ANALYSIS, DETECTION, AND MITIGATION OF ATTACKS IN CYBER-PHYSICAL SYSTEMS

HANXIAO LIU

School of Electrical and Electronic Engineering

A thesis submitted to the Nanyang Technological University and KTH Royal Institute of Technology in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

2021

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

29 September 2021
.....
Date

.....
Hanxiao Liu

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

29 September 2021

Date

Prof. Lihua Xie

29 September 2021

Date

Prof. Karl Henrik Johansson

Authorship Attribution Statement

This thesis contains material from six papers published or under review in the following peer-reviewed journals and conferences and one paper under preparation in which I am listed as an author.

Chapter 2 is based on the following publication:

- Hanxiao Liu, Yilin Mo, and Karl Henrik Johansson, "Active Detection against Replay Attack: a Survey on Watermark Design for Cyber-physical Systems," *Safety, Security, and Privacy for Cyber-physical Systems*, *Springer*, 2020.

The contributions of the co-authors are as follows:

- Hanxiao Liu prepared the manuscript drafts.
- Yilin Mo supervised the research and revised the drafts.
- Karl Henrik Johansson participated in discussions and revised the drafts.

Chapter 3 is based on the following publications:

- Hanxiao Liu, Yuqing Ni, Lihua Xie, and Karl Henrik Johansson, "An Optimal Linear Attack Strategy on Remote State Estimation," in *Proceedings of IFAC World Congress*, 2020.
- Hanxiao Liu, Yuqing Ni, Lihua Xie, and Karl Henrik Johansson, "How Vulnerable is Innovation-based Remote State Estimation: Fundamental Limits under Linear Attacks," *Automatica*, provisionally accepted as Regular Paper.

The contributions of the co-authors are as follows:

- Hanxiao Liu developed the approach, performed the simulations and prepared the manuscript drafts.
- Yuqing Ni participated in the discussion of the proposed approach and revised the drafts.
- Lihua Xie supervised the research and revised the drafts.
- Karl Henrik Johansson participated in discussions and revised the drafts.

Chapter 4 is based on the following publications:

- Hanxiao Liu, Jiaqi Yan, Yilin Mo, and Karl Henrik Johansson, "An Online Design of Physical Watermarks," in *Proceedings of IEEE Conference on Decision and Control*, pp. 440-445, 2018.

- Hanxiao Liu, Yilin Mo, Jiaqi Yan, Lihua Xie, and Karl Henrik Johansson, "An Online Approach to Physical Watermark Design," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3895-3902, 2020.

The contributions of the co-authors are as follows:

- Hanxiao Liu prepared the manuscript drafts and performed the simulations.
- Yilin Mo provided the initial research direction, supervised the research, and revised the drafts.
- Jiaqi Yan participated in discussions and revised the drafts.
- Lihua Xie and Karl Henrik Johansson supervised the research and revised the drafts.

Chapter 5 is based on the following publication:

- Hanxiao Liu, Yuchao Li*[1], Jonas Mårtensson, Lihua Xie, and Karl Henrik Johansson, "Reinforcement Learning Based Approach for Flip Attack Detection," in *Proceedings of the 59th IEEE Conference on Decision and Control*, pp. 3212-3217, 2020.

The contributions of the co-authors are as follows:

- Hanxiao Liu proposed the approach, carried out the analysis, and prepared the manuscript drafts.
- Yuchao Li co-designed the approach, prepared, and revised the drafts.
- Lihua Xie supervised the research and revised the drafts.
- Jonas Mårtensson and Karl Henrik Johansson participated in discussions and revised the drafts.

Chapter 6 is based on the following paper:

- Hanxiao Liu, Yuchao Li, Karl Henrik Johansson, Jonas Mårtensson, and Lihua Xie, "Rollout Approach to Sensor Scheduling for Remote State Estimation under Integrity Attack," under review.

The contributions of the co-authors are as follows:

- Hanxiao Liu proposed the initial direction, performed simulation works, prepared the manuscript drafts.
- Yuchao Li co-designed the approach, prepared and revised the drafts.
- Lihua Xie, Karl Henrik Johansson, and Jonas Mårtensson participated in discussions and revised the drafts.

[1]The superscript $*$ indicates joint first authors.

29 September 2021

Date

Hanxiao Liu

Acknowledgements

The past four years have been one of the best times of my life. There have been many people who have supported and encouraged me throughout this period. Without their support and patience, this thesis would not have been completed. I would like to express my sincere gratitude to them.

First and foremost, I would like to express my deepest indebtedness to my supervisors, Prof. Lihua Xie, Prof. Karl Henrik Johansson and Prof. Yilin Mo, for offering me the opportunity to join the NTU-KTH Joint PhD Programme so that I can adapt to different environments and meet diverse cultures. I believe that this special experience will benefit all my life. I will be forever in their debt.

I am sincerely thankful to Prof. Xie for his excellent guidance, continuous support and positive encouragement throughout my Ph.D. study and research. I am so grateful for the detailed, prompt, and insightful feedback on all manuscripts that I have sent to him. I learn from Prof. Xie not only about control theory but also positive attitude towards research. For all of this, I can't thank him enough.

I would also like to express my deepest gratitude to my co-supervisor, Prof. Johansson, for his insightful, careful, and detailed comments. I learn from Kalle the enthusiastic attitude and how to consider problems from a border perspective in research. It is really a great honor and pleasure to work with him. The almost one and half year at KTH is unforgettable.

Prof. Mo was my supervisor at the beginning of my PhD career. I really appreciate his patient mentoring and I have learned a lot from him: not only on how to find and solve challenging problems, but also how to write and present papers clearly. I am truly thankful to him for contributing his time and thoughts to me. I am very lucky to have Prof. Mo as my supervisor.

I am sincerely thankful to Yuchao Li from KTH for introducing me to the technical details of reinforcement learning and optimal control during the past two and half

years. Thanks for his valuable advice and continuous support. I would also like to thank Yuqing Ni for her encouragement and company. The time in Sweden became more memorable because of her. I am so glad to meet Yuqing at KTH.

Many thanks to all colleagues and friends at NTU. In particular, thanks to all members of Prof. Xie's group, specially to Zhirong Qiu, Han Wang, Juncheng Li, Kemi Ding, for their kind assistance. Especially, I would like to thank Liang Xu and Yulong Gao for revising the slides and providing valuable suggestions on the presentation of my qualification examinations. Many thanks to all members of Robotics I and the Internet of Things lab for providing a friendly environment. Besides, I am grateful to my fiends Yaling Jin, Yi Ding, Xiaoyi Shen, Xiaoyu Tang, and Xue Li for their company and encouragement. I would like to express my gratitude to all my friends and colleagues at KTH. Many thanks to all my colleagues at the division of Division of Decision and Control for creating a friendly environment. In particular, I am grateful to Xiaoqiang Ren for all the time shared with me at the beginning of my time in Sweden. Special thanks to Yuchao Li for proof reading this thesis, and Elis Stefansson for translating the abstract into Swedish. I would also like to thank Xinlei Yi, Yuchao Li, Song Fang, Kuize Zhang, Mingliang Wang, Xiao Tan, and Pian Yu for the warm welcome and kind assistance. A special mention to my office mates Fei Chen, Joana Fonseca, Mina Ferizbegovic, Rodrigo A. González, and Yuchao Li for providing a friendly environment. Many thanks to David Umsonst, Xingkang He, Yu Xing, Yu Wang for the interesting discussions. I thank Péter Várnai, Erik Berglund, and Elis Stefansson for playing table tennis together. I also want to thank the executives Wong Chow Pang, Wei Jiuan and Goh-Fong Lai Peng from NTU, and Felicia Gustafsson, Tord Christer Magnusson, and Emanuel Borg at KTH, for their support and help.

Finally but by no means least, I wish to dedicate this thesis to my family. In particular special and sincere thanks to my parents for their continuous and unconditional support and love throughout my life. Thanks to my companion Ding Lyu for his continuous support and endless sacrifices. In true love, the greatest distance can be bridged. Without his love and understanding, I think I might not finish this thesis.

Hanxiao Liu
September, 2021

Abstract

Cyber-Physical Systems (CPS) offer close integration among computational elements, communication networks, and physical processes. Such systems play an increasingly important role in a large variety of fields, such as manufacturing, health care, environment, transportation, defence, and so on. Due to the wide applications and critical functions of CPS, increasing importance has been attached to their security. In this thesis, we focus on the security of CPS by investigating vulnerability under cyber-attacks, providing detection mechanisms, and developing feasible countermeasures against cyber-attacks.

The first contribution of this thesis is to analyze the performance of remote state estimation under linear attacks. A linear time-invariant system equipped with a smart sensor is studied. The adversary aims to maximize the state estimation error covariance while staying stealthy. The maximal performance degradation that an adversary can achieve with any linear first-order false data injection attack under strict stealthiness for vector systems and ϵ-stealthiness for scalar systems is characterized. We also provide an explicit attack strategy that achieves this bound and compare it with strategies previously proposed in the literature.

The second problem of this thesis is about the detection of replay attacks. We aim to design physical watermark signals and corresponding detector to protect a control system against replay attacks. For a scenario where the system parameters are available to the operator, a physical watermarking scheme to detect the replay attack is introduced. The optimal watermark signal design problem is formulated as an optimization problem, and the optimal watermark signal and detector are derived. Subsequently, for systems with unknown parameters, we provide an online learning mechanism to asymptotically derive the optimal watermarking signal and corresponding detector.

The third problem under investigation is about the detection of false-data injection attacks when the attacker injects malicious data to flip the distribution of the

manipulated sensor measurements. The detector decides to continue taking observations or to stop based on the received signals, and the goal is to have the flip attack detected as fast as possible while trying to avoid terminating the measurements when no attack is present. The detection problem is modeled as a partially observable Markov decision process (POMDP) by assuming an attack probability, with the dynamics of the hidden states of the POMDP characterized by a stochastic shortest path (SSP) problem. The optimal policy of the SSP solely depends on the transition costs and is independent of the assumed attack probability. By using a fixed-length window and suitable feature function of the measurements, a Markov decision process (MDP) is used to approximate the POMDP. The optimal solution of the MDP is obtained by reinforcement learning.

The fourth contribution of this thesis is to develop a sensor scheduler for remote state estimation under integrity attacks. We seek a trade-off between the energy consumption of communications and accuracy of state estimation when the acknowledgment (ACK) information, sent by the remote estimator to the local sensor, is compromised. The sensor scheduling problem is formulated as an infinite horizon discounted optimal control problem with infinite states. We first analyze the underlying MDP and show that the optimal schedule without ACK attack is of threshold type. Thus, we can simplify the problem by replacing the original state space with a finite state space. For the simplified MDP, when ACK is under attack, the problem is modelled as a POMDP. We analyze the induced MDP that uses a belief vector as its state for the POMDP. The properties of the exact optimal solution are studied via contractive models and it is shown that the threshold solution for the POMDP cannot be readily obtained. A suboptimal solution is provided instead via a rollout approach based on reinforcement learning. We present two variants of rollout and provide corresponding performance bounds.

Contents

List of Figures

List of Tables

Symbols and Acronyms

Symbols

$\mathbb{N}$	the set of natural numbers
$\mathbb{Z}$	the set of integers
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^n$	the set of n-dimensional real column vectors
$\mathbb{R}^{m\times n}$	the set of $m \times n$-dimensional real matrices
$\mathbb{S}^n_+$ ($\mathbb{S}^n_{++}$)	the set of $n \times n$ positive semi-definite (definite) matrices
$X \succeq 0$ ($X \succ 0$)	positive semi-definite (definite) matrix
$\lvert\mathcal{S}\rvert$	the cardinally of finite set $\mathcal{S}$
$\mathbf{0}_{m\times n}$	an $m \times n$ matrix with all entries being zero
I_N	the N-by-N identity matrix
$A^\intercal$	the transpose of matrix A
A^{-1}	the inverse of matrix A
A^+	the pseudo-inverse of matrix A
$A_{i,j}$	the ith row, jth column element of matrix A
$\lVert A\rVert_F$	the Frobenius norm of matrix A
$\mathrm{rank}(A)$	the rank of matrix A
$\mathrm{tr}(A)$	the trace of matrix A
$\rho(A)$	the spectral radius of matrix A
$\det(A)$	the determinant of matrix A
$\lfloor a \rfloor$	the floored integer of real number a
$\mathbb{E}\{\cdot\}$	the expectation operator
$\otimes$	the Kronecker product
$\mathcal{N}(\mu, \Sigma)$	the Gaussian distribution with mean μ and covariance Σ
$h^n(x)$	the function composition of functions $h(h^{n-1}(x))$
$x_{k_1}^{k_2}$	the sequence $\{x_{k_1}, x_{k_1+1}, \ldots, x_{k_2}\}$

$\mathbf{1}$	an all-ones column vector with proper dimension
$\text{supp}(\mathbf{v})$	the set of indices of the components of vector $\mathbf{v}$ which are nonzero

Acronyms

ACK	acknowledgement
CPS	cyber-physical systems
DoS	denial of service
KL	Kullback–Leibler
LQG	linear–quadratic–Gaussian
LTI	linear time-invariant
MDP	Markov decision process
MLR	monotone likelihood ratio
MMSE	minimum mean square error
NP	Neyman-Pearson
PDF	probability density function
POMDP	partially observable Markov decision process
QCD	quickest change detection
RL	reinforcement learning
SISO	single input single output
SSP	stochastic shortest path
a.s.	almost sure convergence of a random sequence
i.i.d.	independent and identically distributed

Chapter 1

Introduction

Cyber-physical systems (CPS) integrate computational elements and physical processes closely. Such systems play a critical role in large varieties of fields, such as manufacturing [6–8], health care [9–11], energy [12–14], and transportation [15–17]. Due to their wide applications and critical functions, it is of paramount importance to ensure the secure operation of CPS. However, this is very challenging due to more and more intelligent malicious cyber-attacks. In this thesis, we focus on the security of CPS by investigating the performance of cyber-attacks, providing detection mechanisms, and developing feasible countermeasures against cyber-attacks.

The outline of this chapter is as follows. Section 1.1 provides the motivations and objectives. The main contributions and the organization of the thesis are provided in Section 1.2.

1.1 Motivations and Objectives

CPS are playing a critical role in our life and society. There are various instances of CPS in people's life, such as smart grids, autonomous automobile systems, robotic systems including but not limited to cleaning robots and personal assistance robots, industrial control systems, medical monitoring, and so on [18]. Figure 1.1 shows the interaction of actors in a smart grid through secure communication flows and electrical flows, and Figure 1.2 illustrates an autonomous vehicle.

Because of the wide applications and critical functions of CPS, a successful attack can hamper economy, cause damage to critical infrastructure and even lead to the loss of human life. There have been many security incidents, of which Stuxnet malware is a representative one, as shown in Figure 1.3. It is the first worm known to attack industrial control systems and is believed to have destroyed 984 uranium enriching centrifuges, which constituted around 30% decrease in enrichment efficiency [19, 20]. Similar security incidents include Maroochy water breach in 2000 [21], Ukrainian power outage in 2015 [22], Triton Malware in 2017 [23], Venezuela blackouts in 2019 [24], and Russian power grid attacks in 2019 [25].

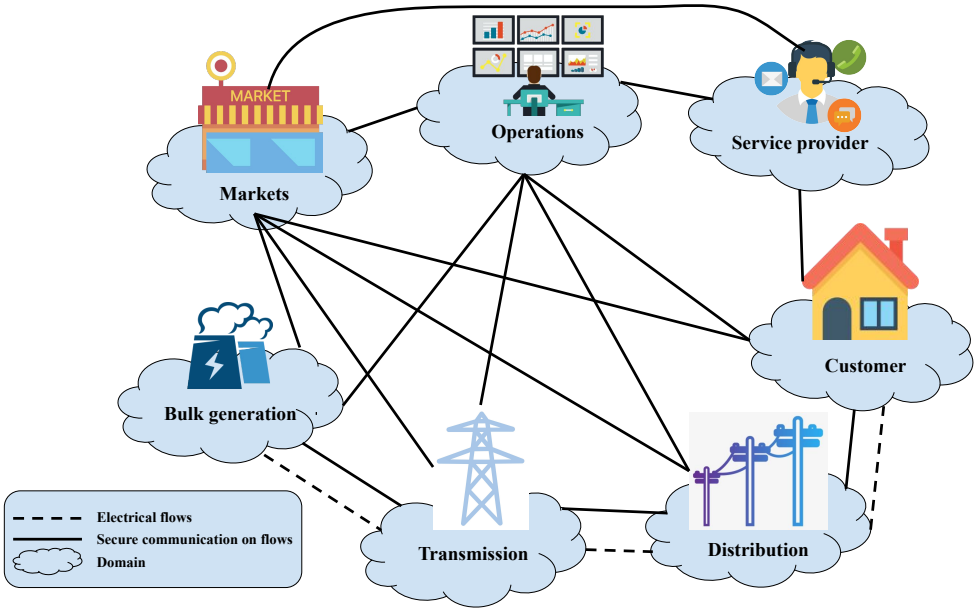

FIGURE 1.1: Illustration of a smart grid system [1].

Securing the normal operation of CPS is very challenging because of more and more intelligent attackers. These attackers may launch stealthy attacks that can cause the system performance degradation. Besides, the complex structure of CPS increases difficulty of analysis, detection, and mitigation. In this thesis, we focus on the security of CPS by investigating the performance of cyber-attacks, providing detection mechanisms, and developing feasible countermeasures against cyber-attacks. The four considered problems from three aspects are discussed next:

- Analysis of cyber-attacks can help us better estimate the system performance degradation that the attack may result in. This is of great importance for the security of CPS.

FIGURE 1.2: Illustration of autonomous vehicles [2].

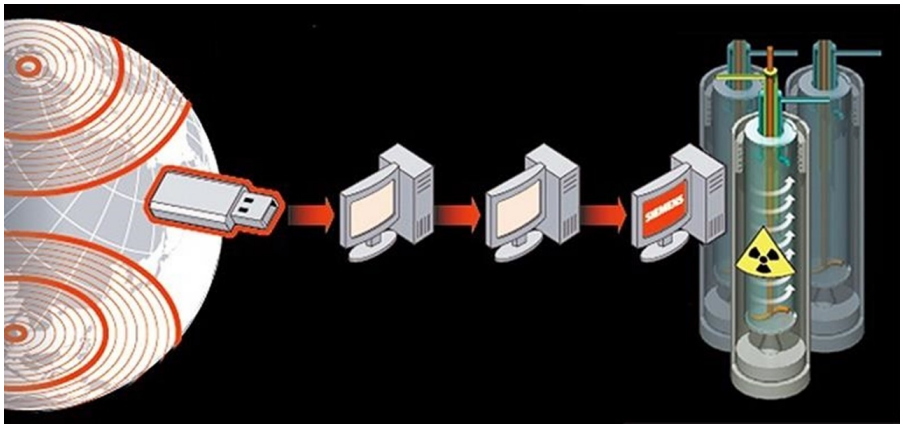

FIGURE 1.3: Stuxnet virus attack principle [3].

In this thesis, the first problem that we consider is to analyze fundamental performance degradation of false data injection attacks, a representative attack type. As a branch of false data injection attack, innovation-based linear attack was first studied in [26]. The authors proposed an optimal linear attack strategy that achieves the maximal performance degradation while not being detected by a χ^2 detector. They also investigated this type of attacks where the stealthiness level was characterized by Kullback–Leibler (KL) divergence [5]. Different from previous attack strategies that have a zero mean random variable, a more general linear attack strategy with an arbitrary mean random variable was studied in [27]. However, the listed works only focus on the attack strategy without memory. In view of this, open questions are *"Is a*

larger performance degradation of the remote state estimation expected when an attacker utilizes both past and present information?" and *"How vulnerable is innovation-based remote state estimation under linear attacks with memory?"* We consider how vulnerable innovation-based remote state estimators are to a linear attack based on both past and present innovations. We derive fundamental vulnerabilities of innovation-based remote estimation. In other words, we seek to evaluate how secure one can make innovation-based remote state estimation under linear attacks.

- Accurate detection of attacks can effectively reduce the risk and loss that a potential attack induces. If the attack can be detected timely and accurately, the defender or operator can take corresponding countermeasures to minimize negative effects. This is very crucial for securing the CPS.

  The second problem we consider is the detection problem of replay attacks, which is motivated by the occurrence of Stuxnet malware. Compared with other attacks, the replay attack needs no knowledge of system parameters and structure. Therefore, it is much simpler to perform this kind of attack from the point of adversaries. The attacker of the replay attack eavesdrops the data from the sensor network, and replays them afterwards. Moreover, χ^2 detector is a kind of detector which is widely employed to detect anomalies in control systems, but this kind of detector is useless for the replay attack [28]. We consider how to design an appropriate detection scheme. It is worth noticing that, a considerable number of published papers assume that the operator or designer has precise knowledge of the system parameters in order to design the corresponding detection scheme. However, acquiring the parameters may be troublesome and costly. Hence, it is beneficial for the system to "learn" the parameters during its operation and automatically design the detector and corresponding detection scheme in real time. Motivated by this idea, we study an online learning scheme to learn the parameters during its operation and automatically design the detector. We develop a data-driven approach to design physical watermark signals to protect systems with unknown parameters, against replay attacks.

  As an important part of CPS, networked embedded sensors are widely used to monitor plants and to detect anomalies. At the same time, due to their vulnerability to malicious attacks, increasing importance has been attached to researches on the security of those systems. From the existing work, it is

shown that the flip attack, where the attacker flips the distribution of the manipulated sensor measurements via injecting specific signals, is optimal to a broad class of problems. Therefore, the third problem considered is how to design detectors for flip attacks. Some researchers captured the properties of the detection problems by the formulation of partially observable Markov decision process (POMDP) and introduce belief states to solve the induced Markov decision processes (MDPs). However, the POMDP framework cannot capture the detection problem with extraneous inputs from attackers. To address this problem, we assume a transition probability to model the attack possibility. However, this assumption makes the solution sensitive to this assumed probability. To circumvent those challenges, Kurt *et al.* [29] applied fixed-length window of observation as the state for online detection in smart grids. Similar idea also appeared in [30] for online learning and attack design. However, unlike the success of RL reported for playing games where the POMDP is given by some simulators with transition probabilities and costs enclosed in the simulator [31], they need to be designed here. It is not clear from those works how the POMDPs shall be designed. In Chapter 5, we will present the design process. Besides, our focus is to determine if there is an attack as quickly as possible. It is somewhat like the quickest change detection (QCD) problem. However, unlike those problems where the probability density functions (PDFs) before and after the change point are known [32, 33], in our work, only a set of possible PDFs are known. We aim to design a flip attack detector to protect a sensor system aiming to estimate a binary state.

- How to mitigate the effects resulting from attacks is equally important for maintaining the normal operation of CPS.

  As the fourth problem of this thesis, we consider the sensor scheduling problem for remote state estimation under attacks. The trade-off between energy consumption and accuracy of remote state estimation has been studied in the literature. To enhance the performance of sensor scheduling, an acknowledgement (ACK) scheme, where the remote estimator or fusion center sends an acknowledgement to the smart sensor to indicate if it receives a packet, is getting more and more attention. Some researchers focus on the effect of ACK-based attacks where the ACK is modified by adversaries. Guo et al. [34] studied the denial of service (DoS) attack on the feedback channel

against ACK-based sensor schedule. They proved that the optimal policy is of threshold type. From the perspective of a defender, an ACK-based deception scheme for sensors was proposed in [35]. The authors used a general asymmetric-information stochastic game to model the interactions between the sensor and the attacker and showed an equivalent belief-based stochastic game to obtain the optimal stationary strategy for each agent [36]. Different from the above works regarding the active deception-based scheme, our work focuses on obtaining an optimal schedule when the ACK received by the sensor is attacked.

1.2 Thesis Outline and Contributions

In this section, we outline the contents of the thesis and present the main contributions.

In Chapter 2, we introduce three kinds of cyber-attacks and give a comprehensive review of the relevant research on analysis, detection, and mitigation of cyber-attacks.

Chapter 3 focuses on the performance analysis of innovation-based remote state estimation under linear attack. Firstly, fundamental bounds on how much the innovation-based remote state estimation performance may be degraded under strictly stealthy attacks for vector systems and ϵ-stealthy attacks for scalar systems are derived. Secondly, explicit expressions are derived for the worst-case attacks and it is shown how the adversary can utilize memory in maximizing the damage of attack. Thirdly, the derived attacks and their impact are compared to other attacks proposed in the literature.

Chapter 4 studies the problem of designing physical watermark signals to detect possible replay attack under the assumption that the system parameters are unknown and need to be identified online. The detection problem of replay attack via physical watermark is discussed, and a countermeasure of designing a watermarking signal that achieves the optimal trade-off between the control performance and detection performance is proposed. The contributions are: firstly, an online learning algorithm is presented to simultaneously infer the parameters of the system based only on the system input and output data and generate the watermark signal as

well as the optimal detector based on the estimated parameters. To the best of our knowledge, it is the first study of detection of replay attacks with unknown system parameters; secondly, we prove that the system parameters which are inferred via our proposed online algorithm converge to the true parameters almost surely even if the input signal asymptotically converges to a degenerate signal; thirdly, we characterize almost sure convergence rate of the estimated system parameters to the true parameters and provide an upper bound for this rate.

Chapter 5 discusses the detection problem of flip attacks. We show how the detection problem can be modeled as a POMDP and how to approximate the behaviour of POMDP via an MDP. Conditions for designing POMDP used to model the flip attacks are given. It is shown that the optimal behavior is independent of the assumed attack possibility. Finally, it is shown that the obtained detector is robust to the assumed attack probability via numerical evaluations.

Chapter 6 investigates the sensor scheduling problem for remote state estimation under integrity attacks. We show that the underlying MDP has a threshold type optimal policy. After it is proved that the exact solution cannot be established via monotone likelihood ratio (MLR) ordering for the simplified problem, we seek an approximate solution and provide two variants of rollout and the corresponding performance guarantees. We prove that the optimal policy of the underlying MDP is of threshold type. We present some properties of the exact optimal solution through contractive models for MDP with belief state and prove that the structural result can not be established through MLR ordering. We provide a suboptimal solution based on an approximation in value space and implement it through rollout with fixed and geometrically distributed truncated steps. The corresponding performance guarantee is provided.

In Chapter 7, we conclude the thesis and discuss directions for future research.

Chapter 2

Background

In this chapter, we review the relevant literature in the cyber-physical systems (CPS) security. Section 2.1 introduces the basic concepts of common cyber-attacks, such as denial of service (DoS), replay, and false data injection attacks. Sections 2.2-2.4 review the existing results regarding the analysis, detection, and mitigation of CPS attacks.

2.1 Attack Scenarios in CPS

There are various cyber-attacks in CPS. Generally speaking, attackers launch attacks by compromising certain sensors, controllers, or communication channels between sensors and controllers or between sensors and remote estimators. The consequence of these attacks on CPS are different from that in traditional computer systems. In traditional computer security, three main concepts are confidentiality, integrity, and availability, which are known as the CIA triad. This triad is used to evaluate the information security [37]. Confidentiality is the ability to protect the information from being accessed by unauthorized parties. Integrity is the ability to prevent the data from being changed by unauthorized parties. Availability refers to the ability that the information can be available when they are needed. Different from traditional cyber systems which mainly focus on the protection of data, and the security of CPS needs to consider the potential attack on the protection of information, physical systems, and communication channels between different devices. Cyber-attacks on CPS may affect physical processes, such as increasing

the state estimation error and degrading the system performance, thus affecting the stabilizability and reliability of systems, and even disrupting the whole physical systems.

In view of this, Teixeira *et al.* [4] proposed three dimensions for the attack space: the adversary's *a priori* system model knowledge, disclosure, and disruption resources, as shown in Figure 2.1. For the dimension "*a priori* system knowledge", the attacker can employ this information to launch more intelligent attacks which may have more complex attack structures and more severe results than the attacker without such knowledge. The second dimension "disclosure resources" can be used by the adversaries to obtain the information of the system and monitor the system in real-time. It is worth noticing that the adversary with only disclosure resources does not disrupt the system. One representative example of this kind of attack is eavesdropping attacks [38]. However, once the attacker has "disruption resources", they are able to affect the normal communication between different devices in the system or between the system and external equipments. At the same time, the integrity and availability of data are violated. A representative form that the attacker employs disruption resources is denial of service (DoS) attacks, for which we give details in later part.

