMA}} = L+1\right) \\ &= \mathbb{P}_0\Bigg(\left[\max_{1\le l\le K}\min_{0\le j\ne l\le K}\left(S_1^L\left(l,j\right)-h\right)<0\right]\bigcap \\ &\qquad \left[\max_{1\le l\le K}\min_{0\le j\ne l\le K}\left(S_2^{L+1}\left(l,j\right)-h\right)\ge 0\right]\Bigg) \\ &\le \mathbb{P}_0\left(\max_{1\le l\le K}\min_{0\le j\ne l\le K}\left(S_2^{L+1}\left(l,j\right)-h\right)\ge 0\right). \end{aligned} \tag{A.112}$$

Similar to the detection problem, the random variables $S_1^L\left(l,j\right)$ and $S_2^{L+1}\left(l,j\right)$, for any $1 \le l \le K$ and $0 \le j \ne l \le K$, have the same distributions. Hence, replace the random variables $S_2^{L+1}\left(l,j\right)$ in (A.112) by the random variables $S_1^L\left(l,j\right)$, we obtain that

$$u_{L+1} \le \mathbb{P}_0\left(\max_{1\le l\le K}\min_{0\le j\ne l\le K}\left(S_1^L\left(l,j\right)-h\right)\ge 0\right) = u_L. \tag{A.113}$$

For $l_0 > L$, we obtain by the same argument that

$$\begin{aligned} u_{l_0+1} &= \mathbb{P}_0\left(T_{\mathrm{FMA}} = l_0+1\right) \\ &= \mathbb{P}_0\Bigg(\bigcap_{k=L}^{l_0}\left[\max_{1\le l\le K}\min_{0\le j\ne l\le K}\left(S_{k-L+1}^k\left(l,j\right)-h\right)<0\right]\bigcap \\ &\qquad \left[\max_{1\le l\le K}\min_{0\le j\ne l\le K}\left(S_{l_0-L+2}^{l_0+1}\left(l,j\right)-h\right)\ge 0\right]\Bigg) \\ &\le \mathbb{P}_0\Bigg(\bigcap_{k=L+1}^{l_0}\left[\max_{1\le l\le K}\min_{0\le j\ne l\le K}\left(S_{k-L+1}^k\left(l,j\right)-h\right)<0\right]\bigcap \\ &\qquad \left[\max_{1\le l\le K}\min_{0\le j\ne l\le K}\left(S_{l_0-L+2}^{l_0+1}\left(l,j\right)-h\right)\ge 0\right]\Bigg) \\ &\le \mathbb{P}_0\Bigg(\bigcap_{k=L}^{l_0-1}\left[\max_{1\le l\le K}\min_{0\le j\ne l\le K}\left(S_{k-L+1}^k\left(l,j\right)-h\right)<0\right]\bigcap \\ &\qquad \left[\max_{1\le l\le K}\min_{0\le j\ne l\le K}\left(S_{l_0-L+1}^{l_0}\left(l,j\right)-h\right)\ge 0\right]\Bigg) = u_{l_0}, \end{aligned} \tag{A.114}$$

where the last inequality comes from the fact that the random variables $S_1^L\left(l,j\right),\cdots,S_{l_0-L+1}^{l_0}\left(l,j\right)$ and $S_2^{L+1}\left(l,j\right),\cdots,S_{l_0-L+2}^{l_0+1}\left(l,j\right)$ have the same distributions, for any $l_0 \ge L$, $1 \le l \le K$ and $0 \le j \ne l \le K$. From the above analysis, we have proved that $u_{l_0} \ge u_{l_0+1}$ for all $l_0 \ge L$. Moreover, we have from the definition of $U_{l_0}$ that

$$\begin{aligned} U_{l_0} - U_{l_0+1} &= \mathbb{P}_0\left(l_0 \le T_{\mathrm{FMA}} < l_0+m_\alpha\right) - \mathbb{P}_0\left(l_0+1 \le T_{\mathrm{FMA}} < l_0+m_\alpha+1\right) \\ &= \left[\sum_{k=l_0}^{l_0+m_\alpha-1}\mathbb{P}_0\left(T_{\mathrm{FMA}}=k\right)\right] - \left[\sum_{k=l_0+1}^{l_0+m_\alpha}\mathbb{P}_0\left(T_{\mathrm{FMA}}=k\right)\right] \\ &= u_{l_0} - u_{l_0+m_\alpha} \ge 0. \end{aligned} \tag{A.115}$$

In other words, the worst-case probability of false alarm of the FMA test corresponds to the first time window, i.e.,

$$\overline{\mathbb{P}}_{\text{fa}}\left(\delta_{\text{FMA}}; m_\alpha\right) = \mathbb{P}_0\left(L \leq T_{\text{FMA}} < L + m_\alpha\right). \tag{A.116}$$

Let us calculate now the upper bound on the worst-case probability of false alarm. It follows from (A.116) that

$$\begin{aligned}
\overline{\mathbb{P}}_{\text{fa}}\left(\delta_{\text{FMA}}; m_\alpha; h\right) &= \mathbb{P}_0\left(L \leq T_{\text{FMA}} < L + m_\alpha\right) \\
&= \mathbb{P}_0\left(\bigcup_{k=L}^{L+m_\alpha-1}\left\{\max_{1\leq l\leq K}\min_{0\leq j\neq l\leq K}\left(S_{k-L+1}^{k}\left(l,j\right)-h\right)\geq 0\right\}\right) \\
&= 1-\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_a-1}\left\{\max_{1\leq l\leq K}\min_{0\leq j\neq l\leq K}\left(S_{k-L+1}^{k}\left(l,j\right)-h\right)< 0\right\}\right) \\
&= 1-\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_a-1}\bigcap_{l=1}^{K}\left\{\min_{0\leq j\neq l\leq K}\left(S_{k-L+1}^{k}\left(l,j\right)-h\right)< 0\right\}\right) \\
&= 1-\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_a-1}\bigcap_{l=1}^{K}\bigcup_{\substack{j=0\\ j\neq l}}^{K}\left\{S_{k-L+1}^{k}\left(l,j\right)< h\right\}\right) \\
&\leq 1-\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\bigcap_{l=1}^{K}\left\{S_{k-L+1}^{k}\left(l,0\right)< h\right\}\right).
\end{aligned}$$

The worst-case probability of false alarm of the FMA detection rule is upper bounded as

$$\overline{\mathbb{P}}_{\text{fa}}\left(\delta_{\text{FMA}}; m_\alpha; h\right) \leq \tilde{\mathbb{P}}_{\text{fa}}\left(\delta_{\text{FMA}}; m_\alpha; h\right) \triangleq 1-\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\bigcap_{l=1}^{K}\left\{S_{k-L+1}^{k}\left(l,0\right)< h\right\}\right),$$

where $\tilde{\mathbb{P}}_{\text{fa}}\left(\delta_{\text{FMA}}; m_\alpha; h\right)$ is the upper bound on the worst-case probability of false alarm $\overline{\mathbb{P}}_{\text{fa}}\left(\delta_{\text{FMA}}; m_\alpha; h\right)$. This upper bound can be estimated by utilizing the numerical method introduced in Proposition 3.1. The proof of part 1 is finished. □.

### A.8.2 Proof of part 2

In the following, we show that the probability of false isolation of the FMA algorithm corresponds to the first time window $[L; 2L-1]$ and its upper bound is obtained for the case of threshold $h \geq 0$. Let $v_{k_0,k}^{l} = \mathbb{P}_{k_0}^{l}\left(T_{\text{FMA}} = k; \nu_{\text{FMA}} \neq l\right)$ be the probability of false isolation at time instant $k$ under the probability measures $\mathcal{P}_{k_0}^{l}$, for $k_0 \leq k \leq k_0 + L - 1$ and $1 \leq l \leq K$. We show in the following that $v_{k_0+1,k+1}^{l} \leq v_{k_0,k}^{l}$, for all $L \leq k_0 \leq k \leq k_0 + L - 1$ and $1 \leq l \leq K$, in considering two scenarios: $k = k_0 = L$ and $L \leq k_0 < k \leq k_0 + L - 1$. In the first scenario, i.e., $k = k_0 = L$,

it is clear that

$$v_{L+1,L+1}^{l} = \mathbb{P}_{L}^{l}\Bigg(\underbrace{\left\{\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}\left(S_{1}^{L}\left(\tilde{l},\tilde{j}\right)-h\right)<0\right\}}_{\text{non-detection at time instant } k=L}\bigcap$$
$$\underbrace{\left\{\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}\left(S_{2}^{L+1}\left(\tilde{l},\tilde{j}\right)-h\right)\geq 0\right\}}_{\text{detection at time instant } k=L+1}\bigcap\underbrace{\left\{\arg\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{2}^{L+1}\left(\tilde{l},\tilde{j}\right)\neq l\right\}}_{\text{false isolation at time instant } k=L+1}\Bigg)$$
$$\leq \mathbb{P}_{L}^{l}\left(\left\{\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}\left(S_{2}^{L+1}\left(\tilde{l},\tilde{j}\right)-h\right)\geq 0\right\}\bigcap\left\{\arg\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{2}^{L+1}\left(\tilde{l},\tilde{j}\right)\neq l\right\}\right).$$

It can be seen that the random variables $S_{1}^{L}\left(\tilde{l},\tilde{j}\right),\cdots,S_{L}^{L}\left(\tilde{l},\tilde{j}\right)$ under the probability measure $\mathcal{P}_{L}^{l}$ have the same distributions as the random variables $S_{2}^{L+1}\left(\tilde{l},\tilde{j}\right),\cdots,S_{L+1}^{L+1}\left(\tilde{l},\tilde{j}\right)$ under the probability measure $\mathcal{P}_{L+1}^{l}$, for all $1\leq l\leq K$, $1\leq\tilde{l}\neq l\leq K$ and $0\leq\tilde{j}\neq\tilde{l}\leq K$. Then, by substituting the random variables $S_{2}^{L+1}\left(\tilde{l},\tilde{j}\right),\cdots,S_{L+1}^{L+1}\left(\tilde{l},\tilde{j}\right)$ under $\mathcal{P}_{L+1}^{l}$ by the random variables $S_{1}^{L}\left(\tilde{l},\tilde{j}\right),\cdots,S_{L}^{L}\left(\tilde{l},\tilde{j}\right)$ under $\mathcal{P}_{L}^{l}$, we obtain

$$v_{L+1,L+1}^{l}\leq\mathbb{P}_{L}^{l}\left(\left\{\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{1}^{L}\left(\tilde{l},\tilde{j}\right)\geq h\right\}\bigcap\left\{\arg\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{1}^{L}\left(\tilde{l},\tilde{j}\right)\neq l\right\}\right)=v_{L,L}^{l}.$$

In the second scenario, i.e., $L\leq k_{0}<k\leq k_{0}+L-1$, we obtain by the same argument that

$$\begin{aligned}
v_{k_{0}+1,k+1}^{l} &= \mathbb{P}_{k_{0}+1}^{l}\left(T_{\mathrm{FMA}}=k+1;\nu_{\mathrm{FMA}}\neq l\right)\\
&= \mathbb{P}_{k_{0}+1}^{l}\Bigg(\underbrace{\bigcap_{\tilde{k}=L}^{k}\left\{\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{\tilde{k}-L+1}^{\tilde{k}}\left(\tilde{l},\tilde{j}\right)<h\right\}}_{\text{non-detection until time instant } k}\bigcap\\
&\qquad\underbrace{\left\{\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{k-L+2}^{k+1}\left(\tilde{l},\tilde{j}\right)\geq h\right\}}_{\text{detection at time instant } k+1}\bigcap\underbrace{\left\{\arg\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{k-L+2}^{k+1}\left(\tilde{l},\tilde{j}\right)\neq l\right\}}_{\text{false isolation at time instant } k+1}\Bigg)\\
&\leq \mathbb{P}_{k_{0}+1}^{l}\Bigg(\bigcap_{\tilde{k}=L+1}^{k}\left\{\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{\tilde{k}-L+1}^{\tilde{k}}\left(\tilde{l},\tilde{j}\right)<h\right\}\bigcap\\
&\qquad\left\{\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{k-L+2}^{k+1}\left(\tilde{l},\tilde{j}\right)\geq h\right\}\bigcap\left\{\arg\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{k-L+2}^{k+1}\left(\tilde{l},\tilde{j}\right)\neq l\right\}\Bigg)\\
&\leq \mathbb{P}_{k_{0}}^{l}\Bigg(\bigcap_{\tilde{k}=L}^{k-1}\left\{\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{\tilde{k}-L+1}^{\tilde{k}}\left(\tilde{l},\tilde{j}\right)<h\right\}\bigcap\\
&\qquad\left\{\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{k-L+1}^{k}\left(\tilde{l},\tilde{j}\right)\geq h\right\}\bigcap\left\{\arg\max_{1\leq\tilde{l}\leq K}\min_{0\leq\tilde{j}\neq\tilde{l}\leq K}S_{k-L+1}^{k}\left(\tilde{l},\tilde{j}\right)\neq l\right\}\Bigg)\\
&\leq \mathbb{P}_{k_{0}}^{l}\left(T_{\mathrm{FMA}}=k;\nu_{\mathrm{FMA}}\neq l\right)=v_{k_{0},k}^{l}. \qquad \text{(A.117)}
\end{aligned}$$

Secondly, we show that the probability of false isolation of the FMA test corresponds to the first time window $[L;2L-1]$. Let $V_{k_{0}}^{l}=\mathbb{P}_{k_{0}}^{l}\left(k_{0}\leq T_{\mathrm{FMA}}\leq k_{0}+L-1;\nu_{\mathrm{FMA}}\neq l\right)$ be the probability

of false isolation of type $l$ under the probability measure $\mathcal{P}^l_{k_0}$, for $1 \leq l \leq K$. Then,

$$\begin{aligned} V^l_{k_0} - V^l_{k_0+1} &= \left[\sum_{k=k_0}^{k_0+L-1} \mathbb{P}^l_{k_0}\left(T_{\text{FMA}} = k; \nu_{\text{FMA}} \neq l\right)\right] - \left[\sum_{k=k_0+1}^{k_0+L} \mathbb{P}^l_{k_0+1}\left(T_{\text{FMA}} = k; \nu_{\text{FMA}} \neq l\right)\right] \\ &= \sum_{k=k_0}^{k_0+L-1} \left[\mathbb{P}^l_{k_0}\left(T_{\text{FMA}} = k; \nu_{\text{FMA}} \neq l\right) - \mathbb{P}^l_{k_0+1}\left(T_{\text{FMA}} = k+1; \nu_{\text{FMA}} \neq l\right)\right] \\ &= \sum_{k=k_0}^{k_0+L-1} \left[v^l_{k_0,k} - v^l_{k_0+1,k+1}\right] \geq 0. \end{aligned} \tag{A.118}$$

Consequently, $\left\{V^l_{k_0}\right\}_{k_0 \geq L}$ is a non-increasing sequence w.r.t. the change-point $k_0$, leading to

$$\overline{\mathbb{P}}_{\text{fi}}\left(\delta_{\text{FMA}}; L; h\right) = \max_{1 \leq l \leq K} V^l_L = \max_{1 \leq l \leq K} \mathbb{P}^l_L\left(L \leq T_{\text{FMA}} < 2L; \nu_{\text{FMA}} \neq l\right). \tag{A.119}$$

In the following, we obtain the upper bound for the worst-case probability of false isolation given in (A.119) for the case of threshold $h \geq 0$. Seeking for simplicity, let us define following event

$$\begin{aligned} A^k_1 &\triangleq \bigcap_{\tilde{k}=L}^{k-1} \left\{\max_{1 \leq \tilde{l} \leq K} \min_{0 \leq \tilde{j} \neq \tilde{l} \leq K} S^{\tilde{k}}_{\tilde{k}-L+1}\left(\tilde{l}, \tilde{j}\right) < h\right\}, \\ A^k_2 &\triangleq \left\{\max_{1 \leq \tilde{l} \leq K} \min_{0 \leq \tilde{j} \neq \tilde{l} \leq K} S^k_{k-L+1}\left(\tilde{l}, \tilde{j}\right) \geq h\right\}, \\ A^k_3 &\triangleq \left\{\arg \max_{1 \leq \tilde{l} \leq K} \min_{0 \leq \tilde{j} \neq \tilde{l} \leq K} S^k_{k-L+1}\left(\tilde{l}, \tilde{j}\right) \neq l\right\}, \end{aligned}$$

where the event $A^k_1$ corresponds to the non-detection until time instant $k-1$, the event $A^k_2$ denotes the detection at time instant $k$ and the event $A^k_3$ stands for the false isolation at time instant $k$.

By assuming that $\mathbb{P}^l_L\left(A^L_1\right) = 1$, the probability of false isolation of type $l$ can be rewritten as

$$V^l_L = \mathbb{P}^l_L\left(\bigcup_{k=L}^{2L-1} \left\{A^k_1 \cap A^k_2 \cap A^k_3\right\}\right) \leq \mathbb{P}^l_L\left(\bigcup_{k=L}^{2L-1} \left\{A^k_2 \cap A^k_3\right\}\right). \tag{A.120}$$

Let us consider the event $\left\{A^k_2 \cap A^k_3\right\}$. It is clear that

$$\begin{aligned} A^k_2 \cap A^k_3 &= \left\{\max_{1 \leq \tilde{l} \leq K} \min_{0 \leq \tilde{j} \neq \tilde{l} \leq K} S^k_{k-L+1}\left(\tilde{l}, \tilde{j}\right) \geq h\right\} \bigcap \left\{\arg \max_{1 \leq \tilde{l} \leq K} \min_{0 \leq \tilde{j} \neq \tilde{l} \leq K} S^k_{k-L+1}\left(\tilde{l}, \tilde{j}\right) \neq l\right\} \\ &= \underbrace{\left\{\left[\max_{1 \leq \tilde{l} \neq l \leq K} \min_{0 \leq \tilde{j} \neq \tilde{l} \leq K} S^k_{k-L+1}\left(\tilde{l}, \tilde{j}\right) \geq h\right] \bigcap \left[\arg \max_{1 \leq \tilde{l} \leq K} \min_{0 \leq \tilde{j} \neq \tilde{l} \leq K} S^k_{k-L+1}\left(\tilde{l}, \tilde{j}\right) \neq l\right]\right\}}_{B_1} \bigcup \\ &= \underbrace{\left\{\left[\min_{0 \leq \tilde{j} \neq l \leq K} S^k_{k-L+1}\left(l, \tilde{j}\right) \geq h\right] \bigcap \left[\arg \max_{1 \leq \tilde{l} \leq K} \min_{0 \leq \tilde{j} \neq \tilde{l} \leq K} S^k_{k-L+1}\left(\tilde{l}, \tilde{j}\right) \neq l\right]\right\}}_{B_2}. \end{aligned} \tag{A.121}$$

Consider now the event $B_1$. Let us assume that there exists an index $j$, for $1 \leq j \neq l \leq K$, satisfying $\min_{0\leq\tilde{j}\neq j\leq K} S^k_{k-L+1}\left(j,\tilde{j}\right) \geq h$. In other words, $S^k_{k-L+1}\left(j,\tilde{j}\right) \geq h$ for all $0 \leq \tilde{j} \neq j \leq K$, leading to $S^k_{k-L+1}(j,l) \geq h$. For threshold $h \geq 0$, we obtain that $S^k_{k-L+1}(l,j) \leq 0$, leading to the fact that $\arg\max_{1\leq\tilde{l}\leq K} \min_{0\leq\tilde{j}\neq\tilde{l}\leq K} S^k_{k-L+1}\left(\tilde{l},\tilde{j}\right) \neq l$. In other words, the event $B_1$ is reduced to

$$B_1 = \left\{ \max_{1\leq\tilde{l}\neq l\leq K} \min_{0\leq\tilde{j}\neq\tilde{l}\leq K} S^k_{k-L+1}\left(\tilde{l},\tilde{j}\right) \geq h \right\}.$$

Consider now the event $B_2$. Let us assume that $\min_{0\leq\tilde{j}\neq l\leq K} S^k_{k-L+1}\left(l,\tilde{j}\right) \geq h$. Then, $S^k_{k-L+1}\left(\tilde{j},l\right) \leq 0$ for $0 \leq \tilde{j} \neq l \leq K$ since $h \geq 0$, leading to $\min_{0\leq\tilde{j}\neq\tilde{l}\leq K} S^k_{k-L+1}\left(\tilde{l},\tilde{j}\right) \leq 0$ for all $1 \leq \tilde{l} \neq l \leq K$. As a result, we obtain that $\arg\max_{1\leq\tilde{l}\leq K} \min_{0\leq\tilde{j}\neq\tilde{l}\leq K} S^k_{k-L+1}\left(\tilde{l},\tilde{j}\right) = l$ and that $B_2 = \emptyset$.

The event $\left\{A^k_2 \cap A^k_3\right\}$ is then reduced to

$$A^k_2 \cap A^k_3 = \left\{ \max_{1\leq\tilde{l}\neq l\leq K} \min_{0\leq\tilde{j}\neq\tilde{l}\leq K} S^k_{k-L+1}\left(\tilde{l},\tilde{j}\right) \geq h \right\}, \quad \forall L \leq k \leq 2L-1, \tag{A.122}$$

and the probability of false isolation of the type $l$ is upper bounded as

$$\begin{aligned}
V^l_L &\leq \mathbb{P}^l_L\left(\bigcup_{k=L}^{2L-1}\left\{ \max_{1\leq\tilde{l}\neq l\leq K} \min_{0\leq\tilde{j}\neq\tilde{l}\leq K} S^k_{k-L+1}\left(\tilde{l},\tilde{j}\right) \geq h \right\}\right) \\
&\leq 1 - \mathbb{P}^l_L\left(\bigcap_{k=L}^{2L-1}\left\{ \max_{1\leq\tilde{l}\neq l\leq K} \min_{0\leq\tilde{j}\neq\tilde{l}\leq K} S^k_{k-L+1}\left(\tilde{l},\tilde{j}\right) < h \right\}\right) \\
&\leq 1 - \mathbb{P}^l_L\left(\bigcap_{k=L}^{2L-1}\bigcap_{\substack{\tilde{l}=1\\ \tilde{l}\neq l}}^{K}\left\{ \min_{0\leq\tilde{j}\neq\tilde{l}\leq K} S^k_{k-L+1}\left(\tilde{l},\tilde{j}\right) < h \right\}\right) \\
&\leq 1 - \mathbb{P}^l_L\left(\bigcap_{k=L}^{2L-1}\bigcap_{\substack{\tilde{l}=1\\ \tilde{l}\neq l}}^{K}\bigcup_{\substack{\tilde{j}=0\\ \tilde{j}\neq\tilde{l}}}^{K}\left\{ S^k_{k-L+1}\left(\tilde{l},\tilde{j}\right) < h \right\}\right) \\
&\leq 1 - \max_{0\leq\tilde{j}\leq K} \mathbb{P}^l_L\left(\bigcap_{k=L}^{2L-1}\bigcap_{\substack{\tilde{l}=1\\ \tilde{l}\neq\tilde{j},l}}^{K}\left\{ S^k_{k-L+1}\left(\tilde{l},\tilde{j}\right) < h \right\}\right).
\end{aligned} \tag{A.123}$$

In other words, the worst-case probability of false isolation is upper bounded by

$$\overline{\mathbb{P}}_{\text{fi}}\left(\delta_{\text{FMA}};L;h\right) \leq \tilde{\mathbb{P}}_{\text{fi}}\left(\delta_{\text{FMA}};L;h\right) \triangleq \max_{1\leq l\leq K}\left[1 - \max_{0\leq\tilde{j}\leq K} \mathbb{P}^l_L\left(\bigcap_{k=L}^{2L-1}\bigcap_{\substack{\tilde{l}=1\\ \tilde{l}\neq\tilde{j},l}}^{K}\left\{ S^k_{k-L+1}\left(\tilde{l},\tilde{j}\right) < h \right\}\right)\right]. \tag{A.124}$$

The upper bound $\tilde{\mathbb{P}}_{\text{fi}}\left(\delta_{\text{FMA}};L;h\right)$ can be evaluated numerically by exploiting the numerical method suggested in Proposition 3.1. The proof of part 2 is finished. □.

### A.8.3 Proof of part 3

The worst-case probability of missed detection is described as

$$\begin{aligned}\overline{\mathbb{P}}_{\mathrm{md}}\left(\delta_{\mathrm{FMA}};L\right) &= \sup_{k_0\geq L}\max_{1\leq l\leq K}\mathbb{P}_{k_0}^{l}\left(T_{\mathrm{FMA}}\geq k_0+L|T_{\mathrm{FMA}}\geq k_0\right)\\ &= \sup_{k_0\geq L}\max_{1\leq l\leq K}\frac{\mathbb{P}_{k_0}^{l}\left(T_{\mathrm{FMA}}\geq k_0+L\right)}{\mathbb{P}_{k_0}^{l}\left(T_{\mathrm{FMA}}\geq k_0\right)}.\end{aligned}\tag{A.125}$$

For $k_0 > L$, we have

$$\begin{aligned}\mathbb{P}_{k_0}^{l}\left(T_{\mathrm{FMA}}\geq k_0+L\right) &= \mathbb{P}_{k_0}^{l}\left(\bigcap_{k=L}^{k_0+L-1}\left[\max_{1\leq\tilde{l}\leq K}\min_{0\leq j\neq\tilde{l}\leq K}\left(S_{k-L+1}^{k}\left(\tilde{l},j\right)-h\right)<0\right]\right),\\ \mathbb{P}_{k_0}^{l}\left(T_{\mathrm{FMA}}\geq k_0\right) &= \mathbb{P}_{k_0}^{l}\left(\bigcap_{k=L}^{k_0-1}\left[\max_{1\leq\tilde{l}\leq K}\min_{0\leq j\neq\tilde{l}\leq K}\left(S_{k-L+1}^{k}\left(\tilde{l},j\right)-h\right)<0\right]\right).\end{aligned}$$

Let us define the events $A_1$, $A_2$ and $A_3$ as follows:

$$\begin{aligned}A_1 &= \bigcap_{k=L}^{k_0-1}\left[\max_{1\leq\tilde{l}\leq K}\min_{0\leq j\neq\tilde{l}\leq K}\left(S_{k-L+1}^{k}\left(\tilde{l},j\right)-h\right)<0\right],\\ A_2 &= \bigcap_{k=k_0}^{k_0+L-2}\left[\max_{1\leq\tilde{l}\leq K}\min_{0\leq j\neq\tilde{l}\leq K}\left(S_{k-L+1}^{k}\left(\tilde{l},j\right)-h\right)<0\right],\\ A_3 &= \left[\max_{1\leq\tilde{l}\leq K}\min_{0\leq j\neq\tilde{l}\leq K}\left(S_{k_0}^{k_0+L-1}\left(\tilde{l},j\right)-h\right)<0\right].\end{aligned}$$

It is worth noting that the event $A_1$ depends on the random vectors $\xi_1^L,\cdots,\xi_{k_0-L}^{k_0-1}$, the event $A_2$ depends on the random vectors $\xi_{k_0-L+1}^{k_0},\cdots,\xi_{k_0-1}^{k_0+L-2}$, and the event $A_3$ depends on the random vector $\xi_{k_0}^{k_0+L-1}$. Moreover, there is no common element between the random vectors $\xi_1^L,\cdots,\xi_{k_0-L}^{k_0-1}$ and the random vector $\xi_{k_0}^{k_0+L-1}$. Therefore, the events $A_1$ and $A_3$ are independent, leading to

$$\begin{aligned}\mathbb{P}_{k_0}^{l}\left(T_{\mathrm{FMA}}\geq k_0+L|T_{\mathrm{FMA}}\geq k_0\right) &= \frac{\mathbb{P}_{k_0}^{l}\left(A_1\cap A_2\cap A_3\right)}{\mathbb{P}_{k_0}^{l}\left(A_1\right)}\\ &\leq \frac{\mathbb{P}_{k_0}^{l}\left(A_1\cap A_3\right)}{\mathbb{P}_{k_0}^{l}\left(A_1\right)}=\frac{\mathbb{P}_{k_0}^{l}\left(A_1\right)\cdot\mathbb{P}_{k_0}^{l}\left(A_3\right)}{\mathbb{P}_{k_0}^{l}\left(A_1\right)}=\mathbb{P}_{k_0}^{l}\left(A_3\right).\end{aligned}\tag{A.126}$$

For $k_0 = L$, we have

$$\mathbb{P}_{k_0}^{l}\left(T_{\mathrm{FMA}}\geq k_0+L|T_{\mathrm{FMA}}\geq k_0\right)=\mathbb{P}_{k_0}^{l}\left(T_{\mathrm{FMA}}\geq k_0+L\right)=\mathbb{P}_{k_0}^{l}\left(A_2\cap A_3\right)\leq\mathbb{P}_{k_0}^{l}\left(A_3\right).\tag{A.127}$$

Then, by replacing the event $A_3$ by its definition, we obtain that

$$\begin{aligned}
\mathbb{P}_{k_0}^{l}\left(T_{\mathrm{FMA}} \geq k_0 + L | T_{\mathrm{FMA}} \geq k_0\right) &\leq \mathbb{P}_{k_0}^{l}\left(\max_{1\leq\tilde{l}\leq K}\min_{0\leq j\neq\tilde{l}\leq K}\left(S_{k_0}^{k_0+L-1}\left(\tilde{l},j\right)-h\right)<0\right)\\
&\leq \mathbb{P}_{k_0}^{l}\left(\bigcap_{\tilde{l}=1}^{K}\left[\min_{0\leq j\neq\tilde{l}\leq K}\left(S_{k_0}^{k_0+L-1}\left(\tilde{l},j\right)-h\right)<0\right]\right)\\
&\leq \mathbb{P}_{k_0}^{l}\left(\min_{0\leq j\neq l\leq K}\left(S_{k_0}^{k_0+L-1}\left(l,j\right)-h\right)<0\right)\\
&\leq \mathbb{P}_{k_0}^{l}\left(\bigcup_{\substack{j=0\\j\neq l}}^{K}\left\{S_{k_0}^{k_0+L-1}\left(l,j\right)<h\right\}\right)\\
&\leq \sum_{\substack{j=0\\j\neq l}}^{K}\mathbb{P}_{k_0}^{l}\left(S_{k_0}^{k_0+L-1}\left(l,j\right)<h\right) \qquad \text{(A.128)}\\
&\leq \sum_{\substack{j=0\\j\neq l}}^{K}\mathbb{P}_{1}^{l}\left(S_{1}^{L}\left(l,j\right)<h\right), \qquad \text{(A.129)}
\end{aligned}$$

Moreover, under the probability measure $\mathcal{P}_1^l$, the LLR $S_1^L(l,j) \sim \mathcal{N}\left(\mu_{S_1^L(l,j)}, \sigma^2_{S_1^L(l,j)}\right)$, where the mean $\mu_{S_1^L(l,j)}$ and the variance $\sigma^2_{S_1^L(l,j)}$ are calculated as follows:

$$\begin{aligned}
\mu_{S_1^L(l,j)} &= \frac{1}{2}\left[\phi_1^L(1,l)-\phi_1^L(1,j)\right]^T\left[\Sigma^{-1}\right]\left[\phi_1^L(1,l)-\phi_1^L(1,j)\right], \qquad \text{(A.130)}\\
\sigma^2_{S_1^L(l,j)} &= \left[\phi_1^L(1,l)-\phi_1^L(1,j)\right]^T\left[\Sigma^{-1}\right]\left[\phi_1^L(1,l)-\phi_1^L(1,j)\right]. \qquad \text{(A.131)}
\end{aligned}$$

Finally, the worst-case probability of missed detection is upper bounded by

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(\delta_{\mathrm{FMA}};L;h\right) \leq \tilde{\mathbb{P}}_{\mathrm{md}}\left(\delta_{\mathrm{FMA}};L;h\right) \triangleq \max_{1\leq l\leq K}\sum_{\substack{j=0\\j\neq l}}^{K}\Phi\left(\frac{h-\mu_{S_1^L(l,j)}}{\sigma_{S_1^L(l,j)}}\right), \qquad \text{(A.132)}$$

where $\tilde{\mathbb{P}}_{\mathrm{md}}\left(\delta_{\mathrm{FMA}};L;h\right)$ is the upper bound on the worst-case probability of missed detection $\overline{\mathbb{P}}_{\mathrm{md}}\left(\delta_{\mathrm{FMA}};L;h\right)$. It can be seen that the upper bound $\tilde{\mathbb{P}}_{\mathrm{md}}\left(\delta_{\mathrm{FMA}};L;h\right)$ can be computed analytically. The proof of part 3 is finished. □.

# Appendix B

# Résumé en Français

## Contents

## B.1 Introduction

Le système de contrôle et d'acquisition de données (*Supervisory Control And Data Acquisition* - SCADA) est un système de télégestion à grande échelle permettant de traiter en temps réel un grand nombre de télémesures et de contrôler à distance des installations techniques. Les systèmes SCADA sont utilisés dans de nombreux secteurs tels que les systèmes de transports, les réseaux de télécommunications, les réseaux électriques, ou les réseaux de distribution de gaz et d'eau. À cause de leur architecture distribuée, les systèmes SCADA sont de plus en plus vulnérables aux cyber-attaques, non seulement au niveau de leurs infrastructures physiques, mais aussi au niveau de leurs réseaux de communication et de leur centre de contrôle. Ayant pénétrés dans un système SCADA, les attaquants peuvent effectuer des activités malveillantes leur permettant de contrôler, au moins partiellement, les processus physiques supervisés. Il est donc nécessaire de mettre en oeuvre des algorithmes de surveillance pour protéger les infrastructures critiques contre des dégâts, des pertes économiques, ou même des pertes humaines.

### B.1.1 Sécurité du système SCADA contre les cyber-attaques

L'architecture typique d'un système SCADA se compose de trois couches principales : la couche de contrôle et de surveillance, la couche de contrôle automatique et la couche physique. La première couche est responsable de contrôler et de surveiller le fonctionnement d'un système SCADA en recueillant des données à partir des appareils de terrain, en effectuant des tâches de surveillance, et en transmettant des commandes de contrôle aux contrôleurs de terrain. La deuxième couche est responsable de réguler le fonctionnement des processus physiques en se basant sur des commandes de contrôle envoyées à partir du centre de contrôle, sur des algorithmes de contrôle, et sur des mesures de capteurs. Finalement, les processus physiques sont équipés d'actionneurs (e.g., des moteurs, des pompes, des vannes), de capteurs (e.g., des capteurs de pression, des capteurs de débit, des capteurs de niveau), et d'autres éléments de protection (e.g., des disjoncteurs, des relais) pour réaliser des procédés technologiques. L'échange de données parmi les composantes du système est réalisé par l'intermédiaire du réseau de communication.

Les systèmes industriels modernes sont devenus vulnérables aux cyber-attaques en raison de l'évolution des technologies d'information et de communication. Plusieurs vulnérabilités d'un système SCADA peuvent être trouvées dans [53]. En exploitant ces vulnérabilités, les attaquants peuvent lancer des actes malveillants sur plusieurs points faibles du système SCADA. Les points d'attaque potentiels peuvent être classifiés en trois catégories [7] : des attaques sur le centre de contrôle, des attaques sur le réseau de communication et des attaques sur les processus physiques. Au cours de ces dernières années, il y a eu de nombreux incidents cyber-physiques survenus dans des infrastructures à sécurité critique telles que la rupture d'eau à Maroochy (2000) [168], l'arrêt d'une centrale nucléaire (2008) [97], le malware Stuxnet (2010) [20], ou la violation d'un site de

traitement d'eau (2011) [213]. Pour ces raisons, une plus grande attention devrait être accordée à la résilience des systèmes SCADA contre des actes malveillants.

### B.1.2 Méthodes de détection et de localisation

De nombreux travaux s'intéressent à l'étude de la sécurité des systèmes SCADA contre des cyber-attaques. Les approches considérées peuvent être classifiées en trois catégories principales [99] : l'approche basée sur la sécurité de l'information, l'approche basée sur la théorie du contrôle sécurisé, et l'approche basée sur la détection et la localisation de défauts. Les méthodes de sécurité de l'information se concentrent principalement sur l'authentification, le contrôle d'accès ou l'intégrité des messages pour assurer la transmission sécurisée de données (e.g., signaux de commande, signaux de contrôle, ou mesures de capteurs) parmi les composantes du réseau. Plusieurs méthodes ont été proposées dans [98] afin d'améliorer la sécurité des systèmes SCADA. Ces méthodes consistent à concevoir des pare-feus spécifiques entre les réseaux de processus et les réseaux d'entreprise, à utiliser des zones démilitarisées pour isoler les réseaux de processus et les réseaux d'entreprise, et à développer des réseaux privés virtuels pour transmettre des données sur des réseaux publics. Cependant, les techniques basées sur la sécurité de l'information semblent insuffisante pour la défense en profondeur des systèmes SCADA [25, 26, 28], notamment contre des attaques internes ciblant la dynamique du système [141].

L'approche de la théorie du contrôle sécurisé, de l'autre côté, est consacrée principalement à l'étude de la la sécurité des systèmes de contrôle en réseau contre plusieurs types d'attaques. Plus précisément, ces méthodes consistent à examiner des vulnérabilités des systèmes de contrôle en réseau, à concevoir des attaques furtives qui peuvent partiellement ou complètement contourner des détecteurs d'anomalies traditionnels, et à proposer des contre-mesures pour révéler de telles attaques. Les cyber-attaques sur les systèmes SCADA peuvent être classifiées en deux catégories principales [99, 141] : l'attaque par déni de service (DoS) et l'attaque sur l'intégrité des données. Les attaques DoS visent à perturber temporairement ou indéfiniment l'échange de données parmi les composantes du réseau, par exemple, par le brouillage des canaux de communication ou des protocoles de routage [99]. Les attaques d'intégrité, d'autre part, visent à modifier l'intégrité des paquets de données (signaux de commande, signaux de contrôle ou mesures de capteurs). Elles sont effectuées en modifiant le comportement des actionneurs et des capteurs ou en pénétrant aux réseaux de communication entre la couche physique et le centre de contrôle [141]. Les stratégies d'attaque d'intégrité peuvent être divisées encore en deux sous-catégories : l'attaque d'intégrité simple et l'attaque d'intégrité furtive. Les attaques d'intégrité simples [80] peuvent être conçues sans connaissance sur le modèle du système. Au contraire, les attaques d'intégrité furtives exigent la connaissance sur le modèle du système et les capacités de perturbation pour contourner des algorithmes de détection classiques. Quelques exemples d'attaques d'intégrité furtives sont la stratégie de rediffusion de données [120], la stratégie d'injection de fausses données [121], la stratégie d'attaque zéro-dynamique [186], et la stratégie d'attaque secrète [169].

Il a été montré que la détection et l'identification d'attaques sont étroitement liées au problème de détection et de localisation de défauts (FDI) [30, 35]. Pour cette raison, les techniques de FDI ont été utilisées pour détecter et identifier des cyber-attaques sur les systèmes SCADA. Par exemple, les auteurs de [27] ont formulé le problème de détection des cyber-attaques sur les systèmes de contrôle de procédé comme un problème de diagnostic de défauts. L'algorithme de la Somme Cumulée (CUSUM) non-paramétrique a été utilisé pour détecter les attaques. En outre, la sécurité des systèmes d'irrigation d'eau a été considérée dans [6, 7]. Dans ce travail, les

auteurs ont démontré que le problème de détection et de localisation de cyber-attaques pourrait être résolu en utilisant une banque d'observateurs d'entrées inconnus [30]. De plus, un traitement global du problème de détection et d'identification d'attaques sur les systèmes cyber-physiques a été donné dans [140,141]. Le modèle d'espace d'état est utilisé pour décrire les systèmes SCADA et les cyber-attaques sont modélisées par des changements additifs de l'équation d'états ainsi que l'équation de mesures. Plusieurs algorithmes centralisés et distribués sont proposés pour détecter et localiser les attaques. Cependant, les travaux mentionnés ont été formulés dans le carde déterministe (sans bruit aléatoire).

La première tâche d'un problème de FDI consiste à déterminer un ensemble d'équations mathématiques qui régissent le système. Le modèle paramétrique du système en régime nominal ainsi qu'en régime anormal est extrêmement important lors de la conception des algorithmes de diagnostic. La deuxième tâche consiste à proposer des algorithmes de détection et de localisation en s'appuyant sur les modèles développés. La conception des algorithmes de diagnostic est, généralement, résolue par l'approche des redondances analytiques qui se compose de deux étapes : la génération de résidus et l'évaluation de résidus. Les résidus sont d'abord générés en exploitant des techniques développées par la communauté de diagnostic de défauts (e.g., le filtre de Kalman ou l'espace de parité) et ils sont ensuite évalués en utilisant des méthodes introduites dans la théorie de la décision statistique (e.g., des tests non-séquentiels, des tests séquentiels, la détection séquentielle de changements brusques) [10,30,35,54,81,206].

Cette thèse se concentre sur la surveillance des systèmes SCADA contre des cyber-attaques. Il est donc nécessaire de proposer des algorithmes de surveillance qui sont capable de détecter et de localiser des actes malveillantes en temps réel. En outre, la conception des algorithmes de surveillance devrait être prise en compte des états inconnus (les paramètres de nuisance) ainsi que des bruits stochastiques. Afin d'éliminer l'impact négatif des paramètres de nuisance pendant la la prise de décision, nous utilisons dans cette thèse l'approche du filtre de Kalman en régime permanent et l'approche par projection dans l'espace de parité de taille fixe. Les résidus générés par les techniques mentionnées contiennent toujours des bruits aléatoires. Donc, ils doivent être évalués en utilisant les résultats de la théorie de la détection séquentielle de changements brusques (ou « ruptures ») [10,175].

La théorie de la détection séquentielle de changements brusques s'intéresse à la détection d'un changement (ou rupture) dans une séquence d'observations qui contiennent des transitions rapides et éventuellement d'identifier le type de ces transitions. Pour le problème classique, la période après le changement est supposée être infiniment longue (voir [105] et aussi [147,175] pour plus de détails). Le problème de détection (pure) de changements brusques entre deux lois de probabilité (une hypothèse de base et une hypothèse concurrente) consiste à calculer l'instant d'arrêt $T$ auquel la présence de la rupture est déclarée. Cet instant de changement doit respecter certains critères d'optimalité. Par exemple, le retard moyen de détection devrait être aussi faible que possible pour une valeur donnée de fausses alarmes. Plusieurs algorithmes optimaux par rapport à différents critères d'optimalité dans le cadre de l'approche non-bayésienne (où l'instant de rupture est inconnu mais non-aléatoire) sont introduits dans [103,113,146]. Les résultats essentiels dans le cadre de l'approche bayésienne (où l'instant de rupture est inconnu et aléatoire) peuvent être trouvés dans [142,158,166,177].

Le problème d'identification (détection-localisation) de changements brusques dans un système stochastique est la généralisation du problème de détection de rupture pour des hypothèses multiples (une hypothèse de base et plusieurs hypothèses concurrentes). Le problème consiste à calculer un couple $(T, \nu)$, où $T$ est l'instant d'arrêt auquel la décision finale $\nu$ est décidée. Les

critères d'optimalité devraient favoriser la rapidité de détection et de localisation avec des taux acceptables de fausses alarmes et de fausses localisations. Plusieurs procédures de détection-localisation asymptotiquement optimales par rapport aux différents critères d'optimalité (pour l'approche non-bayésienne ainsi que l'approche bayésienne) ont été proposées dans [104, 128–130, 132, 138, 175].

Malheureusement, les critères d'optimalité classiques ne sont pas appropriés au problème de détection et de localisation d'attaques dans des systèmes SCADA à cause des raisons suivantes. Tout d'abord, l'adversaire préfère effectuer ses actes malveillants pendant une période finie en raison de ses ressources limitées [6, 7, 25, 80]. De telles attaques entraînent des changements transitoires (c-à-d des signaux de durée finie) dans le système attaqué. En outre, pour les systèmes à sécurité critique, il convient de détecter et de localiser les attaques avec un retard inférieur à une constante fixée *a priori* [9, 67, 69, 70]. Donc, il est pertinent de considérer le problème de surveillance des systèmes SCADA contre des actes malveillants comme un problème de détection et de localisation de changements transitoires dans des systèmes stochastiques et dynamiques. Dans cette thèse, nous utilisons le modèle d'espace d'état à temps discret pour décrire les systèmes SCADA. Les bruits gaussiens sont ajoutés à l'équation d'état ainsi qu'à l'équation de mesure afin de modéliser, respectivement, l'incertitude des processus et l'imprécision des appareils de mesure. Les attaques sont modélisées par des signaux additifs de durée finie dans les deux équations. Le problème consiste à détecter l'instant inconnu où surviennent les actes malveillants, et éventuellement à déterminer le type d'attaque en présence des états inconnus (souvent considérés comme des paramètres de nuisance) et des bruits stochastiques.

Le problème de détection de changements transitoires dans un système stochastique avec des mesures indépendantes a été considéré dans [67–70]. Le critère d'optimalité vise à minimiser la pire probabilité de détection manquée sous la contrainte que la pire probabilité de fausse alarme pour une fenêtre de taille donnée est inférieure à une valeur prescrite. Un algorithme sous-optimal par rapport au critère d'optimalité a été proposé pour le cas d'observations gaussiennes indépendantes. L'idée est la suivante. Tout d'abord, un algorithme de la Somme Cumulée à Fenêtre Limitée des Seuils Variables (VTWL CUSUM) a été considéré pour détecter des changements transitoires. Les bornes supérieures pour la pire probabilité de détection manqué ainsi que pour la pire probabilité de fausse alarme pour une fenêtre de taille donnée ont été calculées. Par la suite, le problème d'optimisation a été formulé comme un choix optimal, basé sur une fonction de détection et des seuils variables, visant à minimiser la borne supérieure pour la pire probabilité de détection manquée sous la contrainte que la borne supérieure pour la pire probabilité de fausse alarme soit inférieure à une valeur prescrite. Le choix optimal des seuils conduit au test de la Moyenne Glissante Finie (Finite Moving Average ou FMA). À notre connaissance, le problème de détection-localisation conjointe de changements transitoires n'est pas encore abordé dans la littérature.

### B.1.3 Contribution et organisation

L'objectif finale de cette thèse est de proposer des algorithmes de détection et de localisation d'attaques cyber-physiques dans des systèmes industriels SCADA. À partir des analyses ci-dessus, il convient d'étudier la surveillance en-ligne des infrastructures à sécurité critique par le biais de la détection et la localisation de changements transitoires dans des systèmes stochastiques et dynamiques. En suivant l'approche par redondance analytique classique, le problème est résolu en deux étapes : la génération de résidus et l'évaluation de résidus. Les résidus sont tout d'abord

générés en utilisant deux méthodes conventionnelles : le filtre de Kalman en régime permanent et la projection sur l'espace de parité de taille fixe. Ils sont ensuite évalués en exploitant des techniques de surveillance des systèmes stochastiques afin de détecter l'instant d'attaque, et éventuellement de classifier le type d'attaque (ou scénario d'attaque). Cette thèse se concentre particulièrement sur l'évaluation de résidus.

Pour le problème de détection, nous généralisons les travaux initiés par Guépié [67–70] à la détection des signaux additifs de durée finie au modèle d'espace d'état en présence d'états inconnus et de bruits stochastiques. Le critère d'optimalité pour la détection de changements transitoires, qui a été proposé par Guépié [67–70], est utilisé dans cette thèse afin d'évaluer les performances statistiques des algorithmes de détection. Il est à noter que les résultats obtenus par Guépié [67–70] dépendent fortement du concept des variables associées [46, 110] qui permet d'établir la borne supérieure de la pire probabilité de fausse alarme. Malheureusement, les résidus générés par les deux méthodes mentionnées ne permettent pas d'utiliser cette propriété. Pour cette raison, nous formulons dans cette thèse le problème d'optimisation d'une manière légèrement différent que celle proposée par Guépié [67–70]. La contribution au problème de détection se décompose comme suit :

- Le développement d'un modèle statistique unifié de résidus. Les modèles statistiques de résidus générés par l'approche du filtre de Kalman en régime permanent et par la projection sur l'espace de parité de taille fixe sont calculés. Plus particulièrement, nous intégrons les deux modèles statistiques dans un modèle statistique unifié des résidus.

- La formulation et la solution du problème d'optimisation. D'abord, l'algorithme de la Somme Cumulée à Fenêtre Limitée des Seuils Variables (VTWL CUSUM) est considéré pour détecter des changements transitoires dans une séquence des résidus en s'appuyant sur le modèle statistique unifié. De façon similaire aux travaux de Guépié [67–70], nous calculons une borne supérieure pour la pire probabilité de détection manquée du test VTWL CUSUM. Ensuite, le problème d'optimisation est formulé comme le choix optimal des seuils variables dans la classe des tests VTWL CUSUM. Au contraire des travaux de Guépié [67–70], nous proposons dans cette thèse de minimiser la borne supérieure pour la pire probabilité de détection manquée sous la contrainte que la pire probabilité de fausse alarme pour une fenêtre de taille donnée soit inférieure à une valeur prescrite. Finalement, nous démontrons que l'algorithme VTWL CUSUM optimisé est équivalent à la règle de décision de la Moyenne Glissante Finie (FMA).

- Le calcul numérique des probabilités d'erreurs. Une méthode numérique est proposée afin d'estimer la probabilité de fausse alarme et la probabilité de détection manqué du test FMA et du test VTWL CUSUM. Cette méthode numérique est plus efficace que l'approche de Monte Carlo conventionnelle en terme de temps de calcul.

- L'analyse de sensibilité du test FMA. En utilisant le méthode numérique mentionnée, nous effectuons l'analyse de robustesse du test FMA par rapport à plusieurs paramètres opérationnels, e.g., la durée d'attaque, les profils d'attaque, les covariances des bruits de processus et des bruits de capteurs.

- L'extension au scénario où les profils de changements transitoires sont partiellement connus. En supposant que la « forme » des profils est parfaitement connue mais leur « amplitude » est inconnue, l'approche du rapport de vraisemblance généralisé (GLR) et l'approche de rapport de vraisemblance pondéré (WLR) sont considérées. Le test VTWL GLR

et le test VTWL WLR sont proposés. Le problème d'optimisation est établi et résolu. Il est montré que le test VTWL GLR optimisé et le test VTWL WLR optimisé correspondent au test FMA GLR et au test FMA WLR, respectivement.

Le problème de localisation est beaucoup plus difficile que le problème de détection. Pour cette raison, peu de résultats théoriques sont obtenus. La contribution de cette partie se décompose comme suit :

- Premièrement, le modèle statistique unifié des résidus générés par les deux méthodes susmentionnées est généralisé à la détection-localisation conjointe de changements transitoires au modèle d'espace d'état à temps discret.
- Deuxièmement, un nouveau critère d'optimalité pour la détection-localisation conjointe de changements transitoires est introduit. Le critère vise à minimiser la pire probabilité de détection manquée soumis à des niveaux acceptables pour la pire probabilité de fausse alarme dans une fenêtre de taille donnée et pour la pire probabilité de fausse localisation pendant la fenêtre de changements transitoires.
- Troisièmement, plusieurs algorithmes de détection-localisation conjointe de changements brusques, e.g., le test WL CUSUM généralisé, le test WL CUSUM par matrice et le test WL CUSUM par vecteur, sont considérés pour détecter et localiser des changements transitoires en s'appuyant sur le modèle statistique unifié des résidus. Notamment, la règle de détection-localisation FMA est proposée.
- Finalement, nous calculons des bornes supérieures pour les pires probabilités d'erreurs, c-à-d pour la pire probabilité de détection manquée, pour la pire probabilité de fausse alarme et pour la pire probabilité de fausse localisation.

Ce résumé est organisé comme suit. La formulation du problème est donnée dans section B.2. La génération de résidus par l'approche de filtre de Kalman et par l'approche d'espace de parité est présentée dans la section B.3. Notamment, le modèle statistique unifié des résidus générés par les deux méthodes est développé. En s'appuyant sur ce modèle, nous proposons dans la section B.4 et la section B.5, respectivement, les algorithmes de détection de signaux transitoires exactement connus ou partiellement connus. La conception des algorithmes de détection-localisation conjointe est considérée dans section B.6. Dans la section B.7, nous appliquons les algorithmes de détection et de localisation à la surveillance d'un réseau de distribution d'eau potable. Finalement, quelques conclusions et perspectives sont données dans la section B.8.

## B.2 Formulation du problème

Dans cette section, nous formulons la détection d'attaques aux systèmes SCADA comme un problème de détection de changements transitoires dans des systèmes stochastiques et dynamiques.

### B.2.1 Modèles du système et des attaques cyber-physiques

Le modèle d'espace d'état à temps discret est utilisé pour décrire des systèmes industriels attaqués :

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + Ka_k^x + w_k \\ y_k &= Cx_k + Du_k + Gd_k + Ha_k^x + Ma_k^y + v_k \end{cases}; \quad x_1 = \overline{x}_1, \tag{B.1}$$

où $x_k \in \mathbb{R}^n$ est le vecteur d'états, $u_k \in \mathbb{R}^m$ est le vecteur de signaux de contrôle, $d_k \in \mathbb{R}^q$ est le vecteur de perturbations, $y_k \in \mathbb{R}^p$ est le vecteur de mesures, $a_k^x \in \mathbb{R}^r$ est le vecteur d'attaque sur les états, $a_k^y \in \mathbb{R}^p$ est le vecteur d'attaque sur les mesures de capteurs, $w_k \in \mathbb{R}^n$ est le vecteur de bruits de processus, $v_k \in \mathbb{R}^p$ est le vecteur de bruits de capteurs ; les matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $F \in \mathbb{R}^{n\times q}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, $G \in \mathbb{R}^{p\times q}$, $K \in \mathbb{R}^{n\times r}$, $H \in \mathbb{R}^{p\times r}$ et $M \in \mathbb{R}^{p\times p}$ sont connues. Les vecteurs $w_k$ et $v_k$ sont des vecteurs gaussiens indépendants et identiquement distribués, c-à-d $\mathbb{E}\left[w_k w_l^T\right] = Q\delta_{kl}$, $\mathbb{E}\left[v_k v_l^T\right] = R\delta_{kl}$ et $\mathbb{E}\left[w_k v_l^T\right] = 0$, où $\delta_{kl} = 1$ si $k = l$ et $\delta_{kl} = 0$ autrement, les matrices $Q \in \mathbb{R}^{n\times n}$ et $R \in \mathbb{R}^{p\times p}$ sont connues et $R$ est définie-positive.

**Remarque B.1.** *Les attaquants peuvent construire les vecteurs d'attaque $a_k^x$ et $a_k^y$ pour réaliser leur objectif malveillant. Il a été démontré que les vecteurs d'attaque $a_k^x$ et $a_k^y$ pourraient être coordonnés pour perturber le système tout en contournant les détecteurs d'anomalies traditionnels [141]. Ces attaques furtives peuvent être conçues par la stratégie de rediffusion de données [120], par la stratégie d'injection de fausses données [121], par la stratégie d'attaque zéro-dynamique [186], ou par la stratégie d'attaque secrète [169]. L'analyse de sécurité du système est requise pour révéler ces attaques furtives (voir, par exemple, [121], [186] ou [99]). Pour cette raison, nous considérons dans ce manuscrit seulement des attaques détectables.*

### B.2.2 Modèle des changements transitoires

Pour simplifier les notations, le vecteur d'attaque sur les états $a_k^x$ et le vecteur d'attaque sur les capteurs $a_k^y$ sont regroupés dans un seul vecteur d'attaque $a_k = \left[(a_k^x)^T, (a_k^y)^T\right]^T \in \mathbb{R}^s$, où $s = r+p$. Posons $B_a = [K, 0] \in \mathbb{R}^{n\times s}$ et $D_a = [H, M] \in \mathbb{R}^{p\times s}$. Donc, le modèle (B.1) est simplifié par :

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + B_a a_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + D_a a_k + v_k \end{cases}; \quad x_1 = \overline{x}_1, \tag{B.2}$$

où les matrices d'attaque $B_a$ et $D_a$ dépendent de l'architecture du système et le vecteur d'attaque $a_k$ est conçu par l'attaquant pour réaliser son objectif malveillant. Supposons que les attaques cyber-physiques sont effectuées pendant une période finie $\tau_a = [k_0, k_0 + L - 1]$, où $k_0$ est l'instant d'attaque inconnu et $L$ est la durée d'attaque présumée connue. Le vecteur d'attaque $a_k$ est décrit par :

$$a_k = \begin{cases} 0 & \text{si} \quad k < k_0 \\ \theta_{k-k_0+1} & \text{si} \quad k_0 \le k < k_0 + L \\ 0 & \text{si} \quad k \ge k_0 + L \end{cases}, \tag{B.3}$$

où $\theta_1, \theta_2, \cdots, \theta_L \in \mathbb{R}^s$ sont des profils d'attaque. Il est à noter que les informations des profils $\theta_1, \theta_2, \cdots, \theta_L$ jouent un rôle important dans la performance statistique des algorithmes de détection. Nous considérons dans cette thèse deux scénarios : les profils sont parfaitement connus et les profils sont partiellement connus.