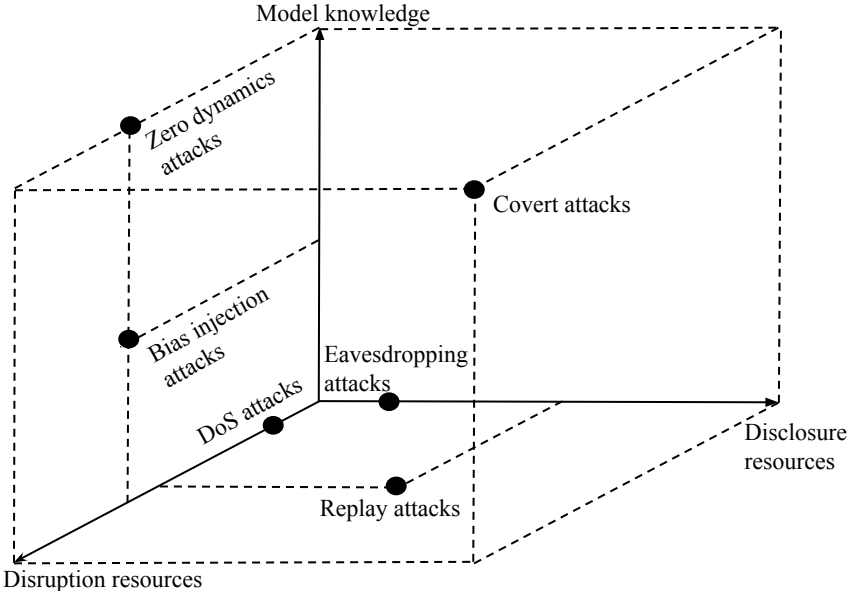

FIGURE 2.1: The cyber-physical attack space [4].

For the introduction of cyber-attacks, we first take a general linear time-invariant (LTI) system as an example to illustrate cyber-attacks.

The system dynamic without attacks is as follows:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k, \\ y_k &= Cx_k + v_k, \end{aligned} \tag{2.1}$$

where $x_k \in \mathbb{R}^n$, $y_k \in \mathbb{R}^m$, and $u_k \in \mathbb{R}^p$ are the vector of state variable, sensor measurement, and the control input at time k, respectively. $w_k \in \mathbb{R}^n$ denotes the process noise and $v_k \in \mathbb{R}^m$ the measurement noise. They are assumed to be mutually independent zero-mean Gaussian variables with covariances $Q \succeq 0$ and $R \succ 0$, i.e., $w_k \sim \mathcal{N}(0, Q)$ and $v_k \sim \mathcal{N}(0, R)$. The systems that we will focus on in later chapters are roughly the same as the above one, and we will give more details when we get to those chapters.

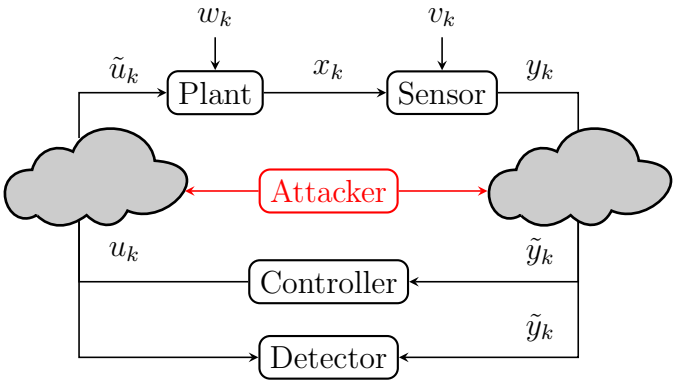

FIGURE 2.2: A schematic of cyber-attacks that can occur in a cyber-physical system.

A schematic of cyber-attacks that can occur in a cyber-physical system is shown in Figure 2.2. In this figure, "Plant" refers to the physical system. "Sensor" is a device used to measure the system variables, and it can be a smart sensor whose functions include signal conditioning, signal processing, and decision making [39]. The systems studied in Chapters 3 and 6 are equipped with smart sensors, and the processed information is sent to the remote estimator (We will give details in Chapters 3 and 6.). "Controller" uses the information from the local sensor to design the control input. "Detector" employs the information from the local sensor and the controller to detect if the system is under normal operation or not. Two cloud shapes denote the communication channels, which can be attacked. In this thesis, we focus on cyber-attacks which usually happen on the communication channel between the sensor and the controller, between the sensor and the detector, and between the controller and the plant. We use the red block "Attacker" and red

arrows to illustrate cyber-attack in the above figure. We denote $\tilde{y}_k$ and $\tilde{u}_k$ as the potentially modified sensor measurement and control input. The different attack scenarios will be discussed later.

Different from the fault studied in the area of fault detection and diagnosis, attacks are usually more complex, more stealthy and they usually lead to more serious consequences. By a stealthy attack, loosely speaking, we mean that this attack can not be detected by the detector through a threshold type method while affecting the normal operation of the system. In this section, we mainly introduce three kinds of cyber-attacks: DoS attack, replay attack, and false data injection attacks. Among them, replay attack and false data injection attacks are our focus in this thesis. DoS attack is also introduced since its scheme is similar to package drops studied in the area of control over communication networks, and the package drop is related to Chapter 6.

2.1.1 DoS Attack

The DoS attack is a kind of cyber-attack that a malicious adversary launches to prevent access to the information for legitimate users. The attacker uses disruption resources and conducts this kind of attack by jamming or flooding the target network with traffic until the target cannot respond or simply crashes [40, 41].

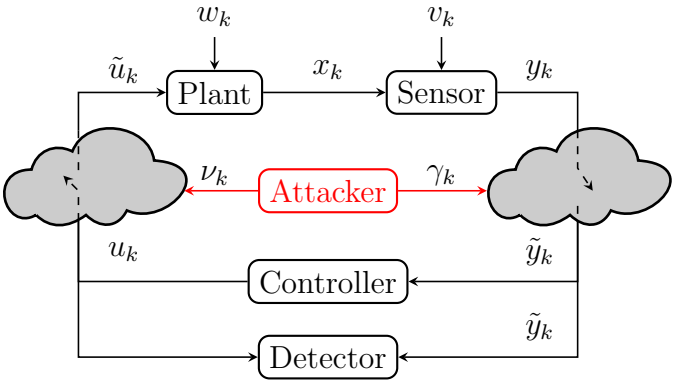

FIGURE 2.3: A schematic of the DoS attack.

Here, we take the system (2.1) as an example to illustrate the model of the DoS attack. In Figure 2.3, the measurement and control input are under DoS attack,

and the system dynamic is as follows:

$$\begin{aligned} x_{k+1} &= Ax_k + B\tilde{u}_k + w_k, \\ \tilde{y}_k &= \gamma_k(Cx_k + v_k), \\ \tilde{u}_k &= \nu_k u_k, \end{aligned}$$

where $\tilde{u}_k$ and u_k denote the actual control input and the desired control input computed by the controller, respectively, and (ν_k, γ_k) are i.i.d. Bernoulli random variables. The stochastic variable ν_k models the packet loss between the controller and the actuator: if the packet is correctly delivered, then $\tilde{u}_k = u_k$, otherwise, $\nu_k = 0$, i.e., u_k is lost and the actuator does nothing. The stochastic variable γ_k models the packet loss between the sensor and the controller: if the packet is correctly delivered, then $\tilde{y}_k = y_k = Cx_k + v_k$, while if the packet is lost, then $\tilde{y}_k = 0$. For more details about similar models, please refer to [42–44].

2.1.2 Replay Attacks

A replay attack is a type of network attack where attackers record systems data from the compromised sensors or actuators for a long enough period and replay these data to the detector afterwards. The disruption resources and disclosure resources are employed by attackers. This kind of attack can bypass the detection of general detectors and secret keys. Furthermore, replay attacks do not need any knowledge of system information, which makes it easy to perform the attack. Similarly, we take the system (2.1) as an example to illustrate the model of the replay attack. Figure 2.4 and Figure 2.5 show the schematic of replay attacks.

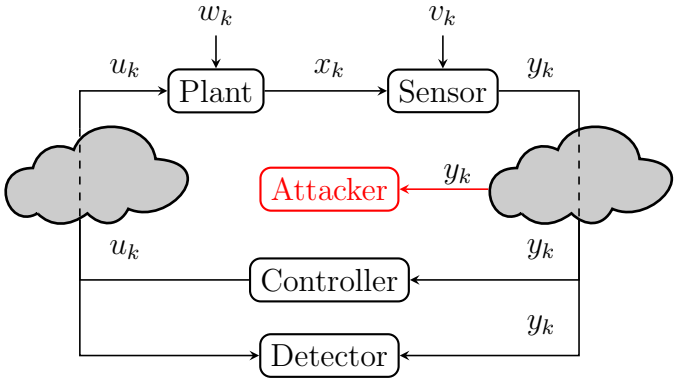

FIGURE 2.4: Phase I of the replay attack.

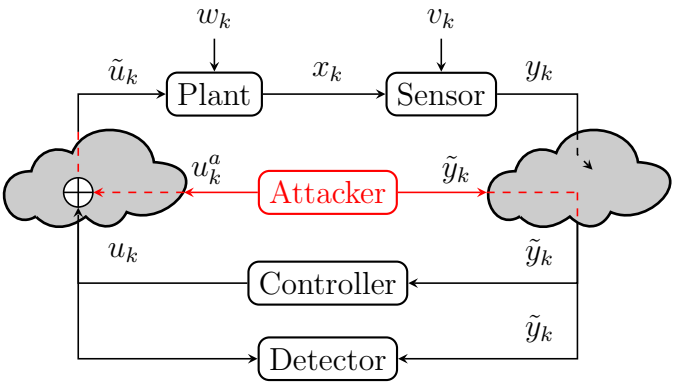

FIGURE 2.5: Phase II of the replay attack.

The system dynamic of Phase I is the same as the one without attack since the attacker only records the system data and does not modify them. For the sake of readability, we rewrite the system model here again:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k, \\ y_k &= Cx_k + v_k. \end{aligned}$$

The system dynamic of Phase II is:

$$\begin{aligned} x_{k+1} &= Ax_k + B\tilde{u}_k + w_k, \\ y_k &= Cx_k + v_k, \\ \tilde{u}_k &= u_k + u_k^a, \\ \tilde{y}_k &= y_{k-\Delta k}, \end{aligned}$$

where u_k^a denotes the control input that the attacker injects, $\tilde{u}_k$ denotes the control signal that the plant receives. $\tilde{y}_k$ denotes the modified sensor measurement that is equal to the real sensor reading at time $k - \Delta k$. For more details about similar models, please refer to [28, 45, 46]. We will study this kind of attack in Chapter 4.

2.1.3 False Data Injection Attacks

The false data injection attack is more complex than the above two attacks. Here, the system knowledge, disruption resources, and disclosure resources are employed by attackers. Generally speaking, attackers can inject malicious data into communication channels including but not limited to between sensors and actuators and

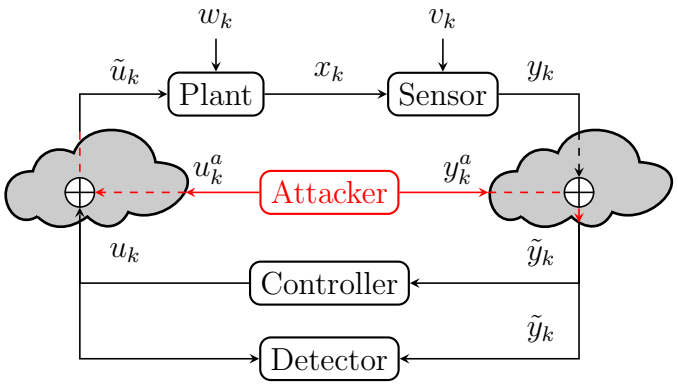

FIGURE 2.6: A schematic of the false data injection attack.

between sensors and detectors, thereby affecting data integrity and degrading the performance of systems.

Similarly, we take the system (2.1) as an example to illustrate the model of the false data injection attack, see Figure 2.6. The dynamic equations are:

$$\begin{aligned} x_{k+1} &= Ax_k + B\tilde{u}_k + w_k, \\ y_k &= Cx_k + v_k, \\ \tilde{u}_k &= u_k + u_k^a, \\ \tilde{y}_k &= y_k + y_k^a, \end{aligned}$$

where u_k^a and y_k^a are the false data injected into actuators and sensors, respectively. In this thesis, Chapters 3, 5, and 6 focus on this type of attack.

2.2 Analysis of Cyber-attacks

Analysis of cyber-attacks can enable us to identify the potential attacks and estimate the effect of the attacks. In this section, we will review the related work regarding how much performance degradation an attack can induce under stealthy attacks. By analyzing the maximal effect of attacks, we expect to get some insights into designing and protecting systems.

The reachable set, as a measure of impact of cyber-attacks, is widely used in the existing literature. Mo and Sinopoli [47] showed how the attacker's strategy can be formulated as a constrained control problem and provided a recursive algorithm

to compute the inner and outer bounds of the reachable set, thus quantifying the maximum perturbation that an attacker can introduce into a cyber-physical system by injecting a designed signal to a subset of the sensors. Similar problem of computing the ellipsoidal bounds was reformulated as a convex optimization problem in terms of linear matrix inequalities in [48]. A synthesis tool for minimizing the reachable set by redesigning controllers and detectors was also proposed in this work. The above works permit a small increase in the alarm rate and the capability of the attack is related with this small value. In [49], the alarm rate was required unchanged before and after the attack, and corresponding synthesis tools were provided to minimizing these bounds by redesigning the observer gain. In [50], the authors assumed that the system disturbance and measurement noise to be unknown-but-bounded and analyzed the reachability issue based on zonotopic set theory. Related works regarding reachability analysis can be found in [51–54].

Worst-case analysis provides another aspect to analyze cyber-attacks. The problem of what is the worst possible attack is getting more and more attention. A linear quadratic function was employed to capture the attacker's control goal and constraints in [55]. In the case that the attacker's probability of being detected is constrained to the false alarm rate of the detector, it is proved that linear feedback is the optimal attack strategy. In the case where there is a bounded bias in the false alarm rate, two algorithms to derive the optimal and suboptimal attack sequence were provided.

Different from the false data injection attack that needs to be stealthy, DoS attack is easily detected. In this context, energy constraint is employed to restrict attack strategies into a bounded set. In [56], the problem of scheduling a DoS attack with limited energy was studied. The optimal attack schedule in a special scenario was proposed and the optimal attack schedule with both energy constrained sensor and attacker was analyzed. A similar problem but with a packet-dropping network was studied in [57].

Different from the stealthiness metric of the attack in the above literature, a metric based on Kullback-Leibler (KL) divergence is employed to characterize the stealthiness level. The concept of δ-marginal stealthiness of the attack was introduced as in [58]. The authors characterized a stealthiness level from the probability of false alarm and investigated the trade-off between the performance degradation of the state estimation and the stealthiness level. Subsequently, a notion of ϵ-stealthiness

based on KL divergence to quantify attack detectability was proposed, and the maximal performance degradation under ϵ-stealthy attack strategy was revealed in [59, 60]. The authors of [61] generalized the above results to vector systems. Furthermore, [62] was devoted to seeking the optimal attack by compromising sensor measurements. In Chapter 3, we adopt the same stealthiness metric as in [59, 62]. Different from these works focusing on designing the optimal attack strategy after deriving the performance degradation bound, we obtain the maximal performance loss under linear attacks.

Among the existing works, the innovation-based attack is getting more and more attention. Innovation-based linear integrity attacks were first studied in [26]. An optimal linear attack strategy was provided to achieve the maximal performance degradation while not being detected by a χ^2 detector. Some extensions of this work can be found in [63, 64]. This type of attacks in the detection framework based on KL divergence was also investigated in [5]. Furthermore, a more general linear attack strategy with an arbitrary mean random noise was studied in [27]. However, all the above papers only consider memoryless attacks. A larger performance degradation of the remote estimator can be expected when the attacker utilizes both past and present information. Motivated by this point, we consider how vulnerable innovation-based remote state estimators are to a linear attack which leverages both past and present innovations in Chapter 3.

2.3 Detection of Cyber-attacks

The detection of cyber-attacks is crucial to the security of CPS. If malicious attacks can be detected accurately and timely, the effect and loss that attacks result in will be effectively decreased. There are several representative detection approaches, such as χ^2-detector based on Kalman filter, Neyman-Pearson (NP) detector, cumulative sum (CUSUM) detector. However, since malicious attacks are usually carefully designed, it is difficult to detect these attacks only using general detection methods. In this section, we will review the existing results on detection strategies from the perspective of control theory and machine learning.

2.3.1 A Control Theory Perspective

A significant amount of research effort has been devoted to intrusion and anomaly detection algorithms to enhance CPS security from a control theory perspective. Zimmer *et al.* [65] presented three mechanisms for time-based intrusion detection. The techniques, through bounds checking, were developed in a self-checking manner by the application and through the operating system scheduler. Mitchell and Chen [66] proposed a hierarchical performance model and techniques for intrusion detection in CPS. They classified the modern CPS intrusion detection system techniques into two design dimensions (detection technique and audit material) and summarized the advantages and disadvantages of different dimension's choices in [67]. Kwon *et al.* [68] discussed necessary and sufficient conditions for when the attacker could be successful without being detected. Their method can be employed to evaluate vulnerability degree of certain CPS. Corresponding detection and defense methodologies against stealthy deception attacks can be developed. In [69], the authors proposed a mathematical framework for CPS and investigated limitations of the monitoring system. Centralized and distributed attack detection and identification monitors were also discussed. By introducing time-varying dynamics into the system that are unknown to the attacker, a moving target approach was proposed to detect attacks in [70]. Similar approach was also studied in [71–73].

Different from the proposed strategies in the above literature, the physical watermarking scheme, which leverages a random variable as a watermark to detect the attack, has been studied to detect replay attacks for the past decades. In this thesis, we also focus on this watermarking scheme in Chapter 4. For the rest of this section, we review some recent physical watermark design approaches for CPS. We focus on how to design physical watermarking to actively detect cyber-attacks, especially replay attacks, thereby securing the CPS.

In [28, 45], a physical watermarking scheme was proposed for control systems. In this scheme, if the system is operating normally, then the effect of the carefully designed watermark signal is present in the sensor measurements. However, if the system is under attack, the introduction of watermarking signals enable the detection of replay attack. Actually, physical watermarking scheme could be considered as an active defense scheme. Mo and Sinopoli [28] investigated the problem of the detection of replay attacks and first proposed the technique of introducing an

authentication signal which is called physical watermark signal in subsequent literature. This approach enables the detection of replay attacks where an adversary can read and modify all sensor data as well as inject a malicious input into the system. Different from false data injection attacks, this type of attack does not need knowledge of the system model to generate stealthy outputs and only replays the recorded sensor measurements to the operator, which leads to that the replayed data and the real data share exactly the same statistics and for which replay attacks cannot be detected efficiently. By injecting a random control signal, the watermark signal, into the control system, it is possible to detect the potential replay attacks so as to secure the system.

The authors of [45, 74] further extended the results of [28] by providing a more general physical authentication scheme to detect the replay attacks. However, the watermark signal may deteriorate the control performance, and therefore it is important to find the optimal trade-off between the control performance and the detection efficiency, which can be cast as an optimization problem. Furthermore, Mo *et al.* [45] also characterized the relationship among the control performance loss, detection rate and the strength of the Gaussian authentication input.

Although we use the term physical watermarking previously, the term physical watermarking was first formally proposed in [46] to authenticate the correct operation of CPS. As a generalization of [28, 45, 74], the technique of designing the optimal watermark signal was to maximize the expected KL divergence between the distributions of the compromised and the healthy residue signals, while guaranteeing a certain maximal control performance loss. The optimization problem is separated into two steps where the optimal direction of the signal for each frequency was first computed and then all possible frequencies were considered to find the optimal watermark signal.

The watermarking approach proposed in [75] was based on an additive watermark signal generated by a dynamical system. Conditions on the parameters of the watermark signal were obtained which ensures that the residue signal of the system under attack is unstable and the attack can be detected. An optimization problem was proposed to give a loss-effective watermark signal with a certain amount of detection rate by adjusting the design parameters. A similar problem was studied for multi-agent systems in [76].

The problem of physical watermark design under packet drops at the control input was analyzed in [77]. It is interesting that Bernoulli packet drops can obtain better detection performance compared with a purely Gaussian watermarking signal. Consequently, a Bernoulli-Gaussian watermark, which incorporates both an additive Gaussian input and a Bernoulli drop process, was jointly designed to achieve the trade-off between detection performance and control performance. The effect of the proposed watermark on closed-loop performance and detection performance was analyzed.

Satchidanandan and Kumar [78] provided a comprehensive procedure for physical watermarking. It suggests a private excitation signal on the control input which can be traced in the system to enable the detection of attacks. Such an active defense technique is used to secure CPS that include single-input-signal output (SISO) systems with Gaussian noise, SISO auto-regressive systems with exogenous Gaussian noise, the SISO autoregressive-moving average systems with exogenous terms, SISO systems with partial observations, multi-input-multi-output systems with Gaussian noise and extension to non-Gaussian systems. In [79], they proposed necessary and sufficient conditions that the statistics of the watermark needs to satisfy in order to achieve security-guaranteeing.

In [80, 81], Rubio-Hernán *et al.* defined cyber adversaries and cyber-physical adversaries and pointed out that the detection schemes proposed by Mo and Sinopoli [28] and Mo *et al.* [46] fails to detect an attack from the latter. Therefore, a multi-watermark-based detection scheme is proposed to overcome the limitation. Furthermore, in [82], a periodic and intermittent event-triggered control watermark detector was presented. The new detector strategy integrates local controllers with remote controller. It was proved that the new detector scheme can detect three adversary models defined in their work.

Different from the additive watermarking schemes, a multiplicative sensor watermarking was proposed in [83]. In this scheme, each sensor output is pre-processed through a watermark generator. The corresponding watermark remover is employed to reconstruct the real sensor measurement from the received watermarked data. This scheme does not degrade the control performance in the absence of attacks and the controller and anomaly detector could be designed independently. Furthermore, it also enables the isolation and identification of the replay attack. A similar scheme was applied to detect cyber sensor routing attacks [84] and false data

injection attacks [85]. The physical sensor re-routing attack and the cyber measurement re-routing attack were considered in [84] and corresponding detectability and isolability of these two attacks are analyzed. In [85], Teixeira and Ferrari showed how to design the watermarking filters to enable the detection of stealthy false data injection attacks and a novel technique is proposed to solve the limitation of single-output systems.

It is worth noticing that in all research discussed above, precise knowledge of the system parameters is required in order to design the watermark signal and the detector. However, acquiring these parameters may be troublesome and costly in practice. Motivated by this, we propose an algorithm that can simultaneously generate the watermarking signal and infer the system parameters to enable the detection of attacks with unknown system parameters in Chapter 4. It is proved that the proposed algorithm converges to the optimal one almost surely.

2.3.2 A Machine Learning Perspective

The above traditional detection methods from a system theory perspective can characterize the qualities of CPS well. However, these methods usually require precise system knowledge. Different from these methods, machine learning based methods usually need a large amount of data for training the detector. Generally, machine learning cannot characterize the attributes of CPS. We do not consider the case where machine learning is used to identify the system model by collecting the data. It is worth noticing that as CPS become more complicated and attacks are more intelligent, control theory-based approaches have difficulties to capture new characteristics of CPS. Hence, more and more researchers explore the possibility of application of machine learning to security of CPS, especially detection of cyber-attacks in CPS.

In [86], the author explored the feasibility of applying machine learning to discriminate power system disturbance, especially on detecting cyber-attacks, and evaluating the performance of different machine learning methods on the disturbance classification for a smart power grid framework. Goh *el at.* provided an unsupervised learning approach to detect cyber-attacks in [87]. Based on the prediction of the long short-term memory recurrent neural network, the difference

between the predicted and the actual sensor data can be used for anomaly detection. A systematic survey about deep learning based anomaly detection methods was presented in [88].

As an important part of CPS, networked embedded sensors are widely used to monitor plants and to detect anomaly. At the same time, due to their vulnerability to malicious attacks, increasing importance has been attached to researches on the security of those systems. [89–91] employed game-theoretic approaches to analyze the attacker's behavior for detection purpose. It is shown that the flip attack, where the attacker flips the distribution of manipulated sensors' measurements, is optimal from the attackers' perspective to a broad class of problems. Therefore, it is well-worth some attention to design detectors for flip attacks. Some researchers attempted to capture the properties of the detection problems by the formalism of partially observable Markov decision process (POMDP). This is achieved by assuming an attack possibility [29, 92] in various forms, which is somewhat similar to the attack tree models used to analyze system reliability [93]. With such a modeling approach, the much celebrated reinforcement learning (RL) methods [94, 95] can be applied to solve the problem. Among all those methods, a theoretically sound approach is to introduce the belief states and to solve in turn the induced Markov decision processes (MDPs) with the belief states as its states [96]. However, there are two major drawbacks of this approach, one of which is generally true for all POMDP problems, while the other is particularly damaging to the detection problems studied here. First, regardless of the size of the state space of the POMDP, the belief state can take infinite number of possible values, which makes the induced MDP infinite dimensional and therefore challenging to solve [95]. Second, the detection problems studied here involve extraneous inputs from attackers, which cannot be captured by the POMDP framework. To address the issue, a transition probability is assumed to model the attack possibility, which is a major approximation, as the true attack probability varies due to various reasons and the transition probability used in POMDP may be different from the true attack probability. When solving a POMDP using belief states, a so-called state estimator is used to compute its value online. It relies explicitly on the transition probability, and makes the solution sensitive to the assumed transition probability. To circumvent those challenges, Kurt *et al.* [29] applied fixed-length window of observation as the state for online detection in smart grids. Similar idea also appeared in [30] for online learning and attack design. However, unlike the success

of RL reported for playing games where the POMDP is given by some simulators with transition probabilities and costs enclosed in the simulator [31], they are here part of the design task. It is not clear from those works how the POMDPs shall be designed.

The focus in Chapter 5 is somewhat like the quickest change detection (QCD) problem, as we aim to determine if there is an attack at every time step as quickly as possible. However, unlike those problems where the probability density functions (PDFs) before and after the change point are known [32, 33], in this thesis, only a set of possible PDFs are known, which is another challenge.

2.4 Mitigation of Cyber-attacks

It is difficult and not realistic to prevent cyber-attacks from happening. Hence, the effective mitigation of attacks is of utmost importance in security of CPS. In this section, we will review the relevant works regarding the mitigation of cyber-attacks.

Among the existing literature, secure state estimation, the problem of estimating the state of a dynamic system from a set of corrupted measurements, has attracted great attention. In [97], Mo and Sinopoli considered the secure estimation of a scalar state based on m measurements in which up to l of m can be potentially manipulated by an adversary. It was shown that the optimal worst-case estimator should be based only on the *a priroi* information under the scenario in which the attacker can manipulate at least half of m measurements ($l \geq m/2$). If the attacker can manipulate less than half of m measurements ($l < m/2$), the optimal estimator is shown to be based on $\binom{m}{2l}$ local estimators. Fawzi *et al.* characterized the maximum number of attacks that can be detected and corrected in [98]. Accurately reconstructing the state of a system if more than half the sensors are attacked, which is somewhat similar to the results in [45], was also shown. Besides, an efficient algorithm was proposed to estimate the state of the dynamic system under attacks when the number of attacks is less than a threshold. The problem of designing output-feedback controllers that stabilize the system under sensor attacks was also considered. A principle of separation between estimation and control was shown to hold and the design of resilient output feedback controllers can be reduced to the design of resilient state estimators. Similar results can also

be found in [99, 100]. However, the solutions proposed in the last three works are computationally intensive. In light of this, an efficient state reconstruction algorithm in the sense of being implementable on computationally limited platforms was provided in [101, 102]. The idea of event-triggered control was utilized on the design of Luenberger-like observer and the proposed observer was shown to be computationally more efficient than previous solutions to the secure state reconstruction problem. In order to harness the combinatorial complexity of the secure state estimation problem under sensor attacks and in the presence of noise, a novel algorithm that employs a satisfiability modulo theory was proposed in [103].

Different from the above literature regarding secure state estimation, a large number of works focus on sensor scheduling for remote state estimation under cyber-attacks, which is another perspective of the mitigation of attacks. In this thesis, we consider how to determine a schedule of sensors in remote state estimation under cyber-attack. Due to the limited energy of sensors, the trade-off between energy consumption and accuracy of remote state estimation has become important. A periodic scheduling scheme under a communication energy constraint was proposed in [104]. To enhance the performance, Mo *et al.* proposed an acknowledgement (ACK) protocol, where the fusion center sends an acknowledgement whenever it receives a packet in [105]. The authors studied sensor scheduling for both no-ACK and ACK protocols and provided some properties of the optimal schedule. The ACK-based sensor scheduling was proved to outperform the one without ACK, i.e., offline schedule, under the same energy constraint in [106]. However, the above works do not consider the effect of cyber-attacks. Motivated by this, Guo et al. [34] studied the denial of service (DoS) attack on feedback channel against the ACK-based sensor schedule from the perspective of attacker, and proved that the optimal policy has a threshold structure. From the perspective of defender, an ACK-based deception scheme for sensors was first proposed in [35]. Furthermore, considering the existence of more intelligent attackers and more complicated attacks, the authors used a general asymmetric-information stochastic game to model the interactions between the sensor and the attacker and presented an equivalent belief-based stochastic game to obtain the optimal stationary strategy for each agent in [36]. Different from the above works regarding the active deception-based scheme, our work focuses on obtaining an optimal schedule when the ACK received by the sensor is attacked, i.e., we are concerned with attack mitigation. Here, by

ACK-based attacks, we mean the ACK is modified by the attacker. It is a class of false data injection attacks.