### B.2.3 Critère d'optimalité

Le critère d'optimalité la détection de changements transitoires, introduit la première fois dans [67, 69], est utilisé dans cette thèse afin d'évaluer les performances statistiques des algorithmes

de détection. Le critère vise à minimiser la pire probabilité de détection manquée :

$$\inf_{T\in C_\alpha}\left\{\overline{\mathbb{P}}_{\mathrm{md}}\left(T;L\right)=\sup_{k_0\geq L}\mathbb{P}_{k_0}\left(T-k_0+1>L|T\geq k_0\right)\right\}, \tag{B.4}$$

parmi tous les instants d'arrêt $T\in C_\alpha$ satisfaisant :

$$C_\alpha=\left\{T:\overline{\mathbb{P}}_{\mathrm{fa}}\left(T;m_\alpha\right)=\sup_{l\geq L}\mathbb{P}_0\left\{l\leq T<l+m_\alpha\right\}\leq\alpha\right\}, \tag{B.5}$$

où $\overline{\mathbb{P}}_{\mathrm{md}}$ est la pire probabilité détection manquée et $\overline{\mathbb{P}}_{\mathrm{fa}}$ est la pire probabilité de fausse alarme pour une fenêtre de taille $m_\alpha$.

## B.3 Méthodes de génération des résidus

Dans cette section, nous considérons deux approches de génération des résidus : le filtre de Kalman en régime permanent et la projection sur l'espace de parité de taille fixe. Nous développons également le modèle statistique unifié de résidus générés par les deux méthodes.

### B.3.1 Approche avec filtre de Kalman en régime permanent

Supposons que le filtre de Kalman est utilisé pour générer une séquence d'innovations. Le gain de Kalman en régime permanent $K_\infty$ est calculé par :

$$K_\infty=P_\infty C^T\left(CP_\infty C^T+R\right)^{-1}, \tag{B.6}$$

où la matrice de covariance de l'erreur d'estimation d'états $P_\infty$ peut être calculée en résolvant l'équation algébrique de Riccati à temps discret suivante :

$$P_\infty=AP_\infty A^T-AP_\infty C^T\left(CP_\infty C^T+R\right)^{-1}CP_\infty A^T+Q. \tag{B.7}$$

Donc, l'opération du filtre de Kalman en régime permanent est décrite comme :

$$\begin{cases}\hat{x}_{k+1|k} &= A\hat{x}_{k|k-1}+Bu_k+Fd_k+AK_\infty\left(y_k-\hat{y}_{k|k-1}\right)\\ \hat{y}_{k|k-1} &= C\hat{x}_{k|k-1}+Du_k+Gd_k\end{cases},\quad \hat{x}_{1|0}=\overline{x}_1, \tag{B.8}$$

où $\hat{x}_{k|k-1}\in\mathbb{R}^n$ est l'estimation des états et $\hat{y}_{k|k-1}\in\mathbb{R}^p$ l'estimation des sorties.

Soit $r_k=y_k-\hat{y}_{k|k-1}\in\mathbb{R}^p$ un vecteur d'innovations. Il a été démontré [10,116] que les innovations $\{r_k\}_{k\geq 1}$ sont des vecteurs gaussiens indépendants de matrice de covariance $J\triangleq CP_\infty C^T+R$. Soit $\varrho_1,\varrho_2,\cdots\in\mathbb{R}^p$ la séquence des vecteurs aléatoires indépendants et identiquement distribués (i.i.d.) suivant une loi normale multidimensionnelle de covariance $J$, c-à-d $\varrho_k\sim\mathcal{N}\left(0,J\right)$. Donc, le modèle statistique des innovations est décrit par :

$$r_k=\begin{cases}\varrho_k & \text{si}\quad k<k_0\\ \psi_{k-k_0+1}+\varrho_k & \text{si}\quad k_0\leq k<k_0+L\\ \tilde{\psi}_k+\varrho_k & \text{si}\quad k\geq k_0+L\end{cases}, \tag{B.9}$$

où $\psi_1, \psi_2, \cdots, \psi_L$ sont des profils de changements transitoires, étant calculés à partir des profils d'attaque $\theta_1, \theta_2, \cdots, \theta_L$ par l'équation suivante :

$$\begin{cases} \epsilon_{k+1} & = (A - AK_\infty C)\,\epsilon_k + (B_a - AK_\infty D_a)\,\theta_k \\ \psi_k & = C\epsilon_k + D_a\theta_k \end{cases}; \quad \epsilon_1 = 0, \tag{B.10}$$

et les profils après les changements $\tilde{\psi}_k$ (pour $k \geq k_0 + L$) ne présentent aucun intérêt.

Soit $r^k_{k-L+1} = \left[r^T_{k-L+1}, \cdots, r^T_k\right]^T \in \mathbb{R}^{Lp}$ le vecteur concaténé des résidus, $\varrho^k_{k-L+1} = \left[\varrho^T_{k-L+1}, \cdots, \varrho^T_k\right]^T \in \mathbb{R}^{Lp}$ le vecteur concaténé des bruits, et $\psi^k_{k-L+1}(k_0) \in \mathbb{R}^{Lp}$ le vecteur des signaux transitoires. Le vecteur $\psi^k_{k-L+1}(k_0)$ dépend de la position relative de l'instant de rupture $k_0$ dans la fenêtre $[k-L+1, k]$ via l'équation suivante :

$$\psi^k_{k-L+1}(k_0) = \begin{cases} [0] & \text{si} \quad k < k_0 \\ \begin{bmatrix} [0] \\ \psi_1 \\ \vdots \\ \psi_{k-k_0+1} \end{bmatrix} & \text{si} \quad k_0 \leq k < k_0 + L\,, \\ \tilde{\psi}^k_{k-L+1}(k_0) & \text{si} \quad k \geq L \end{cases} \tag{B.11}$$

où $[0]$ est un vecteur nul de dimension appropriée et les profils après les changements $\tilde{\psi}^k_{k-L+1}(k_0) \in \mathbb{R}^{Lp}$ ne présentent pas d'intérêt.

En regroupant (B.9)–(B.11), le modèle statistique des innovations $r^k_{k-L+1}$ générées par le filtre de Kalman est exprimé par :

$$r^k_{k-L+1} = \psi^k_{k-L+1}(k_0) + \varrho^k_{k-L+1}, \tag{B.12}$$

où $\varrho^k_{k-L+1} \sim \mathcal{N}(0, \Sigma_\varrho)$ et $\Sigma_\varrho = \mathrm{diag}(J) \in \mathbb{R}^{Lp \times Lp}$ est la matrice diagonale par blocs $J$.

### B.3.2 Approche par projection sur un espace de parité de taille fixe

Dans cette section, nous développons le modèle statistique des résidus générés par projection sur un espace de parité de taille fixe. Les vecteurs $u_k$ et $d_k$ sont éliminés du modèle (B.2) puisqu'ils sont connus. En regroupant les dernières $L$ observations, le modèle d'observation se simplifie :

$$\begin{aligned} \underbrace{\begin{bmatrix} z_{k-L+1} \\ z_{k-L+2} \\ \vdots \\ z_k \end{bmatrix}}_{z^k_{k-L+1}} &= \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{L-1} \end{bmatrix}}_{\mathcal{C}} x_{k-L+1} + \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ C & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2} & CA^{L-3} & \cdots & 0 \end{bmatrix}}_{\mathcal{H}} \underbrace{\begin{bmatrix} w_{k-L+1} \\ w_{k-L+2} \\ \vdots \\ w_k \end{bmatrix}}_{w^k_{k-L+1}} + \\ &\underbrace{\begin{bmatrix} D_a & 0 & \cdots & 0 \\ CB_a & D_a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2}B_a & CA^{L-3}B_a & \cdots & D_a \end{bmatrix}}_{\mathcal{M}} \underbrace{\begin{bmatrix} a_{k-L+1} \\ a_{k-L+2} \\ \vdots \\ a_k \end{bmatrix}}_{\theta^k_{k-L+1}(k_0)} + \underbrace{\begin{bmatrix} v_{k-L+1} \\ v_{k-L+2} \\ \vdots \\ v_k \end{bmatrix}}_{v^k_{k-L+1}}, \end{aligned} \tag{B.13}$$

ou dans une forme plus simple :

$$z_{k-L+1}^{k} = \mathcal{C}x_{k-L+1} + \mathcal{M}\theta_{k-L+1}^{k}(k_0) + \mathcal{H}w_{k-L+1}^{k} + v_{k-L+1}^{k}, \tag{B.14}$$

où $z_{k-L+1}^{k} \in \mathbb{R}^{Lp}$ est le vecteur concaténé d'observations simplifiées, $w_{k-L+1}^{k} \in \mathbb{R}^{Ln}$ est le vecteur concaténé de bruits de processus, $v_{k-L+1}^{k} \in \mathbb{R}^{Lp}$ le vecteur concaténé de bruits de capteurs, $\theta_{k-L+1}^{k}(k_0) \in \mathbb{R}^{Ls}$ le vecteur concaténé de profils d'attaque ; les matrices $\mathcal{C} \in \mathbb{R}^{Lp\times n}$, $\mathcal{M} \in \mathbb{R}^{Lp\times Ls}$ et $\mathcal{H} \in \mathbb{R}^{Lp\times Ln}$. Les bruits de processus $w_{k-L+1}^{k} \sim \mathcal{N}(0, \mathcal{Q})$ et les bruits de capteurs $v_{k-L+1}^{k} \sim \mathcal{N}(0, \mathcal{R})$, où $\mathcal{Q} = \mathrm{diag}(Q) \in \mathbb{R}^{Ln\times Ln}$ et $\mathcal{R} = \mathrm{diag}(R) \in \mathbb{R}^{Lp\times Lp}$ sont des matrices diagonales par blocs $Q$ et $R$, respectivement.

De façon similaire à la définition du vecteur $\psi_{k-L+1}^{k}(k_0)$ dans (B.11) par l'approche de filtre de Kalman, le vecteur $\theta_{k-L+1}^{k}(k_0)$ est exprimé par :

$$\theta_{k-L+1}^{k}(k_0) = \begin{cases} [0] & \text{si} \quad k < k_0 \\ \begin{bmatrix} [0] \\ \theta_1 \\ \vdots \\ \theta_{k-k_0+1} \end{bmatrix} & \text{si} \quad k_0 \le k < k_0 + L \\ \left[\tilde{\theta}_{k-L+1}^{k}(k_0)\right] & \text{si} \quad k \ge k_0 + L \end{cases}, \tag{B.15}$$

où $[0]$ est un vecteur nul de dimension appropriée et les profils après des changements $\tilde{\theta}_{k-L+1}^{k}(k_0) \in \mathbb{R}^{Ls}$ ne présentent pas d'intérêt.

Il est à noter que le paramètre de nuisance $x_{k-L+1}$ doit être éliminé de (B.14) afin d'éviter son impact négatif lors de la prise de décision. La réjection du paramètre de nuisance a été discutée dans [52] en appliquant la théorie des tests invariants. La méthode considérée dans [52] coïncide avec l'approche par espace de parité dans la communauté du diagnostic de défauts [30, 35]. L'idée est comme la suivante. Le vecteur $z_{k-L+1}^{k}$ dans (B.14) est projeté sur le complément orthogonal $R(\mathcal{C})^{\perp}$ de l'espace engendré par les colonnes $R(\mathcal{C})$ de la matrice $\mathcal{C}$ qui est supposé de rang plein. Le vecteur de résidus est calculé par $r_{k-L+1}^{k} = \mathcal{W}z_{k-L+1}^{k}$, où les rangs de la matrice $\mathcal{W} \in \mathbb{R}^{(Lp-n)\times Lp}$ se composent des vecteurs propres de la matrice de projection $\mathcal{P}_{\mathcal{C}}^{\perp} = \mathcal{I} - \mathcal{C}\left(\mathcal{C}^T\mathcal{C}\right)^{-1}\mathcal{C}^T$ correspondants aux valeurs propres 1, où $\mathcal{I}$ est la matrice d'identité de dimension appropriée. La matrice de réjection $\mathcal{W}$ satisfait des propriétés suivantes : $\mathcal{W}\mathcal{C} = 0$, $\mathcal{W}^T\mathcal{W} = \mathcal{P}_{\mathcal{C}}^{\perp}$ et $\mathcal{W}\mathcal{W}^T = \mathcal{I}$. Le modèle de résidus générés par l'approche d'espace de parité est donné par :

$$r_{k-L+1}^{k} = \mathcal{W}z_{k-L+1}^{k} = \mathcal{W}\mathcal{M}\theta_{k-L+1}^{k}(k_0) + \mathcal{W}\left(\mathcal{H}w_{k-L+1}^{k} + v_{k-L+1}^{k}\right). \tag{B.16}$$

Afin de développer le modèle statistique ressemblant à celui de (B.12), définissons le vecteur de profils transitoires $\varphi_{k-L+1}^{k}(k_0) = \mathcal{W}\mathcal{M}\theta_{k-L+1}^{k}(k_0) \in \mathbb{R}^{Lp-n}$ et le vecteur de bruits $\varsigma_{k-L+1}^{k} = \mathcal{W}\left(\mathcal{H}w_{k-L+1}^{k} + v_{k-L+1}^{k}\right) \in \mathbb{R}^{(Lp-n)\times(Lp-n)}$, respectivement. Le modèle statistique (B.16) se réduit à

$$r_{k-L+1}^{k} = \varphi_{k-L+1}^{k}(k_0) + \varsigma_{k-L+1}^{k}, \tag{B.17}$$

où le vecteur $\varsigma_{k-L+1}^{k} \sim \mathcal{N}(0, \Sigma_{\varsigma})$, où la matrice de covariance $\Sigma_{\varsigma} = \mathcal{W}\left(\mathcal{H}\mathcal{Q}\mathcal{H}^T + \mathcal{R}\right)\mathcal{W}^T$.

### B.3.3 Modèle statistique unifié des résidus

Dans cette section, nous développons le modèle statistique unifié de résidus générés par l'approche de filtre de Kalman et par l'approche par projection sur 'espace de parité. En intégrant (B.12)–(B.17), nous obtenons le modèle statistique unifié des résidus suivant :

$$r_{k-L+1}^{k} = \phi_{k-L+1}^{k}\left(k_0\right) + \xi_{k-L+1}^{k}, \tag{B.18}$$

où $\phi_{k-L+1}^{k}\left(k_0\right)$ est le vecteur de signaux transitoires et $\xi_{k-L+1}^{k} \sim \mathcal{N}\left(0, \Sigma\right)$ est le vecteur de bruits aléatoires. Pour l'approche du filtre de Kalman, les profils transitoires sont $\phi_{k-L+1}^{k}\left(k_0\right) = \psi_{k-L+1}^{k}\left(k_0\right)$ et les bruits aléatoires sont $\xi_{k-L+1}^{k} = \varrho_{k-L+1}^{k}$ avec la matrice de covariance $\Sigma = \Sigma_{\varrho}$. Pour l'approche avec l'espace de parité, les profils transitoires sont $\phi_{k-L+1}^{k}\left(k_0\right) = \varphi_{k-L+1}^{k}\left(k_0\right)$ et les bruits aléatoires sont $\xi_{k-L+1}^{k} = \varsigma_{k-L+1}^{k}$ avec la matrice de covariance $\Sigma = \Sigma_{\varsigma}$.

Dans ce manuscrit, nous proposons d'utiliser la distance de Kullback-Leibler (K-L) pour comparer les deux approches de génération des résidus. Il est bien connu [10] que les résidus avec la plus grande distance de K-L offrait de meilleures performances statistiques que ceux avec une plus petite distance de K-L.

Désignons par $\mathcal{P}_{k_0}$ (resp. $\mathcal{P}_0 \triangleq \mathcal{P}_{\infty}$) la distribution conjointe des résidus $r_1^L, r_2^{L+1}, \cdots, r_{k-L+1}^{k}, \cdots$ lorsqu'ils suivent le modèle statistique unifié (B.18). Désignons aussi par $\mathbb{E}_{k_0}$ (resp. $\mathbb{E}_0 \triangleq \mathbb{E}_{\infty}$) l'espérance mathématique correspondante. Dans le cas gaussien, les distances de K-L sont calculées par [10] :

$$\rho_{\mathrm{KF}} = \frac{1}{2}\left[\psi_1^L\left(1\right)\right]^T\left[\Sigma_{\varrho}^{-1}\right]\left[\psi_1^L\left(1\right)\right], \tag{B.19}$$

$$\rho_{\mathrm{PS}} = \frac{1}{2}\left[\varphi_1^L\left(1\right)\right]^T\left[\Sigma_{\varsigma}^{-1}\right]\left[\varphi_1^L\left(1\right)\right], \tag{B.20}$$

où $\rho_{\mathrm{KF}}$ et $\rho_{\mathrm{PS}}$ sont des distances de K-L des résidus générés, respectivement, par l'approche du filtre de Kalman et par l'approche avec l'espace de parité.

Nous considérons maintenant le problème de choix de la matrice de réjection $\mathcal{W}$ pour l'espace de parité. Les résultats principaux sont donnés dans Lemme (B.1).

**Lemme B.1.** *(Choix de la matrice de réjection). Soit $\mathcal{W} \in \mathbb{R}^{(Lp-n)\times n}$ la matrice de réjection telle que ses colonnes constituent une base (non nécessairement une base orthonormale) pour l'espace nul à gauche $R\left(\mathcal{C}\right)^{\perp}$ de la matrice $\mathcal{C}$, satisfaisant ainsi $\mathcal{W}\mathcal{C} = 0$. La distance de K-L*

$$\rho_{\mathrm{PS}} = \frac{1}{2}\left[\mathcal{M}\theta_1^L\left(1\right)\right]^T\left[\mathcal{W}^T\left(\mathcal{W}\mathcal{S}\mathcal{W}^T\right)^{-1}\mathcal{W}\right]\left[\mathcal{M}\theta_1^L\left(1\right)\right] \tag{B.21}$$

*ne dépend pas du choix de la matrice de réjection $\mathcal{W}$.*

*Démonstration.* La preuve de ce lemme peut être trouvée dans la version anglaise du manuscrit. □

Dans [72, 73], Gustafsson a proposé de rejeter les états inconnus du système par la méthode d'estimation des moindres carrés. La matrice de réjection $\mathcal{W}$ est choisie en tenant compte des matrices de covariance des bruits (de processus et de capteurs). Il a été discuté dans [72, 73] que cette méthode offrait des résidus avec une covariance minimale. Cependant, la covariance minimale ne garantit pas les performances statistiques de la procédure de détection en raison de

la projection. Les résultats du Lemme B.1 montrent que la méthode d'estimation des moindres carrés est autant efficace que l'approche par projection sur l'espace de parité proposée dans cette thèse.

Dans le cadre des tests statistiques, un problème analogue de détection optimale de défauts a été traité dans [51]. Un modèle linéaire avec des paramètres de nuisance et une matrice de covariance générale (pas nécessairement diagonale) a été considéré dans le contexte des paramètres de nuisance inconnus mais non-aléatoires. Deux tests invariants différents ont été conçus dans un tel cas. Le premier invariant statistique a été basé sur la connaissance de la matrice d'observation et de la matrice de covariance. Par contre, le deuxième invariant statistique a été conçu en considérant la matrice d'observation seulement. Il a été démontré dans [51] que les deux méthodes sont égales. Cette conclusion est cohérente avec les résultats du Lemme B.1.

## B.4 Algorithmes de détection pour des paramètres complètement connus

Cette section est organisée comme suit. L'algorithme de la Somme Cumulée à Fenêtre Limitée et Seuils Variables (VTWL CUSUM) est conçu dans la sous-section B.4.1. Ensuite, le problème d'optimisation est formulé et résolu dans la sous-section B.4.2. Il est démontré que le choix optimal des seuils conduit à la règle de détection de la Moyenne Glissante Finie (Finite Moving Average ou FMA). En outre, une méthode numérique est proposée dans la sous-section B.4.3 pour estimer la pire probabilité de fausse alarme et la pire probabilité de détection manquée. Finalement, la robustesse du test FMA par rapport à quelques paramètres est examinée dans la sous-section B.4.4.

### B.4.1 Algorithme de Somme Cumulée à Fenêtre Limitée et Seuils Variables

Dans cette sous-section, nous adaptons l'algorithme VTWL CUSUM, qui a été proposé par Guépié [67, 69] pour détecter des changements transitoires dans une séquence des variables gaussiennes indépendantes, au modèle statistique unifié (B.18). L'instant d'arrêt $T_{\text{VTWL}}$ du test VTWL CUSUM est défini directement comme suit :

$$T_{\text{VTWL}} = \inf \left\{ k \geq L : \max_{k-L+1 \leq i \leq k} \left( S_i^k - h_{k-i+1} \right) \geq 0 \right\}, \tag{B.22}$$

où $h_1, h_2, \cdots, h_L$ sont des seuils variables et $S_i^k$ est le logarithme du rapport de vraisemblance (LLR) qui est calculé dans le cas gaussien par :

$$S_i^k = \left[ \phi_{k-L+1}^k (i) \right]^T \left[ \Sigma^{-1} \right] \left[ r_{k-L+1}^k - \frac{1}{2} \phi_{k-L+1}^k (i) \right]. \tag{B.23}$$

L'algorithme VTWL CUSUM se déroule comme suit. Pour chaque instant $k \geq L$, l'algorithme utilise les dernières mesures $y_{k-L+1}, \cdots, y_k$ pour la prise de décision. Tout d'abord, les LLRs $S_i^k$ sont calculés à partir de (B.23) pour chaque indice $i$ de $k - L + 1$ à $k$. Ensuite, chaque LLR $S_i^k$ est comparé au seuil $h_{k-i+1}$ et l'instant d'alarme $T_{\text{VTWL}}$ est déclaré si l'un des LLRs est supérieur ou égal à son seuil correspondant. Les seuils variables $h_1, h_2, \cdots, h_L$ sont considérés comme les paramètres de réglage pour optimiser l'algorithme VTWL CUSUM par rapport au critère d'optimalité (B.4)–(B.5).

### B.4.2 Étude des performances statistiques du VTWL CUSUM

Cette sous-section est consacrée à l'étude des propriétés statistiques de l'algorithme VTWL CUSUM (B.22)–(B.23). Les propriétés de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ et de la pire probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ sont résumées dans le Théorème B.1.

**Théorème B.1.** *Considérons l'algorithme VTWL CUSUM défini par* (B.22)–(B.23)*. Alors,*

*1. La pire probabilité de fausse alarme pour une fenêtre de taille donnée $m_\alpha$ correspond à la première fenêtre $[L; L+m_\alpha-1]$, c-à-d*

$$\overline{\mathbb{P}}_{\text{fa}}\left(T_{\text{VTWL}}; m_\alpha; h_1, h_2, \cdots, h_L\right) = \mathbb{P}_0\left(L \leq T_{\text{VTWL}} \leq L+m_\alpha-1\right). \tag{B.24}$$

*2. La pire probabilité de détection manquée est bornée supérieurement par*

$$\overline{\mathbb{P}}_{\text{md}}\left(T_{\text{VTWL}}; h_1, h_2, \cdots, h_L\right) \leq \tilde{\mathbb{P}}_{\text{md}}\left(T_{VTWL}; h_L\right) \triangleq \Phi\left(\frac{h_L - \mu_{S_1^L}}{\sigma_{S_1^L}}\right), \tag{B.25}$$

*où $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}t^2\right\} dt$ est la fonction de répartition de la loi normale centrée réduite, $\tilde{\mathbb{P}}_{\text{md}}\left(T_{VTWL}; h_L\right)$ est la borne supérieure proposée pour la pire probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$, et les paramètres $\mu_{S_1^L}$ et $\sigma_{S_1^L}$ sont calculés par :*

$$\mu_{S_1^L} = \frac{1}{2}\left[\phi_1^L(1)\right]^T \left[\Sigma^{-1}\right]\left[\phi_1^L(1)\right], \tag{B.26}$$

$$\sigma^2_{S_1^L} = \left[\phi_1^L(1)\right]^T \left[\Sigma^{-1}\right]\left[\phi_1^L(1)\right]. \tag{B.27}$$

*Démonstration.* La preuve de ce théorème peut être trouvée dans la version anglaise. □

Il est à noter que la minimisation simultanée de la pire probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ et de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ est contradictoire. En outre, leur expression analytique n'est pas disponible en raison de la complexité mathématique. Pour ces raisons, nous proposons de minimiser la borne supérieure $\tilde{\mathbb{P}}_{\text{md}}\left(T_{\text{VTWL}}; h_L\right)$ sous la contrainte que la pire probabilité de fausse alarme de taille $m_\alpha$ soit bornée par une valeur prescrite $\alpha \in (0,1)$. Avant d'examiner le problème d'optimisation, nous imposons l'hypothèse suivante sur les profils de changement transitoire $\phi_1^L(1)$ qui sont définis dans (B.18).

**Hypothèse B.1.** *Supposons que le vecteur des profils $\phi_1^L(1)$ défini dans* (B.18) *est non-nul (c-à-d $\psi_1^L(1) \neq 0$ pour l'approche avec le filtre de Kalman et $\varphi_1^L(1) \neq 0$ pour l'approche avec projection sur l'espace de parité).*

L'hypothèse B.1 joue un rôle essentiel dans le choix des seuils de l'algorithme VTWL CUSUM. Cette hypothèse fournit la condition suffisante pour le lemme suivant.

**Lemme B.2.** *Soit $\mathcal{S} \in \mathbb{R}^{m_\alpha}$ le vecteur gaussien multidimensionnel qui se compose de $m_\alpha$ LLRs $S_1^L, S_2^{L+1}, \cdots, S_{m_\alpha}^{L+m_\alpha-1}$. Si l'hypothèse B.1 est satisfaite, la matrice de covariance $\Sigma_{\mathcal{S}} \in \mathbb{R}^{m_\alpha \times m_\alpha}$ du vecteur aléatoire $\mathcal{S}$ est définie positive.*

*Démonstration.* La preuve de ce lemme peut être trouvée dans la version anglaise du manuscrit. □

En exploitant les résultats du Lemme B.2, le choix optimal des seuils variables par rapport au critère d'optimalité (B.4)–(B.5) est formulé et résolu dans le Théorème B.2.

**Théorème B.2.** *Considérons l'algorithme VTWL CUSUM défini par* (B.22)–(B.23)*. Alors,*

*1. Le choix optimal des seuils variables $h_1, h_2, \cdots, h_L$ conduit au problème d'optimisation suivant :*

$$\begin{cases} \inf_{h_1,h_2,\cdots,h_L} & \tilde{\mathbb{P}}_{\mathrm{md}}\left(T_{VTWL}; h_L\right) \\ \textit{subject to} & \overline{\mathbb{P}}_{\mathrm{fa}}\left(T_{\mathrm{VTWL}}; m_\alpha; h_1, h_2, \cdots, h_L\right) \leq \alpha \end{cases}, \tag{B.28}$$

*où $\alpha \in (0,1)$ est une valeur prescrite pour la pire probabilité de fausse alarme dans une durée de taille $m_\alpha$. Le problème d'optimisation* (B.28) *possède la solution unique $(h_1^*, h_2^*, \cdots, h_L^*)$ pour une valeur donnée $\alpha \in (0,1)$, où $h_1^*, h_2^*, \cdots, h_L^* \to \infty$ et $h_L^*$ est calculé par l'équation suivante :*

$$\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\left\{S_{k-L+1}^k < h_L^*\right\}\right) = 1-\alpha. \tag{B.29}$$

*2. L'algorithme VTWL CUSUM optimisé est équivalent au test de la Moyenne Glissante Finie (FMA) suivante :*

$$T_{\mathrm{FMA}}\left(\tilde{h}_L\right) = \inf\left\{k \geq L : \left[\phi_1^L(1)\right]^T \left[\Sigma^{-1}\right] r_{k-L+1}^k \geq \tilde{h}_L\right\}, \tag{B.30}$$

*avec le seuil $\tilde{h}_L = h_L^* + \mu_{S_1^L}$. La borne supérieure pour la pire probabilité de détection manquée du test FMA* (B.30) *est calculée par :*

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(T_{FMA}; \tilde{h}_L\right) \leq \tilde{\mathbb{P}}_{\mathrm{md}}\left(T_{FMA}; \tilde{h}_L\right) \triangleq \Phi\left(\frac{\tilde{h}_L - 2\mu_{S_1^L}}{\sigma_{S_1^L}}\right). \tag{B.31}$$

*Démonstration.* La preuve de ce théorème peut être trouvée dans la version anglaise du manuscrit. □

### B.4.3 Calcul numérique des probabilités d'erreurs

Dans cette sous-section, nous proposons une méthode numérique pour estimer la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\mathrm{fa}}$ et la pire probabilité de détection manquée $\overline{\mathbb{P}}_{\mathrm{md}}$ du test FMA et du test VTWL CUSUM. La méthode proposée est basée sur le calcul numérique de la fonction de répartition d'une distribution multidimensionnelle introduite dans [63]. Notamment, cet algorithme a été mis en oeuvre dans « Matlab Statistics Toolbox » par la fonction MVNCDF.

**Proposition B.1.** *La pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\mathrm{fa}}$ et la pire probabilité de détection manquée $\overline{\mathbb{P}}_{\mathrm{md}}$ du test VTWL CUSUM* (B.22)–(B.23) *et du test FMA* (B.30) *sont calculées numériquement par les formules suivantes :*

*1. La pire probabilité de fausse alarme pour une fenêtre de taille $m_\alpha$ est calculée par :*

$$\overline{\mathbb{P}}_{\mathrm{fa}}\left(T_{\mathrm{VTWL}}; m_\alpha; h_1, h_2, \cdots, h_L\right) = 1 - \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\bigcap_{i=k-L+1}^{k}\left\{S_i^k < h_{k-i+1}\right\}\right), \tag{B.32}$$

$$\overline{\mathbb{P}}_{\mathrm{fa}}\left(T_{\mathrm{FMA}}; m_\alpha; \tilde{h}_L\right) = 1 - \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\left\{S_{k-L+1}^k < \tilde{h}_L - \mu_{S_1^L}\right\}\right). \tag{B.33}$$

*2. La pire probabilité de détection manquée est calculée par :*

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{VTWL}}; h_1, h_2, \cdots, h_L\right) = \sup_{k_0 \geq L} \frac{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0+L-1} \bigcap_{i=k-L+1}^{k} \left\{S_i^k < h_{k-i+1}\right\}\right)}{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0-1} \bigcap_{i=k-L+1}^{k} \left\{S_i^k < h_{k-i+1}\right\}\right)}, \tag{B.34}$$

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(T_{\mathrm{FMA}}; \tilde{h}_L\right) = \sup_{k_0 \geq L} \frac{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0+L-1} \left\{S_{k-L+1}^k < \tilde{h}_L - \mu_{S_1^L}\right\}\right)}{\mathbb{P}_{k_0}\left(\bigcap_{k=L}^{k_0-1} \left\{S_{k-L+1}^k < \tilde{h}_L - \mu_{S_1^L}\right\}\right)}. \tag{B.35}$$

*Démonstration.* La preuve de cette proposition peut être trouvée dans la version anglaise du manuscrit. □

**Remarque B.2.** *Les vecteurs des seuils, les vecteurs des moyennes et les matrices de covariance sont formulés dans la Proposition B.1 pour le calcul numérique des probabilités d'erreurs. Afin d'utiliser la fonction* MVNCDF *de Matlab, il est nécessaire de calculer les espérances mathématiques* $\mathbb{E}_0\left[S_i^k\right]$ *et* $\mathbb{E}_{k_0}\left[S_i^k\right]$ *et les covariances cov* $\left(S_{i_1}^{k_1}, S_{i_2}^{k_2}\right)$*. En outre, la méthode numérique nous permet d'estimer les probabilités d'erreurs au lieu de la méthode de simulation de Monte Carlo traditionnelle. Il est à noter que la méthode proposée est plus efficace que la simulation de Monte Carlo concernant le temps de calcul. En outre, cette méthode numérique sera exploitée pour étudier la robustesse du test FMA dans la sous-section B.4.4.*

**Remarque B.3.** *Le Théorème B.1 a montré que la pire probabilité de fausse alarme* $\overline{\mathbb{P}}_{\mathrm{fa}}$ *pour une durée donnée de taille* $m_\alpha$ *correspond exactement à la première fenêtre* $[L; L + m_\alpha - 1]$*. Donc, la pire probabilité de fausse alarme* $\overline{\mathbb{P}}_{\mathrm{fa}}$ *du test VTWL CUSUM* (B.22)–(B.23) *et du test FMA* (B.30) *peut être calculé en utilisant les équations* (B.32)–(B.33)*. En revanche, la pire probabilité de détection manquée* $\overline{\mathbb{P}}_{\mathrm{md}}$ *concerne l'opération « supremum » sur tous les points de changement* $k_0 \geq L$*. Autrement dit, la pire probabilité de détection manquée ne correspond pas à la première fenêtre* $[L; 2L-1]$*. Heureusement, les résultats des simulation montrent que la pire probabilité de détection manquée* $\mathbb{P}_{k_0}\left(T \geq k_0 + L | \, T \geq k_0\right)$ *tend vers les premières fenêtres, où* $T$ *est l'instant d'arrêt du test VTWL CUSUM et du test FMA. Pour cette raison, nous remplaçons l'opération « supremum » dans les équations* (B.34)–(B.35) *par l'operation « maximum » sur quelques premiers instants de changement* $k_0 \in [L, L + \delta L]$*, où* $\delta L \in \mathbb{N}^+$*, pour estimer la pire probabilité de détection manquée* $\overline{\mathbb{P}}_{\mathrm{md}}$*.*

### B.4.4 Analyse de sensibilité du test FMA

Dans cette sous-section, nous effectuons l'analyse de la sensibilité du test FMA (B.30) afin d'évaluer sa robustesse par rapport à plusieurs paramètres opérationnels : la durée d'attaque $L$, les profils d'attaque $\theta_1, \theta_2, \cdots, \theta_L$, la matrice de covariance $Q$ et la matrice de covariance $R$. Cette analyse de sensibilité est extrêmement importante dans des circonstances pratiques puisque ces paramètres opérationnels ne sont pas exactement connus.

Soient $\overline{L}$, $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$, $\overline{Q}$ et $\overline{R}$, respectivement, les vraies valeurs de la durée d'attaque, des profils d'attaque, de la covariance des bruits du processus et de la covariance des bruits de

capteurs. Il est à noter que les paramètres putatifs correspondants $L$, $\theta_1, \theta_2, \cdots, \theta_L$, $Q$ et $R$ restent intactes. La différence entre les vrais paramètres et les paramètres putatifs entraîne un changement dans le modèle statistique unifié (B.18). Heureusement, la méthode numérique proposée dans la sous-section B.4.3 peut être utilisée pour examiner la robustesse du test FMA par rapport aux paramètres opérationnels.

La pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}\left(T_{\text{FMA}}; m_\alpha; \tilde{h}_L\right)$ et la pire probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}\left(T_{\text{FMA}}; \tilde{h}_L\right)$ peuvent être estimées par les formules (B.33) et (B.35), respectivement. À cause de la différence entre les vrais paramètres et les paramètres putatifs (c-à-d $\overline{L} \neq L$, $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L \neq \theta_1, \theta_2, \cdots, \theta_L$, $\overline{Q} \neq Q$ et $\overline{R} \neq R$), les espérances mathématiques $\mathbb{E}_0\left[S_i^k\right]$ et $\mathbb{E}_{k_0}\left[S_i^k\right]$ et les covariances $\text{cov}\left(S_{i_1}^{k_1}, S_{i_2}^{k_2}\right)$ doivent être recalculées. Les espérances mathématiques $\mathbb{E}_0\left[S_i^k\right]$ et $\mathbb{E}_{k_0}\left[S_i^k\right]$ ne dépendent que des vraies valeurs de la durée d'attaque $\overline{L}$ et des profils d'attaque $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$. En revanche, les covariances $\text{cov}\left(S_{i_1}^{k_1}, S_{i_2}^{k_2}\right)$ dépendent seulement des matrices $\overline{Q}$ et $\overline{R}$. Ces calculs sont détaillés dans la version anglaise du manuscrit.

**Remarque B.4.** *Pour les scénarios où les vraies covariances des bruits sont différentes de leurs valeurs putatives, les innovations générées par le filtre de Kalman ne sont plus indépendantes. Le modèle statistique des résidus n'est plus valable. Pour cette raison, il est nécessaire de recalculer la covariance entre deux innovations dans tels scénarios que $\overline{Q} \neq Q$ et/ou $\overline{R} \neq R$. Cette tâche est réalisée par un algorithme récursif détaillé dans la version anglaise du manuscrit.*

## B.5 Algorithmes de détection pour des paramètres partiellement connus

Dans cette section, nous considérons un scénario plus réaliste où les profils d'attaque sont partiellement connus. Plus précisément, la « forme » des profils est connue mais la « magnitude » des profils est inconnue. Soient $\theta_1, \theta_2, \cdots, \theta_L$ les profils putatifs et $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$ les vrais profils. Ces derniers peuvent être exprimés en fonction des premiers par $\overline{\theta}_j = \gamma\theta_j$, où les profils putatifs sont connus mais le coefficient $\gamma$ est inconnu. L'approche du rapport de vraisemblance généralisé (GLR) et l'approche du rapport de vraisemblance pondéré (WLR) sont envisagées pour résoudre le problème.

### B.5.1 Approche du Rapport de Vraisemblance Généralisé

L'approche du rapport de vraisemblance généralisé (GLR) consiste à remplacer le paramètre inconnu $\gamma$ par son estimation du maximum de vraisemblance. Le logarithme du rapport de vraisemblance (LLR) généralisé $\hat{S}_i^k$ est calculé par :

$$\hat{S}_i^k = \sup_\gamma \left[\gamma\phi_{k-L+1}^k(i)\right]^T \left[\Sigma^{-1}\right] \left[r_{k-L+1}^k - \frac{1}{2}\gamma\phi_{k-L+1}^k(i)\right]. \tag{B.36}$$

Le LLR généralisé $\hat{S}_i^k$ peut être calculé, après quelques transformations simples, comme suit :

$$\hat{S}_i^k = \left[r_{k-L+1}^k\right]^T \left[\overline{\Sigma}(i)\right] \left[r_{k-L+1}^k\right], \tag{B.37}$$

où la matrice $\overline{\Sigma}(i)$, qui dépend de l'indice $i$, est calculée par :

$$\overline{\Sigma}(i) = \frac{\left[\Sigma^{-1}\right]\left[\phi_{k-L+1}^{k}(i)\right]\left[\phi_{k-L+1}^{k}(i)\right]^{T}\left[\Sigma^{-1}\right]}{2\left[\phi_{k-L+1}^{k}(i)\right]^{T}\left[\Sigma^{-1}\right]\left[\phi_{k-L+1}^{k}(i)\right]}. \tag{B.38}$$

L'algorithme VTWL GLR, qui utilise le LLR généralisé $\hat{S}_i^k$, est décrit par :

$$\hat{T}_{\mathrm{GLR}} = \inf\left\{k \geq L : \max_{k-L+1\leq i\leq k}\left(\hat{S}_i^k - h_{k-i+1}\right) \geq 0\right\}, \tag{B.39}$$

où les seuils variables $h_1, h_2, \cdots, h_L$ sont considérés comme les paramètres de réglage pour optimiser l'algorithme VTWL GLR.

### B.5.2 Approche du Rapport de Vraisemblance Pondéré

L'approche du rapport de vraisemblance pondéré (WLR) s'appuie sur l'hypothèse que le paramètre inconnu $\gamma$ est aléatoire et suit une distribution *a priori*. Le logarithme du rapport de vraisemblance (LLR) pondérée $\check{S}_i^k$ est calculé par :

$$\check{S}_i^k = \log\frac{\int\left[p_{\phi_{k-L+1}^{k}(i)}\left(r_{k-L+1}^{k}\right)\right]p_\gamma d\gamma}{p_0\left(r_{k-L+1}^{k}\right)}, \tag{B.40}$$

où $p_\gamma$ est la fonction de densité de paramètre inconnu $\gamma$.

Dans un souci de simplicité, supposons que le paramètre inconnu $\gamma$ suit la distribution uniforme $\mathcal{U}(\gamma_0, \gamma_1)$, où les bornes $0 < \gamma_0 < \gamma_1$ sont connues. Donc, la fonction de densité est donnée par $p_\gamma = {}^{1}/_{(\gamma_1-\gamma_0)}$. D'après quelques transformations, le LLR pondéré $\check{S}_i^k$ est donné par :

$$\begin{aligned}\check{S}_i^k &= \left[r_{k-L+1}^{k}\right]^{T}\left[\overline{\Sigma}(i)\right]\left[r_{k-L+1}^{k}\right] + \log\left[\frac{\sqrt{2\pi}}{b(i)(\gamma_1-\gamma_0)}\right] + \\ &\quad \log\left[\Phi\left(b(i)\gamma_1 - \frac{a(i)}{b(i)}\right) - \Phi\left(b(i)\gamma_0 - \frac{a(i)}{b(i)}\right)\right],\end{aligned} \tag{B.41}$$

où les coefficients $a(i)$ et $b(i)$ sont calculés par :

$$a(i) = \left[\phi_{k-L+1}^{k}(i)\right]^{T}\left[\Sigma^{-1}\right]\left[r_{k-L+1}^{k}\right], \tag{B.42}$$

$$b(i)^2 = \left[\phi_{k-L+1}^{k}(i)\right]^{T}\left[\Sigma^{-1}\right]\left[\phi_{k-L+1}^{k}(i)\right]. \tag{B.43}$$

La règle de détection VTWL WLR, qui utilise le LLR pondéré $\check{S}_i^k$, est décrite par :

$$\check{T}_{\mathrm{WLR}} = \inf\left\{k \geq L : \max_{k-L+1\leq i\leq k}\left(\check{S}_i^k - h_{k-i+1}\right) \geq 0\right\}, \tag{B.44}$$

où les seuils variables $h_1, h_2, \cdots, h_L$ sont considérés comme les paramètres de réglage pour optimiser l'algorithme VTWL WLR.

### B.5.3 Étude des performances statistiques du VTWL GLR et du VTWL WLR

Dans cette sous-section, nous examinons les performances statistiques du test VTWL GLR (B.39) et du test VTWL WLR (B.39). Les résultats principaux sont présentés dans le Théorème B.3 et le Théorème B.4.

**Théorème B.3.** *Considérons le test VTWL GLR défini dans* (B.39) *et le test VTWL WLR défini dans* (B.44)*, respectivement. Alors,*

*1. La pire probabilité de fausse alarme pour une durée de taille $m_\alpha$ dépend de la première fenêtre $[L; L+m_\alpha-1]$, c-à-d*

$$\overline{\mathbb{P}}_{\mathrm{fa}}\left(\hat{T}_{\mathrm{GLR}}\right) = \mathbb{P}_0\left(L \leq \hat{T}_{\mathrm{GLR}} \leq L+m_\alpha-1\right), \tag{B.45}$$

$$\overline{\mathbb{P}}_{\mathrm{fa}}\left(\check{T}_{\mathrm{WLR}}\right) = \mathbb{P}_0\left(L \leq \check{T}_{\mathrm{WLR}} \leq L+m_\alpha-1\right). \tag{B.46}$$

*2. La pire probabilité de détection manquée est bornée supérieurement par :*

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(\hat{T}_{\mathrm{GLR}}\right) \leq \tilde{\mathbb{P}}_{\mathrm{md}}\left(\hat{T}_{GLR}; h_L\right) = \mathbb{P}_1\left(\hat{S}_1^L < h_L\right), \tag{B.47}$$

$$\overline{\mathbb{P}}_{\mathrm{md}}\left(\check{T}_{\mathrm{WLR}}\right) \leq \tilde{\mathbb{P}}_{\mathrm{md}}\left(\check{T}_{WLR}; h_L\right) = \mathbb{P}_1\left(\check{S}_1^L < h_L\right), \tag{B.48}$$

*où $\tilde{\mathbb{P}}_{\mathrm{md}}\left(\hat{T}_{GLR}; h_L\right)$ et $\tilde{\mathbb{P}}_{\mathrm{md}}\left(\check{T}_{WLR}; h_L\right)$ sont les bornes supérieures pour la pire probabilité de détection manquée du test VTWL GLR et du test VTWL WLR, respectivement.*

*Démonstration.* La preuve de ce théorème peut être trouvée dans la version anglaise du manuscrit. □

Nous souhaitons minimiser la borne supérieure $\tilde{\mathbb{P}}_{\mathrm{md}}\left(\hat{T}_{\mathrm{GLR}}; h_L\right)$ (resp. $\tilde{\mathbb{P}}_{\mathrm{md}}\left(\check{T}_{\mathrm{WLR}}; h_L\right)$) sous la contrainte que la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\mathrm{fa}}\left(\hat{T}_{\mathrm{GLR}}\right)$ (resp. $\overline{\mathbb{P}}_{\mathrm{fa}}\left(\check{T}_{\mathrm{WLR}}\right)$) est bornée par une valeur prescrite $\alpha \in (0,1)$.

**Théorème B.4.** *Considérons le test VTWL GLR défini dans* (B.39) *et le test VTWL WLR défini dans* (B.44)*, respectivement. Alors,*

*1. Le choix optimal des seuils $h_1, h_2, \cdots, h_L$ se produit au problème d'optimisation suivant :*

$$\inf_{h_1,\cdots,h_L} \tilde{\mathbb{P}}_{\mathrm{md}}\left(\hat{T}_{GLR}; h_L\right) \text{ soumis à } \overline{\mathbb{P}}_{\mathrm{fa}}\left(\hat{T}_{\mathrm{GLR}}; m_\alpha; h_1, h_2, \cdots, h_L\right) \leq \alpha, \tag{B.49}$$

$$\inf_{h_1,\cdots,h_L} \tilde{\mathbb{P}}_{\mathrm{md}}\left(\check{T}_{WLR}; h_L\right) \text{ soumis à } \overline{\mathbb{P}}_{\mathrm{fa}}\left(\check{T}_{\mathrm{WLR}}; m_\alpha; h_1, h_2, \cdots, h_L\right) \leq \alpha, \tag{B.50}$$

*où $\alpha \in (0,1)$ est une valeur prescrite pour le taux de fausse alarme. Soient $\hat{h}_L^*$ et $\check{h}_L^*$, respectivement, les numéros réels minimum satisfaisant les inégalités suivantes :*

$$\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\left\{\hat{S}_{k-L+1}^k < \hat{h}_L^*\right\}\right) \geq 1-\alpha, \tag{B.51}$$

$$\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\left\{\check{S}_{k-L+1}^k < \check{h}_L^*\right\}\right) \geq 1-\alpha. \tag{B.52}$$

*Donc, les seuils optimaux du problème d'optimisation* (B.49) *(resp.* (B.50)*) sont* $\hat{h}_1^*, \cdots, \hat{h}_{L-1}^* \to +\infty$ *(resp.* $\check{h}, \cdots, \check{h}_{L-1}^* \to +\infty$*) et* $\hat{h}_L^*$ *(resp.* $\check{h}_L^*$*).*

*2. Le test VTWL GLR et le test VTWL WLR optimaux conduisent aux règles de décision FMA correspondantes :*

$$\hat{T}_{\mathrm{FMA}} = \inf\left\{k \geq L : \hat{S}_{k-L+1}^k \geq \hat{h}_L^*\right\}, \tag{B.53}$$

$$\check{T}_{\mathrm{FMA}} = \inf\left\{k \geq L : \check{S}_{k-L+1}^k \geq \check{h}_L^*\right\}, \tag{B.54}$$

*où* $\hat{T}_{\mathrm{FMA}}$ *est l'instant d'arrêt du test FMA GLR et* $\check{T}_{FMA}$ *l'instant d'arrêt du test FMA WLR, et les seuils* $\hat{h}_L^*$ *et* $\check{h}_L^*$ *sont choisis pour assurer des niveaux acceptables de fausses alarmes.*

*Démonstration.* La preuve de ce théorème peut être trouvée dans la version anglaise du manuscrit. □

**Remarque B.5.** *Permettons-nous d'ajouter quelques commentaires sur les résultats du Théorème B.3 et du Théorème B.4. L'estimation numérique de la probabilité de fausse alarme et de la probabilité de détection manquée du test FMA GLR donnée dans* (B.53) *et du test FMA WLR donnée dans* (B.54) *sont difficilement calculables sous forme analytiques. Pour cette raison, nous examinons les performances statistiques du test FMA GLR et du test FMA WLR, en se basant sur une simulation de Monte Carlo, dans la section B.7.*

## B.6 Extension au problème de localisation

Dans cette section, nous formulons le problème d'identification d'attaques cyber-physiques dans les systèmes SCADA comme un problème de détection-localisation conjointe de changements transitoires dans des systèmes stochastiques et dynamiques. Cette section est organisée comme suit. La formulation du problème est présentée dans la sous-section B.6.1. Dans la sous-section B.6.2, nous développons le modèle statistique unifié pour le problème de détection-localisation conjointe de changements transitoires. En s'appuyant sur ce modèle, quelques algorithmes de détection-localisation sont proposés dans la sous-section B.6.3. Finalement, nous étudions dans la sous-section B.6.4 les performances statistiques du test FMA.

### B.6.1 Formulation du problème

De façon similaire au problème de détection, le modèle d'espace d'état à temps discret suivant est employé pour décrire les systèmes SCADA attaqués :

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + B_a a_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + D_a a_k + v_k \end{cases}; \quad x_1 = \overline{x}_1, \tag{B.55}$$

où $x_k \in \mathbb{R}^n$ est le vecteur d'états, $u_k \in \mathbb{R}^m$ est le vecteur de signaux de contrôle, $d_k \in \mathbb{R}^q$ est le vecteur des perturbations, $y_k \in \mathbb{R}^p$ est le vecteur des mesures des capteurs, $a_k \in \mathbb{R}^s$ est le vecteur d'attaque, $w_k \in \mathbb{R}^n$ est le vecteur des bruits de processus, et $v_k \in \mathbb{R}^p$ est le vecteur des bruits des capteurs ; les matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $F \in \mathbb{R}^{n\times q}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, $G \in \mathbb{R}^{p\times q}$, $B_a \in \mathbb{R}^{n\times s}$ et $D_a \in \mathbb{R}^{p\times s}$ sont connues. Les signaux de contrôle $u_k$ et les perturbations

$d_k$ sont connus également. Les bruits des processus $w_k$ et les bruits des capteurs $v_k$ sont des vecteurs gaussiens multidimensionnels centrés réduits indépendants, c-à-d $\mathrm{cov}\,(w_k, w_l) = Q\delta_{kl}$, $\mathrm{cov}\,(v_k, v_l) = R\delta_{kl}$ et $\mathrm{cov}\,(w_k, v_l) = 0$, où $\delta_{kl} = 1$ si $k = l$ et $\delta_{kl} = 0$ autrement.

Supposons que les actes malveillants sont effectuées pendant une période finie $\tau_a = [k_0, k_0 + L - 1]$, où $k_0$ est l'instant d'attaque inconnu et $L$ est la durée d'attaque connue. Pour le problème de détection et de localisation, nous avons $K$ profils d'attaque différents où chaque profil d'attaque est associé à un scénario d'attaque spécifique. Donc, le vecteur d'attaque $a_k$ s'écrit :

$$a_k = \begin{cases} 0 & \text{si} \quad k < k_0 \\ \theta_{k-k_0+1}\,(l) & \text{si} \quad k_0 \leq k < k_0 + L\,, \\ 0 & \text{si} \quad k \geq k_0 + L \end{cases} \tag{B.56}$$

où $l$, for $1 \leq l \leq K$, est le type d'attaque et $K$ est le nombre d'hypothèses. Les profils d'attaque $\theta_1\,(l)\,, \theta_2\,(l)\,, \cdots, \theta_L\,(l)$ du type $l$, pour $1 \leq l \leq K$, sont connus.

**Définition B.1.** *Un algorithme de détection et de localisation de changements doit calculer un couple $(T, \nu)$ en s'appuyant sur des observations $y_1, y_2, \cdots$, où $T > 0$ est l'instant d'arrêt auquel la décision finale $\nu$, pour $1 \leq \nu \leq K$, est décidée.*

Le problème est de proposer des algorithmes pour détecter et localiser un changement transitoire dans le modèle (B.55)–(B.56) en satisfaisant certains critères d'optimalité. Plusieurs critères d'optimalité ont été proposés pour évaluer la performance statistique d'un algorithme de détection-localisation de changements brusques dans un système stochastique. Les critères classiques visent à minimiser le retard moyen pour détection-localisation soumis aux niveaux acceptables de fausses alarmes et de fausses localisations (voir, par exemple, [104,128–130,132]). Pour les infrastructures à sécurité critique [127], il est essentiel de minimiser la pire probabilité de détection-localisation manquée pour des valeurs acceptables de fausse alarmes/localisations.

Dans ce manuscrit, nous proposons un nouveau critère d'optimalité pour le problème de détection et de localisation de changements transitoires ainsi que pour la surveillance en-ligne de infrastructures à sécurité critique. Le critère d'optimalité consiste à minimiser la pire probabilité de détection-isolation manquée soumis à des niveaux acceptables sur la pire probabilité de fausse alarme pour une fenêtre de taille donnée et sur la pire probabilité de fausse localisation pour la fenêtre transitoire. Ce critère d'optimalité est parfaitement approprié au problème de détection et de localisation d'actes malveillantes dans des systèmes SCADA.

Soient $\overline{\mathbb{P}}_{\mathrm{md}}\,(T; L)$ la pire probabilité de détection manquée, $\overline{\mathbb{P}}_{\mathrm{fa}}\,(T; m_\alpha)$ la pire probabilité de fausse alarme pour une fenêtre de taille $m_\alpha$ et $\overline{\mathbb{P}}_{\mathrm{fi}}\,(T; L)$ la pire probabilité de fausse localisation pendant la durée transitoire. La probabilité de fausse alarme et la probabilité de fausse localisation sont définies, respectivement, par :

$$\overline{\mathbb{P}}_{\mathrm{fa}}\,(T; m_\alpha) = \sup_{l_0 \geq L} \mathbb{P}_0\,(l_0 \leq T < l_0 + m_\alpha)\,, \tag{B.57}$$

$$\overline{\mathbb{P}}_{\mathrm{fi}}\,(T; L) = \sup_{k_0 \geq L} \max_{1 \leq l \leq K} \mathbb{P}^l_{k_0}\,(k_0 \leq T < k_0 + L; \nu \neq l)\,, \tag{B.58}$$

où $\mathbb{P}_0$ est la probabilité correspondante au mode de fonctionnement normal du système et $\mathbb{P}^l_{k_0}$ représente la probabilité correspondante à l'instant de changement $k_0$ et au type de changement $l$. Le critère d'optimalité vise à minimiser la pire probabilité de détection manquée :

$$\inf_{T \in C_\alpha} \left\{ \overline{\mathbb{P}}_{\mathrm{md}}\,(T; L) = \sup_{k_0 \geq L} \max_{1 \leq l \leq K} \mathbb{P}^l_{k_0}\,(T - k_0 + 1 > L | T \geq k_0) \right\} \tag{B.59}$$

parmi tous les instants d'arrêt $T$ dans la classe $C_\alpha$ satisfaisant :

$$C_\alpha = \left\{ T : \overline{\mathbb{P}}_{\text{fa}}\left(T; m_\alpha\right) \leq \alpha; \overline{\mathbb{P}}_{\text{fi}}\left(T; L\right) \leq \alpha \right\}, \tag{B.60}$$

où $\alpha \in (0,1)$ est une valeur prescrite.

### B.6.2 Modèle statistique unifié pour le problème de localisation

Dans cette sous-section, nous développons le modèle statistique unifié des résidus générés avec l'approche de filtre de Kalman et avec l'approche par projection sur l'espace de parité pour le problème de détection et de localisation.