For most existing works without the ACK-based attack, the sensor scheduling problem was formulated as an optimal control problem with system dynamics model given a MDP [107, 108]. When an ACK-based attack is present, it is most conveniently formulated by POMDP whose structural results could be obtained via some stochastic ordering [109]. In Chapter 6, we focus on the sensor scheduling problem for remote state estimation under the presence of the ACK-based attack. We aim to obtain scheduling rules that minimize the expectation of an infinite horizon discounted accumulated cost. We investigate the possibility to derive a structural result via Monotone likelihood Ratio (MLR) ordering. It is proved that a threshold type of solution for POMDP cannot be readily available. In view of this, we seek to explore a suboptimal solution approach via rollout is then proposed. "Rollout" was first proposed in [110], which starts with a base policy and produces an improvement by limited lookahead minimization with the use of heuristic at the end. It can be considered as single policy improvement [111] and provides an online approach for solving stochastic scheduling, combinatorial optimization, sequential fault diagnosis, vehicle routing with stochastic demands, and sequential repairing problems [112–116]. In Chapter 6, the rollout is used to estimate the approximation in value space and the corresponding algorithm is provided. Besides, we also provide the performance bounds for the proposed two variants of rollout.

Chapter 3

Performance Analysis of Innovation-based Remote State Estimation under Linear Attack

The last chapter reviewed the related works regarding the security of CPS. From this chapter, we will introduce the main results of the thesis. In this chapter, we analyze the performance of remote state estimation under cyber-attacks, which helps us to better identify potential attacks and estimate the effect of cyber-attacks.

This chapter is concerned with the problem of how secure the innovation-based remote state estimation can be under linear attacks. A linear time-invariant system equipped with a smart sensor is studied. A metric based on Kullback-Leibler divergence is adopted to characterize the stealthiness of the attack. The adversary aims to maximize the state estimation error covariance while stay stealthy. The performance bounds that an adversary can achieve with any linear first-order attack under strict stealthiness for vector systems and ϵ-stealthiness for scalar systems are characterized. We also provide explicit attack strategies that achieve these bounds and compare this attack strategy with strategies previously proposed in the literature. Finally, some numerical examples are given to illustrate the results.

The rest of the chapter is organized as follows. Section 3.1 formulates the problem by introducing the system model, attacks model as well as two metrics. Some preliminaries are introduced in Section 3.1.5. We present the worst-case performance degradation bounds for remote state estimation under strictly stealthy attacks for

vector systems and ϵ-stealthy attacks for scalar systems in Sections 3.2 and 3.3, respectively. In Section 3.4, some numerical examples are provided to verify the performance of the proposed strategies. Conclusions are provided in Section 3.5. For the sake of legibility, some proofs are included in Section 3.6.

3.1 Problem Formulation

In this section, the system and attack models are introduced together with the stealthiness and performance degradation metrics. Finally, the problem of interest is formulated. The diagram for the considered system is illustrated in Figure 3.1. A smart sensor measures the output of a physical plant and transmits the innovation to a remote estimator via a wireless network. The attacker attempts to modify the transmission data, which are received by a remote estimator and a detector. The detailed system model is presented in this section.

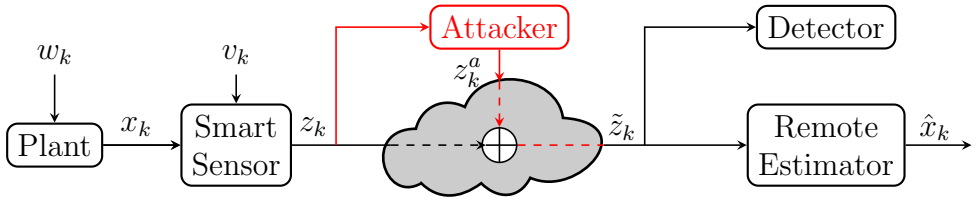

FIGURE 3.1: The system diagram.

3.1.1 System Model

Consider a linear time-invariant (LTI) system described by the following equations:

$$x_{k+1} = Ax_k + w_k, \tag{3.1}$$

$$y_k = Cx_k + v_k, \tag{3.2}$$

where $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}^m$ are the vector of state variables and sensor measurements at time k, respectively, $w_k \in \mathbb{R}^n$ denotes the process noise and $v_k \in \mathbb{R}^m$ the measurement noise. They are assumed to be mutually independent zero-mean Gaussian variables with covariances $Q \succeq 0$ and $R \succ 0$, i.e., $w_k \sim \mathcal{N}(0, Q)$ and $v_k \sim \mathcal{N}(0, R)$. We further assume that x_0 is a zero mean Gaussian random vector independent of the process noise and the measurement noise, and with covariance

Σ. We focus on stable systems and the need for the following assumption is explained in Section 3.2.

Assumption 3.1. The system is stable.

The system is equipped with a local smart sensor. In our work, the smart sensor employs the Kalman filter to process measurement and transmit the innovation to the remote estimator:

$$\begin{aligned}
\hat{x}^s_{k+1|k} &= A\hat{x}^s_k, \\
P_{k+1|k} &= AP_{k|k}A^\intercal + Q, \\
K_k &= P_{k|k-1}C^\intercal(CP_{k|k-1}C^\intercal + R)^{-1}, \\
\hat{x}^s_k &= \hat{x}^s_{k|k-1} + K_k(y_k - C\hat{x}^s_{k|k-1}), \\
P_{k|k} &= P_{k|k-1} - K_kCP_{k|k-1},
\end{aligned}$$

with initialization $\hat{x}_{0|-1} = x_0$.

Under Assumption 3.1, the Kalman gain will converge exponentially. Therefore, we consider a steady-state Kalman filter with gain K and a priori minimum mean square error (MMSE) P:

$$P \triangleq \lim_{k\to\infty} P_{k|k-1}, \tag{3.3}$$

$$K \triangleq PC^\intercal(CPC^\intercal + R)^{-1}. \tag{3.4}$$

As a result, the Kalman filter can be rewritten as:

$$\begin{aligned}
\hat{x}^s_{k+1|k} &= A\hat{x}^s_k, \\
\hat{x}^s_k &= \hat{x}^s_{k|k-1} + Kz_k,
\end{aligned}$$

where $z_k \triangleq y_k - C\hat{x}^s_{k|k-1}$ is the innovation at time k, which is transmitted to the remote estimator. Recall that $z_k \sim \mathcal{N}(0, \Sigma_z)$, where $\Sigma_z \triangleq CPC^\intercal + R \succ 0$. Since $y_k = z_k + C\hat{x}^s_{k|k-1}$, one can argue that z_k contains the same information about the uncertainty in the process as y_k. It is convenient to analyze the filtering performance since z_k follows a Gaussian distribution. Often in the literature [26, 63, 117, 118], similar setups have been considered. In this work, we do not consider the presence of feedback. Since the control input is usually known to the system

operator or detector, it can be separated from the measurement. The results in this chapter can be easily extended to the closed-loop system.

3.1.2 Attack Model

The adversary is assumed to have the following capabilities:

1. The attacker has access to all innovations from the smart sensor, i.e., it knows the innovations $z_1, z_2, \ldots, z_k$ at time k.
2. The attacker can modify the innovations to arbitrary values.
3. The attacker has knowledge of the system matrix A, the measurement matrix C, as well as the covariances of the noises, i.e., Q and R.

Remark 3.1. The third capability can be relaxed. If the attacker does not have access to A but it can access the input and output, it can identify the system parameters. The accuracy of the identification will affect the attack performance. This will be illustrated in Section 3.4.

The attacker injects the false data z_k^a and modifies the innovations in real-time as:

$$\tilde{z}_k = T\tilde{z}_{k-1} + Sz_k + \phi_k, \tag{3.5}$$

where $T \in \mathbb{R}^{m\times m}$ and $S \in \mathbb{R}^{m\times m}$ are matrices to be chosen by the attacker, and $\phi_k \sim \mathcal{N}(0, \Phi)$ is an i.i.d. Gaussian random variable with covariance $\Phi \in \mathbb{S}_+^m$, which is independent of z_k. The attack model (3.5) suggests that the attacker can generate the false data injection attack based on filtering the innovation sequence from the smart sensor with a linear type potentially driven by noise.

Remark 3.2. Observe that the works [5, 26, 63, 64] consider memoryless attacks, i.e., the attack is only based on the current innovation. Here, we seek to explore the possibility of a larger performance degradation for the remote estimator when the attacker utilizes both past and present information. More specifically, we focus on a linear time-invariant first order attack model and characterize the maximal performance degradation that an adversary can achieve. We also provide an explicit attack strategy that achieves this bound. It is hoped that our work can provide some insight into other more general attack models such as linear-time varying and nonlinear attack models.

The remote estimator receives $\tilde{z}_k$ so the remote state estimation follows:

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1}, \tag{3.6}$$

$$\hat{x}_k = \hat{x}_{k|k-1} + K\tilde{z}_k. \tag{3.7}$$

Here, we initialize $\hat{x}_{1|0} = \hat{x}^s_{1|0}$ and $\tilde{z}_k = 0$ when $k \leq 0$.

3.1.3 Detector and Stealthiness Metric

The attacker wants to be stealthy, otherwise the system will take countermeasures to keep a safe operation. We employ a metric based on KL divergence to quantify stealthiness, as first proposed in [59].

The attack detection problem is posed as sequential hypothesis testing. The detector uses the received sequence to carry out the following binary hypothesis testing:

$\mathcal{H}_0$: There is no attack in process. (The remote estimator receives z_1^k).

$\mathcal{H}_1$: There is an attack in process. (The remote estimator receives $\tilde{z}_k^1$).

In testing $\mathcal{H}_0$ versus $\mathcal{H}_1$ there are two types of possible errors: the first type is called "false alarm", which denotes that the detector decides $\mathcal{H}_1$ given $\mathcal{H}_0$, and the second type is called "miss detection", which represents that the detector decides $\mathcal{H}_0$ when $\mathcal{H}_1$ is correct. Here, we denote the probability of miss detection at time k as p_k^M, and the probability of false alarm as p_k^F. Furthermore, the probability of correct detection is p_k^D, which denotes that the detector decides $\mathcal{H}_1$ given $\mathcal{H}_1$. Obviously, $p_k^D + p_k^M = 1$. We provide two definitions for attack stealthiness:

Definition 3.1 (Strictly stealthy attack [62])**.** The attack is strictly stealthy if $p_k^F \geq p_k^D$, $(k \geq 0)$, holds for any detector.

Definition 3.2 (ϵ-stealthy attack [62])**.** The attack is ϵ-stealthy if

$$\limsup_{k\to\infty} -\frac{1}{k}\log p_k^F \leq \epsilon \tag{3.8}$$

holds for any detector that satisfies $0 < p_k^M < \delta$ for all $k \geq 0$ and any value of $\delta \in (0, 1)$.

The definition of ϵ-stealthy attack is motivated by Chernoff-Stein Lemma [119]. If there is not a detector that satisfies $0 < p_k^M < \delta$ and p_k^F converges to zero exponentially with a greater than ϵ rate as k goes to infinity, the corresponding attack is ϵ-stealthy. About the details of the above two definitions, please refer to [62].

3.1.4 Performance Degradation Metric

We employ the ratio of the trace of the covariance of the state estimation error $\tilde{P}$ and P to quantify the performance degradation introduced by the attacker, i.e.,

$$\eta = \frac{\mathrm{tr}(\tilde{P})}{\mathrm{tr}(P)}, \tag{3.9}$$

where P is defined in (3.3) and $\tilde{P}$ is defined as follows:

$$\tilde{P} \triangleq \limsup_{k\to\infty} \frac{1}{k} \sum_{l=1}^{k} \tilde{P}_l, \tag{3.10}$$

where $\tilde{P}_l = E[(x_l - \hat{x}_{l|l-1})(x_l - \hat{x}_{l|l-1})^\intercal]$. When there is no attack, $\tilde{z}_k = z_k$. As $\hat{x}_{1|0} = \hat{x}^s_{1|0}$, one can derive that $\hat{x}_{k|k-1} = \hat{x}^s_{k|k-1}$. Hence, $\tilde{P} = P$ and $\eta = 1$. In other words, the performance will not be degraded without attacks.

3.1.5 KL Divergence

In order to quantify the stealthiness level of attacks, we need to employ the KL divergence [119, 120], which is defined as:

Definition 3.3 (KL divergence)**.** Let $\tilde{z}_1^k$ and z_1^k be two random sequences with joint probability density functions $f_{\tilde{z}_1^k}$ and $f_{z_1^k}$, respectively. The KL divergence between $\tilde{z}_1^k$ and z_1^k equals

$$D(\tilde{z}_1^k \| z_1^k) = \int_{-\infty}^{+\infty} \left(\log \frac{f_{\tilde{z}_1^k}(\xi_1^k)}{f_{z_1^k}(\xi_1^k)} f_{\tilde{z}_1^k}(\xi_1^k) \right) d\xi_1^k. \tag{3.11}$$

One can see that $D(\tilde{z}_1^k \| z_1^k) \geq 0$, and $D(\tilde{z}_1^k \| z_1^k) = 0$ if and only if $f_{\tilde{z}_1^k} = f_{z_1^k}$. Generally, KL divergence is asymmetric, i.e., $D(\tilde{z}_1^k \| z_1^k) \neq D(z_1^k \| \tilde{z}_1^k)$.

Necessary and sufficient conditions for strictly stealthy attacks and ϵ-stealthy attacks are provided in [62]:

Lemma 3.1 (Strictly stealthy attacks [62]). *An attack sequence $\tilde{z}_1^k$ is strictly stealthy if and only if $\{\tilde{z}_1, \tilde{z}_2, \dots\}$ is a sequence of i.i.d. Gaussian random variables with zero mean and covariance $\mathrm{Cov}(\tilde{z}_k) = \Sigma_z = CPC^\intercal + R$.*

Lemma 3.2 (ϵ-stealthy attacks [62]). *If an attack $\tilde{z}_1^k$ is ϵ-stealthy, then*

$$\limsup_{k\to\infty} \frac{1}{k} D(\tilde{z}_1^k \| z_1^k) \leq \epsilon. \tag{3.12}$$

Conversely, if an attack sequence $\tilde{z}_1^k$ is ergodic and satisfies

$$\lim_{k\to\infty} \frac{1}{k} D(\tilde{z}_1^k \| z_1^k) \leq \epsilon, \tag{3.13}$$

then the attack is ϵ-stealthy.

3.1.6 Problem of Interest

We aim to derive fundamental vulnerabilities of innovation-based remote estimation. In other words, we seek to obtain how secure one can make innovation-based remote state estimation under linear attacks.

Given the innovation-based remote state estimator system in Figure 3.1, how vulnerable is such a system under attack (3.5)? The vulnerability is measured by the worst performance degradation (3.9). For strictly stealthy and ϵ-stealthy attacks, the worst performance degradation can be formulated as the following optimization problems:

- The attack is strictly stealthy:

$$\begin{aligned} \arg\max_{T,S,\Phi} \quad & \eta_s \triangleq \limsup_{k\to\infty} \frac{\frac{1}{k}\sum_{l=1}^{k} \mathrm{tr}(\tilde{P}_l)}{\mathrm{tr}(P)}, \\ \text{s.t.} \quad & \text{The attack is strictly stealthy.} \end{aligned} \tag{3.14}$$

- The attack is ϵ-stealthy:

$$\begin{aligned} \arg\max_{T,S,\Phi} \quad & \eta_\epsilon \triangleq \limsup_{k\to\infty} \frac{\frac{1}{k}\sum_{l=1}^{k} \mathrm{tr}(\tilde{P}_l)}{\mathrm{tr}(P)}, \\ \text{s.t.} \quad & \text{The attack is } \epsilon\text{-stealthy.} \end{aligned} \tag{3.15}$$

We seek to obtain the optimal attack tuple (T^*, S^*, Φ^*) to maximize the performance degradation, while guaranteeing the prespecified stealthiness level.

3.2 Strictly Stealthy Attacks

The following theorem characterizes the maximal performance degradation ratio under a strictly stealthy attack. We also specify the optimal attack strategy.

For the simplicity of notations, we define

$$\mathcal{P}_1 \triangleq K\Sigma_z K^\intercal.$$

Theorem 3.1. *Consider system (3.1)-(3.2) satisfying Assumption 3.1. For strictly stealthy attacks of the form (3.5), the worst performance degradation ratio for the estimation error covariance is*

$$\eta_s = 1 + 4\frac{\mathrm{tr}(\mathcal{X})}{\mathrm{tr}(P)},$$

where $\mathcal{X} = \mathcal{X}_1 - \mathcal{P}_1$ and $\mathcal{X}_1$ is the solution to the Lyapunov equation: $\mathcal{X}_1 = A\mathcal{X}_1 A^\intercal + \mathcal{P}_1$. The corresponding attack strategy is $(T^*, S^*, \Phi^*) = (\boldsymbol{0}_{m\times m}, -I_m, \boldsymbol{0}_{m\times m})$.

Proof. Rewrite (3.5) as follows:

$$\begin{aligned} \tilde{z}_l &= T\tilde{z}_{l-1} + Sz_l + \phi_l \\ &= \sum_{i=1}^{l} T^{l-i} S z_i + \sum_{i=1}^{l} T^{l-i}\phi_i. \end{aligned} \tag{3.16}$$

By Lemma 3.1, the covariance of $\tilde{z}_l$ $(l = 1, 2, \dots)$ needs to satisfy

$$\mathrm{Cov}(\tilde{z}_l) = \sum_{i=0}^{l-1} T^i \left(S\Sigma_z S^\intercal + \Phi\right)(T^i)^\intercal = \Sigma_z.$$

A feasible solution thus belongs to one of the three cases:

- $(T, S, \Phi) = (\mathbf{0}_{m\times m}, I_m, \mathbf{0}_{m\times m})$: $\tilde{z}_l = z_l$, i.e., the attacker is not launching an attack.
- $(T, S, \Phi) = (\mathbf{0}_{m\times m}, -I_m, \mathbf{0}_{m\times m})$: $\tilde{z}_l = -z_l$, i.e., the attacker flips the sign of the innovation.
- $(T, S, \Phi) = (\mathbf{0}_{m\times m}, S, \Sigma_z - S\Sigma_z S^\intercal)$, where $S \neq \pm I_m$.

Now we derive the corresponding ratio η_s for cases 2 and 3. Let us rewrite $\tilde{P}_l$ as follows:

$$\begin{aligned}\tilde{P}_l =&\mathbb{E}[(x_l - \hat{x}_{l|l-1})(x_l - \hat{x}_{l|l-1})^\intercal] \\ =&\mathbb{E}[(x_l - \hat{x}^s_{l|l-1})(x_l - \hat{x}^s_{l|l-1})^\intercal] + \mathbb{E}[(\hat{x}^s_{l|l-1} - \hat{x}_{l|l-1})(\hat{x}^s_{l|l-1} - \hat{x}_{l|l-1})^\intercal] \\ &+2\mathbb{E}[(x_l - \hat{x}^s_{l|l-1})(\hat{x}^s_{l|l-1} - \hat{x}_{l|l-1})^\intercal] \\ =&P + \mathbb{E}[(\hat{x}^s_{l|l-1} - \hat{x}_{l|l-1})(\hat{x}^s_{l|l-1} - \hat{x}_{l|l-1})^\intercal],\end{aligned} \tag{3.17}$$

where the last equality holds due to the orthogonality principle, i.e., all the random variables generated by the smart sensor are independent of the estimation error $x_l - \hat{x}^s_{l|l-1}$ of the MMSE estimate $\hat{x}^s_{l|l-1}$ [62]. More specifically, $\hat{x}^s$ is the state estimate of the smart sensor. $\hat{x}$ is the state estimate of the remote estimator and it is updated by the modified innovation $\tilde{z}_k$, where $\tilde{z}_k$ is linear with the innovation of z_k. Since the error vector $x_l - \hat{x}^s_{l|l-1}$ is orthogonal to the innovation z_k, the last equality holds. Define $\tilde{e}_l \triangleq \hat{x}^s_{l|l-1} - \hat{x}_{l|l-1}$, where

$$\hat{x}^s_{l|l-1} = A\hat{x}^s_{l-1|l-2} + AKz_{l-1} = A^{l-1}\hat{x}_{1|0} + \sum_{i=1}^{l-1} A^i K z_{l-i},$$

and

$$\hat{x}_{l|l-1} = A\hat{x}_{l-1|l-2} + AK\tilde{z}_{l-1} = A^{l-1}\hat{x}_{1|0} + \sum_{i=1}^{l-1} A^i K \tilde{z}_{l-i}. \tag{3.18}$$

Since $\hat{x}^s_{1|0} = \hat{x}_{1|0}$, which implies that $\tilde{e}_1 = \mathbf{0}_{m\times 1}$, we have

$$\mathbb{E}[\tilde{e}_l \tilde{e}_l^\intercal] = \sum_{i=1}^{l-1} A^i K \mathbb{E}\left[(z_{l-i} - \tilde{z}_{l-i})\,(z_{l-i} - \tilde{z}_{l-i})^\intercal\right](A^i K)^\intercal. \tag{3.19}$$

For case 2, $\tilde{z}_l = -z_l$, so

$$\begin{aligned}
&\lim_{l\to\infty} \mathbb{E}[\tilde{e}_l \tilde{e}_l^\intercal] \\
=&\lim_{l\to\infty} \sum_{i=1}^{l-1} A^i K \mathbb{E}\left[(z_{l-i} - (-z_{l-i}))\,(z_{l-i} - (-z_{l-i}))^\intercal\right](A^i K)^\intercal \\
=&4 \lim_{l\to\infty} \sum_{i=1}^{l-1} A^i K \Sigma_z K^\intercal (A^i)^\intercal = 4\mathcal{X},
\end{aligned} \tag{3.20}$$

where $\mathcal{X} = \mathcal{X}_1 - \mathcal{P}_1$ and $\mathcal{X}_1$ is the solution to $\mathcal{X}_1 = A\mathcal{X}_1 A^\intercal + \mathcal{P}_1$.

Hence, the performance degradation ratio for case 2 can be calculated as

$$\eta_{s,2} = \lim_{k\to\infty} \frac{\frac{1}{k}\sum_{l=1}^{k} \mathrm{tr}(\tilde{P}_l)}{\mathrm{tr}(P)} = 1 + 4\frac{\mathrm{tr}(\mathcal{X})}{\mathrm{tr}(P)} > 1.$$

For case 3, $\tilde{z}_l = Sz_l + \phi_l$, the covariance of ϕ_l is $\Sigma_z - S\Sigma_z S^\intercal$. We take the limit of $\mathbb{E}[\tilde{e}_l \tilde{e}_l^\intercal]$,

$$\begin{aligned}
&\lim_{l\to\infty} \mathbb{E}[\tilde{e}_l \tilde{e}_l^\intercal] \\
=&\lim_{l\to\infty} \sum_{i=1}^{l-1} A^i K \mathbb{E}\left[[(I_m - S)z_{l-i} - \phi_{l-i}]\,[(I_m - S)z_{l-i} - \phi_{l-i}]^\intercal\right](A^i K)^\intercal \\
=&\lim_{l\to\infty} \sum_{i=1}^{l-1} A^i K \left[(I_m - S)\Sigma_z (I_m - S)^\intercal + \Sigma_z - S\Sigma_z S^\intercal\right] K^\intercal (A^i)^\intercal.
\end{aligned} \tag{3.21}$$

Similarly, for the simplicity of notations, we define

$$\mathcal{P}_2 \triangleq K\left[(I_m - S)\Sigma_z (I_m - S)^\intercal + \Sigma_z - S\Sigma_z S^\intercal\right] K^\intercal,$$

Since $\mathcal{P}_2$ is positive semi-definite and A is stable, (3.21) can be simplified as

$$\lim_{l\to\infty} \mathbb{E}[\tilde{e}_l \tilde{e}_l^\intercal] = \mathcal{Y},$$

where $\mathcal{Y} = \mathcal{Y}_1 - \mathcal{P}_2$ and $\mathcal{Y}_1$ is the solution to the discrete Lyapunov equation $\mathcal{Y}_1 = A\mathcal{Y}_1 A^\intercal + \mathcal{P}_2$.

The performance degradation ratio η_3 for case 3) is as follows:

$$\eta_{s,3} = \lim_{k\to\infty} \frac{\frac{1}{k}\sum_{l=1}^{k} \mathrm{tr}(\tilde{P}_l)}{\mathrm{tr}(P)} = \frac{\mathrm{tr}(P) + \mathrm{tr}(\mathcal{Y})}{\mathrm{tr}(P)} \leq \eta_{s,2},$$

where inequality holds since

$$\begin{aligned}
&4\,\mathrm{tr}(\mathcal{X}) - \mathrm{tr}(\mathcal{Y}) \\
=&\sum_{i=1}^{\infty} \mathrm{tr}\left(A^i K\left(2\Sigma_z + S\Sigma_z + \Sigma_z S^\intercal\right)\left(A^i K\right)^\intercal\right) \\
=&\sum_{i=1}^{\infty} \mathrm{tr}\left(A^i K\left((I_m + S)\Sigma_z (I_m + S)^\intercal - S\Sigma_z S^\intercal + \Sigma_z\right)\left(A^i K\right)^\intercal\right) \\
\geq& 0.
\end{aligned}$$

Hence, the worst performance degradation ratio is $\eta_s = 1 + 4\dfrac{\mathrm{tr}(\mathcal{X})}{\mathrm{tr}(P)}$ with the corresponding attack strategy $(T^*, S^*, \Phi^*) = (\mathbf{0}_{m\times m}, -I_m, \mathbf{0}_{m\times m})$. □

Remark 3.3. In Theorem 3.1, the linear first-order attack model (3.5) is considered. The same result can be easily extended to a general linear time-invariant attack model of the form:

$$\begin{aligned}
c_k &= M c_{k-1} + N z_{k-1}, \\
\tilde{z}_k &= W c_k + G z_k,
\end{aligned}$$

where $c_k \in \mathbb{R}^p, M \in \mathbb{R}^{p\times p}, N \in \mathbb{R}^{p\times m}, W \in \mathbb{R}^{m\times p}, G \in \mathbb{R}^{m\times m}$. That is, under strictly stealthy attacks of the above form, the worst case performance remains the same as that in Theorem 3.1 and the optimal attack strategy is that $G = -I_p$ and the rest parameters M, N and W need to satisfy $WM^iN = 0$ $(i = 0, 1, \dots)$.

Remark 3.4. For scalar systems, the worst performance degradation ratio is

$$\eta_s = 1 + \frac{4A^2K^2(C^2P + R)}{(1 - A^2)P}$$

and the corresponding attack strategy is $(T^*, S^*, \Phi^*) = (0, -1, 0)$. Hence, the degradation is worse for systems with a higher Kalman filter gain. Note also that the worst case attack simply flips the sign of the innovation sequence.

Remark 3.5. Under the strict stealthiness metric, the optimal attack strategy in our work is aligned with the result about the worst-case linear attack under the χ^2 false alarm detector obtained in [26]. The reason why the optimal attack policies are the same for the different problem settings is that the modified innovation needs to preserve the statistics of the attack-free innovation, which lead to that $T = \mathbf{0}_{m\times m}$. However, since we consider a more general model that utilizes both past

and current information, the derivation of the optimal attack strategy is different from that in [26]. Note that $\eta = 1$ when $A = \mathbf{0}_{n\times n}$.

Remark 3.6. Under Assumption 3.1, i.e. A is stable, Theorem 3.1 provides a closed-form solution for the performance degradation ratio. If A is not stable, (3.21) will diverge, which is not interesting. Besides, although we mainly study the scenario under strictly stealthy attacks in this section, the strictly stealthy attack can be considered as a special case of ϵ-stealthy attacks with $\epsilon = 0$. In other words, a strictly stealthy attack strategy should be feasible when considering an ϵ-stealthy attack.

3.3 ϵ-stealthy Attacks

In this section, we will characterize the maximal performance degradation under an ϵ-stealthy attack. The memoryless attacker $T = \mathbf{0}_{m\times m}$ was studied in [5, 27]. We focus on the attacker with memory, i.e., $T \neq \mathbf{0}_{m\times m}$. For the sake of analysis, we will focus on scalar systems, i.e., $m = n = 1$ in the following analysis. The vector case will be a potential future work. In order to differentiate between scalar and vector systems, we use σ_z^2 to replace Σ_z to represent the covariance of z_k, i.e., $\sigma_z^2 = C^2P + R$.

For the simplicity of notation, define

$$q \triangleq \frac{\Phi}{\sigma_z^2}, \quad q \geq 0.$$

Then we have the following lemma, the proof of which is reported in Section 3.6.