#### Approche avec le filtre de Kalman en régime permanent

Considérons le filtre de Kalman en régime permanent pour générer une séquence des résidus. Le filtre de Kalman est présenté dans (B.6)–(B.8). Soit $\{\varrho_k\}_{k\geq 1} \in \mathbb{R}^p$ une séquence des vecteurs gaussiens multidimensionnels centrés réduits indépendants avec la matrice de covariance $J \triangleq CP_\infty C^T + R$. Le modèle statistique de résidus est donc décrit par :

$$r_k = \begin{cases} \varrho_k & \text{si} \quad k < k_0 \\ \psi_{k-k_0+1}\left(l\right) + \varrho_k & \text{si} \quad k_0 \leq k < k_0 + L \\ \tilde{\psi}_k\left(l\right) + \varrho_k & \text{si} \quad k \geq k_0 + L \end{cases}, \tag{B.61}$$

où les profils transitoires $\psi_1\left(l\right), \psi_2\left(l\right), \cdots, \psi_L\left(l\right) \in \mathbb{R}^p$ sont calculés à partir des profils d'attaque $\theta_1\left(l\right), \theta_2\left(l\right), \cdots, \theta_L\left(l\right)$ du type $l$ avec l'équation suivante :

$$\begin{cases} \epsilon_{k+1} & = \left(A - AK_\infty C\right)\epsilon_k + \left(B_a - AK_\infty D_a\right)\theta_k\left(l\right) \\ \psi_k\left(l\right) & = C\epsilon_k + D_a\theta_k\left(l\right) \end{cases}; \quad \epsilon_1 = 0, \tag{B.62}$$

et les profils après les changements $\tilde{\psi}_k\left(l\right)$ (c-à-d pour $k \geq k_0 + L$) ne présentent pas d'intérêt.

De façon similaire au problème de détection, soit $r_{k-L+1}^k = \left[r_{k-L+1}^T, \cdots, r_k^T\right]^T \in \mathbb{R}^{Lp}$ le vecteur concaténé des innovations, $\varrho_{k-L+1}^k = \left[\varrho_{k-L+1}^T, \cdots, \varrho_k^T\right] \in \mathbb{R}^{Lp}$ le vecteur concaténé des bruits aléatoires, et $\psi_{k-L+1}^k\left(k_0, l\right) \in \mathbb{R}^{Lp}$ le vecteur des changements transitoires. Le vecteur $\psi_{k-L+1}^k\left(k_0, l\right)$ dépend de la position relative entre l'instant de changement $k_0$ dans la fenêtre $[k-L+1, k]$ et du type de changement $l$ par la relation suivante :

$$\psi_{k-L+1}^k\left(k_0, l\right) = \begin{cases} [0] & \text{si} \quad k < k_0 \\ \begin{bmatrix} [0] \\ \psi_1\left(l\right) \\ \vdots \\ \psi_{k-k_0+1}\left(l\right) \end{bmatrix} & \text{si} \quad k_0 \leq k < k_0 + L \\ \left[\tilde{\psi}_{k-L+1}^k\left(k_0, l\right)\right] & \text{si} \quad k \geq L \end{cases}, \tag{B.63}$$

où les profils après les changements $\tilde{\psi}_{k-L+1}^k\left(k_0, l\right) \in \mathbb{R}^{Lp}$ ne présentent aucun intérêt. Le modèle statistique de résidus est décrit par :

$$r_{k-L+1}^k = \psi_{k-L+1}^k\left(k_0, l\right) + \varrho_{k-L+1}^k, \tag{B.64}$$

où les bruits $\varrho_{k-L+1}^{k} \sim \mathcal{N}(0, \Sigma_{\varrho})$, où $\Sigma_{\varrho} = \mathrm{diag}(J) \in \mathbb{R}^{Lp \times Lp}$ est la matrice diagonale par blocs $J$.

### Approche par projection sur l'espace de parité de taille fixe

Considérons maintenant l'approche basée sur l'espace de parité pour générer une séquence des résidus. De façon similaire au problème de détection, le modèle statistique d'observations simplifiées est donné par :

$$z_{k-L+1}^{k} = \mathcal{C} x_{k-L+1} + \mathcal{M} \theta_{k-L+1}^{k}(k_0, l) + \mathcal{H} w_{k-L+1}^{k} + v_{k-L+1}^{k}, \tag{B.65}$$

où $z_{k-L+1}^{k} \in \mathbb{R}^{Lp}$ est le vecteur concaténé des mesures de capteurs, $w_{k-L+1}^{k} \in \mathbb{R}^{Ln}$ est le vecteur concaténé des bruits de processus, $v_{k-L+1}^{k} \in \mathbb{R}^{Lp}$ est le vecteur concaténé des bruits de capteurs, $\theta_{k-L+1}^{k}(k_0, l) \in \mathbb{R}^{Ls}$ est le vecteur concaténé des signaux transitoires ; les matrices $\mathcal{C} \in \mathbb{R}^{Lp \times n}$, $\mathcal{M} \in \mathbb{R}^{Lp \times Ls}$ et $\mathcal{H} \in \mathbb{R}^{Lp \times Ln}$. Les bruits des processus $w_{k-L+1}^{k} \sim \mathcal{N}(0, \mathcal{Q})$ et les bruits des capteurs $v_{k-L+1}^{k} \sim \mathcal{N}(0, \mathcal{R})$, où $\mathcal{Q} = \mathrm{diag}(Q) \in \mathbb{R}^{Ln \times Ln}$ et $\mathcal{R} = \mathrm{diag}(R) \in \mathbb{R}^{Lp \times Lp}$ sont les matrices diagonales par blocs $Q$ et $R$, respectivement.

Pareillement au vecteur $\psi_{k-L+1}^{k}(k_0, l)$ défini dans (B.64), le vecteur de profils transitoires $\theta_{k-L+1}^{k}(k_0, l)$ dépend de la position relative de l'instant de rupture $k_0$ dans la fenêtre $[k-L+1, k]$ et du type de changement $l$ par la relation suivante :

$$\theta_{k-L+1}^{k}(k_0, l) = \begin{cases} [0] & \text{si} \quad k < k_0 \\ \begin{bmatrix} [0] \\ \theta_1(l) \\ \vdots \\ \theta_{k-k_0+1}(l) \end{bmatrix} & \text{si} \quad k_0 \leq k < k_0 + L \\ \left[\tilde{\theta}_{k-L+1}^{k}(k_0, l)\right] & \text{si} \quad k \geq k_0 + L \end{cases}, \tag{B.66}$$

où les profils après les changements $\tilde{\theta}_{k-L+1}^{k}(k_0, l) \in \mathbb{R}^{Ls}$ ne présentent aucun intérêt.

Pareillement au problème de détection, le vecteur de résidus est obtenu par projection du vecteur d'observations simplifiées $z_{k-L+1}^{k}$ sur l'espace orthogonal $R(\mathcal{C})^{\perp}$ aux colonnes $R(\mathcal{C})$ de la matrice $\mathcal{C}$. Le modèle statistique des résidus est :

$$r_{k-L+1}^{k} = \varphi_{k-L+1}^{k}(k_0, l) + \varsigma_{k-L+1}^{k}, \tag{B.67}$$

avec les profils transitoires $\varphi_{k-L+1}^{k}(k_0, l) = \mathcal{W}\mathcal{M}\theta_{k-L+1}^{k}(k_0, l)$, les bruits aléatoires $\varsigma_{k-L+1}^{k} = \mathcal{W}\left(\mathcal{H} w_{k-L+1}^{k} + v_{k-L+1}^{k}\right)$ avec la matrice $\Sigma_{\varsigma} = \mathcal{W}\left(\mathcal{H}\mathcal{Q}\mathcal{H}^{T} + \mathcal{R}\right)\mathcal{W}^{T}$. La matrice de réjection $\mathcal{W}$ satisfait les conditions suivantes : $\mathcal{W}\mathcal{C} = 0$, $\mathcal{W}^{T}\mathcal{W} = \mathcal{P}_{\mathcal{C}}^{\perp}$ et $\mathcal{W}\mathcal{W}^{T} = \mathcal{I}$.

### Modèle statistique unifié des résidus

En combinant (B.64) et (B.67), nous obtenons le modèle statistique unifié des résidus générés par les deux approches, c-à-d l'approche de filtre de Kalman et l'approche d'espace de parité, comme suit :

$$r_{k-L+1}^{k} = \phi_{k-L+1}^{k}(k_0, l) + \xi_{k-L+1}^{k}, \tag{B.68}$$

où $\phi_{k-L+1}^{k}(k_0,l)$ est le vecteur des profils transitoires, et $\xi_{k-L+1}^{k} \sim \mathcal{N}(0,\Sigma)$ est le vecteur des bruits aléatoires. Pour l'approche avec le filtre de Kalman, les profils transitoires sont $\phi_{k-L+1}^{k}(k_0,l) = \psi_{k-L+1}^{k}(k_0,l)$ et les bruits aléatoires sont $\xi_{k-L+1}^{k} = \varrho_{k-L+1}^{k}$ avec $\Sigma = \Sigma_{\varrho}$. En revanche, les profils transitoires sont $\phi_{k-L+1}^{k}(k_0,l) = \varphi_{k-L+1}^{k}(k_0,l)$ et les bruits aléatoires sont $\xi_{k-L+1}^{k} = \varsigma_{k-L+1}^{k}$ avec $\Sigma = \Sigma_{\varsigma}$ pour l'approche par projection sur l'espace de parité.

Nous utilisons la distance de Kullback-Leibler (K-L) pour comparer les méthodes de génération de résidus. Soient $\mathcal{P}_{k_0}^{l}$ (resp. $\mathcal{P}_0 \triangleq \mathcal{P}_\infty \triangleq \mathcal{P}_{k_0}^{0}$) la distribution conjointe des résidus $r_1^L, r_2^{L+1}, \cdots, r_{k-L+1}^{k}, \cdots$ lorsqu'ils suivent le modèle statistique unifié (B.68), et $\mathbb{E}_{k_0}^{l}$ (resp. $\mathbb{E}_0 \triangleq \mathbb{E}_\infty \triangleq \mathbb{E}_{k_0}^{0}$) l'espérance mathématique correspondante. Dans le cas gaussien, les distances de K-L sont calculées, respectivement, pour l'approche avec le filtre de Kalman et pour l'approche basée sur l'espace de parité :

$$\rho_{\mathrm{KF}}(j,l) = \frac{1}{2}\left[\psi_1^L(1,l) - \psi_1^L(1,j)\right]^T \left[\Sigma_{\varrho}^{-1}\right] \left[\psi_1^L(1,l) - \psi_1^L(1,j)\right], \tag{B.69}$$

$$\rho_{\mathrm{PS}}(j,l) = \frac{1}{2}\left[\varphi_1^L(1,l) - \varphi_1^L(1,j)\right]^T \left[\Sigma_{\varsigma}^{-1}\right] \left[\varphi_1^L(1,l) - \varphi_1^L(1,j)\right], \tag{B.70}$$

où $\rho_{\mathrm{KF}}(j,l)$ et $\rho_{\mathrm{PS}}(j,l)$ sont les distances de K-L entre $\mathcal{P}_1^j$ et $\mathcal{P}_1^l$ des résidus générés par l'approche avec le filtre de Kalman et l'approche basée sur l'espace de parité, respectivement.

### B.6.3 Algorithmes de détection-localisation conjointe

Dans cette section, nous considérons plusieurs procédures pour la détection-localisation conjointe des changements transitoires en nous basant sur modèle statistique unifié (B.68).

#### Algorithme WL CUSUM généralisé

Pour le problème de détection et de localisation de changements brusques dans un système stochastique, Nikiforov [130] et Lai [104] ont proposés, respectivement, l'algorithme CUSUM généralisé et l'algorithme WL CUSUM généralisé. Pour la surveillance en-ligne, l'algorithme WL CUSUM peut être adapté au modèle statistique unifié (B.68). Définissons directement l'algorithme WL CUSUM généralisé $\delta_{\mathrm{GWL}} = (T_{\mathrm{GWL}}, \nu_{\mathrm{GWL}})$, qui utilise les dernières $L$ observations à chaque instant $k \geq L$, comme suit :

$$T_{\mathrm{GWL}} = \inf\left\{k \geq L : \max_{1\leq l\leq K} \max_{k-L+1\leq i\leq k} \min_{0\leq j\neq l\leq K} \left(S_i^k(l,j) - h\right) \geq 0\right\}, \tag{B.71}$$

$$\nu_{\mathrm{GWL}} = \arg \max_{1\leq l\leq K} \max_{T_{\mathrm{GWL}}-L+1\leq i\leq T_{\mathrm{GWL}}} \min_{0\leq j\neq l\leq K} S_i^{T_{\mathrm{GWL}}}(l,j), \tag{B.72}$$

où $h$ est le seuil et $S_i^k(l,j)$, pour $k-L+1 \leq i \leq k$, $1 \leq l \leq K$ et $0 \leq j \neq l \leq K$, est le logarithme du rapport de vraisemblance (LLR), qui est calculé dans le cas gaussien par :

$$S_i^k(l,j) = \left[\left(\phi_{k-L+1}^k(i,l) - \phi_{k-L+1}^k(i,j)\right)\right]^T \left[\Sigma^{-1}\right] \left[r_{k-L+1}^k - \frac{\phi_{k-L+1}^k(i,l) + \phi_{k-L+1}^k(i,j)}{2}\right]. \tag{B.73}$$

Le test WL CUSUM généralisé (B.71)–(B.72) fonctionne de la façon suivante. À chaque instant $k \geq L$, le test WL CUSUM généralisé (B.71)–(B.72) utilise les dernières $L$ observations pour la prise de décision. Tout d'abord, le modèle statistique unifié (B.68) est développé en s'appuyant

sur la génération des résidus. Ensuite, pour chaque indice $i$ de $k-L+1$ à $k$, les LLRs $S_i^k(l,j)$, pour $1 \leq l \leq K$ et $0 \leq j \neq l \leq K$, sont calculés. L'instant d'arrêt $T_{\mathrm{GWL}}$ est déclaré s'il existe $l$, pour $1 \leq l \leq K$, et qu'il existe au moins un indice $i \in [k-L+1, k]$ tel que tous les LLRs $S_i^k(l,j)$, pour $0 \leq j \neq l \leq K$, sont supérieurs ou égaux au seuil $h$.

### Algorithme WL CUSUM par matrice

L'algorithme CUSUM par matrice a été proposé dans [138] en modifiant l'algorithme CUSUM généralisé pour obtenir une forme récursive. L'algorithme WL CUSUM par matrice $\delta_{\mathrm{MWL}} = (T_{\mathrm{MWL}}, \nu_{\mathrm{MWL}})$, qui utilise les dernières $L$ observations à chaque instant $k \geq L$, est défini par :

$$T_{\mathrm{MWL}} = \inf\left\{k \geq L : \max_{1\leq l\leq K} \min_{0\leq j\neq l\leq K} \max_{k-L+1\leq i\leq k} \left(S_i^k(l,j) - h\right) \geq 0\right\}, \tag{B.74}$$

$$\nu_{\mathrm{MWL}} = \arg \max_{1\leq l\leq K} \min_{0\leq j\neq l\leq K} \max_{T_{\mathrm{MWL}}-L+1\leq i\leq T_{\mathrm{MWL}}} S_i^{T_{\mathrm{MWL}}}(l,j), \tag{B.75}$$

où $h$ est le seuil et les LLRs $S_i^k(l,j)$, pour $k-L+1 \leq i \leq k$, $1 \leq l \leq K$ et $0 \leq j \neq l \leq K$, sont calculés dans (B.73).

**Remarque B.6.** *L'algorithme WL CUSUM par matrice* (B.74)–(B.75) *fonctionne de la même façon que l'algorithme WL CUSUM généralisé* (B.71)–(B.72) *à l'exception du remplacement de l'opération « max-min » dans* (B.71)–(B.72) *par l'opération « min-max » dans* (B.74)–(B.75).

### Algorithme WL CUSUM par vecteur

L'algorithme WL CUSUM par vecteur est obtenu en remplaçant la statistique $\max_{k-L+1\leq i\leq k} S_i^k(l,j)$ dans l'algorithme WL CUSUM par matrice (B.74)–(B.75) par la statistique suivante :

$$g_k(l,j) = \max_{k-L+1\leq i\leq k} S_i^k(l,0) - \max_{k-L+1\leq i\leq k} S_i^k(j,0). \tag{B.76}$$

L'algorithme WL CUSUM par vecteur $\delta_{\mathrm{VWL}} = (T_{\mathrm{VWL}}, \nu_{\mathrm{VWL}})$ est donc défini comme suit :

$$T_{\mathrm{VWL}} = \inf\left\{k \geq L : \max_{1\leq l\leq K} \min_{0\leq j\neq l\leq K} \left(g_k(l,j) - h\right) \geq 0\right\}, \tag{B.77}$$

$$\nu_{\mathrm{VWL}} = \arg \max_{1\leq l\leq K} \min_{0\leq j\neq l\leq K} g_{T_{\mathrm{VWL}}}(l,j), \tag{B.78}$$

où $h$ est le seuil et les LLRs $S_i^k(l,j)$, pour $k-L+1 \leq i \leq k$, $1 \leq l \leq K$ et $0 \leq j \neq l \leq K$, sont calculés dans (B.73).

### Algorithme à Moyenne Glissante Finie (FMA)

La version FMA du test WL CUSUM généralisé, du test WL CUSUM par matrice et du test WL CUSUM par vecteur, est décrite par :

$$T_{\mathrm{FMA}} = \inf\left\{k \geq L : \max_{1\leq l\leq K} \min_{0\leq j\neq l\leq K} \left(S_{k-L+1}^k(l,j) - h\right) \geq 0\right\}, \tag{B.79}$$

$$\nu_{\mathrm{FMA}} = \arg \max_{1\leq l\leq K} \min_{0\leq j\neq l\leq K} S_{T_{\mathrm{FMA}}-L+1}^{T_{\mathrm{FMA}}}(l,j), \tag{B.80}$$

où $h$ est le seuil et les LLRs $S_{k-L+1}^k(l,j)$, pour $1 \leq l \leq K$ et $0 \leq j \neq l \leq K$, sont calculés dans (B.73).

**Remarque B.7.** *Il est à noter que la règle de décision FMA* (B.79)–(B.80) *est la généralisation du test FMA* (B.30) *pour le problème de détection. Elle est également la version FMA du test WL CUSUM généralisé* (B.71)–(B.72)*, du test WL CUSUM par matrice* (B.74)–(B.75) *et du test WL CUSUM par vecteur* (B.77)–(B.78)*. Les performances statistiques du test FMA* (B.79)–(B.80) *seront examinées dans la sous-section suivante.*

### B.6.4 Étude des performances statistiques du FMA

Dans cette sous-section, nous étudions les performances statistiques de la règle de détection FMA (B.79)–(B.80). Surtout, nous calculons les bornes supérieures pour la pire probabilité de fausse alarme, pour la pire probabilité de fausse localisation et pour la pire probabilité de détection-localisation manquée. Les résultats principaux sont présentés dans le Théorème B.5.

**Théorème B.5.** *Considérons le test FMA* (B.79)–(B.80)*. Soient* $\tilde{\mathbb{P}}_{\text{fa}}(\delta_{\text{FMA}})$*,* $\tilde{\mathbb{P}}_{\text{fi}}(\delta_{\text{FMA}})$*, et* $\tilde{\mathbb{P}}_{\text{md}}(\delta_{\text{FMA}})$*, respectivement, les bornes supérieures pour* $\overline{\mathbb{P}}_{\text{fa}}(\delta_{\text{FMA}})$*,* $\overline{\mathbb{P}}_{\text{fi}}(\delta_{\text{FMA}})$*, et* $\overline{\mathbb{P}}_{\text{md}}(\delta_{\text{FMA}})$*. Alors,*

*1. La pire probabilité de fausse alarme pour une fenêtre de taille* $m_\alpha$ *dépend de la première fenêtre* $[L; L+m_\alpha-1]$*, c-à-d*

$$\overline{\mathbb{P}}_{\text{fa}}(\delta_{\text{FMA}}; m_\alpha; h) = \mathbb{P}_0(L \leq T_{\text{FMA}} \leq L+m_\alpha-1), \tag{B.81}$$

*et elle est bornée supérieurement par :*

$$\overline{\mathbb{P}}_{\text{fa}}(\delta_{\text{FMA}}; m_\alpha; h) \leq \tilde{\mathbb{P}}_{\text{fa}}(\delta_{FMA}; m_\alpha; h) \triangleq 1 - \mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\bigcap_{l=1}^{K}\left\{S^k_{k-L+1}(l,0) < h\right\}\right). \tag{B.82}$$

*2. La pire probabilité de localisation pendant la fenêtre de changement dépend de la première fenêtre* $[L; 2L-1]$*, c-à-d*

$$\overline{\mathbb{P}}_{\text{fi}}(\delta_{\text{FMA}}; L; h) = \max_{1\leq l\leq K} \mathbb{P}^l_L(L \leq T_{\text{FMA}} < 2L; \nu_{\text{FMA}} \neq l), \tag{B.83}$$

*et elle est bornée supérieurement dans le cas du seuil* $h \geq 0$ *par :*

$$\overline{\mathbb{P}}_{\text{fi}}(\delta_{\text{FMA}}; L; h) \leq \tilde{\mathbb{P}}_{\text{fi}}(\delta_{FMA}; L; h) \triangleq \max_{1\leq l\leq K}\left[1 - \max_{0\leq \tilde{j}\leq K}\mathbb{P}^l_L\left(\bigcap_{k=L}^{2L-1}\bigcap_{\substack{j=1\\ j\neq\tilde{j},l}}^{K}\left\{S^k_{k-L+1}\left(j,\tilde{j}\right) < h\right\}\right)\right]. \tag{B.84}$$

*3. La pire probabilité de détection-localisation manquée est bornée supérieurement par :*

$$\overline{\mathbb{P}}_{\text{md}}(\delta_{\text{FMA}}; L; h) \leq \tilde{\mathbb{P}}_{\text{md}}(T_{FMA}; L; h) \triangleq \max_{1\leq l\leq K}\sum_{\substack{j=0\\ j\neq l}}^{K}\Phi\left(\frac{h-\mu_{S^L_1(l,j)}}{\sigma_{S^L_1(l,j)}}\right), \tag{B.85}$$

*où* $\mu_{S^L_1(l,j)}$ *et* $\sigma_{S^L_1(l,j)}$ *sont calculés par :*

$$\mu_{S^L_1(l,j)} = \frac{1}{2}\left[\phi^L_1(1,l) - \phi^L_1(1,j)\right]^T\left[\Sigma^{-1}\right]\left[\phi^L_1(1,l) - \phi^L_1(1,j)\right], \tag{B.86}$$

$$\sigma^2_{S^L_1(l,j)} = \left[\phi^L_1(1,l) - \phi^L_1(1,j)\right]^T\left[\Sigma^{-1}\right]\left[\phi^L_1(1,l) - \phi^L_1(1,j)\right]. \tag{B.87}$$

*Démonstration.* La preuve de ce théorème peut être trouvée dans la version anglaise du manuscrit. □

**Remarque B.8.** *Ajoutons quelques commentaires sur les résultats du Théorème B.5. La borne supérieure $\tilde{\mathbb{P}}_{md}$ pour la pire probabilité de détection-localisation manquée peut être calculée analytiquement. En revanche, la borne supérieure pour la pire probabilité de fausse alarme et la borne supérieure pour la pire probabilité de fausse localisation peuvent être estimées numériquement en utilisant la méthode numérique proposée dans la Proposition B.1.*

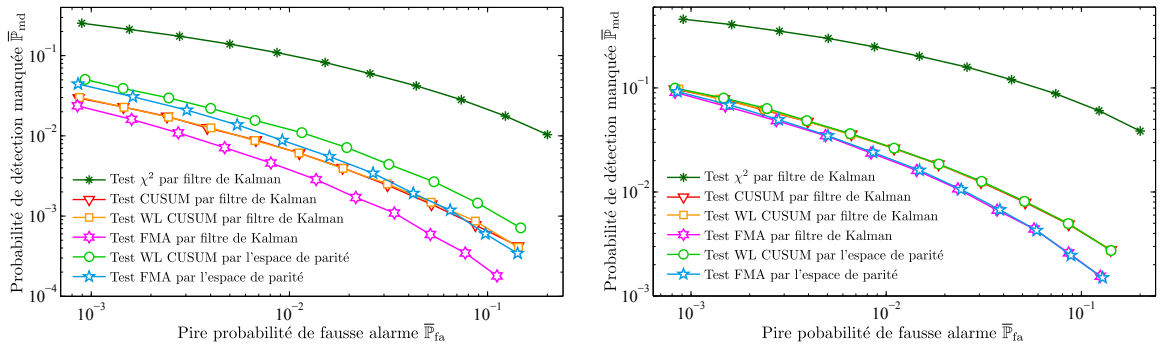
## B.7 Exemples numériques

Dans cette section, nous appliquons les algorithmes développés dans les sections ci-dessus au problème de détection et de localisation des attaques cyber-physiques dans un réseau de distribution d'eau potable simple. Les lecteurs intéressés peuvent consulter la version anglaise du manuscrit pour l'architecture du réseau d'eau et les paramètres de simulation. Nous présentons dans cette section seulement les principaux résultats de simulation. Les résultats de simulation pour les paramètres parfaitement connus sont donnés dans la sous-section B.7.1. Dans la sous-section B.7.2, nous effectuons l'analyse de sensibilité du test FMA par rapport à plusieurs paramètres. Les performances statistiques des algorithmes de détection pour le cas où les paramètres sont partiellement connus sont présentées dans la section B.7.3. Finalement, nous comparons dans la sous-section B.7.4 les performances statistiques de quelques algorithmes de localisation.

### B.7.1 Résultats de simulation pour des paramètres parfaitement connus

Dans cette sous-section, nous présentons les résultats de simulation dans un contexte idéale où les paramètres sont parfaitement connus.

#### Comparaison entre le test FMA et des tests classiques

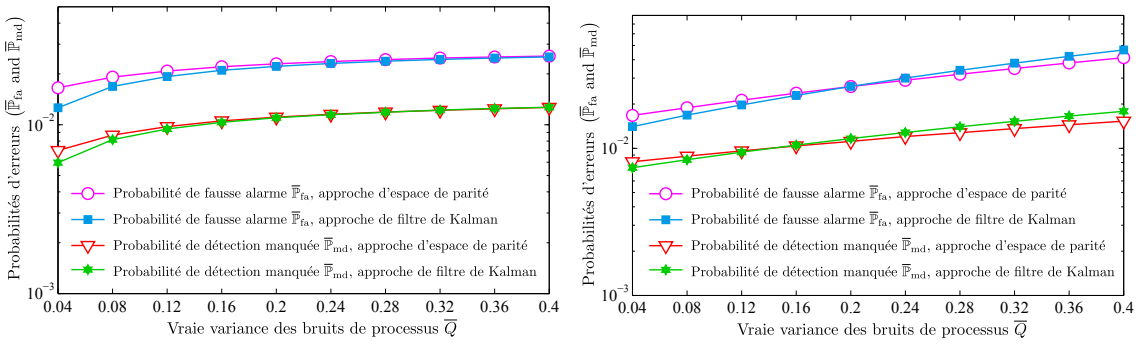



(a) Variance des bruits des processus $Q = 0.02$. (b) Variance des bruits des processus $Q = 0.2$.

FIGURE B.1 – Comparaison des performances statistiques de plusieurs détecteurs. La probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ est décrite comme la fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$.

Dans la figure B.1, nous comparons les performances statistiques de plusieurs règles de détection avec la simulation de Monte Carlo de $10^6$ répétitions. La probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ est décrite comme la fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$. L'instant de rupture est $k_0 = L + 1 = 9$. Le test WL CUSUM est un cas spécial du test VTWL CUSUM avec les seuils égaux, c-à-d $h_1 = h_2 = \cdots = h_L$. Les remarques suivantes peuvent être déduites des résultats de simulation. Tout d'abord, les algorithmes proposés (test CUSUM, test WL CUSUM, et test FMA) sont meilleurs que le test $\chi^2$ qui s'appuie sur une statistique non-paramétrique. Ce phénomène peut être expliqué par le fait que le test $\chi^2$ n'exploite pas les informations sur les profils de changements transitoires. Deuxièmement, étant donné un niveau acceptable sur la probabilité de fausse alarme, la probabilité de détection manquée des tests FMA proposés est beaucoup plus petite que celle des tests CUSUM et WL CUSUM, pour l'approche avec le filtre de Kalman et l'approche basée sur l'espace de parité. En d'autres termes, les tests FMA sont meilleurs que les tests traditionnels par rapport au critère d'optimalité adapté à la détection de signaux transitoires. Ces résultats de simulation sont obtenus du fait que l'optimisation de l'algorithme VTWL CUSUM conduit à la règle de détection FMA. Enfin, les performances statistiques des algorithmes basés sur l'approche avec le filtre de Kalman sont meilleures que celles des algorithmes basés sur l'approche par projection sur l'espace de parité lorsque les bruits des processus sont petits (voir la différence dans la sous-figure B.1a pour $Q = 0.02$ et la sous-figure B.1b pour $Q = 0.2$).