Lemma 3.3. *Consider the scalar system (3.1)-(3.2), the optimization problem (3.15) is equivalent to the following problem:*

$$\begin{aligned} \arg\max_{T,S,q} \quad & J(T,S,q), \\ \text{s. t.} \quad & -\frac{1}{2} - \frac{1}{2}\log(S^2+q) + \frac{S^2+q}{2(1-T^2)} = \epsilon, \\ & -S_{oq\max} < S \leq -\sqrt{e^{-2\epsilon}-q}, \end{aligned} \tag{3.22}$$

where

$$J(T,S,q) = (1-S)^2 + q + \frac{T^2(S^2+q)}{1-T^2} - 2AT\frac{S-S^2-ST^2-q}{(1-T^2)(1-AT)}.$$

For a given q, denote

$$J_{q1}(T,S) \triangleq J(T,S,q). \tag{3.23}$$

From the constraint function of (3.22), one can obtain

$$T = f_q(S) \triangleq \sqrt{1 - \frac{S^2+q}{2\epsilon+1+\log(S^2+q)}}. \tag{3.24}$$

By substituting (3.24) into (3.23), we have

$$\begin{aligned} J_{q1}(f_q(S),S) = J_{q2}(S) \triangleq & -(2\epsilon+\log(S^2+q)) - \frac{2S}{1-Af_q(S)} \\ & + \frac{2(2\epsilon+1+\log(S^2+q))}{1-Af_q(S)}. \end{aligned} \tag{3.25}$$

The following lemma characterizes the worst performance ratio for the estimation error covariance and gives the corresponding attack strategy to achieve this performance bound for a given q, the proof of which is reported in Section 3.6.

Lemma 3.4. *Consider scalar system (3.1)-(3.2) satisfying Assumption 3.1 and linear attack of the form (3.5). Given $q \geq 0$ and $\epsilon > 0$, under the ϵ-stealthy attacks, the worst performance degradation ratio for the estimation error covariance is*

$$\eta_\epsilon = 1 + \frac{J_{q\,\mathrm{opt}} A^2 K^2 \sigma_z^2}{(1-A^2)P},$$

where $J_{q\,\mathrm{opt}} = J_{q2}(S_q)$ with S_q being such that $J'_{q2}(S_q) = 0$. The corresponding attack strategy is (T_q, S_q), where $T_q = f_q(S_q)$.

Next, we will first prove that the optimal strategy requires $q = 0$. Then, we provide the optimal attack strategy and the corresponding worst case performance. Finally, we show that our proposed attack strategy can achieve a better attack performance than that of the existing work in [5] under the same ϵ-stealthy attacks.

Similarly, for the sake of readability, the proof of the following lemma is reported in Section 3.6.

Lemma 3.5. *The solution to the original optimization problem (3.15) requires $q = 0$. Hence, the optimization problem (3.22) can be transformed into the following*

problem:

$$\begin{aligned} \arg\max_{S,T} \quad & J(S,T,0), \\ \text{s.t.} \quad & -\frac{1}{2}-\frac{1}{2}\log(S^2)+\frac{S^2}{2(1-T^2)}=\epsilon, \\ & 0<|T|\leq\sqrt{1-e^{-2\epsilon}}. \end{aligned}$$

Before we give the theorem regarding ϵ-stealthy attacks, we define the following equations for the simplicity of notations:

$$\begin{aligned} f_0(S) &\triangleq \sqrt{1-\frac{S^2}{2\epsilon+1+\log(S^2)}}, \\ J_0(S) &\triangleq -(2\epsilon+\log(S^2))-\frac{2S}{1-Af_0(S)}+\frac{2(2\epsilon+1+\log(S^2))}{1-Af_0(S)}. \end{aligned}$$

The following theorem characterizes the maximal performance degradation ratio under an ϵ-stealthy attack. We also provide the attack strategy to achieve the maximum.

Theorem 3.2. *Consider the scalar system (3.1)-(3.2) satisfying Assumption 3.1 and linear attack of the form (3.5). Given $\epsilon>0$, under the ϵ-stealthy attacks, the worst performance degradation ratio for the estimation error covariance is*

$$\eta_\epsilon = 1+\frac{J_{\text{opt}}A^2K^2\sigma_z^2}{(1-A^2)P},$$

where $J_{\text{opt}}=J_0(S_{\text{opt}})$. The corresponding attack strategy is $(T_{\text{opt}},S_{\text{opt}},0)$, where S_{opt} satisfies $J_0'(S_{\text{opt}})=0$ and $T_{\text{opt}}=f_0(S_{\text{opt}})$.

Proof. The results follows from Lemma 3.4 and Lemma 3.5.

Remark 3.7. The attack policy in [5] is given by $\tilde{z}_k=\sqrt{X}z_k$, where X is the largest solution of the equation $X=2\epsilon+1+\log X$. It corresponds to our model with $q_g=0,T_g=0,S_g=-\sqrt{X}$. The corresponding performance degradation ratio is

$$\eta_{\epsilon,g}=1+\frac{(1-S_g)^2A^2K^2\sigma_z^2}{(1-A^2)P}.$$

Hence, the difference of performance degradation between our approach and that in [5] is given by:

$$\eta_s - \eta_{\epsilon,g} = \frac{(J_{\text{opt}} - (1 - S_g)^2)A^2K^2\sigma_z^2}{(1 - A^2)P},$$

where J_{opt} is defined in Theorem 3.2. Note that the optimal parameter S for our proposed approach is between $-S_{oq\max}$ and $-e^{-\epsilon}$ while the existing linear attack strategy takes $-S_{oq\max}$, where $S_{oq\max}$ is defined in the proof of Lemma 3.4 and $J_{opt} \geq (1 - S_g)^2$. Hence, it is clear that that the performance degradation ratio bound for the estimation error covariance induced by the proposed attack strategy is larger than or equal to the existing linear attack strategy in [5] under ϵ-stealthy attacks with the same ϵ.

Remark 3.8. We focus on the scalar case in this section. For the vector case, Lemma 3.7 needs to be rewritten as "If an attacker employs an ϵ-stealthy attack in the form of (3.5), then $\rho(T) < 1$." In Lemma 3.8, the derivation of the objective function involves the sum of a geometric sequence. For the vector case, we need to use Proposition 1.5.31 in [121], "Let A be a square matrix. If $\rho(A) < 1$, the series $S = I + A + A^2 + \cdots$ converges to $(I - A)^{-1}$." Reconsider Lemma 3.8, since $\rho(A) < 1$ and $\rho(T) < 1$, the above proposition can be directly applied. Then, the optimization problem for the vector case can be obtained by using a similar method. However, it is difficult to obtain a closed-form solution since the optimization problem is not convex and involves more than one parameter. Further studies will be carried out in the future.

3.4 Simulation

In this section, we provide some numerical examples to evaluate the performance of the proposed attack strategies.

3.4.1 Vector Case under Strictly Stealthy Attacks

In this subsection, we set the system parameters as follows:

$$A = \begin{bmatrix} 0.5 & 0.2 \\ 0.1 & 0.8 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 0.6 & 0 \\ 0 & 0.3 \end{bmatrix}, \quad R = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.6 \end{bmatrix}.$$

We can obtain that

$$K = \begin{bmatrix} 0.3583 & -0.2866 \\ 0.2374 & 0.2027 \end{bmatrix}, \quad P = \begin{bmatrix} 0.6833 & -0.0302 \\ -0.0302 & 0.3548 \end{bmatrix},$$

and from Theorem 3.1, the optimal attack performance degradation ratio is $\eta = 4.6017$. Assume that the attack starts at time $k = 53$. We run 10000 simulations. The ratio of the state estimation error covariance $\tilde{P}$ to P v.s. time k is shown in Figure 3.2. The parameter S_i $(i = 1, 2)$ for attack 1-2 and S^* for strictly stealthy attack and S_{normal} for normal operation are as follows:

$$S^* = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad S_{\text{normal}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad S_1 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}, \quad S_2 = \begin{bmatrix} -0.5 & 0 \\ 0.1 & -0.8 \end{bmatrix},$$

and the corresponding covariance of the added noise is derived by $\Phi = \Sigma_z - S\Sigma_z S^\intercal$.

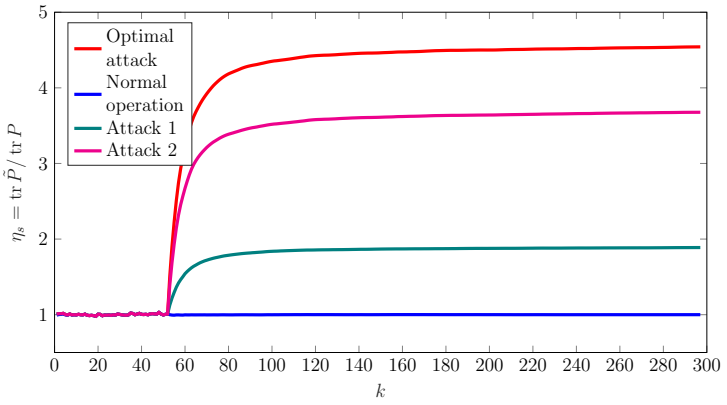

FIGURE 3.2: The ratio of the error covariance $\tilde{P}$ to P v.s. time k. The red line is the ratio of simulation under strictly stealthy attack. The blue line is the ratio of the simulation under normal operation. The teal and magenta lines denote the corresponding ratio under different attack type 1 to attack type 2, respectively.

From this figure, there is an obvious difference of the performance degradation between the normal operation and an attack operation. It is easy to see that the

error covariance ratio under the optimal attack is larger than the one under normal operation and other attacks with different attack parameters. Besides, we can also see that the optimal simulation value is almost the same as the theoretical value.

3.4.2 Different ϵ-stealthy Level

In this subsection, we consider an LTI system with scalars and set $A = 0.4, C = 1, Q = 0.2$, and $R = 0.5$. One can compute that $K = 0.3102$, and $P = 0.2248$.

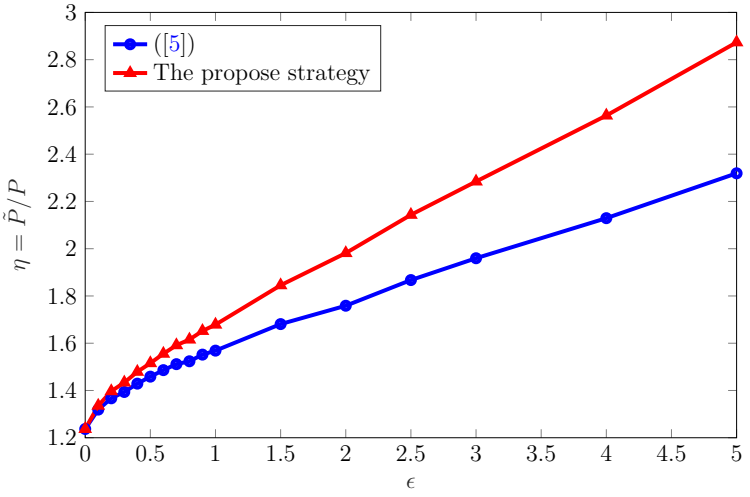

FIGURE 3.3: The ratio of the error covariance $\tilde{P}$ to P v.s. stealthiness level ϵ. The blue line with circle markers is the ratio obtained from the existing work [5]. The red line with upward-pointing triangle markers denotes the ratio in our work.

The ratio of the state estimation error covariance $\tilde{P}$ to P v.s. stealthiness level ϵ is shown in Figure 3.3. From this figure, one could see that the error covariance obtained in our work is equal to the one obtained in the existing work [5] when $\epsilon = 0$. And the error covariance obtained in our work is larger than the one derived in [5] when $\epsilon > 0$. Furthermore, the difference of the error covariances between our work and [5] is becoming larger as ϵ grows.

The values of T, S and T_k (which is used in [5]) v.s. the stealthiness level ϵ are shown in Figure 3.4. From Figure 3.4, one can see that as ϵ grows, the absolute values of T and S are becoming larger. It means that as the stealthiness level ϵ increases, the attacker employs more past attack information and current innovation.

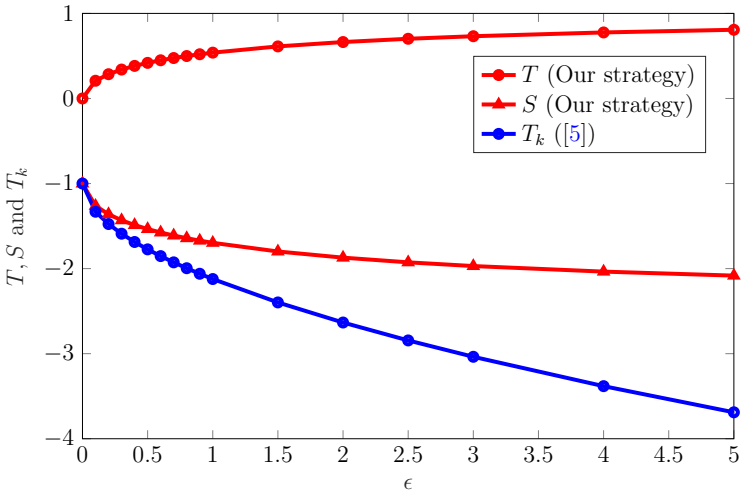

FIGURE 3.4: The values of T, S and T_k (which is used in [5]) v.s. the stealthiness level ϵ. The red lines with circle markers and triangle markers are the value of T and S in our proposed strategy, respectively. The blue line with circle markers denotes the value of T_k from the existing work [5].

3.4.3 Different System Parameter A

In this subsection, we consider an LTI system with scalars and set $\epsilon = 0.8, C = 1, Q = 0.2$, and $R = 0.5$. We study the difference induced by different system parameter A.

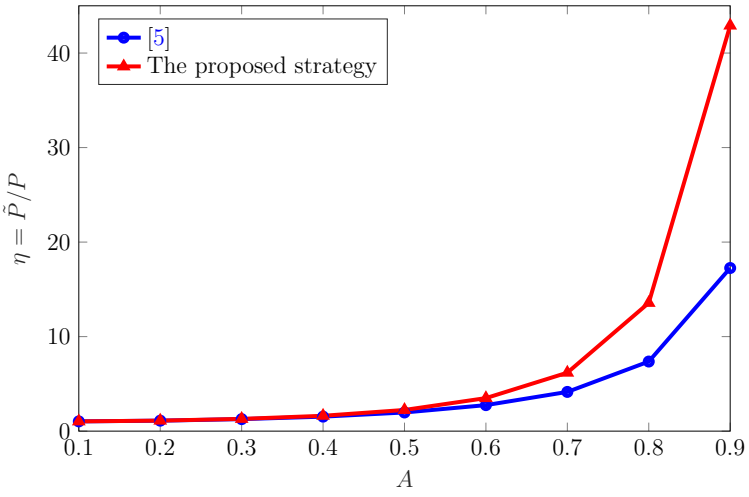

FIGURE 3.5: The ratio of the error covariance $\tilde{P}$ to P v.s. A. The blue line with circle markers is the ratio obtained in the existing work [5]. The red line with upward-pointing triangle markers denotes the ratio in our work.

The ratio of the error covariance $\tilde{P}$ to P v.s. A is shown in Figure 3.5. From this figure, one could see that the error covariance obtained in our work is larger than

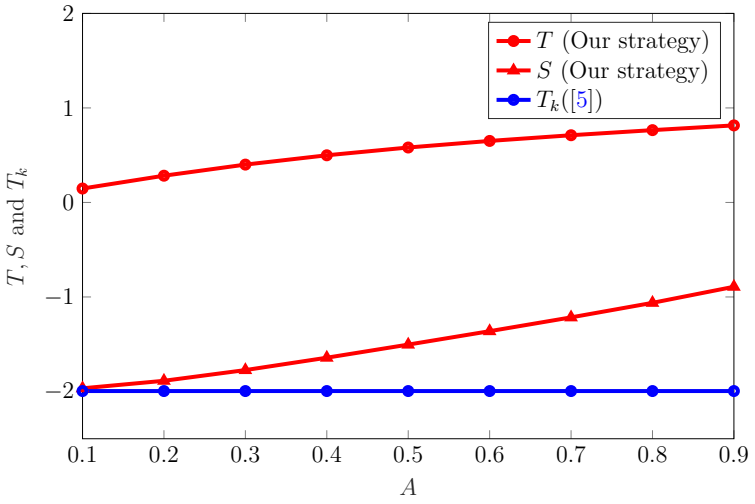

FIGURE 3.6: The values of T, S and T_k (which is used in [5]) v.s. A. The red lines with circle markers and triangle markers are the value of T and S in our proposed strategy, respectively. The blue line with circle markers denotes the value of T_k from the existing work [5].

the one derived in [5]. Furthermore, the difference of the error covariances between our work and [5] is becoming larger with A increasing.

The values of T, S and T_k (which is used in [5]) v.s. the system matrix A are shown in Figure 3.6. From this figure, one can see that as A increases, the absolute value of T is becoming larger and the absolute value of S is becoming smaller. It implies that as the system parameter A increases, the remote state estimator will attach more importance to the priori state estimate by (3.6) and (3.7). Correspondingly, the attacker will employ the past information more and use current innovation less in order to maximize the attack performance. Since the proposed approach in [5] is only related with the stealthiness level, the value of T_k keeps constant.

3.5 Conclusion

In this chapter, we characterized the fundamental limits for innovation-based remote state estimation under linear attacks. The attacker was constrained to follow a linear attack type based on the past attack signal, the latest innovation and an additive random variable. We obtained optimal attack strategies to achieve maximal performance degradation under a given stealthiness requirement. Then we provided the maximal performance degradation ratio and the corresponding optimal

attack strategy to achieve this maximum under strictly stealthy attacks for vector systems, which is a generalization of the previous work. For ϵ-stealthy attacks on scalar systems, the optimal attack strategy with an additive random noise was also presented. It was proven that the maximal performance degradation ration can be achieved without additive noise and the proposed strategy performs better than the existing linear attack strategies in terms of performance degradation. Simulation results were presented to support the theoretical results.

3.6 Proofs of Lemmas

3.6.1 Proof of Lemma 3.3

The whole section is devoted to proving Lemma 3.3. We shall present several lemmas and then proceed with the proof of Lemma 3.3.

First, we give the following lemma to characterize the property of the modified innovation sequence, which will be used to simplify the constraint condition of the optimization problem (3.15). The following lemma is for a vector case, and the scalar case follows as a special case.

Lemma 3.6. *If an attacker employs an attack in the form of* (3.5)*, the differential entropy of the compromised innovation sequence* $\tilde{z}_1^k$ *is equal to* $\frac{k}{2}\log\left((2\pi e)^m \det(\mathcal{S})\right)$*, where* $\mathcal{S} \triangleq S\Sigma_z S^\intercal + \Phi$*.*

Proof. Here, we use the notation $h(\tilde{z}_1^k)$ to represent the differential entropy:

$$h(\tilde{z}_1^k) = -\int f_{\tilde{z}_1^k}(\xi)\log f_{\tilde{z}_1^k}(\xi)d\xi,$$

where $f_{\tilde{z}_1^k}$ is the probability density function.

By (3.16), $\tilde{z}_1^k$ follows a multivariate Gaussian distribution. We have:

$$h(\tilde{z}_1^k) = \frac{1}{2}\log\left((2\pi e)^{mk}\det(\Sigma)\right), \tag{3.26}$$

where

$$\begin{aligned}&\Sigma \triangleq \operatorname{Cov}\left([\tilde{z}_1^\intercal, \tilde{z}_2^\intercal, \cdots, \tilde{z}_k^\intercal]^\intercal\right)\\&= \begin{bmatrix} \mathcal{S} & \mathcal{S}T^\intercal & \cdots & \mathcal{S}(T^{k-1})^\intercal \\ T\mathcal{S} & T\mathcal{S}T^\intercal + \mathcal{S} & \cdots & T\mathcal{S}(T^{k-1})^\intercal + \mathcal{S}(T^{k-2})^\intercal \\ \vdots & \vdots & \ddots & \vdots \\ T^{k-1}\mathcal{S} & T^{k-1}\mathcal{S}T^\intercal + T^{k-2}\mathcal{S} & \cdots & T^{k-1}\mathcal{S}(T^{k-1})^\intercal + \cdots + \mathcal{S} \end{bmatrix},\end{aligned} \tag{3.27}$$

and $\mathcal{S} \triangleq S\Sigma_z S^\intercal + \Phi$. One can perform an elementary row transformation on the matrix Σ and obtain $\det(\Sigma) = (\det(\mathcal{S}))^k$.

Hence, for any T, the differential entropy can be obtained as follows:

$$h(\tilde{z}_1^k) = \frac{k}{2}\log\left((2\pi e)^m \det(\mathcal{S})\right). \tag{3.28}$$

The proof is completed. □

Lemma 3.7. *If an attacker employs an ϵ-stealthy attack in the form of (3.5), then $|T| < 1$.*

Proof. From Lemma 3.6, it is easy to obtain

$$\begin{aligned}&\frac{1}{k}D(\tilde{z}_1^k \| z_1^k)\\&= -\frac{1}{k}h(\tilde{z}_1^k) + \frac{1}{2}\log(2\pi\sigma_z^2) + \frac{1}{k}\sum_{l=1}^{k}\frac{E[(\tilde{z}_l)^2]}{2\sigma_z^2}\\&= -\frac{1}{2}\log\left(2\pi e(S^2\sigma_z^2 + \Phi)\right) + \frac{1}{2}\log(2\pi\sigma_z^2) + \frac{1}{k}\sum_{l=1}^{k}\frac{E[(\tilde{z}_l)^2]}{2\sigma_z^2}\\&= -\frac{1}{2} - \frac{1}{2}\log(\frac{S^2\sigma_z^2 + \Phi}{\sigma_z^2}) + \frac{1}{k}\sum_{l=1}^{k}\frac{E[(\tilde{z}_l)^2]}{2\sigma_z^2}.\end{aligned}$$

Let us consider the sufficient condition of ϵ-stealthy:

$$\lim_{k\to\infty}\frac{1}{k}D(\tilde{z}_1^k \| z_1^k) \le \epsilon, \tag{3.29}$$

which implies

$$\lim_{k\to\infty} -\frac{1}{2} - \frac{1}{2}\log(\frac{S^2\sigma_z^2 + \Phi}{\sigma_z^2}) + \frac{1}{k}\sum_{l=1}^{k}\frac{E[(\tilde{z}_l)^2]}{2\sigma_z^2} \le \epsilon,$$

where

$$\mathbb{E}[(\tilde{z}_l)^2] = \sum_{i=0}^{l-1} T^{2i} \left(S^2\sigma_z^2 + \Phi\right).$$

Similarly, we divide four cases ($0 < |T| < 1$, $|T| = 1$ and $|T| > 1$) to compute $\mathbb{E}[(\tilde{z}_l)^2]$:

- $0 < |T| < 1$:

$$\mathbb{E}[(\tilde{z}_l)^2] = \frac{1 - T^{2l}}{1 - T^2} \left(S^2\sigma_z^2 + \Phi\right), \tag{3.30}$$

  then we have:

$$\lim_{k\to\infty} \frac{1}{k} \sum_{l=1}^{k} \frac{E[(\tilde{z}_l)^2]}{2\sigma_z^2} = \frac{S^2\sigma_z^2 + \Phi}{2(1 - T^2)\sigma_z^2},$$

  which could satisfy the requirement of ϵ-stealthiness.

- $|T| = 1$:

$$\mathbb{E}[(\tilde{z}_l)^2] = l \left(S^2\sigma_z^2 + \Phi\right),$$

  then,

$$\lim_{k\to\infty} \frac{1}{k} \sum_{l=1}^{k} \frac{\mathbb{E}[(\tilde{z}_l)^2]}{2\sigma_z^2} \to \infty,$$

  which contradicts the requirement of ϵ-stealthiness.

- $|T| > 1$: the sum is expressed as (3.30). It is easy to check that $\lim_{k\to\infty} \frac{1}{k} \sum_{l=1}^{k} \frac{E[(\tilde{z}_l)^2]}{2\sigma_z^2}$ will diverge, which also contradicts the requirement of ϵ-stealthiness.

As a result, T must satisfy that $0 < |T| < 1$. □

Lemma 3.8. *The optimization problem (3.15) is equivalent to the following problem:*

$$\begin{aligned} \arg\max_{T,S,q} \quad & (1-S)^2+q+\frac{T^2(S^2+q)}{1-T^2}-2AT\frac{S-S^2-ST^2-q}{(1-T^2)(1-AT)}, \\ \text{s.t.} \quad & -\frac{1}{2}-\frac{1}{2}\log(S^2+q)+\frac{S^2+q}{2(1-T^2)}\leq\epsilon, \\ & 0<|T|<1. \end{aligned} \tag{3.31}$$

Proof. From equation (3.17), one can see that the error covariance between the state estimate and the real state, $\tilde{P}_l$, can be split into two parts, one is the minimum mean square error P which is constant, and the other is the error covariance of $\tilde{e}_{l+1}=\tilde{x}^s_{l+1|l}-\tilde{x}_{l+1|l}$. Note that

$$\begin{aligned} \tilde{e}_{k+1} =& \hat{x}^s_{k+1|k}-\hat{x}_{k+1|k} \\ =& A\hat{x}^s_{k|k-1}+AKz_k-(A\hat{x}_{k|k-1}+AK\tilde{z}_k) \\ =& A(\hat{x}^s_{k|k-1}-\hat{x}_{k|k-1})-AK(T\tilde{z}_{k-1}+Sz_k+\phi_k)+AKz_k \\ =& A\tilde{e}_k+AK(1-S)z_k-AKT\tilde{z}_{k-1}-AK\phi_k. \end{aligned} \tag{3.32}$$

From (3.32), one can know that $\mathbb{E}[\tilde{e}_k]=0$. Hence, the covariance of $\tilde{e}_k$ is

$$\begin{aligned} & \mathbb{E}[(\tilde{e}_{k+1})^2] \\ =& A^2\mathbb{E}[(\tilde{e}_k)^2]+[AK(1-S)]^2\sigma_z^2+2A^2K(1-S)\mathbb{E}[\tilde{e}_kz_k] \\ & +(AKT)^2\mathbb{E}[(\tilde{z}_{k-1})^2]+A^2K^2q\sigma_z^2-2A^2KT\mathbb{E}[\tilde{e}_k\tilde{z}_{k-1}] \\ \overset{(a)}{=}& A^2\mathbb{E}[(\tilde{e}_k)^2]+[AK(1-S)]^2\sigma_z^2+(AKT)^2\mathbb{E}[(\tilde{z}_{k-1})^2] \\ & +A^2K^2q\sigma_z^2-2A^2KT\mathbb{E}[\tilde{e}_k\tilde{z}_{k-1}] \\ =& A^2\mathbb{E}[(\tilde{e}_k)^2]+[AK(1-S)]^2\sigma_z^2-2A^2KT\mathbb{E}[\tilde{e}_k\tilde{z}_{k-1}] \\ & +A^2K^2q\sigma_z^2+(AKT)^2\frac{1-T^{2(k-1)}}{1-T^2}(S^2+q)\sigma_z^2, \end{aligned} \tag{3.33}$$

where

$$\begin{aligned} \tilde{e}_k =& A\tilde{e}_{k-1}+AK(1-S)z_{k-1}-AKT\tilde{z}_{k-2}-AK\phi_{k-1} \\ =& A^{k-1}\tilde{e}_1+AK\left(\sum_{i=1}^{k-1}A^{k-1-i}(1-S)z_i\right) \\ & -AK\left(\sum_{i=0}^{k-2}A^{k-2-i}T\tilde{z}_i\right)-AK\left(\sum_{i=1}^{k-1}A^{k-1-i}\phi_i\right), \end{aligned}$$

and (a) holds due to the independence of $\tilde{e}_k$ and z_k.

To simplify the notations, we define

$$\begin{aligned}
\Xi_1 \triangleq & AK(1-S)\mathbb{E}\left[\left(\sum_{i=1}^{k-1} A^{k-1-i} z_i\right)\tilde{z}_{k-1}\right], \\
\Xi_2 \triangleq & AK\mathbb{E}\left[\left(\sum_{i=0}^{k-2} A^{k-2-i} T\tilde{z}_i\right)\tilde{z}_{k-1}\right], \\
\Xi_3 \triangleq & AK\mathbb{E}\left[\left(\sum_{i=1}^{k-1} A^{k-1-i}\phi_i\right)\tilde{z}_{k-1}\right],
\end{aligned}$$

then we have

$$\begin{aligned}
\Xi_1 =& \frac{1-(AT)^{k-1}}{1-AT} AK(1-S)S\sigma_z^2, \\
\Xi_2 =& \frac{AT\left[1-(AT)^{k-2}\right]}{1-AT}\frac{KT(S^2+q)\sigma_z^2}{1-T^2} - \frac{AT^k(T^{k-2}-A^{k-2})}{T-A}\frac{KT(S^2+q)\sigma_z^2}{1-T^2} \\
\Xi_3 =& \frac{1-(AT)^{k-1}}{1-AT} AKq\sigma_z^2,
\end{aligned}$$

Reconsider the third term of (3.33), we have

$$\begin{aligned}
& \mathbb{E}[\tilde{e}_k\tilde{z}_{k-1}] \\
=& \Xi_1 - \Xi_2 - \Xi_3 \\
=& \frac{1-(AT)^{k-1}}{1-AT} AK(1-S)S\sigma_z^2 - \frac{1-(AT)^{k-1}}{1-AT} AKq\sigma_z^2 \\
& + \frac{AT^k(T^{k-2}-A^{k-2})}{T-A}\frac{KT(S^2+q)\sigma_z^2}{1-T^2} - \frac{AT\left[1-(AT)^{k-2}\right]}{1-AT}\frac{KT(S^2+q)\sigma_z^2}{1-T^2}.
\end{aligned}$$

Consider the asymptotic behavior for (3.33) and take the limit for the above equation, one can obtain equation (3.34) (see the next page). From (3.33), it is easy to obtain (3.35) (see the next page) since $\lim_{k\to\infty}\frac{1}{k}\mathbb{E}[(\tilde{e}_1)^2] = 0$ and $\lim_{k\to\infty}\frac{1}{k}\mathbb{E}[(\tilde{e}_{k+1})^2] = 0$. Hence, the optimization problem can be rewritten as

$$\begin{aligned}
\arg\max_{T,S,q}\ & \sigma_z^2 A^2K^2\left[\left[(1-S)^2+q+T^2\frac{(S^2+q)}{1-T^2}\right]\right. \\
& \left. - 2AT\left[\frac{(1-S)S}{1-AT} - \frac{T^2(S^2+q)}{(1-T^2)(1-AT)} - \frac{q}{1-AT}\right]\right], \\
\text{s.t.}\ & -\frac{1}{2} - \frac{1}{2}\log(S^2+q) + \frac{S^2+q}{2(1-T^2)} \le \epsilon, \\
& 0 < |T| < 1.
\end{aligned}$$

$$
\begin{aligned}
&\lim_{k\to\infty}\frac{1}{k}\sum_{l=1}^{k}\mathbb{E}[(\tilde{e}_{l+1})^2]\\
=&\lim_{k\to\infty}\frac{1}{k}\sum_{l=1}^{k}A^2\mathbb{E}[(\tilde{e}_n)^2]+A^2K^2(1-S)^2\sigma_z^2+A^2K^2q\sigma_z^2+(AKT)^2\frac{(S^2+q)\sigma_z^2}{1-T^2}\\
&-2A^2KT\frac{1}{k}\sum_{l=1}^{k}\mathbb{E}[\tilde{e}_n\tilde{z}_{l-1}]\\
=&\lim_{k\to\infty}\frac{1}{k}\sum_{l=1}^{k}A^2\mathbb{E}[(\tilde{e}_n)^2]+A^2K^2\left[(1-S)^2+q+T^2\frac{(S^2+q)}{1-T^2}\right]\sigma_z^2\\
&-2A^2KT\frac{1}{k}\sum_{l=1}^{k}\mathbb{E}[\tilde{e}_n\tilde{z}_{l-1}]\\
=&\lim_{k\to\infty}\frac{1}{k}\sum_{l=1}^{k}A^2\mathbb{E}[(\tilde{e}_n)^2]+A^2K^2\left[(1-S)^2+q+T^2\frac{(S^2+q)}{1-T^2}\right]\sigma_z^2\\
&-2A^2KT\left[\frac{AK(1-S)S}{1-AT}-\frac{AKT^2(S^2+q)}{(1-T^2)(1-AT)}-\frac{AKq}{1-AT}\right]\sigma_z^2,
\end{aligned}
\tag{3.34}
$$

$$
\begin{aligned}
&\lim_{k\to\infty}\frac{1-A^2}{k}\sum_{l=1}^{k}E[(\tilde{e}_{l+1})^2]\\
=&\lim_{k\to\infty}\frac{A^2}{k}E\left[(\tilde{e}_1)^2-(\tilde{e}_{k+1})^2\right]+A^2K^2\left[(1-S)^2+q+T^2\frac{(S^2+q)}{1-T^2}\right]\sigma_z^2\\
&-2A^2KT\left[\frac{AK(1-S)S}{1-AT}-\frac{AKT^2(S^2+q)}{(1-T^2)(1-AT)}-\frac{AKq}{1-AT}\right]\sigma_z^2\\
=&A^2K^2\left[(1-S)^2+q+T^2\frac{(S^2+q)}{1-T^2}\right]\sigma_z^2\\
&-2A^2KT\left[\frac{AK(1-S)S}{1-AT}-\frac{AKT^2(S^2+q)}{(1-T^2)(1-AT)}-\frac{AKq}{1-AT}\right]\sigma_z^2,
\end{aligned}
\tag{3.35}
$$

Since $\sigma_z^2>0$ and $A^2K^2>0$, the optimization problem can be simplified as follows:

$$
\begin{aligned}
&\arg\max_{T,S,q}\ (1-S)^2+q+\frac{T^2(S^2+q)}{1-T^2}-\frac{2AT(S-ST^2-S^2-q)}{(1-T^2)(1-AT)},\\
&\text{s.t. }-\frac{1}{2}-\frac{1}{2}\log(S^2+q)+\frac{S^2+q}{2(1-T^2)}\le\epsilon,\\
&0<|T|<1.
\end{aligned}
$$

The proof is completed. □

Lemma 3.9. *When S is negative, q is fixed and the absolute value of T is fixed, $J(T,S,q) \geq J(-T,S,q)$, where the sign of T is the same as the sign of A.*

Proof. Consider the objective function J, one has

$$\begin{aligned}
&J(T,S,q) - J(-T,S,q) \\
=&(1-S)^2 + q + \frac{T^2(S^2+q)}{1-T^2} - \frac{2AT(S-ST^2-S^2-q)}{(1-T^2)(1-AT)} \\
&- \left[(1-S)^2 + q + \frac{T^2(S^2+q)}{1-T^2} + \frac{2AT(S-ST^2-S^2-q)}{(1-T^2)(1+AT)}\right] \\
=&\frac{-2AT(S-ST^2-S^2-q)}{1-T^2}\left(\frac{1}{1-AT} + \frac{1}{1+AT}\right).
\end{aligned}$$

When the sign of T is the same as the sign of A, i.e., $AT{>}0$, the above equation is non-negative, which implies $J(T,S,q) \geq J(-T,S,q)$. □

Lemma 3.10. *The attack tuple (T^*, S^*, q^*) that maximizes the performance degradation ratio for the estimation error covariance satisfies $-\frac{1}{2} - \frac{1}{2}\log(S^{*2}+q^*) + \frac{S^{*2}+q^*}{2(1-T^{*2})} = \epsilon$, where $S^* < 0$.*

Proof. First, we assume that there exists an attack tuple (T_e, S_e, q_e) such that $J(T_e,S_e,q_e) > J(T^*,S^*,q^*)$, where

$$-\frac{1}{2} - \frac{1}{2}\log(S_e^2+q_e) + \frac{S_e^2+q_e}{2(1-T_e^2)} < \epsilon. \tag{3.36}$$

Let S_e^* denote the corresponding smallest solution to the equation

$$-\frac{1}{2} - \frac{1}{2}\log(S_e^{*2}+q_e) + \frac{S_e^{*2}+q_e}{2(1-{T_e}^2)} = \epsilon.$$

Considering the derivative of J with respect to S and the property of the constraint, one can verify that $J(T_e,S_e^*,q_e) > J(T_e,S_e,q_e)$. Since among all the attack tuples satisfying the constraint equality, (T^*,S^*,q^*) is the optimal one that achieves the maximum value of J, we have $J(T^*,S^*,q^*) \geq J(T_e,S_e^*,q_e)$. Hence, $J(T^*,S^*,q^*) > J(T_e,S_e,q_e)$ is a contradiction to the early assumption. The proof is completed. □

For the simplicity of analysis, we only consider $T > 0$ and $A > 0$. Hence, T is non-negative in the above equation. The case when $T < 0$ and $A < 0$ is essentially the same.

Reconsider the constraint function of (3.31). Define $\mathscr{S} \triangleq S^2 + q$ and

$$\mathscr{C} \triangleq -\frac{1}{2} - \frac{1}{2}\log(\mathscr{S}) + \frac{\mathscr{S}}{2(1-T^2)} - \epsilon. \tag{3.37}$$

It is easy to obtain that $\mathscr{C}$ takes the minimum value at $\mathscr{S} = 1 - T^2$. Since $\mathscr{C}$ must satisfy $\mathscr{C} \leq 0$, $\mathscr{S} \geq e^{-2\epsilon}$ should hold. Hence, the range of S is $-S_{oq\max} < S \leq -\sqrt{e^{-2\epsilon} - q}$, where $-S_{oq\max}$ is the smaller solution to the equation $S^2 + q = 1 + \log(S^2 + q) + 2\epsilon$, which implies the critical solution when $T = 0$. One can prove Lemma 3.3 by the above lemmas.

3.6.2 Proof of Lemma 3.4

Compute the derivative of J_{q2}:

$$\begin{aligned} & J'_{q2}(S) \\ =& (-2)\frac{SA^2 f_q^2(S) + S^2 + q - A(S^2+q)f_q(S) - S}{(S^2+q)(1 - Af_q(S))^2} \\ & - 2\frac{\left[S(S^2+q) - (S^2+q)(2\epsilon + 1 + \log(S^2+q))\right] Af'_q(S)}{(S^2+q)(1 - Af_q(S))^2}, \end{aligned} \tag{3.38}$$

where

$$f'_q(S) = -\frac{\frac{S(2\epsilon+1+\log(S^2+q)) - S}{(2\epsilon+1+\log(S^2+q))^2}}{\sqrt{1 - \frac{S^2+q}{2\epsilon+1+\log(S^2+q)}}}.$$

First we consider the left boundary. Since there is no derivative of J_{q2} at $S = -S_{oq\max}$, we consider the local property near $S = -S_{oq\max}$. Let us take $S = S_\delta$, where $\frac{S_\delta^2+q}{2\epsilon+1+\log(S_\delta^2+q)} = 1 - \delta$ $(0 < \delta < 1)$. When $\delta \to 0$, we have

$$f_q(S_\delta) = \sqrt{1 - \frac{S_\delta^2 + q}{2\epsilon + 1 + \log(S_\delta^2 + q)}} = \sqrt{\delta}.$$

Hence, we rewrite the numerator of (3.38) as follows:

$$\begin{aligned}
&\lim_{\delta\to 0} S_\delta A^2 f_q^2(S_\delta) + S_\delta^2 + q - A(S_\delta^2+q) f_q(S_\delta) - S_\delta \\
&+ \left[S_\delta(S_\delta^2+q) - (S^2+q)(2\epsilon+1+\log(S_\delta^2+q))\right] A f_q'(S_\delta) \\
=&\lim_{\delta\to 0} S_\delta A^2 \delta + S_\delta^2 + q - A(S_\delta^2+q)\sqrt{\delta} - S_\delta \\
&+ \left[S_\delta(S_\delta^2+q) - (S_\delta^2+q)(2\epsilon+1+\log(S_\delta^2+q))\right] A f_q'(S_\delta), \\
=&\lim_{\delta\to 0} S_\delta^2 + q - S_\delta + (S_\delta^2+q)\left(S_\delta - \frac{S_\delta^2+q}{1-\delta}\right) A f_q'(S_\delta),
\end{aligned} \tag{3.39}$$

where

$$f_q'(S_\delta) = -\frac{\frac{S_\delta(2\epsilon+1+\log(S_\delta^2)+q)-S_\delta}{(2\epsilon+1+\log(S_\delta^2+q))^2}}{\sqrt{1-\frac{S_\delta^2+q}{2\epsilon+1+\log(S_\delta^2+q)}}} = -\frac{S_\delta(\frac{S_\delta^2+q}{1-\delta}-1)}{(\frac{S_\delta^2+q}{1-\delta})^2\sqrt{\delta}}.$$

Hence, as δ approaches to 0, (3.39) is given by

$$\lim_{\delta\to 0} S_\delta^2 + q - S_\delta - (S_\delta^2+q)\left(S_\delta - \frac{S_\delta^2+q}{1-\delta}\right) A \frac{S_\delta(\frac{S_\delta^2+q}{1-\delta}-1)}{(\frac{S_\delta^2+q}{1-\delta})^2\sqrt{\delta}}. \tag{3.40}$$

Since $\lim_{\delta\to 0}\left[-(S_\delta^2+q)\left(S_\delta - \frac{S_\delta^2+q}{1-\delta}\right) A \frac{S_\delta}{(\frac{S_\delta^2+q}{1-\delta})^2\sqrt{\delta}}\right] = -\infty$, the sign of (3.40) is determined by the sign of $S_\delta^2 + q - 1 + \delta$. Hence, we have

$$\begin{aligned}
&\lim_{\delta\to 0} S_\delta^2 + q - 1 + \delta = \lim_{S_\delta \to -S_{oq\max}} S_\delta^2 + q - 1 + \delta \\
=&S_{oq\max}^2 + q - 1 > 0.
\end{aligned}$$

Hence, when $S_\delta \to -S_{oq\max}^+$, the derivative of J_{q2} is positive.

When $S_\epsilon = -\sqrt{e^{-2\epsilon}-q}$, we have:

$$f_q(S_\epsilon) = \sqrt{1 - \frac{e^{-2\epsilon}}{2\epsilon+1+\log(e^{-2\epsilon})}} = \sqrt{1-e^{-2\epsilon}},$$

and

$$f_q'(S_\epsilon) = -\frac{\frac{-\sqrt{e^{-2\epsilon}-q}(2\epsilon+1+\log(e^{-2\epsilon}))+\sqrt{e^{-2\epsilon}-q}}{(2\epsilon+1+\log(e^{-2\epsilon}))^2}}{\sqrt{1-\frac{e^{-2\epsilon}}{2\epsilon+1+\log(e^{-2\epsilon})}}} = 0.$$

$$
\begin{aligned}
&S_\epsilon A^2 f_q^2(S_\epsilon) + S_\epsilon^2 + q - A(S_\epsilon^2 + q) f_q(S_\epsilon) - S_\epsilon \\
&\quad + \left[S_\epsilon(S_\epsilon^2 + q) - (S_\epsilon^2 + q)(2\epsilon + 1 + \log(S_\epsilon^2 + q))\right] A f_q'(S_\epsilon) \\
=&S_\epsilon A^2(1 - e^{-2\epsilon}) + e^{-2\epsilon} - A e^{-2\epsilon}\sqrt{1 - e^{-2\epsilon}} - S_\epsilon \\
=&S_\epsilon[A^2(1 - e^{-2\epsilon}) - 1] + e^{-2\epsilon}(1 - A\sqrt{1 - e^{-2\epsilon}}) \\
\overset{(b)}{>}&0,
\end{aligned}
$$

where inequality (b) holds since $A^2 < 1$, $1 - e^{-2\epsilon} \leq 1$ and $S_\epsilon < 0$. Hence, the derivative of J_{q2} at $S = -\sqrt{e^{-2\epsilon} - q}$ is negative.

Since the function J_1 is continuous, there must be at least one maximum point where its first derivative is zero. Hence, $\eta = 1 + \dfrac{J_{qopt} A^2 K^2 \sigma_z^2}{(1 - A^2)P}$, where $J_{qopt} = J_{q2}(S_q)$. □

3.6.3 Proof of Lemma 3.5

By analyzing the derivative of J with respect to S, combining (3.24), (3.25), and Lemma 3.4, we know that when S takes its minimum value, J obtains the maximum. Hence, $q = 0$ performs better than $q > 0$. In other words, the solution to the optimization problem (3.15) requires $q = 0$. □

Chapter 4

An Online Approach to Physical Watermark Design against Replay Attack

The last chapter analyzed the performance of remote state estimation under cyber attacks. As an important part of study on security of CPS, the detection of cyber attacks is getting more and more attention. How to detect potential attacks timely and accurately is the main concern for the following two chapters.

In this chapter, we consider the problem of designing physical watermark signals to optimally detect the replay attack on a linear time-invariant (LTI) system, under the assumption that the system parameters are unknown. Firstly, a replay attack model, where an attacker records the sensor measurements and replays them in order to fool the system, is provided. Then, we introduce a physical watermark scheme that leverages a random variable input as a watermark signal to detect possible replay attacks. The optimal watermark signal design problem is cast as an optimization problem. We aim to achieve the optimal trade-off between detection performance and control performance. We provide an online watermarking design algorithm to deal with cases with unknown system parameters. It is proved that the proposed algorithm converges to the optimal one and the corresponding convergence rate is characterized. Finally, we provide two examples to verify the effectiveness of the proposed technique.

The rest of this chapter is organized as follows. In Section 4.1, we formulate the problem by introducing the system and the attack model. In Section 4.2, we introduce the physical watermarking scheme. An online algorithm is presented in Section 4.3. This algorithm can simultaneously infer the necessary system parameters and design the watermark signal as well as the detector. Furthermore, the almost sure convergence of the watermark signal to the optimal one is proved and the corresponding convergence rate is characterized. In Section 4.4, we provide a numerical example and an industrial process example to illustrate the effectiveness of the proposed approach. Conclusions are given in Section 4.5. In Section 4.6, we give a detailed proof of Theorem 4.3.

4.1 Problem Formulation

In this section, we introduce an LTI model of CPS and a replay attack model, which will be employed in the rest of this chapter.

Consider an LTI system described by the following equations:

$$\begin{aligned} x_k &= Ax_{k-1} + B\phi_k + w_k, \\ y_k &= Cx_k + v_k, \end{aligned} \tag{4.1}$$

where $x_k \in \mathbb{R}^n$ is the vector of state variables, $y_k \in \mathbb{R}^m$ is the vector of sensors' measurements, $w_k \in \mathbb{R}^n$ is a zero mean independently and identically distributed (i.i.d.) Gaussian process noise with covariance $Q \succeq 0$, $v_k \in \mathbb{R}^m$ is a zero mean i.i.d. Gaussian measurement noise with covariance $R \succeq 0$, and $\phi_k \in \mathbb{R}^p$ means the watermark signal and we will give more details in Section 4.2.

Remark 4.1. For the simplicity of notations, we consider a stable open-loop system. However, the proposed framework in this chapter can be easily extended to a closed-loop system with an unstable plant but a stabilizing controller, which will be discussed in Section 4.2.

It is worth noticing that the introduction of the watermark signal is for intrusion detection and not for stabilization. Hence, we only consider stable systems or systems that have been pre-stabilized by some controller.

It is assumed that the process noise $w_0, w_1, \cdots$ and the measurement noise $v_0, v_1, \cdots$ are independent of each other. Moreover, since CPS usually operate for an extended period of time, we assume that the system is already in the steady state, which means that the initial condition x_{-1} is a zero mean Gaussian random vector independent of the process noise and the measurement noise and with covariance Σ, where Σ satisfies:

$$\Sigma = A\Sigma A^\intercal + Q. \tag{4.2}$$

We further make the following assumption regarding the system parameters:

Assumption 4.1. The system is strictly stable. Furthermore, (A, C) is observable and (A, B) is controllable.

Remark 4.2. Since we can perform a Kalman decomposition [122] and only work with the observable and controllable subspace, the observability and controllability assumption is without loss of generality.

Next, we introduce a replay attack model. We make the following assumptions regarding the attacker's knowledge and resources:

1. The adversary has the knowledge of all the real-time sensor measurements. In other words, it knows the sensor's measurement $y_0, \ldots, y_k$ at time k.
2. The adversary can violate the integrity of all sensory data. Specifically, the adversary can modify the real sensor signals y_k to arbitrary sensor signals y'_k.

Given the above knowledge and resources, the attacker can employ the following replay attack strategy:

1. The adversary records a sequence of sensor measurements y_ks from time k_1 to $k_1 + T$, where T is large enough to ensure that the attacker can replay the sequence for an extended period of time during the attack.
2. The adversary modifies the sensor measurements y_k to the recorded signals from time k_2 to $k_2 + T$, i.e.,

$$y'_k = y_{k-\Delta k},\ \forall\ k_2 \leq k \leq (k_2 + T),$$

where $\Delta k = k_2 - k_1$.

Note that since the system is already in the steady state, both the replayed signal y'_k and the real signal y_k from the sensors will share exactly the same statistics. Hence, if no physical watermark signal is introduced, i.e. $\phi_k = 0$, replay attacks can be stealthy for a large class of linear systems. That is the reason why we adopt the physical watermarking scheme. For more detailed discussion on the detectability of the replay attack, please refer to [28].

Consider the system architecture in Figure 4.1.

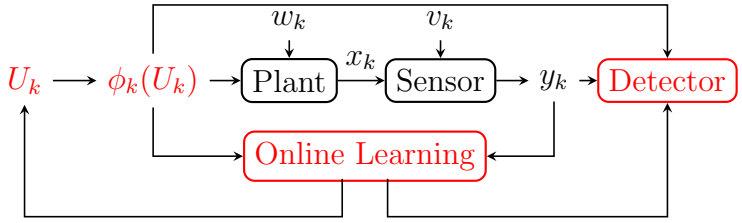

FIGURE 4.1: The system diagram.

Our goal is to design an online learning algorithm for the optimal replay attack detector and the optimal parameters U_k of the physical watermark signals, based on the collected input ϕ_k and output y_k. The physical watermark scheme will be introduced in detail in Section 4.3. Based on this scheme, we develop an approach to infer the necessary system parameters based only on the injected data ϕ_k and output data y_k, and design the parameters marked in red in the above figure: the covariance U_k of the watermark signal ϕ_k and the optimal detector based on the estimated parameters.

4.2 Physical Watermark for Systems with Known Parameters

In this section, we will introduce the physical watermark, which enables the detection of replay attacks. The optimal physical watermark is derived via solving an optimization problem that aims to achieve the optimal trade-off between control performance and detection performance. At the end of this section, the extension to closed-loop systems will be presented.

4.2.1 Physical Watermark Scheme

The main idea of physical watermark is to inject a random noise ϕ_k, which is called the watermark signal, into the system (4.1) to excite the system and check whether the system responds to the watermark signal in accordance to the dynamical model of the system. In this section, we will restrict the watermark signal ϕ_k to be zero mean i.i.d. Gaussian random variables and its covariance is denoted as U. We choose it to be zero mean because we do not wish to introduce any bias to the system state.

When no attack is present, y_k can be written as:

$$y_k = \sum_{t=0}^{k} CA^t B\phi_{k-t} + \sum_{t=0}^{k} CA^t w_{k-t} + v_k + CA^{k+1}x_{-1}. \tag{4.3}$$

To simplify notations, we define

$$\varphi_k \triangleq \sum_{\tau=0}^{k} H_\tau \phi_{k-\tau}, \tag{4.4}$$

$$\vartheta_k \triangleq \sum_{t=0}^{k} CA^t w_{k-t} + v_k + CA^{k+1}x_{-1}, \tag{4.5}$$

where H_τ is defined as

$$H_\tau \triangleq CA^\tau B. \tag{4.6}$$

Hence, y_k can be rewritten as:

$$y_k = \varphi_k + \vartheta_k. \tag{4.7}$$

From (4.4), we can know that φ_k follows a Gaussian distribution with mean zero and its covariance converges to $\mathcal{U}$, where

$$\mathcal{U} \triangleq \sum_{\tau=0}^{\infty} H_\tau U H_\tau^\intercal. \tag{4.8}$$

Similarly, from (4.5), it is easy to see that ϑ_k follows a Gaussian distribution with mean zero and covariance $\mathcal{W} = C\Sigma C^\intercal + R$, where Σ is defined in (4.2).

When the replay attack is present, the replayed y'_k can be represented as

$$y'_k = y_{k-\Delta k} = \varphi_{k-\Delta k} + \vartheta_{k-\Delta k}.$$

Since φ_k follows a Gaussian distribution with mean zero, $\varphi_{k-\Delta k}$ is a Gaussian random variable with mean zero and covariance $\mathcal{U}$. Therefore, y'_k follows a Gaussian distribution with mean zero and covariance $\mathcal{U}+\mathcal{W}$. As a result, in order to detect replay attack, we need a detector to differentiate the distribution of y_k under the following two hypotheses:

$\mathcal{H}_0$: y_k follows a Gaussian distribution with mean φ_k and covariance $\mathcal{W}$, i.e., $y_k \sim \mathcal{N}_0(\varphi_k, \mathcal{W})$.

$\mathcal{H}_1$: y_k follows a Gaussian distribution with mean zero and covariance $\mathcal{U}+\mathcal{W}$, i.e., $y_k \sim \mathcal{N}_1(0, \mathcal{U}+\mathcal{W})$.

Remark 4.3. Notice that the watermark signal $\phi_0, \cdots, \phi_k$ are known to the system operator and detector. The distribution of y_k is conditioned on $\{\phi_k\}_k$ and this distribution converges to a Gaussian distribution with mean φ_k and covariance $\mathcal{W}$.

The Neyman-Pearson (NP) detector [123] for hypothesis $\mathcal{H}_0$ versus hypothesis $\mathcal{H}_1$ is formalized by the following lemma:

Lemma 4.1. *At time k, the NP detector rejects $\mathcal{H}_0$ in favor of $\mathcal{H}_1$ if*

$$g_k = \big(y_k - \varphi_k\big)^\intercal \mathcal{W}^{-1} \big(y_k - \varphi_k\big) - y_k^\intercal \left(\mathcal{W}+\mathcal{U}\right)^{-1} y_k \geq \eta, \tag{4.9}$$

where η is a threshold chosen by the system operator. Otherwise, hypothesis $\mathcal{H}_0$ is accepted.

Remark 4.4. For the sake of simplicity, we only consider detecting the replay attack based on the current measurement y_k. In principle, by considering joint distribution of y_k, $y_{k-1}, \cdots, y_{k-\Delta t}$, one may take a moving horizon approach to design a detector. However, the proposed framework in this chapter can be easily extended to a multiple y_k's case by stacking the state vector.

Remark 4.5. Notice that since hypothesis $\mathcal{H}_0$ is time-varying due to the φ_k term, the threshold η needs to be time-varying to ensure that the false alarm rate is constant. If η is still chosen as a constant instead, then the system operator

could calculate the expected false alarm rate by numerical integration, as φ_k is a stationary process.

The performance of the detector is quantified by the expected KL divergence between distributions $\mathcal{N}_0$ and $\mathcal{N}_1$, which is formalized by the following theorem:

Theorem 4.1. *The expected KL divergence of distribution $\mathcal{N}_0$ and $\mathcal{N}_1$ is*

$$\mathbb{E}\ D_{KL}\left(\mathcal{N}_1 \| \mathcal{N}_0\right) = \operatorname{tr}\left(\mathcal{U}\mathcal{W}^{-1}\right) - \frac{1}{2}\log\det\left(I + \mathcal{U}\mathcal{W}^{-1}\right). \tag{4.10}$$

Furthermore, the expected KL divergence satisfies the following inequality

$$\begin{aligned} \frac{1}{2}\operatorname{tr}\left(\mathcal{U}\mathcal{W}^{-1}\right) &\leq \mathbb{E}\ D_{KL}\left(\mathcal{N}_1 \| \mathcal{N}_0\right) \\ &\leq \operatorname{tr}\left(\mathcal{U}\mathcal{W}^{-1}\right) - \frac{1}{2}\log\left[1 + \operatorname{tr}\left(\mathcal{U}\mathcal{W}^{-1}\right)\right]. \end{aligned} \tag{4.11}$$

Proof. The proof is essentially the same as the proof in [46]. □

Remark 4.6. Notice that the expected KL divergence is a convex function of $\mathcal{U}$ and hence U. However, both the upper and lower bounds of it are increasing functions of $\operatorname{tr}(\mathcal{U}\mathcal{W}^{-1})$. Therefore, we could maximize $\operatorname{tr}(\mathcal{U}\mathcal{W}^{-1})$, which is linear with respect to U, instead of directly maximizing the detection performance, i.e., the expected KL divergence, which is computationally difficult.

It is worth noticing that although the watermark signal enables the detection of replay attacks, it also degrades the system control performance. Hence, it is of great importance to design the watermark signal to achieve the optimal trade-off between intrusion detection performance and system control performance. In order to quantify the control performance loss, the following linear-quadratic-Gaussian (LQG) metric is used:

$$J = \lim_{T\to+\infty} \mathbb{E}\left(\frac{1}{T}\sum_{k=0}^{T-1}\begin{bmatrix} y_k \\ \phi_k \end{bmatrix}^{\mathsf{T}} X \begin{bmatrix} y_k \\ \phi_k \end{bmatrix}\right), \tag{4.12}$$

where

$$X = \begin{bmatrix} X_{yy} & X_{y\phi} \\ X_{\phi y} & X_{\phi\phi} \end{bmatrix} \succ 0$$

is the weight matrix for the LQG control and it is chosen by the system operator.