**Comparaison entre l'approche avec le filtre de Kalman et l'approche avec l'espace de parité**




(a) Condition parfaite : $\overline{Q} = Q$. La vraie valeur $\overline{Q}$ et la valeur putative $Q$ varient de $Q = \overline{Q} = 0.02$ à $Q = \overline{Q} = 0.4$.

(b) Condition imparfaite : $\overline{Q} \neq Q$. La valeur putative est fixée à $Q = 0.1$ tandis que la vraie valeur varie de $\overline{Q} = 0.02$ à $\overline{Q} = 0.4$.

FIGURE B.2 – Comparaison entre deux méthodes de génération de résidus : approche avec le filtre de Kalman et approche avec l'espace de parité. La probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ et la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ sont décrites comme la fonction de la vraie variance des bruits des processus $\overline{Q}$.

La comparaison entre les méthodes de génération de résidus, l'approche avec le filtre de Kalman en régime permanent et l'approche par projection sur l'espace de parité de taille fixe, est montrée dans la figure B.2. La probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ et la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ sont affichées en fonction de la vraie variance des bruits des processus $\overline{Q}$ qui varie de $\overline{Q} = 0.02$ à $\overline{Q} = 0.4$. Deux scénarios sont considérés : $\overline{Q} = Q$ et $\overline{Q} \neq Q$.

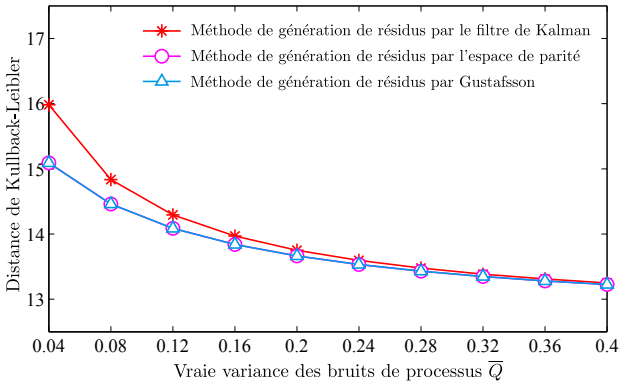



FIGURE B.3 – Distance de K-L des résidus par rapport à la vraie variance des bruits des processus $\overline{Q}$.

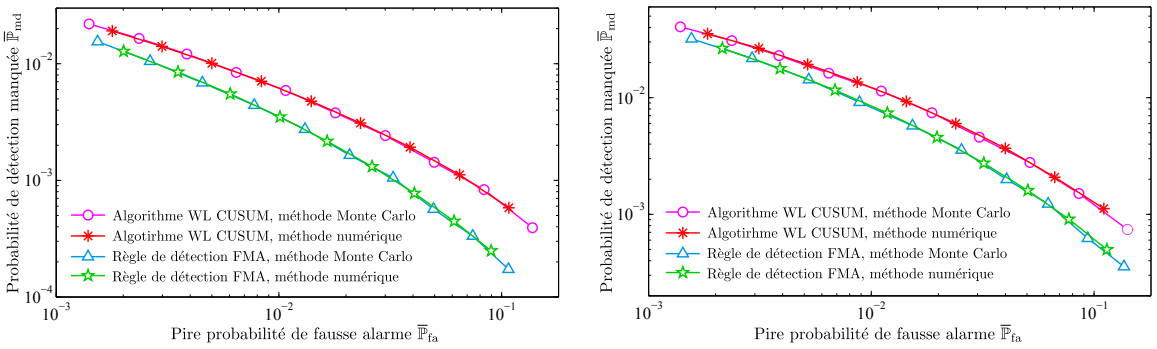
Considérons maintenant la condition parfaite où la variance des bruits des processus est connue exactement (c-à-d $\overline{Q} = Q$). À partir de la figure B.2a, nous pouvons constater que l'approche avec le filtre de Kalman en régime permanent est plus performante que l'approche avec l'espace de parité de taille fixe, en particulier lorsque les bruits de processus sont faibles. Ce phénomène est expliqué dans la figure B.3 où les distances de Kullback-Leibler (K-L) des résidus générés par les deux approches sont calculées et comparées. Le filtre de Kalman génère des résidus avec une distance de K-L plus grande que celle obtenue avec l'espace de parité. La différence devient significative dans de tels scénarios quand les bruits des processus sont extrêmement petites. En revanche, la différence est négligeable lorsque les bruits des processus sont importants. Ce phénomène est expliqué par le rapprochement de l'approche bayésienne (e.g., le filtre de Kalman) avec l'approche minimax (e.g., l'espace de parité) qui produit une erreur significative seulement si les bruits de processus sont faibles et que, par conséquent, l'information *a priori* joue un rôle important.

Considérons maintenant le scénario pratique où la vraie valeur de la variance des bruits de processus est différente de sa valeur putative (c-à-d $\overline{Q} \neq Q$). La valeur putative est choisie telle que $Q = 0.1$ et la vraie valeur varie de $\overline{Q} = 0.02$ à $\overline{Q} = 0.4$. Les performances statistiques du test FMA en se basant sur l'approche avec le filtre de Kalman et avec l'approche par projection sur l'espace de parité sont données dans la figure B.2b. Nous pouvons constater que l'approche avec le filtre de Kalman est plus sensitive aux bruits que processus que l'approche avec l'espace de parité. Ce phénomène peut être expliqué par le fait que le filtre de Kalman, lorsqu'il dispose d'informations erronées sur des bruits de processus, peut produire une erreur cumulée sur l'estimation des états, notamment dans des scénarios où la vraie matrice de covariance des bruits des processus est plus grande que sa valeur putative. Par conséquent, la performance statistique d'un algorithme basé sur cette approche se réduit significativement.

### Comparaison entre la méthode numérique et la simulation de Monte Carlo

La comparaison entre la méthode numérique proposée et la simulation Monte Carlo est donnée dans la figure B.4.

La simulation de Monte Carlo est réalisée avec $10^6$ répétitions tandis que la méthode numérique est effectuée avec une précision de $10^{-5}$. La probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ est décrite en




(a) Approche avec le filtre de Kalman régime permanent.

(b) Approche par projection sur l'espace de parité de taille fixe.

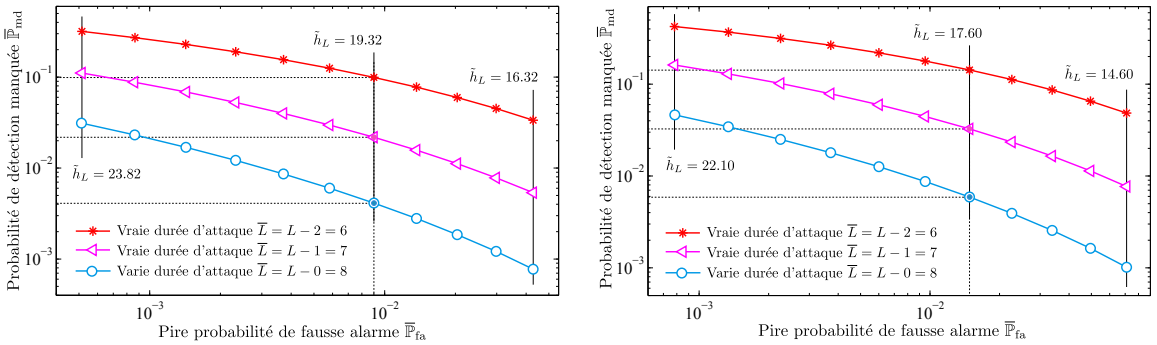
FIGURE B.4 – Comparaison entre la méthode numérique et la simulation de Monte Carlo.

fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$, pour les deux méthodes de génération des résidus (le filtre de Kalman dans la sous-figure B.4a et l'espace de parité dans la sous-figure B.4b). À partir des résultats de simulation, nous pouvons constater que les courbes numériques coïncident parfaitement avec les courbes de Monte Carlo, ce qui confirme la qualité de la méthode numérique proposée.

### B.7.2 Analyse de sensibilité du test FMA

Cette sous-section est consacrée à l'analyse de robustesse du test FMA par rapport à plusieurs paramètres opérationnels tels que la durée d'attaque, les profils d'attaque, la covariance des bruits des processus et la covariance des bruits des capteurs.

#### Sensibilité du FMA par rapport à la durée d'attaque




(a) Approche avec le filtre de Kalman en régime permanent.

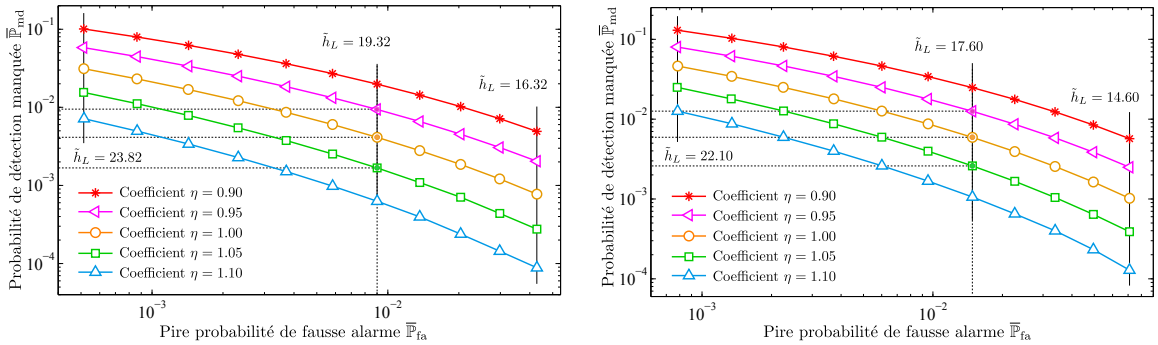
(b) Approche avec l'espace de parité de taille fixe.

FIGURE B.5 – Sensibilité du test FMA par rapport à la durée d'attaque.

La sensibilité du test FMA par rapport à la durée d'attaque est illustrée dans la figure B.5, pour l'approche de filtre de Kalman (sous-figure B.5a) et l'approche d'espace de parité (sous-figure B.5b). La probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ est tracée en fonction de la pire probabilité

de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ pour différentes valeurs de la vraie durée d'attaque $\overline{L} = \{6, 7, 8\} \leq L = 8$. Si la vraie durée d'attaque $\overline{L}$ est supérieure à sa valeur putative $L$, la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ reste intacte puisque toutes les détections avec un retard supérieur à $L$ sont considérées comme manquées. En revanche, pour $\overline{L} \leq L$, la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ dépend fortement de la vraie durée d'attaque $\overline{L}$. Ce phénomène est expliqué par le fait qu'une petite durée d'attaque $\overline{L}$ entraîne un petit changement dans la distribution des observations, augmentant ainsi la probabilité de détection manquée. D'autre part, la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ est insensible à la vraie durée d'attaque $\overline{L}$. Ce phénomène est due au fait que, dans le cas d'une fausse alarme, toutes les observations sont générées à partir du mode de fonctionnement normal du système.

### Sensibilité du FMA par rapport aux profils d'attaque

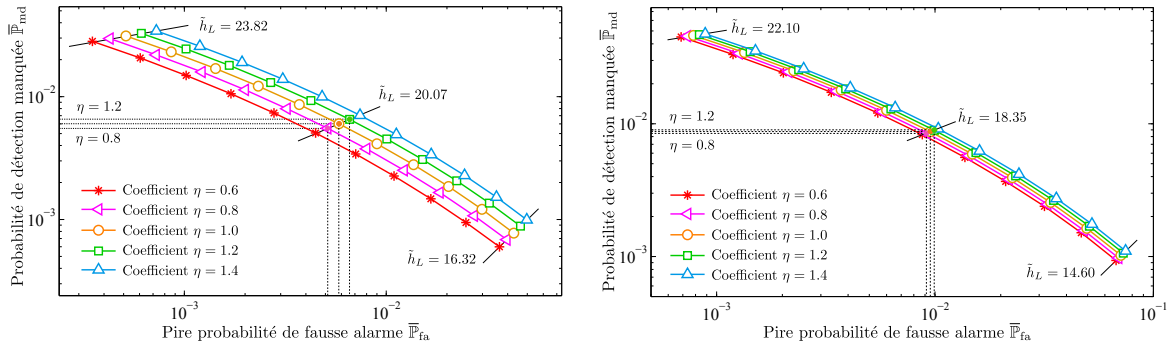



(a) Approche avec le filtre de Kalman en régime permanent.

(b) Approche avec l'espace de parité de taille fixe.

FIGURE B.6 – Sensibilité du test FMA par rapport aux profils d'attaque. La probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ est tracée comme fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ pour différentes valeurs de $\eta = \{0.90, 0.95, 1.00, 1.05, 1.10\}$. Les vrais profils d'attaque sont liés aux profils putatifs par $\overline{\theta}_j = \eta\theta_j$, pour $1 \leq j \leq L$.

La sensibilité du test FMA par rapport aux profils d'attaque est illustrée dans la figure B.6, pour l'approche avec le filtre de Kalman (sous-figure B.6a) et l'approche avec l'espace de parité (sous-figure B.6b). Les vrais profils d'attaque sont choisis tels que $\overline{\theta}_j = \eta\theta_j$, pour $1 \leq j \leq L$, où $\eta = \{0.90, 0.95, 1.00, 1.05, 1.10\}$. En d'autres termes, l'amplitude des profils varie de 90% à 110%, mais la « forme » des profils reste inchangé. De façon similaire au cas précédant, la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ est insensible aux vrais profils d'attaque puisque toutes les observations sont générées à partir du mode de fonctionnement normal du système. En revanche, la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ dépend fortement des vrais profils d'attaque $\overline{\theta}_j$, pour $1 \leq j \leq L$. Plus petits sont les vrais profils $\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_L$, plus grande est la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$. Ce phénomène peut être expliqué par le fait que les petits profils d'attaque conduisent à des petits changements dans la distribution des observations, augmentant ainsi la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$.




(a) Approche avec le filtre de Kalman en régime permanent.

(b) Approche avec l'espace de parité de taille fixe.

FIGURE B.7 – Sensibilité du test FMA par rapport aux bruits des processus. La probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ est tracée en fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ pour différentes valeurs de $\eta = \{0.90, 0.95, 1.00, 1.05, 1.10\}$. La vraie variance des bruits de processus est liée à sa valeur putative par $\overline{Q} = \eta Q$.

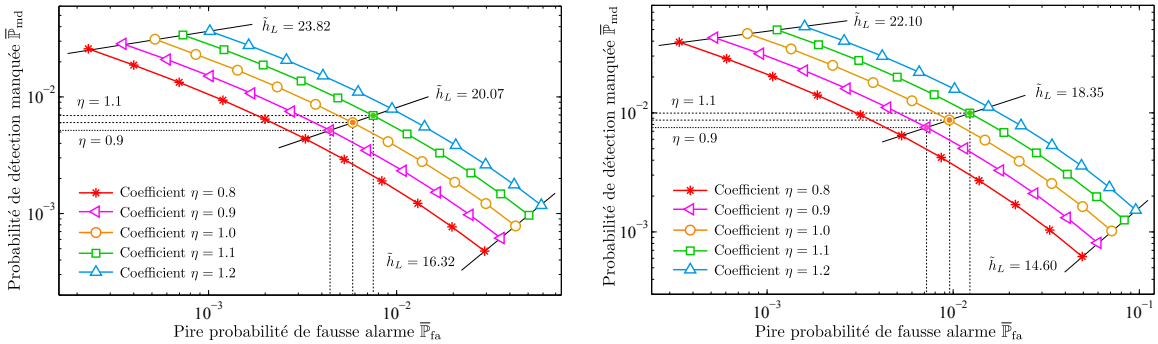
### Sensibilité du FMA par rapport aux bruits des processus

La sensibilité du test FMA par rapport aux bruits de processus est tracée dans la figure B.7, pour l'approche avec le filtre de Kalman (sous-figure B.7a) et l'approche avec l'espace de parité (sous-figure B.7b). Dans chaque sous-figure, la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ est exprimée comme une fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ pour différentes valeurs du coefficient $\eta = \{0.6, 0.8, 1.0, 1.2, 1.4\}$. La vraie variance des bruits de processus est reliée à sa valeur putative par $\overline{Q} = \eta Q$. Dans ce cas, la différence $\overline{Q} - Q$ influe toutes les probabilités de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ et de détection manquée $\overline{\mathbb{P}}_{\text{md}}$. Nous pouvons constater, à partir des sous-figures B.7a et B.7b, que l'approche basée sur le filtre de Kalman est beaucoup plus sensible aux bruits des processus que l'approche basée sur l'espace parité. Cette conclusion est cohérente avec celle tirée dans la sous-section B.7.1.

Pour simplifier l'explication, trois isolignes de seuil constant $\tilde{h}_L$ sont ajoutés à la sous-figure B.7a (pour le filtre de Kalman) et à la sous-figure B.7b (pour l'espace de parité). Le paramètre de réglage $\tilde{h}_L$ est fixé par la sélection d'un point de la courbe correspondant à $\eta = 1.0$. La pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ et la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ sont déterminés en dessinant, respectivement, des lignes pointillées verticales et horizontales à partir du point sélectionné. Les variations des probabilités d'erreurs ($\overline{\mathbb{P}}_{\text{fa}}$ et $\overline{\mathbb{P}}_{\text{md}}$) en raison de la différence entre la vraie covariance des bruits de processus et sa valeur putative peuvent être estimées en utilisant l'isoligne croisant le point sélectionné. Par exemple, deux isolignes pour $\tilde{h}_L = 20.7$ dans la sous-figure B.7a et pour $\tilde{h}_L = 18.35$ dans la sous-figure B.7b sont utilisées pour déterminer les variations des probabilités d'erreurs.

### Sensibilité du FMA par rapport aux bruits de capteurs

La sensibilité du test FMA par rapport aux bruits de capteurs est illustrée dans la figure B.8, pour l'approche avec le filtre de Kalman (sous-figure B.8a) et l'approche avec l'espace de parité (sous-figure B.8b). Dans chaque sous-figure, la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ est décrite comme une fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ pour différentes valeurs du




(a) Approche avec le filtre de Kalman en régime permanent.

(b) Approche avec l'espace de parité de taille fixe.

FIGURE B.8 – Sensibilité du test FMA par rapport aux bruits de capteurs. La probabilité de détection manquée $\overline{\mathbb{P}}_{md}$ est tracée en fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{fa}$ pour différentes valeurs de $\eta = \{0.90, 0.95, 1.00, 1.05, 1.10\}$. La vraie variance des bruits de capteurs est liée à sa valeur putative par $\overline{R} = \eta R$.

coefficient $\eta = \{0.8, 0.9, 1.0, 1.1, 1.2\}$. La vraie covariance des bruits de capteurs est reliée à sa valeur putative par $\overline{R} = \eta R$. Il est à noter que la performance statistique du test FMA est inversement proportionnelle aux bruits des capteurs. Les variations des probabilités d'erreurs en raison de la différence entre la vraie covariance des bruits de capteurs et sa valeur putative ($\overline{R} \neq \overline{R}$) peuvent être traiter de la même façon que dans la cas précédent.

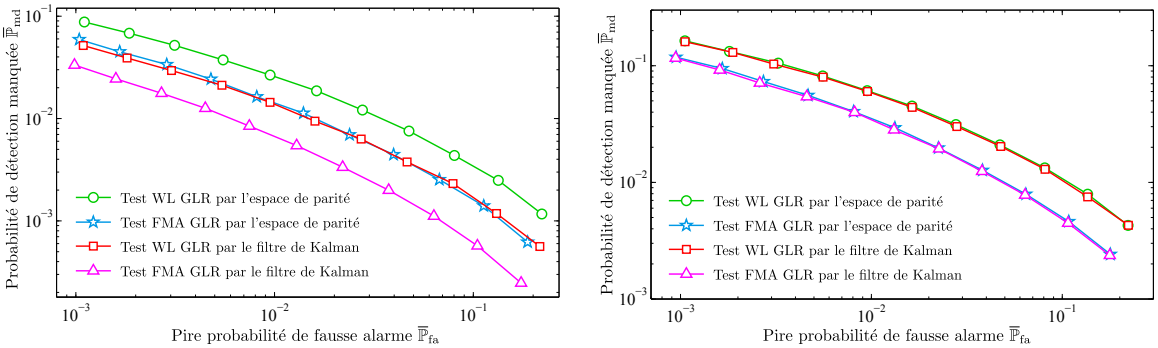
### B.7.3 Résultats de simulation pour les paramètres partiellement connus

Dans cette sous-section, nous examinons les performances statistiques de plusieurs algorithmes de détection dans le scénario ou les paramètres sont partiellement connus. Plus précisément, la « forme » des profils est connue mais leur amplitude est inconnue.

#### Comparaison entre le test FMA GLR et le test WL GLR

La comparaison entre le test FMA GLR et le test WL GLR, pour l'approche avec le filtre de Kalman et l'approche avec l'espace de parité, est représenté dans la figure B.9. Deux valeurs de variance des bruits de processus sont considérées : $Q = 0.02$ (dans la sous-figure B.9a) et $Q = 0.2$ (dans la sous-figure B.9b). Dans chaque sous-figure, la probabilité de détection manquée $\overline{\mathbb{P}}_{md}$ est décrite comme une fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{fa}$.

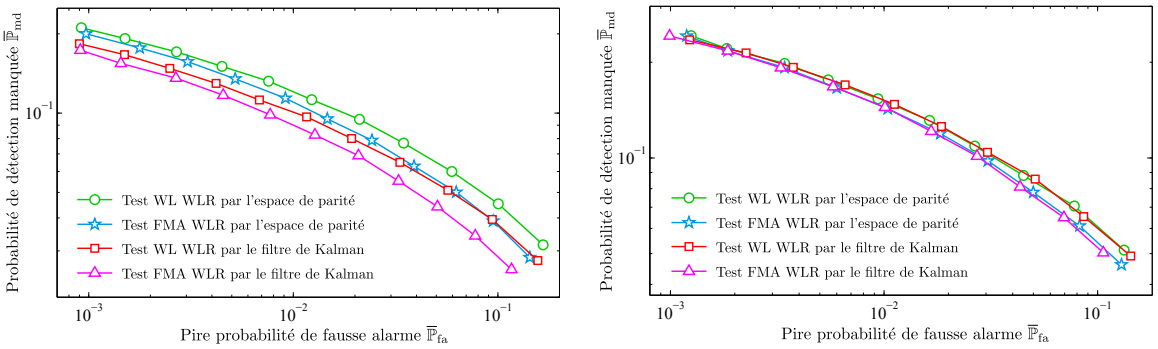
À partir des résultats de simulation, nous pouvons constater que, pour une valeur donnée sur la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{fa}$, la probabilité de détection manquée $\overline{\mathbb{P}}_{md}$ du test FMA GLR est inférieure à celle du test WL GLR, pour l'approche avec le filtre de Kalman ainsi que pour l'approche basée sur l'espace de parité. En d'autres termes, le test FMA GLR donne des meilleures performances statistiques que le test WL GLR par rapport au critère de détection des signaux transitoires. En outre, l'approche avec le filtre de Kalman est plus efficace que l'approche avec l'espace de parité, particulièrement lorsque les bruits de processus sont petits (voir la différence dans la sous-figure B.9a et dans la sous-figure B.9b).




(a) Variance des bruits des processus $Q = 0.02$. (b) Variance des bruits des processus $Q = 0.2$.

FIGURE B.9 – Comparaison entre le test FMA GLR et le test WL GLR. La probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ est exprimée comme une fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$. Deux valeurs de la variance des bruits de processus sont considérées : $Q = 0.02$ et $Q = 0.2$.

### Comparaison entre le test FMA WLR et le test WL WLR




(a) Variance des bruits des processus $Q = 0.02$. (b) Variance des bruits des processus $Q = 0.2$.

FIGURE B.10 – Comparaison entre le test FMA WLR et le test WL WLR. La probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ est exprimée en fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$. Deux valeurs de la variance des bruits de processus sont considérées : $Q = 0.02$ et $Q = 0.2$.

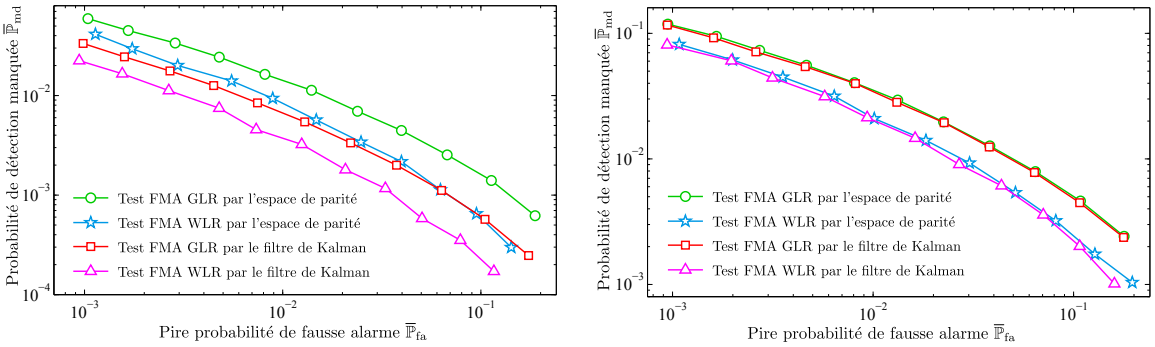
La comparaison entre le tes FMA WLR et le test WL WLR, pour l'approche avec le filtre de Kalman et l'approche avec l'espace de parité, est représentée dans la figure B.10. Deux valeurs de la variance des bruits de processus sont considérées : $Q = 0.02$ (dans la sous-figure B.10a) et $Q = 0.2$ (dans la sous-figure B.10b). Dans chaque sous-figure, la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ est décrite comme une fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$.

La distribution *a priori* du paramètre $\gamma$ est choisie par $\gamma \sim \mathcal{U}(\gamma_0, \gamma_1)$, où $\gamma_0 = 0.5$ et $\gamma_1 = 1.5$. La simulation est effectuée de la manière suivante. Pour chaque exécution de Monte Carlo, le paramètre $\gamma$ est généré à partir de la distribution uniforme $\gamma \sim \mathcal{U}(0.5, 1.5)$. Les vrais profils d'attaque sont ensuite calculés à partir de leurs valeurs putatives par $\overline{\theta}_j = \gamma\theta_j$, pour $1 \leq j \leq L$. Finalement, les algorithmes qui se basent sur l'approche de WLR sont exécutés afin d'obtenir la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ et la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$.

À partir des résultats donnés dans la figure B.10, nous pouvons conclure que le test FMA WLR fonctionne beaucoup mieux que le test WL WLR pour les deux approches de génération de

résidus. Comme précédemment, l'approche avec le filtre de Kalman offre des meilleures performances statistiques que l'approche avec l'espace de parité, en particulier pour des petits bruits des processus. Ce phénomène peut être constaté dans la sous-figure B.10a (pour $Q = 0.02$) et dans la sous-figure B.10b (pour $Q = 0.2$).

#### Comparaison entre le test FMA GLR et le test FMA WLR




(a) Variance des bruits des processus $Q = 0.02$.

(b) Variance des bruits des processus $Q = 0.2$.

FIGURE B.11 – Comparaison entre le test FMA GLR et le test FMA WLR. La probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$ est exprimée en fonction de la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$. Deux valeurs de la variance des bruits de processus sont considérées : $Q = 0.02$ et $Q = 0.2$.