Remark 4.7. In this chapter, we choose the LQG metric to quantify the control performance of a system running in the steady state. It should be noted that other performance metrics can also be incorporated into our framework, as long as they can be computed from the Markov parameters H_τ. This is due to the fact that H_τ can be inferred using our proposed online approach and is one of the most important parameters in online learning process.

Since y_k and ϕ_k converge to a stationary process, J can be written as

$$J = \lim_{k\to} \operatorname{tr}\left(X \operatorname{Cov}\left(\begin{bmatrix} y_k \\ \phi_k \end{bmatrix}\right)\right) = \operatorname{tr}\left(X \begin{bmatrix} \mathcal{W}+\mathcal{U} & H_0 U \\ U H_0^\intercal & U \end{bmatrix}\right).$$

Hence, J is an affine function with respect to U, which can be represented as

$$J = J_0 + \Delta J = \operatorname{tr}(X_{yy}\mathcal{W}) + \operatorname{tr}(XS),$$

where J_0 denotes the optimal LQG cost, and S is linear with respect to U, being defined as

$$S \triangleq \begin{bmatrix} \mathcal{U} & H_0 U \\ U H_0^\intercal & U \end{bmatrix}.$$

Hence, in order to obtain the optimal trade-off between the detection performance and control performance, we formulate the following optimization problem:

$$\begin{aligned} U_* = \arg\max_{U\geq 0} \quad & \operatorname{tr}(\mathcal{U}\mathcal{W}^{-1}) \\ \text{subject to} \quad & \operatorname{tr}(XS) \leq \delta, \end{aligned} \tag{4.13}$$

where δ is a design parameter depending on how much control performance loss is tolerable.

An important property of the optimization problem (4.13) is that the optimal solution is usually a rank-1 matrix, which is formalized by the following theorem:

Theorem 4.2. *The optimization problem* (4.13) *is equivalent to*

$$\begin{aligned} U_* = \arg\max_{U\geq 0} \quad & \operatorname{tr}(U\mathcal{P}) \\ \textit{subject to} \quad & \operatorname{tr}(U\mathcal{X}) \leq \delta, \end{aligned} \tag{4.14}$$

where

$$\mathcal{P} \triangleq \sum_{\tau=0}^{\infty} H_\tau^\intercal \mathcal{W}^{-1} H_\tau, \tag{4.15}$$

$$\mathcal{X} \triangleq \left(\sum_{\tau=0}^{\infty} H_\tau^\intercal X_{yy} H_\tau \right) + H_0^\intercal X_{y\phi} + X_{\phi y} H_0 + X_{\phi\phi}. \tag{4.16}$$

The optimal solution to (4.14) *is*

$$U_* = zz^\intercal,$$

where z is the eigenvector corresponding to the maximum eigenvalue of the matrix $\mathcal{X}^{-1}\mathcal{P}$ and $z^\intercal \mathcal{X} z = \delta$. Furthermore, the solution is unique if $\mathcal{X}^{-1}\mathcal{P}$ has only one maximum eigenvalue.

Proof. By the definition of $\mathcal{U}$, we know that

$$\operatorname{tr}(\mathcal{U}\mathcal{W}^{-1}) = \sum_{\tau=0}^{\infty} \operatorname{tr}\left(H_\tau U H_\tau^\intercal \mathcal{W}^{-1}\right) = \sum_{\tau=0}^{\infty} \operatorname{tr}\left(U H_\tau^\intercal \mathcal{W}^{-1} H_\tau\right) = \operatorname{tr}\left(U\mathcal{P}\right).$$

Following similar steps as in the above proof, we have that $\operatorname{tr}(XS) = \operatorname{tr}(U\mathcal{X})$. Moreover, since $X \succ 0$, we have that

$$\mathcal{X} \succeq H_0^\intercal X_{yy} H_0 + H_0^\intercal X_{y\phi} + X_{\phi y} H_0 + X_{\phi\phi} \succeq X_{\phi\phi} - X_{\phi y} X_{yy}^{-1} X_{y\phi} \succ 0.$$

The proof about the optimal solution is similar to the proof of Theorem 7 in [45] and we omit it here.

□

4.2.2 Extension to Closed-loop Systems

In this section, we discuss how to generalize the problem formulation for a closed-loop system with a stabilizing controller. Let us consider the following system discussed in [28]:

$$x_{k+1} = Ax_k + B(u_k + \phi_k) + w_k,$$
$$y_k = Cx_k + v_k,$$

with the following estimator and controller:

$$\hat{x}_{k+1} = A\hat{x}_k + K(y_{k+1} - CA\hat{x}_k),$$
$$u_k = L\hat{x}_k,$$

and LQG cost as

$$J = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{k=0}^{T-1} y_k^\intercal X_{yy} y_k + (u_k + \phi_k)^\intercal X_{\phi\phi}(u_k + \phi_k)\right],$$

where u_k is the optimal LQG control signal.

The state $\tilde{x}_k$ and output $\tilde{y}_k$ can be redefined as:

$$\tilde{x}_k = \begin{bmatrix} x_k \\ \hat{x}_k \end{bmatrix}, \text{ and } \tilde{y}_k = \begin{bmatrix} y_k \\ u_k \end{bmatrix},$$

and the design of watermark signal in a closed-loop system can be converted to the open-loop formulation.

It should be noted that precise knowledge of the system parameters is needed in order to design the detector and the optimal watermark signal. However, acquiring the parameters may be costly and troublesome. Moreover, there may be unforeseen changes in the model of the system, such as topological changes in power systems. Hence, the identified system model may change during the system operation. As a result, it is beneficial for the system to "learn" the parameters and design the detector and watermark signal in real-time, which is our focus in the next section.

4.3 Physical Watermark for Systems with Unknown Parameters

In this section, we focus on designing an online "learning" approach to infer the system parameters, based on which, we show how to design watermark signals and the optimal detector and prove that the covariance of the physical watermark and the detector asymptotically converge to the optimal ones with known system parameters.

Throughout the section, we make the following assumptions:

Assumption 4.2. 1. A is diagonalizable.

2. The maximum eigenvalue of $\mathcal{X}^{-1}\mathcal{P}$ is unique.

3. The system is not under attack during the learning phase.

4. The number of distinct eigenvalues of A, which is denoted as $\tilde{n}$, is known.

5. The LQG weight matrix X and the largest tolerable LQG loss δ are known.

Remark 4.8. The first and second assumptions are required in order to ensure that the optimal covariance of the watermark signal is a differentiable function of H_τ, i.e., the problem is not ill-conditioned. The third assumption is needed since there is no way to do system identification if there is no (real) sensory data and it is also necessary to prove the asymptotic convergence of our proposed algorithm to the optimal one without attack as this cannot be achieved in finite time due to the existence of the inherent process and measurement noise. Nevertheless, we illustrate through simulation, that after a certain period of learning phase, our algorithm can approximate the optimal solution with reasonable accuracy and the system can detect replay attack. The fourth assumption is also needed in order to prove convergence, although we shall demonstrate in the simulation that we can use a reduced model to approximate the system with good accuracy. The fifth assumption should hold for all practical cases as X and δ are design parameters chosen by the system operator.

For the sake of legibility, we will first introduce our algorithm and then present the theorem on the correctness of our approach.

4.3.1 An Online Algorithm

In this subsection, we will present the complete algorithm in a pseudo-code form. After that, the online "learning" scheme will be introduced in detail.

Algorithm 1 describes our proposed online watermarking algorithm. The notations are described later in the subsection.

First, we initialize some parameters which will be used later. In each round of the **while** iteration, the optimal covariance of the watermarking $U_{k,*}$ based on

current knowledge is computed firstly. Based on the derived covariance, one can update the covariance U_k by combining "exploration" and "exploitation" term which will be described in detail later. According to the updated covariance, we generate the watermarking signals ϕ_k and inject them to the plant. Then we collect the sensory data y_k and employ them and watermarking signals to infer necessary system parameters $H_{k,\tau}, \mathcal{P}_k, \mathcal{X}_k$. Based on the estimated parameters, one can update the NP detector $\hat{g}_k$. Then one can repeat the above process to identify system parameters and design the watermarking signals and the detector.

A pseudo-code form for Algorithm 1 is given by:

Algorithm 1 Online Watermarking Design

Require: $\mathcal{P}_{-1} \leftarrow I$, $\mathcal{X}_{-1} \leftarrow X_{\phi\phi}, k \leftarrow 0$

Ensure:

1: **while** true **do**
2: $U_{k,*} \leftarrow \arg\max_{U \geq 0, \operatorname{tr}(U\mathcal{X}_{k-1}) \leq \delta} \operatorname{tr}(U\mathcal{P}_{k-1})$
3: $U_k \leftarrow U_{k,*} + (k+1)^{-\beta}\delta I$
4: Generate random variable $\zeta_k \sim \mathcal{N}(0, I)$
5: Apply watermark signal $\phi_k \leftarrow U_k^{1/2}\zeta_k$
6: Collect sensory data y_k
7: $H_{k,\tau} \leftarrow \frac{1}{k-\tau+1}\sum_{t=\tau}^{k} y_t \phi_{t-\tau}^{\intercal} U_{t-\tau}^{-1}$
8: Compute the coefficient of $p_k(x)$ by solving (4.23)
9: **if** $p_k(x)$ is Schur stable **then**
10: Update $\mathcal{P}_k, \mathcal{X}_k$ from (4.24)-(4.29)
11: **end if**
12: Update $\hat{g}_k$ from (4.30)
13: $k \leftarrow k+1$
14: **end while**

Remark 4.9. For Algorithm 1, $\mathcal{P}_k, \mathcal{X}_k$ are defined in (4.18), U is the covariance of watermarking signal, and $H_{k,\tau}$ is defined in (4.20). Step 3 is the update of the covariance of the physical watermark in (4.17). All parameters will be illustrated in the following subsections.

Next, we give details of this algorithm.

Generation of the Watermark Signal ϕ_k

We design U_k, which can be viewed as an approximation for the optimal covariance of the watermark signal U, as

$$U_k = U_{k,*} + \frac{\delta}{(k+1)^\beta} I, \tag{4.17}$$

where β is the decay rate and $0 < \beta < 1$, δ is the maximum tolerable LQG loss defined in (4.13), and $U_{k,*}$ is the solution of the following optimization problem

$$\begin{aligned} U_{k,*} = \arg\max_{U \geq 0} \quad & \operatorname{tr}(U\mathcal{P}_{k-1}), \\ \text{subject to} \quad & \operatorname{tr}(U\mathcal{X}_{k-1}) \leq \delta, \end{aligned} \tag{4.18}$$

where $\mathcal{P}_{k-1}$ is the estimate of $\mathcal{P}$ matrix and $\mathcal{X}_{k-1}$ is the estimate of $\mathcal{X}$ matrix, based on $y_0, \ldots, y_{k-1}$, $\phi_0, \ldots, \phi_{k-1}$, both of which are initialized as:

$$\mathcal{P}_{-1} = I,\ \mathcal{X}_{-1} = X_{\phi\phi}.$$

The inference procedure of $\mathcal{P}_k$ and $\mathcal{X}_k$ for $k \geq 0$ will be provided in the further subsections.

Remark 4.10. It is worth noticing that the second term $\frac{1}{(k+1)^\beta} I$ on the right hand side of (4.17) is important for parameter identification. This is because that $U_{k,*}$ is in general a rank-1 matrix and thus it does not provide persistent excitation to the system to identify the necessary system parameters. Conceptually, one can interpret the $\frac{1}{(k+1)^\beta} I$ term as an "exploration" term, as it provides necessary excitation to the system in order to infer the parameters. The first term $U_{k,*}$ can be interpreted as an "exploitation" term since it is optimal under the current knowledge of the system parameters.

At each time k, the watermark signal is chosen to be

$$\phi_k = U_k^{1/2} \zeta_k, \tag{4.19}$$

where ζ_ks are i.i.d. Gaussian random vectors with covariance I.

Inference on H_τ

In this subsection, we focus on inferring the necessary system parameters from the collected sensory data and watermark signals. We will first identify the Markov parameters H_τ of the system.

Define $H_{k,\tau}$, where $0 \leq \tau \leq 3\tilde{n} - 2$, as

$$\begin{aligned} H_{k,\tau} &\triangleq \frac{1}{k-\tau+1} \sum_{t=\tau}^{k} y_t \phi_{t-\tau}^{\intercal} U_{t-\tau}^{-1} \\ &= H_{k-1,\tau} + \frac{1}{k-\tau+1} \left(y_k \phi_{k-\tau}^{\intercal} U_{k-\tau}^{-1} - H_{k-1,\tau} \right), \end{aligned} \tag{4.20}$$

where $H_{k,\tau}$ is an estimate of H_τ at time k.

Remark 4.11. Notice that other methods, such as subspace identification, may perform better for classical system identification tasks than our proposed method. However, since the covariance of our watermark signal converges to a degenerate matrix (of rank 1), it is non-trivial to analyze the convergence properties for more advanced system identification methods, such as subspace identification, which we shall leave as our future work.

Note that the calculation of the matrices $\mathcal{U}$, $\mathcal{W}$, $\mathcal{P}$ and $\mathcal{X}$ requires H_τ for all $\tau \geq 0$. Next, we will show that in fact only finitely many H_τs are needed to compute those matrices, which requires the following lemma:

Lemma 4.2. *Assume that the matrix A is diagonalizable and $\lambda_1, \ldots, \lambda_{\tilde{n}}$ are its distinct eigenvalues, then there exist unique $\Omega_1, \cdots, \Omega_{\tilde{n}}$, such that*

$$H_\tau = \sum_{i=1}^{\tilde{n}} \lambda_i^\tau \Omega_i. \tag{4.21}$$

Proof. Without loss of generality, it is assumed that A is a diagonal matrix. We have

$$A^\tau = \operatorname{diag}(\lambda_1^\tau I_1, \lambda_2^\tau I_2, \cdots, \lambda_{\tilde{n}}^\tau I_{\tilde{n}}),$$

where λ_i is the ith distinct eigenvalue of A, λ_i^τ is λ_i to the power of τ, and I_i denotes the identity matrix of size n_i by n_i with n_i the multiplicity of λ_i. Therefore, H_τ

can be written as

$$H_\tau = CA^\tau B = \sum_{i=1}^{\tilde{n}} \lambda_i^\tau \Omega_i,$$

with $\Omega_i = C\,\mathrm{diag}(0,\dots,0,I_i,0,\dots,0)B$. □

Since A satisfies its own minimal polynomial $p(x) = \prod_{i=1}^{\tilde{n}}(x-\lambda_i) = x^{\tilde{n}} + \alpha_{\tilde{n}-1}x^{\tilde{n}-1} + \dots + \alpha_0$, we know that for arbitrary $i \geq 0$:

$$H_{i+\tilde{n}} + \alpha_{\tilde{n}-1}H_{i+\tilde{n}-1} + \cdots + \alpha_0 H_i = CA^i p(A)B = 0. \tag{4.22}$$

By (4.22), one can use $H_0, H_1, \cdots, H_{3\tilde{n}-2}$ to estimate λ_is and Ω_is and thus H_τ for any τ. To this end, we define:

$$\begin{bmatrix} \alpha_{k,0} \\ \vdots \\ \alpha_{k,\tilde{n}-1} \end{bmatrix} \triangleq -\Xi_k^{-1} \begin{bmatrix} \mathrm{tr}(\mathcal{H}_{k,0}^\intercal \mathcal{H}_{k,\tilde{n}}) \\ \vdots \\ \mathrm{tr}(\mathcal{H}_{k,\tilde{n}-1}^\intercal \mathcal{H}_{k,\tilde{n}}) \end{bmatrix}, \tag{4.23}$$

where

$$\Xi_k \triangleq \begin{bmatrix} \mathrm{tr}(\mathcal{H}_{k,0}^\intercal \mathcal{H}_{k,0}) & \cdots & \mathrm{tr}(\mathcal{H}_{k,0}^\intercal \mathcal{H}_{k,\tilde{n}-1}) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(\mathcal{H}_{k,\tilde{n}-1}^\intercal \mathcal{H}_{k,0}) & \cdots & \mathrm{tr}(\mathcal{H}_{k,\tilde{n}-1}^\intercal \mathcal{H}_{k,\tilde{n}-1}) \end{bmatrix},$$

and

$$\mathcal{H}_{k,i} \triangleq \begin{bmatrix} H_{k,i} \\ H_{k,i+1} \\ \vdots \\ H_{k,i+2\tilde{n}-2} \end{bmatrix}.$$

Remark 4.12. We can prove that $\alpha_{k,i}$ from (4.23) is the solution of the following problem:

$$\min \|\mathcal{H}_{k,\tilde{n}} + \alpha_{\tilde{n}-1}\mathcal{H}_{k,\tilde{n}-1} + \cdots + \alpha_0 \mathcal{H}_{k,0}\|_F,$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix.

Denote the roots of the polynomial $p_k(x) = x^{\tilde{n}} + \alpha_{k,\tilde{n}-1}x^{\tilde{n}-1} + \cdots + \alpha_{k,0}$ to be $\lambda_{k,1}, \cdots, \lambda_{k,\tilde{n}}$. Define a Vandermonde like matrix V_k as

$$V_k \triangleq \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_{k,1} & \lambda_{k,2} & \cdots & \lambda_{k,\tilde{n}} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{k,1}^{3\tilde{n}-2} & \lambda_{k,2}^{3\tilde{n}-2} & \cdots & \lambda_{k,\tilde{n}}^{3\tilde{n}-2} \end{bmatrix},$$

where $\lambda_{k,i}$ denotes an estimate of λ_i at time k and $\lambda_{k,i}^{\tau}$ denotes $\lambda_{k,i}$ to the power of τ, and we estimate Ω_i as

$$\begin{bmatrix} \Omega_{k,1} \\ \vdots \\ \Omega_{k,\tilde{n}} \end{bmatrix} = (V_k \otimes I_m)^{+} \begin{bmatrix} H_{k,0} \\ \cdots \\ H_{k,3\tilde{n}-2} \end{bmatrix}. \tag{4.24}$$

Inference on φ_k, ϑ_k and $\mathcal{W}$

In this subsection, we focus on the inference of ϑ_k and φ_k defined in (4.4), which corresponds to the parts of y_k generated by the noise and watermark signals respectively. We will further infer the covariance $\mathcal{W}$ of ϑ_k.

Define $\hat{\varphi}_k$ as

$$\hat{\varphi}_k \triangleq \sum_{i=1}^{\tilde{n}} \hat{\varphi}_{k,i}, \tag{4.25}$$

where $\hat{\varphi}_{k,i} = \lambda_{k,i}\hat{\varphi}_{k-1,i} + \Omega_{k,i}\phi_k$ and $\hat{\varphi}_{-1,i} = 0$. Then, one can estimate ϑ_k as

$$\hat{\vartheta}_k \triangleq y_k - \hat{\varphi}_k, \tag{4.26}$$

and estimate the covariance of ϑ_k as

$$\mathcal{W}_k \triangleq \frac{1}{k+1}\sum_{t=0}^{k} \hat{\vartheta}_t\hat{\vartheta}_t^{\intercal}. \tag{4.27}$$

Inference on $\mathcal{P}$, $\mathcal{X}$, $\mathcal{U}$ and g_k

Finally we can derive an estimation of the $\mathcal{P}$ and $\mathcal{X}$ matrices, which are required to compute the optimal covariance U of the watermark signal, given by

$$\begin{aligned}
\mathcal{P}_k &= \sum_{\tau=0}^{\infty} \left(\sum_{i=1}^{\tilde{n}} \lambda_{k,i}^{\tau} \Omega_{k,i} \right)^{\intercal} \mathcal{W}_k^{-1} \left(\sum_{i=1}^{\tilde{n}} \lambda_{k,i}^{\tau} \Omega_{k,i} \right) \\
&= \sum_{\tau=0}^{\infty} \left(\sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{n}} \lambda_{k,i}^{\tau} \lambda_{k,j}^{\tau} \Omega_{k,i}^{\intercal} \mathcal{W}_k^{-1} \Omega_{k,j} \right) \\
&= \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{n}} \left(\sum_{\tau=0}^{\infty} (\lambda_{k,i} \lambda_{k,j})^{\tau} \right) \Omega_{k,i}^{\intercal} \mathcal{W}_k^{-1} \Omega_{k,j} \\
&= \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{1}{1 - \lambda_{k,i} \lambda_{k,j}} \Omega_{k,i}^{\intercal} \mathcal{W}_k^{-1} \Omega_{k,j}, \qquad (4.28)
\end{aligned}$$

where (28) is derived from the summation of geometric series, and

$$\begin{aligned}
\mathcal{X}_k &= \sum_{\tau=0}^{\infty} \left(\sum_{i=1}^{\tilde{n}} \lambda_{k,i}^{\tau} \Omega_{k,i} \right)^{\intercal} X_{yy} \left(\sum_{i=1}^{\tilde{n}} \lambda_{k,i}^{\tau} \Omega_{k,i} \right) + \sum_{i=1}^{\tilde{n}} \Omega_{k,i}^{\intercal} X_{y\phi} + X_{\phi y} \sum_{i=1}^{\tilde{n}} \Omega_{k,i} + X_{\phi\phi} \\
&= \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{1}{1 - \lambda_{k,i} \lambda_{k,j}} \Omega_{k,i}^{\intercal} X_{yy} \Omega_{k,j} + \sum_{i=1}^{\tilde{n}} \Omega_{k,i}^{\intercal} X_{y\phi} + X_{\phi y} \sum_{i=1}^{\tilde{n}} \Omega_{k,i} + X_{\phi\phi}. \qquad (4.29)
\end{aligned}$$

The NP detection statistics g_k can be approximated by

$$\hat{g}_k = (y_k - \hat{\varphi}_k)^T \mathcal{W}_k^{-1} (y_k - \hat{\varphi}_k) - y_k^T (\mathcal{W}_k + \mathcal{U}_k)^{-1} y_k, \qquad (4.30)$$

with

$$\begin{aligned}
\mathcal{U}_k &= \sum_{\tau=0}^{\infty} \left(\sum_{i=1}^{\tilde{n}} \lambda_{k,i}^{\tau} \Omega_{k,i} \right) U_{k,*} \left(\sum_{i=1}^{\tilde{n}} \lambda_{k,i}^{\tau} \Omega_{k,i} \right)^{\intercal} \\
&= \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{1}{1 - \lambda_{k,i} \lambda_{k,j}} \Omega_{k,i} U_{k,*} \Omega_{k,j}^{\intercal}. \qquad (4.31)
\end{aligned}$$

Remark 4.13. For our proposed algorithm, the system identification and watermark design are tightly coupled. As is discussed in Remark 4.10, the injection of a rank-1 watermark signal (we assume that it is performed optimally) is required for the watermarking-based replay attack detection. On the other hand, system

identification requires persistency of excitation, i.e., the injected signal needs to be full rank. Therefore, the covariance of the injected signal is carefully designed to be the sum of the "optimal" rank-1 covariance matrix on our current system knowledge and a diminishing factor $\frac{1}{(k+1)^\beta} I$, and we will further prove that although this additional term vanishes asymptotically, it provides enough information to perfectly identify the necessary system parameters.

4.3.2 Algorithm Properties

The following theorem establishes the convergence of $U_{k,*}$ and g_k, the proof of which is provided in Section 4.6.

Theorem 4.3. *Assume that A is strictly stable and Assumption 4.2 holds. If $0 < \beta < 1$, then for any $\epsilon > 0$, the following limits hold almost surely:*

$$\lim_{k\to\infty} \frac{U_{k,*} - U_*}{k^{-\gamma+\epsilon}} = 0, \lim_{k\to\infty} \frac{\hat{g}_k - g_k}{k^{-\gamma+\epsilon}} = 0, \tag{4.32}$$

where $\gamma = \frac{1-\beta}{2} > 0$. In particular, $U_{k,*}$ and $\hat{g}_k$ almost surely converge to U_* and g_k, respectively.

Combining the definition of $U_k = U_{k,*} + \frac{\delta}{(k+1)^\beta} I$, we have the following corollary:

Corollary 4.1. *Assume that A is strictly stable and Assumption 2 holds. If $0 < \beta < 1$, then for any $\epsilon > 0$, the following limit holds almost surely:*

$$\lim_{k\to\infty} \frac{U_k - U_*}{k^{-\min\{\gamma,\beta\}+\epsilon}} = 0. \tag{4.33}$$

Remark 4.14. Notice that (4.32) implies that both $U_{k,*} - U_*$ and $\hat{g}_k - g_k$ are of the order $O(k^{-\gamma+\epsilon})$ as k approaches infinity. Therefore, the convergence rate γ is maximized when $\beta \to 0^+$, which corresponds to the case where the exploration term $\frac{\delta}{(k+1)^\beta} I$ in U_k remains constant. However, although this will maximize the performance for the inference algorithm, the covariance U_k of the watermark signal ϕ_k will not converge to the true optimal U_*. In order to achieve "fastest" convergence rate of U_k, we need to choose the decay rate to be $\beta = \arg\max_\beta \{\gamma, \beta\} = 1/3$.

It is worth noticing that Theorem 4.3 only provides an upper bound for the almost sure convergence rate and we shall leave the study on the exact convergence rate as a further research direction. It is also of interest to explore if faster convergence can be achieved by using more advanced system identification techniques.

4.4 Simulation

In this section, the performance of the proposed technique is verified. We will apply the online "learning" algorithm to a numerical example and an industrial process example.

4.4.1 A Numerical Example

First we choose $m = 3, n = 5, p = 2$ and A, B, C are all randomly generated, with A being stable. Here, they are chosen as follows:

$$A = \begin{bmatrix} -0.2679 & -0.0051 & -0.3906 & 0.0236 & 0.1320 \\ 0.2384 & -0.0025 & 0.2226 & -0.9777 & -0.3195 \\ 0.4326 & 0.0046 & 0.6363 & 0.02755 & -0.0349 \\ -0.1930 & -0.0872 & -0.1145 & 0.3041 & 0.0685 \\ 0.3989 & 0.0316 & 0.4282 & -0.2710 & -0.1041 \end{bmatrix},$$

$$B = \begin{bmatrix} 0.8967 & -2.0470 \\ 1.2113 & 0.9807 \\ 0.9199 & -1.6308 \\ 2.0321 & 0.0613 \\ -0.1976 & -0.5899 \end{bmatrix},$$

$$C = \begin{bmatrix} -0.7265 & -1.1636 & 0.8859 & -0.5425 & -0.2432 \\ -2.1019 & 0.0029 & 1.2457 & -0.2289 & 1.9068 \\ -1.1004 & -0.2283 & 0.7216 & 0.1878 & -0.3110 \end{bmatrix}.$$

It is assumed that X in (4.12), the covariance matrices Q and R are all identity matrices with proper dimensions. It is assumed that δ in (4.14) is equal to 20% of optimal LQG cost J_0. Figure 4.2 shows relative error $\|U_{k,*} - U_*\|_F / \|U_*\|_F$ of the estimated $U_{k,*}$ v.s. time k for different βs.

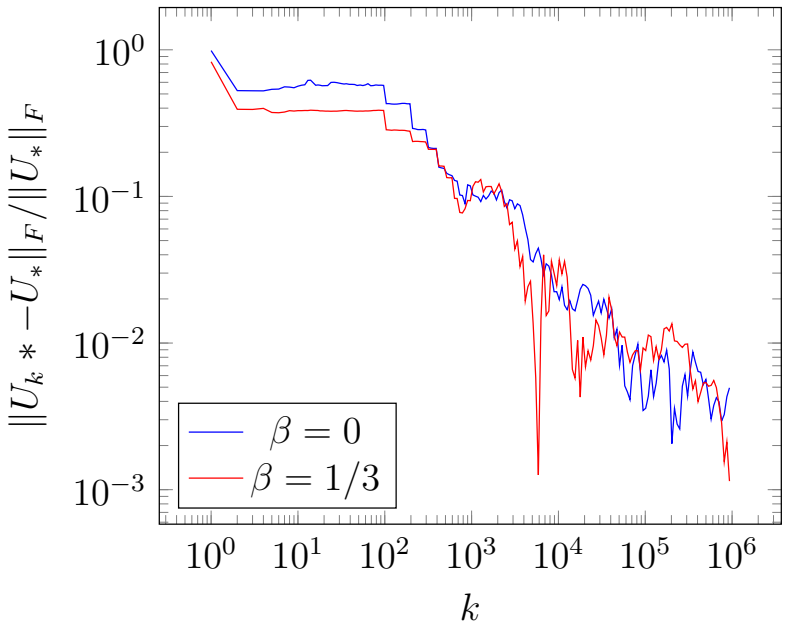

FIGURE 4.2: Relative error of $U_{k,*}$ for different β. The blue line is the relative error of $U_{k,*}$ when $\beta = 0$. The red line denotes the relative error of $U_{k,*}$ when $\beta = 1/3$.

From Figure 4.2, one can see that the estimator error converges to 0 as time k goes to infinity and the convergence approximately follows a power law. From Theorem 4.3, we know that $U_{k,*} - U_* \sim O(k^{-\gamma+\epsilon})$, where $\gamma = (1-\beta)/2$. However, from Figure 4.2, it seems that the convergence speed of the error for different β is comparable.

Next, we consider the detection performance of the online watermarking design, after an initial inference period without attack. We assume that the adversary records the sensor data from $k = 10^4 + 1$ to $k = 10^4 + 100$ and replays them from $k = 10^4 + 101$ to $k = 10^4 + 200$. Figure 4.3 shows the trajectory of the NP statistic g_k and our estimate $\hat{g}_k$ of g_k for one simulation. We can see that $\hat{g}_k$ can track g_k with a high accuracy. Moreover, both $\hat{g}_k$ and g_k are significantly larger when the system is under the replay attack ($k > 10^4 + 101$). Hence, it is not difficult to conclude that although without the knowledge of system parameters, we can still successfully estimate g_k and detect the replay attack.

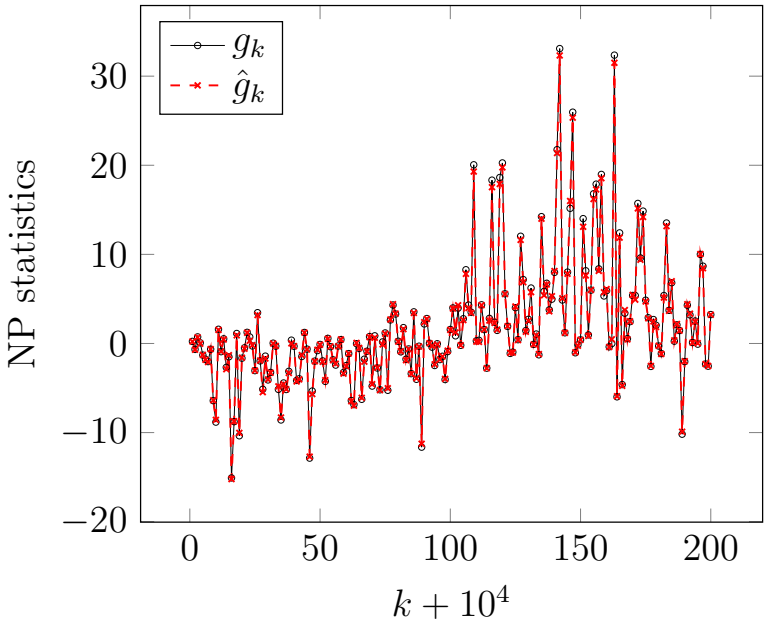

FIGURE 4.3: The NP statistics v.s. time. The black solid line with circle markers is the true NP statistics g_k, assuming full system knowledge. The red dashed line with cross markers denotes our estimated $\hat{g}_k$.

4.4.2 TEP Example

Tennessee Eastman Process (TEP) is a commonly used process control system proposed by Downs and Vogel in [124]. In this simulation, we adopt a simplified version of TEP from [125], as follows:

$$\dot{x} = Ax + Bu,$$
$$y = Cx,$$

where A, B and C are constant matrices [1].

This system simulates a multiple-input and multiple-output system of order $n = 8$ with $p = 4$ inputs and $m = 10$ outputs. We discretize the system using the control system toolbox in MATLAB, by selecting a sample time of 0.6s. We choose X in (4.12), the covariance matrices Q and R to be identity matrices with proper dimensions. It is assumed that δ in (4.14) is equal to 2% of J_0, and $\beta = 1/3$. In this simulation, we assume that we do not know the dimension of the state space, which is 8, and instead we underestimate it by assuming that A only has $\tilde{n} = 5$ distinct eigenvalues.

[1] For more details about this dynamic model, please refer to Appendix I in [125].

Figure 4.4 illustrates the relative error $\|U_{k,*}-U_*\|_F/\|U_*\|_F$ after running the system for roughly 1 week ($10^6 \times 0.6s \approx 0.992$week). Figure 4.5 illustrates the NP statistics g_k and the estimated NP statistics $\hat{g}_k$, assuming that the attacker collects the measurement from $k = 10^6 + 1$ to $k = 10^6 + 100$ and replays them to the system from $k = 10^6 + 101$ to $k = 10^6 + 200$. We can see that although we underestimate the dimensions of the system, our algorithm can still achieve a high accuracy.

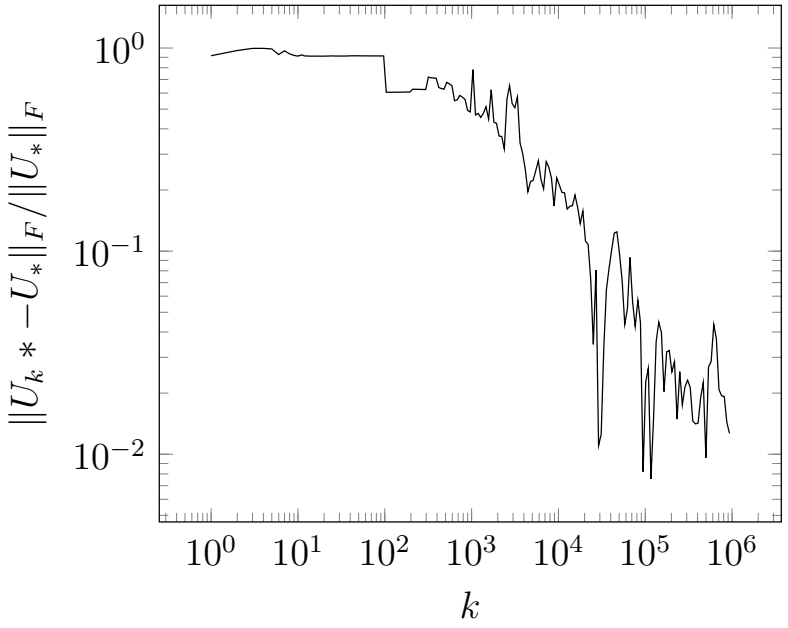

FIGURE 4.4: Relative error of $U_{k,*}$.

4.5 Conclusion

In this chapter, we proposed an online algorithm that can simultaneously generate the watermark signals and infer the necessary system parameters. It was proved that our algorithm converges to the optimal one with known system parameters and an upper bound for the almost surely convergence rate was characterized. A numerical example and TEP example were provided to verify the effectiveness of our proposed approach.

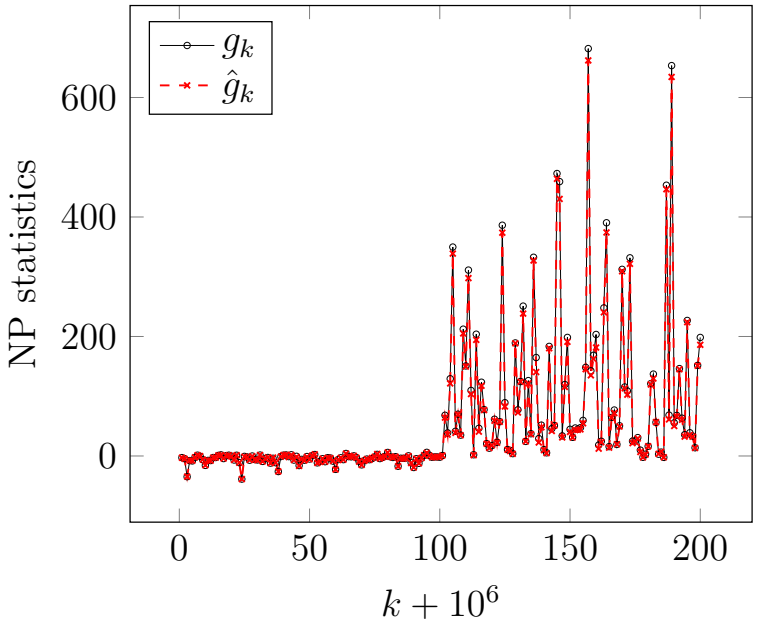

FIGURE 4.5: The NP statistics v.s. time. The black solid line with circle markers is the true NP statistics g_k, assuming full system knowledge. The red dashed line with cross markers denotes our estimated $\hat{g}_k$.

4.6 Proof of Theorem 4.3

In this section, we present the proof of Theorem 4.3. We will first provide some preliminary results and then proceed with the proof of Theorem 4.3.

Preliminary Results

For the simplicity of notations, for a random variable (vector, or matrices) x_k, we denote that $x_k \sim \mathcal{C}(\alpha)$ if for all $\epsilon > 0$, $x_k \sim O(k^{\alpha+\epsilon})$, i.e.,

$$\lim_{k\to\infty} \frac{\|x_k\|}{k^{\alpha+\epsilon}} \overset{a.s.}{=} 0.$$

Note that $x_k \sim O(k^\alpha)$ implies that $x_k \sim \mathcal{C}(\alpha)$, but the reverse is not necessarily true[2]. Some basic properties of $\mathcal{C}(\alpha)$ functions are established by the following lemma:

Lemma 4.3. *Assume that $x_k \sim \mathcal{C}(\alpha)$ and $y_k \sim \mathcal{C}(\beta)$, with $\alpha \geq \beta$, then the following statements hold:*

[2]To see a counterexample, $\log k \sim \mathcal{C}(0)$, but $\log k$ is not of the order $O(k^0)$.

1. $x_k + y_k \sim \mathcal{C}(\alpha)$, $x_k \times y_k \sim \mathcal{C}(\alpha + \beta)$, *and* $(x_k + \Delta x_k)(y_k + \Delta y_k) - x_k y_k \sim \mathcal{C}(\max\{\alpha\beta', \alpha'\beta, \alpha'\beta'\})$, *suppose that* $\Delta x_k \sim \mathcal{C}(\alpha')$ *and* $\Delta y_k \sim \mathcal{C}(\beta')$.

2. $\sum_{t=0}^{k} x_t \sim \mathcal{C}(\alpha + 1)$.

3. *Suppose* f *is differentiable at* 0 *and* $\alpha < 0$, *then* $f(x_k) - f(0) \sim \mathcal{C}(\alpha)$.

4. $s_k \sim \mathcal{C}(\alpha)$, *with* $s_k = \rho s_{k-1} + x_k$, $s_{-1} = 0$, *where* $|\rho| < 1$.

5. *Assume that* X_k *is a matrix and* $X_k \sim \mathcal{C}(\alpha)$. *Let*

$$S_k = AS_{k-1}B + X_k,\ S_{-1} = 0,$$

   where A, B *are matrices of proper dimensions. Then* $S_k \sim \mathcal{C}(\alpha)$ *if* $B^\intercal \otimes A$ *is strictly stable.*

6. $\zeta_k \sim \mathcal{C}(0)$, *where* $\{\zeta_k\}$ *is a sequence of i.i.d. Gaussian random variables, i.e.,* $\zeta_k \sim \mathcal{N}(\bar{\mu}, Z)$.

Proof. It is easy to prove the first three statements and thus we only prove the last three statements.

4. Since $x_k \sim \mathcal{C}(\alpha)$, it is easy to see that for any $\epsilon > 0$,

$$\sup_k \frac{|x_k|}{k^{\alpha+\epsilon}} = M_a(\epsilon) < \infty. \quad a.s.$$

   As a result,

$$\frac{|s_k|}{k^{\alpha+2\epsilon}} \leq \frac{1}{k^\epsilon} \sum_{i=1}^{k} \left|\rho^{i-1}\right| \left|\frac{x_{k-i}}{k^{\alpha+\epsilon}}\right| \leq \frac{1}{k^\epsilon} \frac{M_a(\epsilon)}{1 - |\rho|},$$

   which almost surely converges to 0 as k approaches infinity. Hence, $s_k \sim \mathcal{C}(\alpha)$.

5. For the last statement, note that

$$vec(S_k) = (B^\intercal \otimes A)\, vec(S_{k-1}) + vec(X_k).$$

   Hence, the argument that $S_k \sim \mathcal{C}(\alpha)$ follows the same line of proof as the fourth statement.

6. We only need to prove for the case where ζ_k follows the standard normal distribution. The high dimensional case can then be proved by checking each entry of ζ_k with proper scaling and shifting. For any $\epsilon, \phi > 0$, we have the following equality:

$$P\left(\frac{|\zeta_k|}{k^\epsilon} > \phi\right) = \sqrt{\frac{2}{\pi}} \int_{\phi k^\epsilon}^{\infty} \exp(-x^2/2) dx.$$

Suppose that k is large enough, such that $\phi k^\epsilon > 1$, then we have

$$\begin{aligned}\int_{\phi k^\epsilon}^{\infty} \exp(-x^2/2) dx &\leq \int_{\phi k^\epsilon}^{\infty} \exp(-x^2/2) \times x dx \\ &= \exp(-\phi^2 k^{2\epsilon}/2),\end{aligned}$$

and

$$\lim_{k\to\infty} k^2 \exp(-\phi^2 k^{2\epsilon}/2) = \lim_{x\to\infty} \left(\frac{2x}{\phi^2}\right)^{1/\epsilon} \exp(-x) = 0.$$

Hence, using direct comparison test for infinite series, we can prove that

$$\sum_{k=1}^{\infty} P\left(\frac{|\zeta_k|}{k^\epsilon} > \phi\right) < \infty.$$

By Borel-Cantelli Lemma, it further implies

$$\limsup_{k\to\infty} \frac{|\zeta_k|}{k^\epsilon} \leq \phi, \ a.s.$$

Since ϕ can be arbitrarily small, $\frac{\zeta_k}{k^\epsilon} \to 0$ almost surely.

□

We denote $\{\mathcal{F}_k\}$ as a filtration of sigma algebras and $\{M_k\}$ as a matrix-valued stochastic process that is adapted to the filtration $\{\mathcal{F}_k\}$. If the equality

$$\mathbb{E}\left(M_{k+1}|\mathcal{F}_k\right) = M_k$$

holds for all k, we call $\{M_k\}$ a (matrix-valued) martingale with respect to the filtration $\{\mathcal{F}_k\}$.

For the rest of this section, we shall assume that the filtration $\mathcal{F}_k$ denotes the σ-algebra generated by the random variables $\{x_{-1}, \phi_0, \cdots, \phi_k, w_0, \cdots, w_k, v_0, \cdots, v_k\}$. Now we have the following lemma to establish a strong law for matrix-valued martingale:

Lemma 4.4. *If $M_k = \Phi_0 + \Phi_1 + \cdots + \Phi_k$ is a matrix-valued martingale such that*

$$\mathbb{E}\left\|\Phi_k\right\|^2 \sim \mathcal{C}(\beta),$$

where $0 \leq \beta < 1$, then $\frac{M_k}{k}$ converges to 0 almost surely. Furthermore,

$$\frac{M_k}{k} \sim \mathcal{C}\left(\frac{\beta-1}{2}\right).$$

Proof. Denote $\Phi_{k,ij}$ ($M_{k,ij}$) as the (i,j)-th entry of the matrix Φ_k (M_k). It can be trivially proved that $\{M_{k,ij}\}$ is a scalar martingale and since[3]

$$\Phi_{k,ij}^2 \leq \left\|\Phi_k\right\|^2,$$

one has that $\mathbb{E}\Phi_{k,ij}^2 \sim \mathcal{C}(\beta)$. For simplicity, we define $\kappa \triangleq \frac{\beta+1}{2}$. We can easily verify that for any $\epsilon > 0$ and large enough i, the following equations hold:

$$i^{1-1}\left(k^{-\kappa-\epsilon}\right)^{2-2} = 1,$$

and

$$\sum_{k=i}^{\infty}\left(k^{-\kappa-\epsilon}\right)^2 k^{-1} \leq \int_{i-1}^{\infty} x^{-2\kappa-2\epsilon-1}dx = \frac{1}{2\kappa+2\epsilon}\left(i-1\right)^{-2\kappa-2\epsilon} \leq \frac{1}{\kappa}\left(i^{-\kappa-\epsilon}\right)^2.$$

The last inequality holds since $\epsilon > 0$ and for large enough i, $\frac{i-1}{i} \to 1$.

Finally, we can prove the following equality

$$\left(k^{-\kappa-\epsilon}\right)^2 \mathbb{E}\Phi_{k,ij}^2 = k^{-\beta-2\epsilon-1}\mathbb{E}\Phi_{k,ij}^2 = k^{-1-\epsilon}\frac{\mathbb{E}\Phi_{k,ij}^2}{k^{\beta+\epsilon}} \sim O(k^{-1-\epsilon}),$$

which implies that

$$\sum_{k=1}^{\infty}\left(k^{-\kappa-\epsilon}\right)^2 \mathbb{E}\Phi_{k,ij}^2 < \infty.$$

[3]The reason is the fact that $\|A\| = \sup_{\|u\|=\|v\|=1}|u^\intercal Av| \geq |e_i^\intercal Ae_j|$.

Hence, by Lemma 1 in [126], one can deduce the following equation:

$$\lim_{k\to\infty} \frac{M_{k,ij}/k}{k^{\kappa-1+\epsilon}} = \lim_{k\to\infty} k^{-\kappa-\epsilon} M_{k,ij} \overset{a.s.}{=} 0. \tag{4.34}$$

Note that (4.34) holds for all entries of the matrix M_k. Hence, $\frac{M_k}{k} \sim \mathcal{C}(\kappa-1)$, with $\kappa - 1 = \frac{\beta+1}{2} - 1 = \frac{\beta-1}{2}$. Since $\beta < 1$, $\frac{M_k}{k}$ converges to 0 almost surely. □

Now we start to prove Theorem 4.3, which requires several intermediate steps.

Boundedness of U_k

Lemma 4.5. *U_k is upper and lower bounded by:*

$$\delta \left(\left(X_{\phi\phi} - X_{\phi y} X_{yy}^{-1} X_{y\phi} \right)^{-1} + I \right) \succeq U_k \succeq \frac{\delta}{(k+1)^\beta} I. \tag{4.35}$$

Proof. The first inequality is easily proved since $U_k = U_{k,*} + \frac{\delta}{(k+1)^\beta} I$. For the second one, note that

$$\begin{aligned} \mathcal{X}_k &\succeq \left(\sum_{i=1}^{\tilde{n}} \Omega_{k,i} \right)^\intercal X_{yy} \left(\sum_{i=1}^{\tilde{n}} \Omega_{k,i} \right) + \sum_{i=1}^{\tilde{n}} \Omega_{k,i}^\intercal X_{y\phi} + X_{\phi y} \sum_{i=1}^{\tilde{n}} \Omega_{k,i} + X_{\phi\phi} \\ &\succeq X_{\phi\phi} - X_{\phi y} X_{yy}^{-1} X_{y\phi}. \end{aligned}$$

Hence, $\mathrm{tr}(U_{k,*}\mathcal{X}_k) \leq \delta$ implies that

$$\delta \mathcal{X}_k^{-1} = \delta \left(X_{\phi\phi} - X_{\phi y} X_{yy}^{-1} X_{y\phi} \right)^{-1} \succeq U_{k,*},$$

and

$$U_{k,*} + \delta I = \delta \left(\left(X_{\phi\phi} - X_{\phi y} X_{yy}^{-1} X_{y\phi} \right)^{-1} + I \right) \succeq U_k.$$

□

Convergence of $H_{k,\tau}$

Lemma 4.6. *$H_{k,\tau} - H_\tau \sim \mathcal{C}(-\gamma)$, with $\gamma = (1-\beta)/2$. In particular, $H_{k,\tau}$ converges to H_τ almost surely.*

Proof. It is easy to know that y_k and U_{k+1} are measurable with respect to $\mathcal{F}_k$. Moreover, denote $k_1, k_2 \geq 0$ as two time indices, then we can verify that

$$\mathbb{E}(\phi_{k_1}\phi_{k_2+1}^\intercal|\mathcal{F}_{k_2}) = \begin{cases} U_{k_2+1} & \text{if } k_1 = k_2+1 \\ 0 & \text{otherwise} \end{cases},$$
$$\mathbb{E}(w_{k_1}\phi_{k_2+1}^\intercal|\mathcal{F}_{k_2}) = 0,\ \mathbb{E}(v_{k_1}\phi_{k_2+1}^\intercal|\mathcal{F}_{k_2}) = 0, \tag{4.36}$$

which, combined with (4.3), implies that

$$\mathbb{E}\left(y_{k+\tau}\phi_k^\intercal U_k^{-1}|\mathcal{F}_{k-1}\right) = H_\tau. \tag{4.37}$$

Next, we compute the expectation of $\|y_{k+\tau}\phi_k^\intercal U_k^{-1}\|^2$. Note that from (4.19), $\phi_k = U_k^{1/2}\zeta_k$, where ζ_k follows the standard Gaussian distribution. Therefore,

$$\begin{aligned}\|y_{k+\tau}\phi_k^\intercal U_k^{-1}\|^2 &= \|y_{k+\tau}\phi_k^\intercal U_k^{-2}\phi_k y_{k+\tau}^\intercal\| \\ &\leq \|y_{k+\tau}\|^2\|\zeta_k\|^2\|U_k^{-1}\| \leq \delta(k+1)^\beta\|y_{k+\tau}\|^2\|\zeta_k\|^2.\end{aligned}$$

The last inequality holds due to (4.35). Furthermore, by Cauchy-Schwarz inequality, we have

$$\mathbb{E}\|y_{k+\tau}\phi_k^\intercal U_k^{-1}\|^2 \leq \delta(k+1)^\beta\sqrt{\mathbb{E}\|y_{k+\tau}\|^4}\sqrt{\mathbb{E}\|\zeta_k\|^4}.$$

It is worth noticing that $\|\zeta_k\|$ is χ-distributed with p degree of freedom and thus $\mathbb{E}\|\zeta_k\|^4 = p(p+2)$. On the other hand, we can prove that $\sup_k \mathbb{E}\|y_k\|^4$ is bounded since by (4.35), U_k is upper bounded. Hence, we prove that

$$\mathbb{E}\|y_{k+\tau}\phi_k^\intercal U_k^{-1}\|^2 \sim \mathcal{C}(\beta),$$

which further implies that

$$\begin{aligned}\mathbb{E}\|y_{k+\tau}\phi_k^\intercal U_k^{-1} - H_\tau\|^2 &\leq \mathbb{E}\left(\|y_{k+\tau}\phi_k^\intercal U_k^{-1}\| + \|H_\tau\|\right)^2 \\ &\leq \mathbb{E}\left(2\|y_{k+\tau}\phi_k^\intercal U_k^{-1}\|^2 + 2\|H_\tau\|^2\right) \sim \mathcal{C}(\beta). \end{aligned} \tag{4.38}$$

Therefore, by (4.37), we can prove that the stochastic process

$$\mathcal{S}_{\tau,i}(k+1) = \mathcal{S}_{\tau,i}(k) + \left[y_{(k+1)\tilde{\tau}+i}\phi_{k\tilde{\tau}+i+1}U_{k\tilde{\tau}+i+1}^{-1} - H_\tau\right] \tag{4.39}$$

is a matrix-valued martingale for the filtration $\mathcal{F}_{k\tilde{\tau}+i}$, where $\tilde{\tau} = \tau + 1$, and $0 \leq i \leq \tau$. Then by (4.38) and Lemma 4.4, we can obtain that

$$\frac{\mathcal{S}_{\tau,i}(k)}{k} \sim \mathcal{C}(-\gamma).$$

By the definition of $\mathcal{S}_{\tau,i}(k)$, we can see that for large enough k,

$$H_{k,\tau} - H_\tau = \sum_{i=0}^{\tau} \frac{k_i}{k} \times \frac{\mathcal{S}_{\tau,i}(k_i)}{k_i}. \tag{4.40}$$

where $k_i = \max\{t \in \mathbb{N} : t\tilde{\tau} + i \leq k\}$. Notice that $k_i \geq 0$ and $\sum k_i = k$. As a result, the estimation error of $H_{k,\tau} - H_\tau$ is a convex combination of $\mathcal{S}_{\tau,i}$s. Hence, for any $\epsilon > 0$, we have

$$\|H_{k,\tau} - H_\tau\| \leq \max_{0\leq i\leq\tau} \frac{\|\mathcal{S}_{\tau,i}(k_i)\|}{k_i} \sim O(k_i^{-\gamma+\epsilon}). \tag{4.41}$$

Note that when k is large enough, $\frac{k}{k_i} \to \tau$, which implies that $H_{k,\tau} - H_\tau \sim \mathcal{C}(-\gamma)$. The almost sure convergence can be trivially proved since $\gamma > 0$ is positive. □

Convergence of $\lambda_{k,i}$ and $\Omega_{k,i}$

Due to the convergence of $H_{k,\tau}$ to H_τ, one has that Ξ_k converges to Ξ, where

$$\Xi \triangleq \begin{bmatrix} \operatorname{tr}(\mathcal{H}_0^\intercal \mathcal{H}_0) & \cdots & \operatorname{tr}(\mathcal{H}_0^\intercal \mathcal{H}_{\tilde{n}-1}) \\ \vdots & \ddots & \vdots \\ \operatorname{tr}(\mathcal{H}_{\tilde{n}-1}^\intercal \mathcal{H}_0) & \cdots & \operatorname{tr}(\mathcal{H}_{\tilde{n}-1}^\intercal \mathcal{H}_{\tilde{n}-1}) \end{bmatrix},$$

with

$$\mathcal{H}_i \triangleq \begin{bmatrix} H_i \\ \vdots \\ H_{i+2\tilde{n}-2} \end{bmatrix}.$$

We first prove that the matrix Ξ is invertible. Suppose that there exists $\tilde{\alpha} = [\tilde{\alpha}_0, \ldots, \tilde{\alpha}_{\tilde{n}-1}]^\intercal$, such that $\Xi\tilde{\alpha} = 0$, then

$$0 = \tilde{\alpha}^\intercal \Xi \tilde{\alpha} = \left\| \sum_{i=0}^{\tilde{n}-1} \mathcal{H}_i \tilde{\alpha}_i \right\|_F^2,$$

which further implies that $CA^i\tilde{p}(A)B = 0$ for all $0 \leq i \leq 2n-2$, and $\tilde{p}(x) = \tilde{\alpha}_{\tilde{n}-1}x^{\tilde{n}-1} + \cdots + \tilde{\alpha}_0$. As a result, one knows that

$$\begin{bmatrix} C \\ \vdots \\ CA^{\tilde{n}-1} \end{bmatrix} \tilde{p}(A) \begin{bmatrix} B & \cdots & A^{\tilde{n}-1}B \end{bmatrix} = 0.$$

By the fact that (A, B) is controllable and (A, C) is observable, $\tilde{p}(A)$ must be equal to 0. However, since $p(x)$ is minimal polynomial of A, $\tilde{p}(x)$ must be constantly 0, it is proved that $\tilde{\alpha} = 0$ and Ξ is invertible.

Denote α_is as the coefficients of the minimal polynomial $p(x) = x^{\tilde{n}} + \alpha_{\tilde{n}-1}x^{\tilde{n}-1} + \cdots + \alpha_0$ of A, i.e., the monic polynomial with minimum degree. Hence, we have

$$H_{i+\tilde{n}} + \alpha_{\tilde{n}-1}H_{i+\tilde{n}-1} + \cdots + \alpha_0 H_i = CA^i p(A)B = 0. \tag{4.42}$$

As a result, we can prove that,

$$\begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_{\tilde{n}-1} \end{bmatrix} = -\Xi^{-1} \begin{bmatrix} \mathrm{tr}(\mathcal{H}_0^\intercal \mathcal{H}_{\tilde{n}}) \\ \vdots \\ \mathrm{tr}(\mathcal{H}_{\tilde{n}-1}^\intercal \mathcal{H}_{\tilde{n}}) \end{bmatrix}, \tag{4.43}$$

which, combined with $H_{k,\tau} - H_\tau \sim \mathcal{C}(-\gamma)$ and Lemma 4.3.3, proves that $\alpha_{k,i} - \alpha_k \sim \mathcal{C}(-\gamma)$. Since all the roots of the polynomial $p(x)$ are distinct, we can prove (see [127]) that $\lambda_{k,i}$s are differentiable functions of $\alpha_{k,i}$s at a neighborhood of α_i, which further proves that $\lambda_{k,i} - \lambda_i \sim \mathcal{C}(-\gamma)$.

Now we define V as

$$V \triangleq \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_{\tilde{n}} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{3\tilde{n}-2} & \lambda_2^{3\tilde{n}-2} & \cdots & \lambda_n^{3\tilde{n}-2} \end{bmatrix}.$$

Since $p(x)$ is the minimal polynomial of A and A is diagonalizable, the roots λ_is of $p(x)$ are distinct, which proves that V is of full column rank. Hence,

$$\operatorname{rank}(V \otimes I_m) = \operatorname{rank}(V) \times \operatorname{rank}(I_m) = \tilde{n}m,$$

which implies that $V \otimes I_m$ is of full column rank.