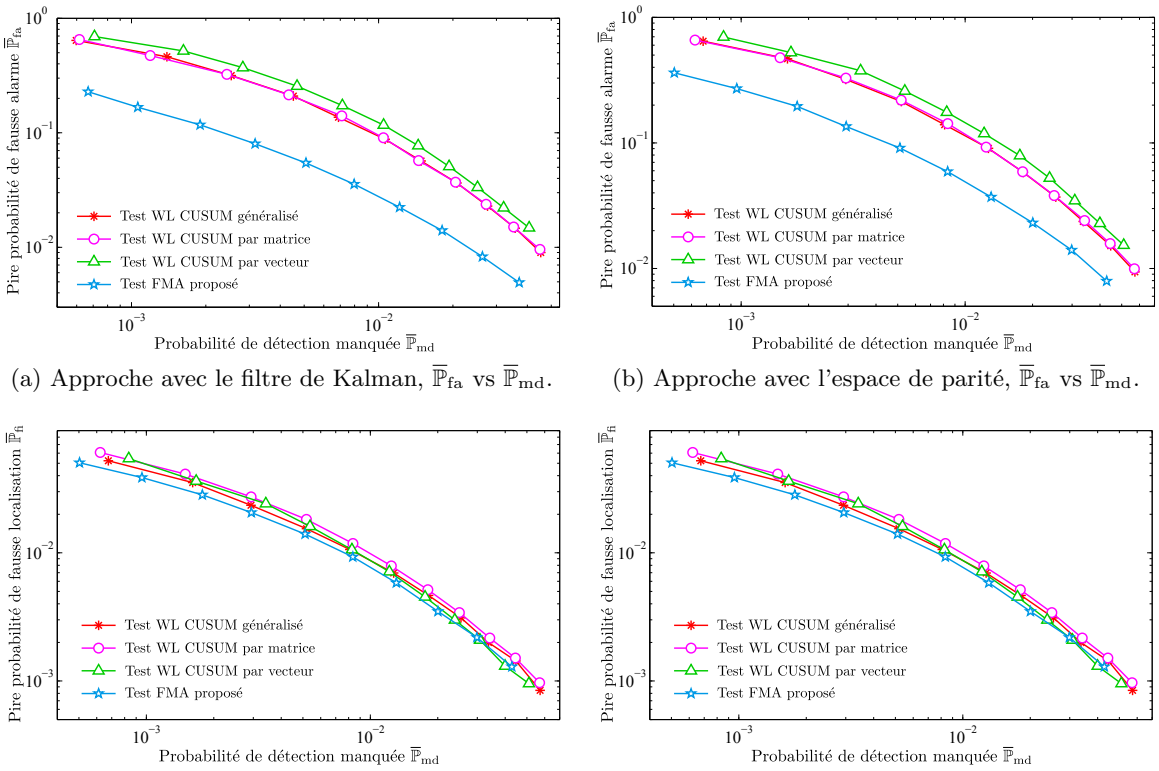
La comparaison entre l'approche GLR et l'approche WLR est illustrée dans la figure B.11 où les performances statistiques du test FMA GLR et celles du test FMA WLR sont présentées. Il est à noter que les performances statistiques du test FMA WLR dépendent fortement du paramètre $\gamma$ puisque les vrais profils d'attaque sont définis par $\overline{\theta}_j = \gamma\theta_j$, pour $1 \leq j \leq L$. Afin de comparer les deux approches, nous fixons le paramètre $\gamma = 1$ pour l'approche WLR. Les paramètres $\gamma_0$ et $\gamma_1$ prennent les valeurs $\gamma_0 = 0.5$ et $\gamma_1 = 1.5$, respectivement. À partir des résultats donnés dans la sous-figure B.11a ($Q = 0.02$) et dans la sous-figure B.11b ($Q = 0.2$), nous pouvons constater que le test FMA WLR offre de meilleures performances statistiques que le test FMA GLR pour tous les deux méthodes de génération de résidus. Cette phénomène peut être expliqué par le fait que l'approche WLR utilise l'information *a priori* sur l'amplitude des profils tandis que l'approche GLR n'exploite pas cette information essentielle.

### B.7.4 Résultats de simulation pour des algorithmes de localisation

Cette sous-section est consacrée à la comparaison des performances statistiques de plusieurs algorithmes de localisation de signaux transitoires. Nous donnons des résultats de simulation seulement dans le scénario où $\rho_{12} \geq \max\{\rho_{01}, \rho_{02}\}$. Les lecteurs intéressés peuvent consulter la version anglaise du manuscrit pour l'autre scénario où $\rho_{12} \leq \min\{\rho_{01}, \rho_{02}\}$. L'instant de rupture $k_0$ est fixé à $k_0 = L + 1 = 9$ dans toutes les simulations suivantes.

#### Comparaison entre le test FMA et les tests classiques

La comparaison entre le test FMA proposé et les tests classiques (WL CUSUM généralisé, WL CUSUM par matrice et WL CUSUM par vecteur) est présentée dans la figure B.12. Les résultats




(a) Approche avec le filtre de Kalman, $\overline{\mathbb{P}}_{\text{fa}}$ vs $\overline{\mathbb{P}}_{\text{md}}$.

(b) Approche avec l'espace de parité, $\overline{\mathbb{P}}_{\text{fa}}$ vs $\overline{\mathbb{P}}_{\text{md}}$.

(c) Approche avec le filtre de Kalman, $\overline{\mathbb{P}}_{\text{fi}}$ vs $\overline{\mathbb{P}}_{\text{md}}$.

(d) Approche avec l'espace de parité, $\overline{\mathbb{P}}_{\text{fi}}$ vs $\overline{\mathbb{P}}_{\text{md}}$.

FIGURE B.12 – Comparaison entre le test FMA proposé et les tests classiques (WL CUSUM généralisé, WL CUSUM par matrice et WL CUSUM par vecteur). La pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ et la pire probabilité de fausse localisation $\overline{\mathbb{P}}_{\text{fi}}$ sont tracées en fonction de la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$.

sont obtenus à l'aide de la simulation de Monte Carlo de $2.10^5$ répétitions. La pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ et la pire probabilité de fausse localisation $\overline{\mathbb{P}}_{\text{fi}}$ sont tracées en fonction de la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$. Les deux méthodes de génération de résidus sont considérées.

Il peut être remarqué, à partir des résultats de simulation, que pour une valeur donnée sur la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$, la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{\text{fa}}$ et la pire probabilité de fausse localisation $\overline{\mathbb{P}}_{\text{fi}}$ du test FMA sont inférieures à celles des tests WL CUSUM. Autrement dit, le test FMA proposé est plus performant que les tests classiques par rapport au critère d'optimalité de détection-localisation des signaux transitoires.

### Comparaison entre l'approche avec le filtre de Kalman et l'approche avec l'espace de parité

La comparaison entre l'approche avec le filtre de Kalman et l'approche avec l'espace de parité est présentée dans la figure B.13. Les probabilités d'erreurs ($\overline{\mathbb{P}}_{\text{fa}}$ et $\overline{\mathbb{P}}_{\text{fi}}$) sont tracées en fonction de la probabilité de détection manquée $\overline{\mathbb{P}}_{\text{md}}$. À partir des résultats de simulation, nous pouvons constater que l'approche avec le filtre de Kalman donne des meilleures performances statistiques que l'approche avec l'espace de parité. Ce phénomène peut être expliqué par le fait

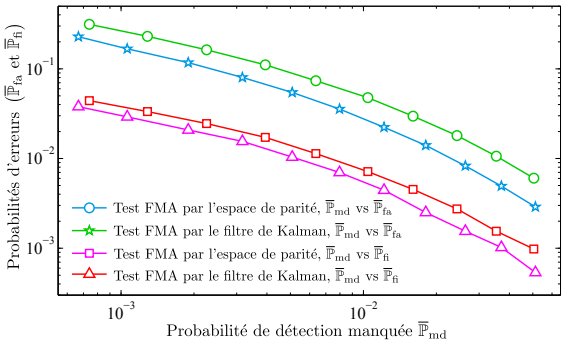
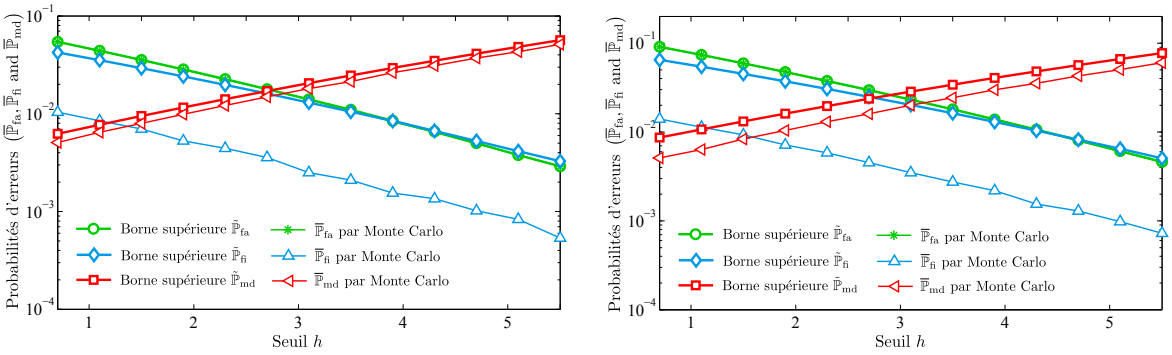



FIGURE B.13 – Comparaison entre l'approche avec le filtre de Kalman et l'approche avec l'espace de parité en utilisant les détecteurs FMA. La pire probabilité de fausse alarme $\overline{\mathbb{P}}_{fa}$ et la pire probabilité de fausse localisation $\overline{\mathbb{P}}_{fi}$ sont tracées en fonction de la probabilité de détection manquée $\overline{\mathbb{P}}_{md}$.

que la première approche génère des résidus avec les distances de K-L plus élevés que la seconde approche.

### Évaluation des bornes supérieurs du test FMA

Les bornes supérieures des probabilités d'erreurs sont évaluées dans la figure B.14, pour l'approche avec le filtre de Kalman (sous-figure B.14a) et l'approche avec l'espace de parité (sous-figure B.14b).




(a) Approche avec le filtre de Kalman en régime permanent.

(b) Approche avec l'espace de parité de taille fixe.

FIGURE B.14 – Évaluation des bornes supérieures pour les probabilités d'erreurs du test FMA. Les bornes supérieures pour $\overline{\mathbb{P}}_{fa}$, $\overline{\mathbb{P}}_{fi}$ et $\mathbb{P}_{md}$ sont tracées en fonction du seuil $h$.

Les bornes proposées sont comparées avec les probabilités d'erreurs correspondantes qui sont obtenues par la simulation de Monte Carlo avec $2.10^5$ répétitions. Nous pouvons constater que la borne supérieure $\tilde{\mathbb{P}}_{fa}$ pour la pire probabilité de fausse alarme $\overline{\mathbb{P}}_{fa}$ est extrêmement précise. En revanche, la borne supérieure $\tilde{\mathbb{P}}_{fi}$ pour la pire probabilité de fausse localisation $\overline{\mathbb{P}}_{fi}$ n'est pas très précise. Enfin, la borne supérieure pour la probabilité de détection manquée $\tilde{\mathbb{P}}_{md}$ semble acceptable.

## B.8 Conclusions et perspectives

Cette thèse s'est intéressée au problème de détection et de localisation d'attaques cyber-physiques sur des systèmes SCADA. Un modèle d'espace d'état à temps discret a été employé pour décrire les processus physiques. Les actes malveillants ont été modélisées comme des signaux additifs de durée finie qui agissent sur les deux équations du système, à savoir l'équation d'état et l'équation d'observations. La prise de décision devait tenir compte des états inconnus (considérés comme des paramètres de nuisance) et des bruits des processus et des capteurs. L'approche traditionnelle FDI (Fault Detection and Isolation) a été utilisée pour résoudre ce problème. Cette approche est composée de deux étapes : la génération des résidus et l'évaluation des résidus. La première étape a pour but de générer une séquence de résidus qui sont indépendants des paramètres de nuisance. Ensuite, la deuxième étape consiste à déterminer une rupture dans la séquence des résidus, et éventuellement à identifier le type de changements.

Dans cette thèse, nous avons utilisé deux méthodes classiques pour générer les résidus : le filtre de Kalman en régime permanent et la projection sur l'espace de parité de taille fixe. Nous avons proposé un modèle statistique unifié des résidus générés par les deux approches mentionnées. Cette thèse s'est particulièrement concentrée sur l'évaluation des résidus en se basant sur le modèle statistique unifié. Nous avons proposé des algorithmes de détection et de localisation des changements transitoires.

Pour le problème de détection, l'algorithme VTWL CUSUM a été adapté au modèle statistique unifié pour détecter une rupture dans la séquence des résidus. Le critère d'optimalité vise à minimiser la pire probabilité de détection manquée sous la contrainte que la pire probabilité de fausse alarme pour une fenêtre de taille donnée soit inférieure à une valeur prescrite. Comme il est difficile de résoudre le problème d'optimisation exact, nous avons minimisé la borne supérieure de la pire probabilité de détection manquée pour une valeur donné de la pire probabilité de fausse alarme dans la classe des tests VTWL CUSUM. Il a été démontré que le test VTWL CUSUM optimisé était équivalent à l'algorithme de la Moyenne Glissante Finie (Finite Moving Average ou FMA). De plus, nous avons proposé une méthode numérique pour estimer les probabilités d'erreurs du test FMA et du test VTWL CUSUM. Surtout, cette méthode numérique a été exploitée pour examiner la robustesse du test FMA par rapport à plusieurs paramètres opérationnels. Finalement, nous avons considéré aussi un scénario plus réaliste où la « forme » des profils est connue exactement mais leur amplitude est inconnue. L'approche du rapport de vraisemblance généralisé (GLR) et l'approche du rapport de vraisemblance pondérée (WLR) ont été envisagées pour résoudre le problème, ce qui a conduit au test VTWL GLR et au test VTWL WLR. Il a été démontré que le test VTWL GLR optimisé et le test VTWL WLR optimisé sont équivalent au test FMA GLR et au test FMA WLR, respectivement.

Pour le problème de détection-localisation conjointe de changements transitoires, un modèle statistique unifié a été développé et un nouveau critère d'optimalité a été proposé. Plusieurs algorithmes classiques de détection-localisation ont été considérés pour détecter l'instant de rupture et identifier le type du changement transitoire. Notamment, nous avons proposé un algorithme basé sur la Moyenne Glissante Finie (FMA) adaptée au problème de localisation de signaux transitoires. Les bornes supérieures pour des probabilités d'erreurs du test FMA ont été obtenues.

Les résultats théoriques sont appliqués à la détection et à la localisation des attaques cyber-physiques dans un réseau SCADA de distribution d'eau potable. Les conclusions suivantes peuvent être déduites des résultats de simulation. Premièrement, les tests FMA (pour le pro-

blème de détection et également celui de localisation) sont nettement plus performants que les tests classiques par rapport au critère d'optimalité de détection des signaux transitoires. Deuxièmement, l'approche avec le filtre de Kalman en régime permanent offre de meilleures performances statistiques que l'approche par projection sur l'espace de parité de taille fixe lorsque les paramètres sont parfaitement connus. Cependant, le filtre de Kalman est plus sensible aux bruits des processus que la projection sur l'espace de parité. Dans les scénarios où la vraie valeur de la covariance des bruits des processus est plus grande que sa valeur putative, la projection sur l'espace de parité peut offrir de meilleurs résultats que le filtre de Kalman. Finalement, une méthode numérique est proposée pour estimer les probabilités d'erreurs ainsi que pour examiner la robustesse du test FMA par rapport aux paramètres opérationnels.

Avant de clore ce manuscrit, nous aimerions suggérer plusieurs points d'approfondissement, pour les perspectives à court terme ainsi que pour celles à long terme. Dans un premier temps, nous pouvons envisager les travaux suivants :

- *Problème de détection* : Le problème de détection de changements transitoires peut être approfondi de la façon suivante. La première tâche consisterait à rechercher le test optimal (ou le test asymptotiquement optimal) par rapport au critère d'optimalité pour la détection de changements transitoires. La deuxième tâche devrait se concentrer sur la détection des signaux transitoires avec des profils variables. La tâche finale consisterait à détecter des changements transitoires complètement inconnus.

- *Problème de localisation* : Nous pouvons poursuivre le problème de localisation de changements transitoires par les travaux suivants. Premièrement, il serait intéressant de rechercher un test sous-optimal ou asymptotiquement optimal par rapport au critère d'optimalité proposé. Deuxièmement, il serait utile d'évaluer la probabilité de fausse alarme et la probabilité de fausse détection séparément. Finalement, un calcul plus précis de la borne supérieure pour la pire probabilité de fausse localisation serait très pertinent.

Dans un deuxième temps, nous pouvons envisager les approches suivantes :

- *Approche non-paramétrique* : Le modèle paramétrique peut être difficile à obtenir dans de nombreuses situations. L'incertitude du modèle peut conduire à une dégradation extrême de la performance statistique des algorithmes de détection et de localisation. Par contre, l'approche non-paramétrique ne nécessite pas de connaître les modèles du système et des attaques. Les méthodes non-paramétriques sont basées sur l'analyse des données observées. Le problème de détection peut être résolu en appliquant des techniques de classification mono-classe alors que les méthodes de classification multi-classes peuvent être utilisées pour le problème de localisation.

- *Approche semi-paramétrique* : L'approche paramétrique utilisée dans ce manuscrit dépend fortement des modèles des systèmes SCADA et des attaques cyber-physiques. Parfois, ces modèles paramétriques sont difficiles à obtenir. D'un autre côté, l'approche non-paramétrique ne s'intéresse pas véritablement au fonctionnement des systèmes SCADA, c-à-d à l'interaction entre les processus physiques et le centre de contrôle via les cyber-composants. Par conséquent, l'approche semi-paramétrique serait une combinaison naturelle de l'approche paramétrique et de l'approche non-paramétrique. En général, un modèle semi-paramétrique se compose de deux parties : une partie paramétrique et une autre partie non-paramétrique. La partie paramétrique contient les phénomènes qui peuvent être

décrits avec un modèle ayant un nombre limité de paramètres inconnus tandis que la partie non-paramétrique comprend les informations sur les phénomènes non-modélisés.

# Bibliography


[1] R. Alamian, M. Behbahani-Nejad, and A. Ghanbarzadeh. A state space model for transient flow simulation in natural gas pipelines. *Journal of Natural Gas Science and Engineering*, 9:51–59, 2012.

[2] Omar AM Alyt, Abbas S Omar, and Atef Z Elsherbeni. Detection and localization of rf radar pulses in noise environments using wavelet packet transform and higher order statistics. *Progress In Electromagnetics Research*, 58:301–317, 2006.

[3] S. Amin, A. Cardenas, and S. Sastry. Safe and secure networked control systems under denial-of-service attacks. *Hybrid Systems: Computation and Control*, pages 31–45, 2009.

[4] Saurabh Amin. *On Cyber Security for Networked Control Systems*. PhD thesis, University of California Berkeley, 2011.

[5] Saurabh Amin, Xavier Litrico, S Shankar Sastry, and Alexandre M Bayen. Stealthy deception attacks on water scada systems. In *Proceedings of the 13th ACM international conference on Hybrid systems: computation and control*, pages 161–170. ACM, 2010.

[6] Saurabh Amin, Xavier Litrico, S Shankar Sastry, and Alexandre M Bayen. Cyber security of water scada systems–part ii: attack detection using enhanced hydrodynamic models. *Control Systems Technology, IEEE Transactions on*, 21(5):1679–1693, 2013.

[7] Saurabh Amin, Xavier Litrico, Shankar Sastry, and Alexandre M Bayen. Cyber security of water scada systems–part i: analysis and experimentation of stealthy deception attacks. *Control Systems Technology, IEEE Transactions on*, 21(5):1963–1970, 2013.

[8] P. Armitage. Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(1):137–144, 1950.

[9] B Bakhache and I Nikiforov. Reliable detection of faults in measurement systems. *International Journal of adaptive control and signal processing*, 14(7):683–700, 2000.

[10] Michèle Basseville and Igor V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Information and System Sciences Series. Prentice Hall, Inc., Englewood Cliffs, N.J., U.S.A., 1993.

[11] C.W. Baum and V.V. Veeravalli. A sequential procedure for multihypothesis testing. *Information Theory, IEEE Transactions on*, 40(6), 1994.

[12] Bulent Baygun and Alfred O Hero. Optimal simultaneous detection and estimation under a false alarm constraint. *Information Theory, IEEE Transactions on*, 41(3):688–703, 1995.

[13] M. Behbahani-Nejad and A. Bagheri. The accuracy and efficiency of a matlab-simulink library for transient flow simulation of gas pipelines and networks. *Journal of Petroleum Science and Engineering*, 70(3):256–265, 2010.

[14] Boldizsár Bencsáth, Gábor Pék, Levente Buttyán, and Márk Félegyházi. The cousins of stuxnet: Duqu, flame, and gauss. *Future Internet*, 4(4):971–1003, 2012.

[15] L Billmann and Rolf Isermann. Leak detection methods for pipelines. *Automatica*, 23(3):381–385, 1987.

[16] Matt Bishop. *Introduction to computer security*. Addison-Wesley Professional, 2004.

[17] Rakesh B Bobba, Katherine M Rogers, Qiyan Wang, Himanshu Khurana, Klara Nahrstedt, and Thomas J Overbye. Detecting false data injection attacks on dc state estimation. In *Preprints of the First Workshop on Secure Control Systems, CPSWEEK 2010*, 2010.

[18] Tomasz Bojdecki. Probability maximizing approach to optimal stopping and its application to a disorder problem. *Stochastics*, 3(1-4):61–71, 1980.

[19] A.A. Borovkov. *Mathematical Statistics*. Gordon and Breach Science Pulishers, Amsterdam, 1998. Translated from Russian by Moullagaliev.

[20] M. Brunner, H. Hofinger, C. Krauss, C. Roblee, P. Schoo, and S. Todt. Infiltrating critical infrastructures with next-generation attacks. *Fraunhofer Institute for Secure Information Technology (SIT), Munich*, 2010.

[21] Jens Burgschweiger, Bernd Gnadig, and Marc C Steinbach. Optimization models for operative planning in drinking water networks. *Optimization and Engineering*, 10(1):43–73, 2009.

[22] Eric Byres, David Leversage, and Nate Kube. Security incidents and trends in scada and process industries. *The Industrial Ethernet Book*, 39(2):12–20, 2007.

[23] Eric J Byres and P Eng. Cyber security and the pipeline control system. *Pipeline & Gas Journal236*, (2), 2009.

[24] Francis Campan, Van Long Do, Patric Nader, Paul Honeine, Pierre Beauseroy, Lionel Fillatre, Philippe Cornu, Igor Nikiforov, Guillaume Prigent, and Jérôme Rouxel. Scala - surveillance continue d'activité et localisation d'agressions. In *Workshop Interdisciplinaire sur la sécurité Globale, WISG 2013*, pages 1–7, 2013.

[25] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry. Challenges for securing cyber physical systems. In *Workshop on Future Directions in Cyber-physical Systems Security. DHS*, 2009.

[26] A.A. Cárdenas, S. Amin, and S. Sastry. Research challenges for the security of control systems. In *Proceedings of the 3rd conference on Hot topics in security*, pages 1–6. USENIX Association, 2008.

[27] Alvaro A Cárdenas, Saurabh Amin, Zong-Syun Lin, Yu-Lun Huang, Chi-Yen Huang, and Shankar Sastry. Attacks against process control systems: risk assessment, detection, and response. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, pages 355–366. ACM, 2011.

[28] Alvaro A Cardenas, Saurabh Amin, and Shankar Sastry. Secure control: Towards survivable cyber-physical systems. In *Distributed Computing Systems Workshops, 2008. ICDCS'08. 28th International Conference on*, pages 495–500. IEEE, 2008.

[29] Biao Chen and P Willett. Detection of hidden markov model transient signals. *Aerospace and Electronic Systems, IEEE Transactions on*, 36(4):1253–1268, 2000.

[30] Jie Chen and Ron J Patton. *Robust model-based fault diagnosis for dynamic systems*. Kluwer academic publishers, 1999.

[31] E Chow and A Willsky. Analytical redundancy and the design of robust failure detection systems. *Automatic Control, IEEE Transactions on*, 29(7):603–614, 1984.

[32] CPNI. Good practice guide, process control and scada security. Centre for the Protection of National Infrastructure (CPNI), 2005.

[33] Mary Jane Credeur. Fbi probes georgia water plant break-in on terror concern. http://www.bloomberg.com/news/2013-04-30/fbi-probes-georgia-water-plant-break-in-on-terror-concern.html/, 30 Avril 2013.

[34] G. Dán and H. Sandberg. Stealth attacks and protection schemes for state estimators in power systems. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 214–219. IEEE, 2010.

[35] Steven X Ding. *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer, 2008.

[36] Van Long Do, Lionel Fillatre, and Igor Nikiforov. A statistical method for detecting cyber/physical attacks on scada systems. In *Control Applications (CCA), 2014 IEEE Conference on*, pages 364–369. IEEE, 2014.

[37] Van Long Do, Lionel Fillatre, and Igor Nikiforov. Sensitivity analysis of the sequential test for detecting cyber-physical attacks. In *23rd European Signal Processing Conference (EUSIPCO 2015)*, September 2015.

[38] Van Long Do, Lionel Fillatre, and Igor Nikiforov. Sequential detection of transient changes in stochastic-dynamical systems. *Journal de la Société Française de Statistique (JSFdS)*, xyz(xyz):xyz–xyz, 2015. In revision.

[39] Van Long Do, Lionel Fillatre, and Igor Nikiforov. Sequential monitoring of scada systems against cyber/physical attacks. In *9th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS 2015)*, Paris, France, September 2015.

[40] Van Long Do, Lionel Fillatre, and Igor Nikiforov. Statistical approaches for detecting cyber-physical attacks on scada systems. *IEEE Transactions on Control Systems Technology*, xyz(xyz):xyz–xyz, 2015. In preparation.

[41] Van Long Do, Lionel Fillatre, and Igor Nikiforov. Two sub-optimal algorithms for detecting cyber/physical attacks on scada systems. In *Proceedings of the X International Conference on System Identification and Control Problems (SICPRO'15)*, 2015.

[42] DOE. 21 steps to improve cyber security of scada networks. Office of Energy Assurance, U.S. Department of Energy, 2002.

[43] V.P. Dragalin, A.G. Tartakovsky, and V.V. Veeravalli. Multihypothesis sequential probability ratio tests. i. asymptotic optimality. *Information Theory, IEEE Transactions on*, 45(7):2448–2461, 1999.

[44] V.P. Dragalin, A.G. Tartakovsky, and V.V. Veeravalli. Multihypothesis sequential probability ratio tests. ii. accurate asymptotic expansions for the expected sample size. *Information Theory, IEEE Transactions on*, 46(4):1366–1383, 2000.

[45] Demetrios G Eliades and Marios M Polycarpou. A fault diagnosis and security framework for water systems. *Control Systems Technology, IEEE Transactions on*, 18(6):1254–1265, 2010.

[46] James D Esary, Frank Proschan, David W Walkup, et al. Association of random variables, with applications. *The Annals of Mathematical Statistics*, 38(5):1466–1474, 1967.

[47] N. Falliere, L.O. Murchu, and E. Chien. W32. stuxnet dossier. *White paper, Symantec Corp., Security Response*, 2011.

[48] Nicolas Falliere. Exploring stuxnet's plc infection process. http://www.symantec.com/connect/blogs/exploring-stuxnet-s-plc-infection-process, 22 September 2010.

[49] T.S. Ferguson. *Mathematical statistics: A decision theoretic approach*, volume 7. Academic Press New York, 1967.

[50] Mitra Fouladirad. *Détection de défaillance dans un système stochastique linéaire en présence de paramètres de nuisance*. PhD thesis, Université de Technologie de Troyes, 2005.

[51] Mitra Fouladirad, L Freitag, and Igor Nikiforov. Optimal fault detection with nuisance parameters and a general covariance matrix. *International journal of adaptive control and signal processing*, 22(5):431–439, 2008.

[52] Mitra Fouladirad and Igor Nikiforov. Optimal statistical fault detection with nuisance parameters. *Automatica*, 41(7):1157–1171, 2005.

[53] Igor Nai Fovino, Alessio Coletta, and Marcelo Masera. *Taxonomy of security solutions for the scada sector*. JRC - Joint Research Centre of the European Commission, 2010.

[54] Paul M Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results. *Automatica*, 26(3):459–474, 1990.

[55] Paul M Frank and X Ding. Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *Journal of process control*, 7(6):403–424, 1997.

[56] Gene F Franklin, J David Powell, and Michael L Workman. *Digital control of dynamic systems*, volume 3. Addison-wesley Menlo Park, 1998.

[57] PA Fridman. A method of detecting radio transients. *Monthly Notices of the Royal Astronomical Society*, 409(2):808–820, 2010.

[58] Benjamin Friedlander and Boaz Porat. Detection of transient signals by the gabor representation. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(2):169–180, 1989.

[59] Benjamin Friedlander and Boaz Porat. Performance analysis of transient detectors based on a class of linear data transforms. *Information Theory, IEEE Transactions on*, 38(2):665–673, 1992.

[60] Mordechai Frisch and Hagit Messer. The use of the wavelet transform in the detection of an unknown transient signal. *Information Theory, IEEE Transactions on*, 38(2):892–897, 1992.

[61] Brendan Galloway and Gerhard P Hancke. Introduction to industrial control networks. *Communications Surveys & Tutorials, IEEE*, 15(2):860–880.

[62] Wei Gao, Thomas Morris, Bradley Reaves, and Drew Richey. On scada control system command and response injection and intrusion detection. In *eCrime Researchers Summit (eCrime), 2010*, pages 1–9. IEEE, 2010.

[63] Alan Genz and Frank Bretz. Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics*, 11(4):950–971, 2002.

[64] J. Gertler. Analytical redundancy methods in fault detection and isolation. In *Proceedings of IFAC/IAMCS symposium on safe process*, volume 1, pages 9–21, 1991.

[65] J.J. Gertler. Survey of model-based failure detection and isolation in complex plants. *Control Systems Magazine, IEEE*, 8(6):3–11, 1988.

[66] S. Gorman. Electricity grid in us penetrated by spies. *Wall Street Journal*, 8, 2009.

[67] Blaise Kévin Guépié. *Détection séquentielle de signaux transitoires : application à la surveillance d'un réseau d'eau potable*. PhD thesis, Université de Technologie de Troyes, 2013.

[68] Blaise Kevin Guepie, Lionel Fillatre, and Igor Nikiforov. Detecting an abrupt change of finite duration. In *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on*, pages 1930–1934. IEEE, 2012.

[69] Blaise Kévin Guépié, Lionel Fillatre, and Igor Nikiforov. Sequential detection of transient changes. *Sequential Analysis*, 31(4):528–547, 2012.

[70] Blaise Kevin Guepie, Lionel Fillatre, and Igor V Nikiforov. Sequential monitoring of water distribution network. In *System Identification*, volume 16, pages 392–397, 2012.

[71] Rachana Ashok Gupta and Mo-Yuen Chow. Networked control system: Overview and research trends. *Industrial Electronics, IEEE Transactions on*, 57(7):2527–2535, 2010.

[72] Fredrik Gustafsson. Stochastic observability and fault diagnosis of additive changes in state space models. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 5, pages 2833–2836. IEEE, 2001.

[73] Fredrik Gustafsson. Stochastic fault diagnosability in parity spaces. In *Proceedings of the 15th IFAC World Congress*, pages 736–736, 2002.

[74] Chunming Han, Peter Willett, Biao Chen, and Douglas Abraham. A detection optimal min-max test for transient signals. *Information Theory, IEEE Transactions on*, 44(2):866–869, 1998.

[75] Chunming Han, Peter K Willett, and Douglas A Abraham. Some methods to evaluate the performance of page's test as used to detect transient signals. *Signal Processing, IEEE Transactions on*, 47(8):2112–2127, 1999.

[76] A Herrán-González, JM De La Cruz, B De Andrés-Toro, and JL Risco-Martín. Modeling and simulation of a gas distribution pipeline network. *Applied Mathematical Modelling*, 33(3):1584–1600, 2009.

[77] Joao P Hespanha, Payam Naghshtabrizi, and Yonggang Xu. A survey of recent results in networked control systems. *Proceedings-IEEE*, 95(1):138, 2007.

[78] Wassily Hoeffding. Lower bounds for the expected sample size and the average risk of a sequential procedure. *The Annals of Mathematical Statistics*, pages 352–368, 1960.

[79] Yi Huang, Husheng Li, Kristy A Campbell, and Zhu Han. Defending false data injection attack on smart grid network using adaptive cusum test. In *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*, pages 1–6. IEEE, 2011.

[80] Y.L. Huang, A.A. Cárdenas, S. Amin, Z.S. Lin, H.Y. Tsai, and S. Sastry. Understanding the physical and economic consequences of attacks on control systems. *International Journal of Critical Infrastructure Protection*, 2(3):73–83, 2009.

[81] Inseok Hwang, Sungwan Kim, Youdan Kim, and Chze Eng Seah. A survey of fault detection, isolation, and reconfiguration methods. *Control Systems Technology, IEEE Transactions on*, 18(3):636–653, 2010.

[82] Rolf Isermann. Process fault detection based on modeling and estimation methods - a survey. *Automatica*, 20(4):387–404, 1984.

[83] Rolf Isermann. *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer, 2006.

[84] ME Kabay. Attacks on power systems: Hackers, malware.

[85] Thomas Kailath. *Linear systems*, volume 1. Prentice-Hall Englewood Cliffs, NJ, 1980.

[86] Noriaki Kamioka. Scada system of tokyo gas for wide-area city gas distribution. In *SICE Annual Conference (SICE), 2012 Proceedings of*, pages 333–336. IEEE, 2012.

[87] SL Ke and HC Ti. Transient analysis of isothermal gas flow in pipeline network. *Chemical Engineering Journal*, 76(2):169–177, 2000.

[88] Jean-Yves Keller and Dominique Sauter. Monitoring of stealthy attack in networked control systems. In *Control and Fault-Tolerant Systems (SysTol), 2013 Conference on*, pages 462–467. IEEE, 2013.

[89] Jack Kiefer and Lionel Weiss. Some properties of generalized sequential probability ratio tests. *The Annals of Mathematical Statistics*, pages 57–74, 1957.

[90] Ryszard Klempous, Barbara Łysakowska, and Jan Nikodem. Supervisory control and data acquisition system for the gas flow networks. In *Computer Aided Systems Theory–EUROCAST'95*, pages 530–538. Springer, 1996.

[91] Karl-Rudolf Koch. *Parameter estimation and hypothesis testing in linear models*. Springer-Verlag Berlin Heidelberg, 1999.

[92] Damien Koenig, Nadia Bedjaoui, and Xavier Litrico. Unknown input observers design for time-delay systems application to an open-channel. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, pages 5794–5799. IEEE, 2005.

[93] Oliver Kosut, Liyan Jia, Robert J Thomas, and Lang Tong. Limiting false data attacks on power system state estimation. In *Information Sciences and Systems (CISS), 2010 44th Annual Conference*, pages 1–6. IEEE, 2010.

[94] Oliver Kosut, Liyan Jia, Robert J Thomas, and Lang Tong. Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 220–225. IEEE, 2010.

[95] Oliver Kosut, Liyan Jia, Robert J Thomas, and Lang Tong. On malicious data attacks on power system state estimation. In *Universities Power Engineering Conference (UPEC), 2010 45th International*, pages 1–6. IEEE, 2010.

[96] Jaroslav Králik, Petr Stiegler, Z Vostry, and Jiří Zavorka. Dynamic modeling of large-scale networks with application to gas distribution. 1988.

[97] B. Krebs. Cyber incident blamed for nuclear power plant shutdown. *Washington Post, June*, 5:2008, 2008.

[98] R.L. Krutz. *Securing SCADA systems*. Wiley Pub., 2006.

[99] Cheolhyeon Kwon, Weiyi Liu, and Inseok Hwang. Security analysis for cyber-physical systems against stealthy deception attacks. In *American Control Conference (ACC), 2013*, pages 3344–3349. IEEE, 2013.

[100] Tze Leung Lai. Optimal stopping and sequential tests which minimize the maximum expected sample size. *The Annals of Statistics*, pages 659–673, 1973.

[101] Tze Leung Lai. Nearly optimal sequential tests of composite hypotheses. *The Annals of Statistics*, pages 856–886, 1988.

[102] Tze Leung Lai. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 613–658, 1995.

[103] Tze Leung Lai. Information bounds and quick detection of parameter changes in stochastic systems. *Information Theory, IEEE Transactions on*, 44(7):2917–2929, 1998.

[104] Tze Leung Lai. Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems. *Information Theory, IEEE Transactions on*, 46(2):595–608, 2000.

[105] Tze Leung Lai. Sequential analysis: some classical problems and new challenges. *Statistica Sinica*, 11(2):303–350, 2001.

[106] Tze Leung Lai et al. Control charts based on weighted sums. *The Annals of Statistics*, 2(1):134–147, 1974.

[107] Tze Leung Lai and Jerry Zhaolin Shan. Efficient recursive algorithms for detection of abrupt changes in signals and control systems. *Automatic Control, IEEE Transactions on*, 44(5):952–966, 1999.

[108] Edward A Lee. Cyber physical systems: Design challenges. In *Object Oriented Real-Time Distributed Computing (ISORC), 2008 11th IEEE International Symposium on*, pages 363–369. IEEE, 2008.

[109] E.L. Lehmann and J.P. Romano. *Testing statistical hypotheses*. Springer, 2005.

[110] Erich Leo Lehmann. Some concepts of dependence. *The Annals of Mathematical Statistics*, pages 1137–1153, 1966.

[111] B Liscouski and W Elliot. Final report on the august 14, 2003 blackout in the united states and canada: Causes and recommendations. *A report to US Department of Energy*, 40(4), 2004.

[112] 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.

[113] G. Lorden. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, pages 1897–1908, 1971.

[114] Gary Lorden. Open-ended tests for koopman-darmois families. *The Annals of Statistics*, pages 633–643, 1973.

[115] Gary Lorden. 2-sprt's and the modified kiefer-weiss problem of minimizing an expected sample size. *The Annals of Statistics*, pages 281–291, 1976.

[116] R.K. Mehra and J. Peschon. An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica*, 7(5):637–640, 1971.

[117] Carl D Meyer. *Matrix analysis and applied linear algebra*. Siam, 2000.

[118] Bill Miller and Dale Rowe. A survey scada of and critical infrastructure incidents. In *Proceedings of the 1st Annual conference on Research in information technology*, pages 51–56. ACM, 2012.

[119] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli. False data injection attacks against state estimation in wireless sensor networks. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 5967–5972. IEEE, 2010.

[120] Y. Mo and B. Sinopoli. Secure control against replay attacks. In *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pages 911–918. IEEE, 2009.

[121] Y. Mo and B. Sinopoli. False data injection attacks in control systems. In *Preprints of the 1st Workshop on Secure Control Systems*, 2010.

[122] Yilin Mo, Rohan Chabukswar, and Bruno Sinopoli. Detecting integrity attacks on scada systems. *IEEE Transactions on Control Systems Technology, 23(4)*, pages 1396–1407, 2014.

[123] George V Moustakides. Multiple optimality properties of the shewhart test. *Sequential Analysis*, 33(3):318–344, 2014.

[124] G.V. Moustakides. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, 14(4):1379–1387, 1986.

[125] Pal-Stefan Murvay and Ioan Silea. A survey on gas leak detection and localization techniques. *Journal of Loss Prevention in the Process Industries*, 25(6):966–973, 2012.

[126] Igor Nikiforov. Eléments de théorie de la décision statistique ii: compléments. In Régis Lengellé, editor, *Décision et reconnaissance des formes en signal*. Lavoisier, 2002.

[127] Igor Nikiforov. Optimal sequential change detection and isolation. In *Proc. 15th IFAC World Congress-b'02, Barcelona, Spain*, 2002.

[128] Igor V Nikiforov. On two new criteria of optimality for the problem of sequential change diagnosis. In *American Control Conference, 1995. Proceedings of the*, volume 1, pages 97–101. IEEE, 1995.

[129] Igor V Nikiforov. A simple recursive algorithm for diagnosis of abrupt changes in random signals. *Information Theory, IEEE Transactions on*, 46(7):2740–2746, 2000.

[130] I.V. Nikiforov. A generalized change detection problem. *Information Theory, IEEE Transactions on*, 41(1):171–187, 1995.

[131] I.V. Nikiforov. Two strategies in the problem of change detection and isolation. *Information Theory, IEEE Transactions on*, 43(2):770–776, 1997.

[132] I.V. Nikiforov. A lower bound for the detection/isolation delay in a class of sequential tests. *Information Theory, IEEE Transactions on*, 49(11):3037–3047, 2003.

[133] James R Norris. *Markov chains*. Number 2. Cambridge university press, 1998.

[134] Albert H Nuttall. Detection performance of power-law processors for random signals of unknown location, structure, extent, and strength. Technical report, DTIC Document, 1994.

[135] Albert H Nuttall. Near-optimum detection performance of power-law processors for random signals of unknown locations, structure, extent, and arbitrary strengths. Technical report, DTIC Document, 1996.

[136] Albert H Nuttall. *Detection capability of linear-and-power processor for random burst signals of unknown location*. Naval Undersea Warfare Center Division, 1997.

[137] Andrzej Osiadacz. Simulation and analysis of gas networks. 1987.

[138] Taragay Oskiper and H Vincent Poor. Online activity detection in a multiuser environment using the matrix cusum algorithm. *Information Theory, IEEE Transactions on*, 48(2):477–493, 2002.

[139] ES Page. Continuous inspection schemes. *Biometrika*, pages 100–115, 1954.

[140] Fabio Pasqualetti. *Secure Control Systems: A Control-Theoretic Approach to Cyber-Physical Security*. PhD thesis, University of California, 2012.

[141] Fabio Pasqualetti, F Dorfler, and Francesco Bullo. Attack detection and identification in cyber-physical systems. *Automatic Control, IEEE Transactions on*, 58(11):2715–2729, 2013.

[142] M. Pollak. Optimal detection of a change in distribution. *The Annals of Statistics*, pages 206–227, 1985.

[143] M. Pollak and D. Siegmund. Approximations to the expected sample size of certain sequential tests. *The Annals of Statistics*, 3(6):1267–1282, 1975.

[144] M. Pollak and A.G. Tartakovsky. Optimality properties of the shiryaev-roberts procedure. *Statistica Sinica*, 2009.

[145] Moshe Pollak and Abba M Krieger. Shewhart revisited. *Sequential Analysis*, 32(2):230–242, 2013.

[146] A.S. Polunchenko and A.G. Tartakovsky. State-of-the-art in sequential change-point detection. *Methodology and computing in applied probability*, 14(3):649–684, 2012.

[147] H Vincent Poor and Olympia Hadjiliadis. *Quickest detection*, volume 40. Cambridge University Press Cambridge, 2009.

[148] Boaz Porat and Benjamin Friedlander. Performance analysis of a class of transient detection algorithms-a unified framework. *Signal Processing, IEEE Transactions on*, 40(10):2536–2546, 1992.

[149] K. Poulsen. Slammer worm crashed ohio nuke plant network. *Security Focus*, 19, 2003.

[150] K Premkumar, Anurag Kumar, and Venugopal V Veeravalli. Bayesian quickest transient change detection. In *Proc. Fifth International Workshop on Applied Probability (IWAP)*, 2010.

[151] Eyal Price and Avi Ostfeld. A successive linear programming scheme for optimal operation of water distribution networks. In *World Environmental and Water Resources Congress 2012@ sCrossing Boundaries*, pages 2964–2970. ASCE, 2012.

[152] Paul Quinn-Judge. Cracks in the system. *TIME Magazine*, January 9, 2002.

[153] Md Ashfaqur Rahman and Hamed Mohsenian-Rad. False data injection attacks with incomplete information against smart power grids. In *Global Communications Conference (GLOBECOM), 2012 IEEE*, pages 3153–3158. IEEE, 2012.

[154] Fahmida Y. Rashid. Telvent hit by sophisticated cyber-attack, scada admin tool compromised. http://www.securityweek.com/telvent-hit-sophisticated-cyber-attack-scada-admin-tool-compromised/, 26 September 2012.

[155] Bradley Reaves and Thomas Morris. Discovery, infiltration, and denial of service in a process control system wireless network. In *eCrime Researchers Summit, 2009. eCRIME'09.*, pages 1–9. IEEE, 2009.

[156] Thomas Reed. *At the abyss: an insider's history of the Cold War*. Random House LLC, 2007.

[157] Y. Ritov. Decision theoretic optimality of the cusum procedure. *The Annals of Statistics*, pages 1464–1469, 1990.

[158] SW Roberts. A comparison of some control chart procedures. *Technometrics*, 8(3):411–430, 1966.

[159] J. David Rogers and Conor M. Watkins. Overview of the taum sauk pumped storage power plant upper reservoir failure, reynolds county, mo. In *The 6th International Conference on Case Histories in Geotechnical Engineering*, Arlington, 2008.

[160] Henrik Sandberg, André Teixeira, and Karl H Johansson. On security indices for state estimators in power networks. In *Preprints of the First Workshop on Secure Control Systems, CPSWEEK 2010, Stockholm, Sweden*, 2010.

[161] Wojciech Sarnowski and Krzysztof Szajowski. Optimal detection of transition probability change in random sequence. *Stochastics An International Journal of Probability and Stochastic Processes*, 83(4-6):569–581, 2011.

[162] Thilo Sauter. The three generations of field-level networks - evolution and compatibility issues. *Industrial Electronics, IEEE Transactions on*, 57(11):3585–3595, 2010.

[163] Luca Schenato, Bruno Sinopoli, Massimo Franceschetti, Kameshwar Poolla, and S Shankar Sastry. Foundations of control and estimation over lossy networks. *Proceedings of the IEEE*, 95(1):163–187, 2007.

[164] Lester W Schmerr Jr. *Fundamentals of ultrasonic nondestructive evaluation: a modeling approach*. Springer, 1998.

[165] Walter Andrew Shewhart. *Economic control of quality of manufactured product*, volume 509. ASQ Quality Press, New York, USA, 1931.

[166] A.N. Shiryaev. On optimum methods in quickest detection problems. *Theory of Probability & Its Applications*, 8(1):22–46, 1963.

[167] Bruno Sinopoli, Luca Schenato, Massimo Franceschetti, Kameshwar Poolla, Michael I Jordan, and Shankar S Sastry. Kalman filtering with intermittent observations. *Automatic Control, IEEE Transactions on*, 49(9):1453–1464, 2004.

[168] J. Slay and M. Miller. Lessons learned from the maroochy water breach. *Critical Infrastructure Protection*, pages 73–82, 2007.

[169] Roy S Smith. A decoupled feedback structure for covertly appropriating networked control systems. *Proc. IFAC World Congress*, pages 90–95, 2011.

[170] Edward H Smyth and ABB Totalflow. Scada and telemetry in gas transmission systems. *ABB White Paper*, 2007.

[171] M. Sobel and A. Wald. A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. *The Annals of Mathematical Statistics*, 20(4):502–522, 1949.

[172] Keith Stouffer, Joe Falco, and Karen Scarfone. Guide to industrial control systems (ics) security. *NIST special publication*, pages 800–82, 2011.

[173] Roy L Streit and Peter K Willett. Detection of random transient signals via hyperparameter estimation. *Signal Processing, IEEE Transactions on*, 47(7):1823–1834, 1999.

[174] Morteza Talebi, Jianan Wang, and Zhihua Qu. Secure power systems against malicious cyber-physical data attacks: Protection and identification. In *International Conference on Power Systems Engineering*, pages 11–12, 2012.

[175] A. Tartakovsky, I. Nikiforov, and M. Basseville. *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 2014.

[176] AG Tartakovsky. Asymptotic performance of a multichart cusum test under false alarm probability constraint. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, pages 320–325. IEEE, 2005.

[177] A.G. Tartakovsky and G.V. Moustakides. State-of-the-art in bayesian changepoint detection. *Sequential Analysis*, 29(2):125–145, 2010.

[178] A.G. Tartakovsky and V.V. Veeravalli. Quickest change detection in distributed sensor systems. In *Proceedings of the 6th International Conference on Information Fusion*, pages 756–763, 2003.

[179] A.G. Tartakovsky and V.V. Veeravalli. Asymptotically optimal quickest change detection in distributed sensor systems. *Sequential Analysis*, 27(4):441–475, 2008.

[180] Alexander G Tartakovsky. Asymptotic optimality of certain multihypothesis sequential tests: Non-iid case. *Statistical Inference for Stochastic Processes*, 1(3):265–295, 1998.

[181] Alexander G Tartakovsky. Multidecision quickest change-point detection: Previous achievements and open problems. *Sequential Analysis*, 27(2):201–231, 2008.

[182] Alexander G Tartakovsky, Boris L Rozovskii, Rudolf B Blazek, and Hongjoong Kim. A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *Signal Processing, IEEE Transactions on*, 54(9):3372–3382, 2006.

[183] A. Teixeira, S. Amin, H. Sandberg, K.H. Johansson, and S.S. Sastry. Cyber security analysis of state estimators in electric power systems. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 5991–5998. IEEE, 2010.

[184] André Teixeira, Daniel Pérez, Henrik Sandberg, and Karl Henrik Johansson. Attack models and scenarios for networked control systems. In *Proceedings of the 1st international conference on High Confidence Networked Systems*, pages 55–64. ACM, 2012.

[185] André Teixeira, Henrik Sandberg, Gyorgy Dan, and Karl H Johansson. Optimal power flow: Closing the loop over corrupted data. In *American Control Conference (ACC), 2012*, pages 3534–3540. IEEE, 2012.

[186] André Teixeira, Iman Shames, Henrik Sandberg, and Karl H Johansson. Revealing stealthy attacks in control systems. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pages 1806–1813. IEEE, 2012.

[187] André Teixeira, Iman Shames, Henrik Sandberg, and Karl Henrik Johansson Johansson. A secure control framework for resource-limited adversaries. *Automatica*, 2014.

[188] R. Tsang. Cyberthreats, vulnerabilities and attacks on scada networks. *University of California, Berkeley, Working Paper, http://gspp. berkeley. edu/iths/Tsang_SCADA% 20Attacks. pdf (as of Dec. 28, 2011)*, 2010.

[189] R.J. Turk. *Cyber incidents involving control systems.* Idaho National Engineering and Environmental Laboratory, 2005.

[190] Maria E Valcher. State observers for discrete-time linear systems with unknown inputs. *Automatic Control, IEEE Transactions on*, 44(2):397–401, 1999.

[191] J. van Helvoirt. Centrifugal compressor surge: Modeling and identification for control, 2007.

[192] V.V. Veeravalli and C.W. Baum. Asymptotic efficiency of a sequential multihypothesis test. *Information Theory, IEEE Transactions on*, 41(6):1994 –1997, nov 1995.

[193] A. Wald. *Sequential analysis.* John Wiley & Sons, Inc., New York, USA, 1947.

[194] A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3):326–339, 1948.

[195] Abraham Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.

[196] Abraham Wald and Jacob Wolfowitz. Bayes solutions of sequential decision problems. *The Annals of Mathematical Statistics*, pages 82–99, 1950.

[197] Tobias Walk. Cyber-attack protection for pipeline scada systems. *Pipelines International Digest*, 2012.

[198] J Wang and P Willett. Detecting transients of unknown length. In *Aerospace Conference, 2005 IEEE*, pages 2236–2247. IEEE, 2005.

[199] J Wang and Peter Willett. A variable threshold page procedure for detection of transient signals. *Signal Processing, IEEE Transactions on*, 53(11):4397–4402, 2005.

[200] Zhen Wang and Peter Willett. A performance study of some transient detectors. *IEEE transactions on signal processing*, 48(9):2682–2685, 2000.

[201] Zhen Wang and Peter Willett. Improved power-law detection of transients. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 5, pages 3181–3184. IEEE, 2001.

[202] Zhen Wang and Peter K Willett. All-purpose and plug-in power-law detectors for transient signals. *Signal Processing, IEEE Transactions on*, 49(11):2454–2466, 2001.

[203] Peter Willett and Biao Chen. A new sequential detector for short-duration signals. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 4, pages 2529–2532. IEEE, 1998.

[204] Peter Willett and Z Jane Wang. The vtp test for transients of equal detectability. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 5, pages V–273. IEEE, 2003.

[205] A. Willsky and H. Jones. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *Automatic Control, IEEE Transactions on*, 21(1):108–112, 1976.

[206] A.S. Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12(6):601–611, 1976.

[207] A.S. Willsky, JJ Deyst, and BS Crawford. Two self-test methods applied to an inertial system problem. *Journal of Spacecraft and Rockets*, 12:434–437, 1975.

[208] A.S. Willsky and H.L. Jones. A generalized likelihood ratio approach to state estimation in linear systems subjects to abrupt changes. In *Decision and Control including the 13th Symposium on Adaptive Processes, 1974 IEEE Conference on*, volume 13, pages 846–853. IEEE, 1974.

[209] J Wunnenberg and PM Frank. Sensor fault detection via robust observers. In *System fault diagnostics, reliability and related knowledge-based approaches*, pages 147–160. Springer, 1987.

[210] Le Xie, Yilin Mo, and Bruno Sinopoli. False data injection attacks in electricity markets. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 226–231. IEEE, 2010.

[211] Le Xie, Yilin Mo, and Bruno Sinopoli. Integrity data attacks in power market operations. *Smart Grid, IEEE Transactions on*, 2(4):659–666, 2011.

[212] Yanling Yuan, Zuyi Li, and Kui Ren. Modeling load redistribution attacks in power systems. *Smart Grid, IEEE Transactions on*, 2(2):382–390, 2011.

[213] Kim Zetter. Attack on city water station destroys pump. http://www.wired.com/threatlevel/2011/11/hackers-destroy-water-pump/, 2011.

[214] B. Zhu, A. Joseph, and S. Sastry. A taxonomy of cyber attacks on scada systems. In *Internet of Things (iThings/CPSCom), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing*, pages 380–388. IEEE, 2011.

[215] B. Zhu and S. Sastry. Scada-specific intrusion detection/prevention systems: a survey and taxonomy. In *Proceedings of the 1st Workshop on Secure Control Systems (SCS)*, 2010.

# Index

## O



## P



## R



## S



## T



## U



## V



## W



## Z

# Van Long DO

## Doctorat : Optimisation et Sûreté des Systèmes

Année 2015


## Détection et localisation séquentielle d'attaques cyber-physiques aux systèmes SCADA


Cette thèse s'inscrit dans le cadre du projet « SCALA » financé par l'ANR à travers le programme ANR-11-SECU-0005. Son objectif consiste à surveiller des systèmes de contrôle et d'acquisition de données (SCADA) contre des attaques cyber-physiques. Il s'agit de résoudre un problème de détection-localisation séquentielle de signaux transitoires dans des systèmes stochastiques et dynamiques en présence d'états inconnus et de bruits aléatoires. La solution proposée s'appuie sur une approche par redondance analytique composée de deux étapes : la génération de résidus, puis leur évaluation. Les résidus sont générés de deux façons distinctes, avec le filtre de Kalman ou par projection sur l'espace de parité. Ils sont ensuite évalués par des méthodes d'analyse séquentielle de rupture selon de nouveaux critères d'optimalité adaptés à la surveillance des systèmes à sécurité critique. Il s'agit donc de minimiser la pire probabilité de détection manquée sous la contrainte de niveaux acceptables pour la pire probabilité de fausse alarme et la pire probabilité de fausse localisation. Pour la tâche de détection, le problème d'optimisation est résolu dans deux cas : les paramètres du signal transitoire sont complètement connus ou seulement partiellement connus. Les propriétés statistiques des tests sous-optimaux obtenus sont analysées. Des résultats préliminaires pour la tâche de localisation sont également proposés. Les algorithmes développés sont appliqués à la détection et à la localisation d'actes malveillants dans un réseau d'eau potable.



Mots clés : analyse séquentielle - détection du signal - rupture (statistique) - modèles linéaires (statistique) - criminalité informatique.


## Sequential Detection and Isolation of Cyber-physical Attacks on SCADA Systems


This PhD thesis is registered in the framework of the project "SCALA" which received financial support through the program ANR-11-SECU-0005. Its ultimate objective involves the on-line monitoring of Supervisory Control And Data Acquisition (SCADA) systems against cyber-physical attacks. The problem is formulated as the sequential detection and isolation of transient signals in stochastic-dynamical systems in the presence of unknown system states and random noises. It is solved by using the analytical redundancy approach consisting of two steps: residual generation and residual evaluation. The residuals are firstly generated by both Kalman filter and parity space approaches. They are then evaluated by using sequential analysis techniques taking into account certain criteria of optimality. However, these classical criteria are not adequate for the surveillance of safety-critical infrastructures. For such applications, it is suggested to minimize the worst-case probability of missed detection subject to acceptable levels on the worst-case probability of false alarm and false isolation. For the detection task, the optimization problem is formulated and solved in both scenarios: exactly and partially known parameters. The sub-optimal tests are obtained and their statistical properties are investigated. Preliminary results for the isolation task are also obtained. The proposed algorithms are applied to the detection and isolation of malicious attacks on a simple SCADA water network.



Keywords: sequential analysis - signal detection - change-point problems - linear models (statistics) – computer crimes.


**Thèse réalisée en partenariat entre :**

Agence Nationale de la Recherche ANR

utt université de technologie Troyes

**Ecole Doctorale** "Sciences et Technologies"