Hence, by Lemma 4.2, we have

$$\begin{bmatrix} \Omega_1 \\ \vdots \\ \Omega_{\tilde{n}} \end{bmatrix} = (V \otimes I_m)^+ \begin{bmatrix} H_0 \\ \cdots \\ H_{3\tilde{n}-2} \end{bmatrix}. \tag{4.44}$$

Hence, by Lemma 4.3.3, $\Omega_{k,i} - \Omega_i \sim \mathcal{C}(-\gamma)$.

Convergence of $\hat{\varphi}_k$, $\hat{\vartheta}_k$ and $\mathcal{W}_k$

First we prove that $\hat{\varphi}_k - \varphi_k \sim \mathcal{C}(-\gamma)$, which holds as long as $\hat{\varphi}_{k,i} - \varphi_{k,i} \sim \mathcal{C}(-\gamma)$ for all i, where

$$\varphi_{k,i} = \lambda_i \varphi_{k-1,i} + \Omega_i \phi_k,\ \varphi_{-1,i} = 0.$$

Notice that the error between $\hat{\varphi}_{k,i}$ and $\varphi_{k,i}$ satisfies the following recursive equation:

$$\varphi_{k+1,i} - \hat{\varphi}_{k+1,i} = (\lambda_i - \lambda_{k,i})\varphi_{k,i} + \lambda_{k,i}(\varphi_{k,i} - \hat{\varphi}_{k,i}) + (\Omega_i - \Omega_{k,i})\phi_k.$$

For any $\epsilon > 0$, one has

$$\frac{\|\varphi_{k+1,i} - \hat{\varphi}_{k+1,i}\|}{(k+1)^{-\gamma+2\epsilon}} \leq |\lambda_{k,i}| \frac{\|\varphi_{k,i} - \hat{\varphi}_{k,i}\|}{k^{-\gamma+2\epsilon}} + \frac{|\lambda_i - \lambda_{k,i}|}{k^{-\gamma+\epsilon}} \frac{\|\varphi_{k,i}\|}{k^{\epsilon}} + \frac{\|\Omega_i - \Omega_{k,i}\|}{k^{-\gamma+\epsilon}} \frac{\|\phi_k\|}{k^{\epsilon}}.$$

Notice that $\phi_k = U_k^{1/2}\zeta_k$. Since $\zeta_k \sim \mathcal{C}(0)$ by Lemma 4.3.6, and U_k is upper bounded by Lemma 4.5, $\phi_k \sim \mathcal{C}(0)$. Thus, $\varphi_{k,i} \sim \mathcal{C}(0)$ by Lemma 4.3.4. Furthermore, since $\lambda_{k,i} - \lambda_i \sim \mathcal{C}(-\gamma)$ and $\Omega_{k,i} - \Omega_i \sim \mathcal{C}(-\gamma)$, for any $\epsilon_1 > 0$, there exists K (possibly random), such that for any $k \geq K$, the following inequalities hold almost surely,

$$|\lambda_i - \lambda_{k,i}| \leq \epsilon_1, \ \frac{|\lambda_i - \lambda_{k,i}|}{k^{-\gamma+\epsilon}} \frac{\|\varphi_{k,i}\|}{k^{\epsilon}} + \frac{\|\Omega_i - \Omega_{k,i}\|}{k^{-\gamma+\epsilon}} \frac{\|\phi_k\|}{k^{\epsilon}} \leq \epsilon_1.$$

Hence, for $k \geq K$, we have

$$\frac{\|\varphi_{k+1,i} - \hat{\varphi}_{k+1,i}\|}{(k+1)^{-\gamma+2\epsilon}} \leq (|\rho| + \epsilon_1) \times \frac{\|\varphi_{k,i} - \hat{\varphi}_{k,i}\|}{k^{-\gamma+2\epsilon}} + \epsilon_1. \ a.s.$$

Since $|\rho| < 1$, we can choose ϵ_1 small enough such that $|\rho| + \epsilon_1 < 1$, as a result,

$$\limsup_{k\to\infty} \frac{\|\varphi_{k,i} - \hat{\varphi}_{k,i}\|}{k^{-\gamma+2\epsilon}} \leq \frac{\epsilon_1}{1 - |\rho| - \epsilon_1}. \ a.s.$$

Hence, $\|\varphi_{k,i} - \hat{\varphi}_{k,i}\|/k^{-\gamma+3\epsilon} \overset{a.s.}{\to} 0$, which proves that $\varphi_{k,i} - \hat{\varphi}_{k,i} \sim \mathcal{C}(-\gamma)$.

Lemma 4.7.

$$\frac{1}{k+1}\sum_{t=0}^{k} \vartheta_t \vartheta_t^{\intercal} - \mathcal{W} \sim \mathcal{C}(-0.5), \tag{4.45}$$

where $\vartheta_k \triangleq \sum_{t=0}^{k} CA^t w_{k-t} + v_k + CA^{k+1}x_{-1}$.

Proof. Define function $\mathcal{A} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$, such that for any symmetric matrix $X \in \mathbb{S}^{n\times n}$, the following equality holds,

$$\mathcal{A}(X) = X + AXA^{\intercal} + A^2XA^{2\intercal} + \cdots,$$

For non-symmetric $X \in \mathbb{R}^{n\times n}$, let us define

$$\mathcal{A}(X) = \mathcal{A}\left(\frac{X + X^{\intercal}}{2}\right).$$

It can be proved that

$$\mathcal{A}(X) - A^k\mathcal{A}(X)A^{kT} = \sum_{i=0}^{k-1} A^i \frac{X + X^{\intercal}}{2} A^{i\intercal}.$$

For the simplicity of notations, we define $w_{-1} = x_{-1}$. By mathematical induction, $\sum_{t=0}^{k} \vartheta_t \vartheta_t^\intercal$ can be represented as

$$\sum_{t=0}^{k} \vartheta_t \vartheta_t^\intercal = \mathcal{M}_k - CA\mathcal{N}_k A^\intercal C^\intercal, \tag{4.46}$$

where

$$\mathcal{M}_k = \mathcal{M}_{k-1} + \Pi_k \tag{4.47}$$

$$\mathcal{N}_k = A\mathcal{N}_{k-1}A^\intercal + 2\mathcal{A}\left(\left(\sum_{t=-1}^{k} A^{k-t} w_t\right) w_k^\intercal\right) - \mathcal{A}(w_k w_k^\intercal). \tag{4.48}$$

with

$$\begin{aligned}\Pi_k = v_k v_k^\intercal + v_k \left(\sum_{t=-1}^{k} CA^{k-t} w_t\right)^\intercal + \left(\sum_{t=-1}^{k} CA^{k-t} w_t\right) v_k^\intercal \\ + 2C\mathcal{A}\left(\left(\sum_{t=-1}^{k} A^{k-t} w_t\right) w_k^\intercal\right) C^\intercal - C\mathcal{A}\left(w_k w_k^\intercal\right) C^\intercal,\end{aligned}$$

and initial condition

$$\mathcal{N}_{-1} = \mathcal{A}\left(x_{-1} x_{-1}^\intercal\right), \ \mathcal{M}_{-1} = CA\mathcal{N}_{-1}A^\intercal C^\intercal. \tag{4.49}$$

Then we can prove that

$$\mathbb{E}(\Pi_k | \mathcal{F}_{k-1}) = \mathcal{W}, \ \mathbb{E}\|\Pi_k - \mathcal{W}\|^2 \sim O(0).$$

Therefore, $\mathcal{M}_k - k\mathcal{W}$ is a martingale and $\mathcal{M}_k/k - \mathcal{W} \sim \mathcal{C}(-0.5)$ by Lemma 4.4. On the other hand, for $\mathcal{N}_k$, since $A \otimes A$ is stable, $\mathcal{N}_k \sim \mathcal{C}(0)$ by Lemma 4.3, which proves that

$$\frac{1}{k+1} \sum_{t=0}^{k} \vartheta_t \vartheta_t^\intercal - \mathcal{W} \sim \mathcal{C}(-0.5).$$

□

Now one can rewrite $\mathcal{W}_k - \mathcal{W}$ as

$$\begin{aligned}
&\mathcal{W}_k - \mathcal{W} \\
&= \left(\frac{1}{k+1} \sum_{t=0}^{k} \vartheta_t \vartheta_t^\intercal - \mathcal{W} \right) - \frac{1}{k+1} \sum_{t=0}^{k} (\vartheta_t (\hat{\varphi}_t - \varphi_t)^\intercal + (\hat{\varphi}_t - \varphi_t) \vartheta_t^\intercal) \\
&\quad + \frac{1}{k+1} \sum_{t=0}^{k} (\hat{\varphi}_t - \varphi_t)(\hat{\varphi}_t - \varphi_t)^\intercal.
\end{aligned}$$

Hence, by Lemma 4.3, $\mathcal{W}_k - \mathcal{W} \sim \mathcal{C}(\max\{-0.5, -\gamma, -2\gamma\}) = \mathcal{C}(-\gamma)$.

Convergence of the Rest

By Lemma 4.3.3, one can prove that $\mathcal{P}_k - \mathcal{P}$, $\mathcal{X}_k - \mathcal{X}$ are all of the class $\mathcal{C}(-\gamma)$, as they are differentiable functions of $\lambda_{k,i}$, $\Omega_{k,i}$ and $\mathcal{W}_k$. Hence, $U_{k,*} - U_* \sim \mathcal{C}(-\gamma)$ since $U_{k,*}$ is a differentiable function of $\mathcal{P}_k$ and $\mathcal{X}_k$ at a neighborhood of $\mathcal{P}$ and $\mathcal{X}$ (please see [127]).

Hence, since $\mathcal{U}_k$ is a differentiable function of $\lambda_{k,i}$, $\Omega_{k,i}$ and $U_{k,*}$, we can prove that $\mathcal{U}_k - \mathcal{U} \sim \mathcal{C}(-\gamma)$.

Finally, one can prove that $\hat{g}_k - g_k \sim \mathcal{C}(-\gamma)$ by Lemma 4.3.1.

Chapter 5

Reinforcement Learning Based Approach for Flip Attack Detection

The last chapter studied an active detection scheme, which enables the detection of replay attacks. However, active detection does not apply under some scenarios and we need to make a decision without modifying input signals. It is more like passive detection, which is also vital to the security of CPS. In this chapter, we focus on this detection approach.

This chapter considers the detection problem of flip attacks to sensor network systems where the adversary flips the distribution of sensor measurements of a binary state. The detector decides to continue taking observations or to stop based on the sensor measurements, and the goal is to have the flip attack detected as fast as possible while trying to avoid terminating the measurements when there is no attack. The detection problem can be modeled as a partially observable Markov decision process (POMDP) by assuming an attack probability, with the dynamics of the hidden states of the POMDP characterized by a stochastic shortest path (SSP) problem. We prove that the optimal policy of the SSP solely depends on the transition costs and is independent of the assumed possibility. By using a fixed-length window and suitable feature function of the measurements, a Markov decision process (MDP) is used to approximate the behavior of the POMDP. The optimal solution of the approximated MDP can then be solved by reinforcement

learning (RL). Finally, we provide numerical evaluations to verify the effectiveness of the proposed approach. It is shown via numerical evaluation that the obtained detector is robust to the change of the assumed attack possibility, and has comparable performance to classical quickest change detection (QCD) method with weaker assumptions.

The rest of this chapter is organized as follows. Section 5.1 introduces the problem formulation, including the overall system architecture of the detection problem as well as problem of interest. The POMDP applied to model the flip attack is presented in Section 5.2. The RL approach applied is detailed in Section 5.3. Section 5.4 gives numerical evaluation and shows the effectiveness of the proposed technique. Concluding remarks are given in Section 5.5.

5.1 Problem Formulation

We consider a flip attack detection problem with the system diagram shown in Figure 5.1.

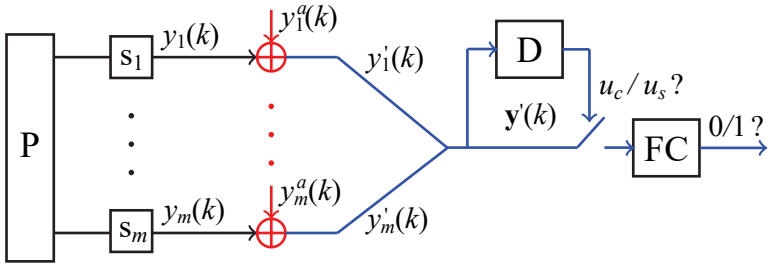

FIGURE 5.1: The system diagram.

In the above figure, a plant P possesses a binary state $\theta \in \{0, 1\}$, measured by sensors $s_1, \cdots, s_m$, whose indices form set $\mathcal{S}$. Let us define the measurement from all m sensors at time k as:

$$\mathbf{y}(k) \triangleq [y_1(k) \quad y_2(k) \quad \cdots \quad y_m(k)]^\intercal \in \mathbb{R}^m. \tag{5.1}$$

All sensors' measurements $\{y_i(k)\}_{j\in\mathcal{S}}$ are independently and identically distributed (i.i.d.) under normal operation. For any Borel-measurable set $B \subset \mathbb{R}$, the probability that $y_i(k) \in B$ is $\kappa_0(B)$ when $\theta = 0$ and equals $\kappa_1(B)$ when $\theta = 1$. Given the measurements of the sensors and the knowledge of the distributions κ_0 and κ_1, a

fusion centre FC is designed to infer the state θ. Naturally, it is assumed that the induced measures κ_0 and κ_1 are different and are absolutely continuous.

However, due to the presence of a malicious adversary that tries to compromise the performance of the fusion centre by attacking some sensors, the fusion centre receives the following manipulated measurements at time k:

$$\mathbf{y}'(k) = \mathbf{y}(k) + \mathbf{y}^a(k), \tag{5.2}$$

where $\mathbf{y}^a(k) \in \mathbb{R}^m$ is the bias vector injected by the attacker at time k. Therefore, a detection unit D forks the measurements $\mathbf{y}'(k)$ and is designed to detect possible attack given received measurements $\mathbf{y}'(k)$ and the distributions κ_0 and κ_1, without knowing the true state θ. The detector can command the fusion centre to continue to process the received measurements or to stop via the switch signal u_c (continue) and u_s (stop).

Regarding the attack type, we make following assumptions. The type of attack studied here is typical in the hypothesis testing and can be found in [89, 91].

Assumption 5.1 (Attacker's knowledge)**.** The attacker has the knowledge of the probability of measures κ_0 and κ_1 and the true state θ.

Assumption 5.2 (l-sparse attack)**.** There exists an index set $\mathcal{L} \subset \mathcal{S} \triangleq \{1, 2, \ldots, m\}$ with $|\mathcal{L}| \leq l$, where $l \leq \lfloor \frac{m}{2} \rfloor$, such that $\cup_{k=1}^{\infty} \text{supp}\{\mathbf{y}^a(k)\} = \mathcal{L}$. Besides, the system knows the number l, but it does not know the set $\mathcal{L}$.

Remark 5.1. When m is large, typically we have $l \ll \frac{m}{2}$.

Assumption 5.3. The compromised sensors are fixed during the whole attack period and the attack will not stop until it is detected or the detector stops detection.

For the type of attacks specified by the assumptions above, the attacker may design the injected signals via various strategies and here we focus on the detection of the flip attack where the attacker flips the distribution of the corrupted sensors' measurements to confuse the fusion centre [91]. This strategy has been shown to be optimal from attacker's perspective when exactly l sensors are compromised. The strategy is when $\theta = 0$, the probability measure generated by $y'_j(k)_{j\in\mathcal{L}}$ is κ_1 and when $\theta = 1$, it is κ_0. Correspondingly, the attacked signal $y^a_j(k)_{j\in\mathcal{L}}$ is derived

as follows:

$$y_j^a(k) = \begin{cases} y_j'(k) - y_j(k) & \text{if } j \in \mathcal{L}, \\ 0 & \text{if } j \notin \mathcal{L}. \end{cases} \tag{5.3}$$

In this chapter, we focus on the design of flip attack detector to protect a sensor system aimed to estimate a binary state. Due to the existence of the attack, there are two operation situations: "normal" and "abnormal". When an attacker launches the attack, the distribution of the compromised sensors' measurements flips to the distribution under the opposite binary state. The detector decides on whether to stop and declare that there is an attack or to continue receiving the observations based on the sensor measurements. The desired behavior should command "continue" if state is "normal" and "stop" if otherwise. The problem of interest is to design the detector given the scope specified here.

5.2 Modeling of Flip Attack via POMDP

In this section, we introduce the POMDP applied to model the flip attack. As opposed to many successful cases reported via the RL methods where there are already simulators available, here the POMDP is needed to serve as the simulator which is part of the challenge. This includes crafting the transition probabilities and the transition costs of the underlying MDP, and the conditional observation probabilities of obtaining various observations. The optimal policy of the designed MDP will give u_c when the state is "normal" and u_s otherwise. We will show that this solely depends on the transition costs and is irrelevant to the assumed attack probability. In addition, we will give some rationales on how the conditional observation probabilities are defined. The POMDP derived here will serve as the simulator used to train the detector.

5.2.1 Modeling the Dynamics of Hidden States as an SSP

The states of the underlying MDP include "normal" state, "abnormal" state, and termination, where the termination is an absorbing state. This type of problems is

widely known as stochastic shortest path (SSP) problem. We will use the semicontractive models introduced in [128] and show that the optimal policy solely depends on the transition costs.

We denote by I the state space of the underlying SSP problem and by U the control space. The state space has two elements 1 and 2, standing for "normal" state and "abnormal" state, respectively, and we will use i or i' to represent the unspecified states in I. The admissible control options are to continue and to stop, denoted as u_c and u_s, respectively, the same for all $i \in I$, and we will use u to represent the unspecified control. We denote by $p_{ii'}(u)$ the transition probability from i to i' under control u and denote by $g(\,\cdot\,, u, \,\cdot\,)$ a deterministic nonnegative function which returns the transition cost. The transition graph is shown in Figure 5.2. It is clear that for the detection problem, we have

$$p_{1i}(u_c) > 0,\ p_{22}(u_c) = 1,\ p_{it}(u_c) = 0,\ p_{it}(u_s) = 1, \tag{5.4}$$

where $t \notin I$ denotes the terminal state, and we require

$$p_{tt}(u) = 1,\ g(t, u, t) = 0,\ \forall u \in U. \tag{5.5}$$

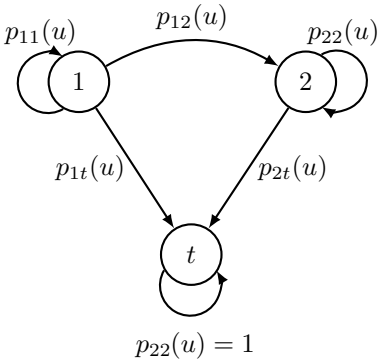

FIGURE 5.2: The transition graph of the SSP model. There are 2 states, plus the termination state t.

A function $\mu : I \to U$ is named as a policy and the set of all policies is denoted by $\mathcal{M}$. It is easy to see that we have $|\mathcal{M}| = 4$. In the context of SSP, a policy is proper if under such a policy, the state is guaranteed to reach t regardless of the initial state; otherwise, it is improper. We denote by $\mathcal{E}(I)$ the set of functions $J : I \to \mathbb{R}^*$ where $\mathbb{R}^* = \mathbb{R} \cup \{\infty, -\infty\}$. We use the mapping $H : I \times U \times \mathcal{E}(I) \to \mathbb{R}^*$

to define the SSP problem as

$$H(i,u,J) \stackrel{\triangle}{=} p_{it}(u)g(i,u,t) + \sum_{i' \in I} p_{ii'}(u)\left(g(i,u,i') + J(i')\right).$$

Then the mappings $T_\mu : \mathcal{E}(I) \to \mathcal{E}(I)$ for every $\mu \in \mathcal{M}$, and $T : \mathcal{E}(I) \to \mathcal{E}(I)$ can be defined in turn as

$$T_\mu J(i) \stackrel{\triangle}{=} H(i,\mu(i),J),\ TJ(i) \stackrel{\triangle}{=} \min_{\mu \in \mathcal{M}} T_\mu J(i),\ \forall i \in I.$$

In addition, the superscript of the operators means composition, viz., $(T^2 J)(i) \stackrel{\triangle}{=} \big(T(TJ)\big)(i)$. Besides, we denote by $J_\mu \in \mathcal{E}(I)$ the cost function of μ defined pointwise by

$$J_\mu(i) \stackrel{\triangle}{=} \limsup_{k \to \infty} (T_\mu^k \bar{J})(i),\ \forall i \in I,$$

where $\bar{J}(i) = 0$ for all i.

Naturally, a desired policy μ_d would be that $\mu_d(1) = u_c$ and $\mu_d(2) = u_s$. Then a plausible choice of stage costs is

$$\begin{aligned} &g(1,u_c,1) = 0,\ g(1,u_c,2) > 0,\ g(2,u_c,2) > 0, \\ &g(1,u_s,t) > 0,\ g(2,u_s,t) = 0. \end{aligned} \tag{5.6}$$

The costs of other situations need not be defined as they have zero transition probability. With the specified problem data, the fundamental questions required to be answered are: 1) is there a fixed point of the corresponding Bellman equation; 2) If so, is the fixed point a cost function of certain policy μ^*; 3) what are the conditions needed in order to have $\mu_d = \mu^*$. We recall the following useful lemma for the answers.

Lemma 5.1 (Proposition 2, [129]). *For any SSP problem defined in the form of the mapping $H(\,\cdot\,,\,\cdot\,,\,\cdot\,)$ with both I and U being finite, assume that there exists at least one proper policy, and the cost functions of all improper policies have value infinity for at least one state. Then there exists $J^* : I \to \mathbb{R}$ such that*

$$J^*(i) = (TJ^*)(i),\ J^*(i) = \min_{\mu \in \mathcal{M}} J_\mu(i), \forall i \in I, \tag{5.7}$$

with the optimal policy that attains the value of J^* denoted as $\mu^* \in \mathcal{M}$. In addition, for every proper μ, it holds that

$$J_\mu(i) = (T_\mu J_\mu)(i),\ \forall i \in I, \tag{5.8}$$

which needs not to be true for the improper ones.

Aided by the above result, we have the following theorem.

Theorem 5.1. *The SSP problem defined by (5.4), (5.5), and (5.6) has the following property: a) the corresponding Bellman equations fulfill (5.7), (5.8), independent of the choice of $p_{12}(u_c)$; b) the attained optimal policy μ^* is the desired policy μ_d if $g(1, u_s, t) > g(1, u_c, 2)$, regardless of the choice of $p_{12}(u_c)$.*

Proof. For part a), one can verify that $\mu(i) = u_s$ is a proper policy and that all improper policies have cost function infinity for state 2. Therefore, a) follows from Lemma 5.1, which does not rely on the specific value of $p_{12}(u_c)$. For b), since J^* is the fixed point of T, we have

$$J^*(1) = \min\Big\{g(1, u_s, t),\ p_{11}(u_c)\big(0 + J^*(1)\big) + p_{12}(u_c)\big(g(1, u_c, 2) + J^*(2)\big)\Big\}, \tag{5.9}$$

$$J^*(2) = \min\Big\{0,\ g(2, u_c, 2) + J^*(2)\Big\}. \tag{5.10}$$

From (5.10), we have $J^*(2) = 0$ and $\mu^*(2) = \mu_d(2) = u_s$. To have $\mu^*(1) = \mu_d(1) = u_c$, one can see that it is required to have $g(1, u_s, t) > g(1, u_c, 2)$. □

From Theorem 5.1, by setting $g(1, u_s, t) > g(1, u_c, 2)$, the SSP can capture the assumed characteristics of the flip attack.

5.2.2 Design of the Conditional Observation Probabilities

The SSP introduced in Section 5.2.1 is used to model the dynamics of hidden states. However, the true state $i = 1$ or 2 is not accessible to the detector. Instead, a measurement $\mathbf{y}'(k) \in \mathbb{R}^m$ defined in (5.2) is available, which makes the environment partially observable and in turn the problem a POMDP. Needless to say, the measurement is conditioned on the state and control of SSP, viz., i and u. However, it is also conditioned on the binary state θ and the compromised sensor

index set $\mathcal{L}$. If one fixes θ and $\mathcal{L}$, then the conditional observation probabilities are fully specified by the attack type and strategy. Under Assumptions 5.2 and 5.3, one can verify that there are in total $|\mathcal{I}|$ different POMDPs induced by the same SSP where $\mathcal{I}$ is an index set defined as

$$\mathcal{I} \triangleq \left\{1,\, 2,\, \ldots,\, 2\sum_{\ell=1}^{l} \frac{m!}{(m-\ell)!}\right\}. \tag{5.11}$$

For every $\ell \in \mathcal{I}$, the remote state θ and compromised sensors $\mathcal{L}$ are fixed and we will name its corresponding POMDP as ℓ-POMDP. To have all those cases covered by one POMDP, we introduce a probability distribution η over $\mathcal{I}$, viz., $\eta(\ell) \geq 0$ $\forall \ell \in \mathcal{I}$ and $\sum_{\ell \in \mathcal{I}} \eta(\ell) = 1$. Such a distribution indicates how likely one particular case ℓ occurs. For example, it is more likely to have one sensor get attacked than to have two, and this is reflected by η where the distribution on cases fewer sensors under attack is higher than those with more. Given the distribution η, when the hidden state is 2, the probability of certain observation $\mathbf{y}'$ is given by the sum of products between the probability of any case ℓ specified by η, and the probability that $\mathbf{y}'$ is observed in ℓ-POMDP.

5.3 RL Approach to the Detection Problem

In principle, the POMDP used to model the flip attack can be solved by introducing the belief states and solving in turn the induced MDP with belief states as its states. However, such an approach relies explicitly on the specific values of transition probabilities in the SSP and the assumed case distribution η. To obtain a detector that is robust to the change of those values, we apply a RL approach to solve the problem. We will show here how the learning problem is formulated, and sketch the procedure to train the detector.

5.3.1 The Target MDP Learned by RL

One of the major challenges of POMDPs is that the Markov property is lost. One well-known example that illustrates the cause is a 1-D gridworld given by [130], and we use a simplified version of it, given in Figure 5.3. In this case, if the percepted information is the number and positions of walls around a state, then the states 1

and 3 are the same. The agent can not differentiate these two states. However, if the system equips a memory that stores a fixed length of past states and actions, it can then distinguish those states.

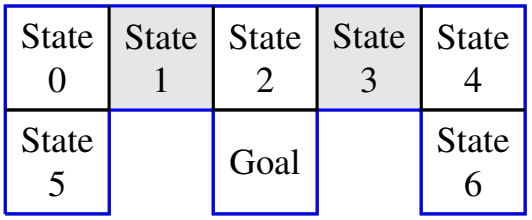

FIGURE 5.3: Aliasing gridworld. The walls around a state are highlighted in blue.

Motivated by this example, we denote an observation at time k as a stored measurement with length $w > 1$, which is given by

$$\mathbf{o}_k \triangleq [\mathbf{y}^\intercal(k-w+1) \ \cdots \ \mathbf{y}^\intercal(k-1) \ \mathbf{y}^\intercal(k)]^\intercal \in \mathbb{R}^{mw}.$$

Here the control need not to be recorded as the only reasonable control is u_c.

Denote by O_ℓ the set of all possible observation $\mathbf{o} \in \mathbb{R}^{mw}$ when the POMDP index is ℓ and define O as $\cup_{\ell\in\mathcal{I}} O_\ell$. Assume that there exists a feature function $\phi : O \to X$ where $|X| < \infty$ and $X \subset \mathbb{R}^n$, and denote $X_\ell \triangleq \phi(O_\ell)$. Here the fundamental assumption we use is that for every $\ell \in \mathcal{I}$, there is a MDP characterized by a mapping $\tilde{H}_\ell : X_\ell \times U \times \mathcal{E}(X_\ell) \to \mathbb{R}^*$ given by

$$\tilde{H}_\ell(\mathbf{x}, u, V_\ell) = \tilde{p}_{\ell,\mathbf{x}t}(u) r_\ell(\mathbf{x}, u, t) + \sum_{\mathbf{z}\in X_\ell} \tilde{p}_{\ell,\mathbf{x}\mathbf{z}}(u)\big(r_\ell(\mathbf{x}, u, \mathbf{z}) + V_\ell(\mathbf{z})\big),$$

where $\tilde{p}_{\ell,\cdot\,\cdot}(\,\cdot\,)$, $r_\ell(\,\cdot\,)$, and $V_\ell(\,\cdot\,)$ are defined accordingly, such that the mean cost of ℓ-POMDP is close to the mean cost of the MDP defined by the above operator after feature transformation. Then the POMDP defining the flip attack can be approximated by $\tilde{H} : X \times U \times \mathcal{E}(X) \to \mathbb{R}^*$ given by

$$\tilde{H}(\mathbf{x}, u, V) = \frac{\sum_{\ell\in\mathcal{I}_\mathbf{x}} \Big(\eta(\ell)\tilde{H}_\ell(\mathbf{x}, u, V_{|X_\ell})\Big)}{\sum_{\ell\in\mathcal{I}_\mathbf{x}} \eta(\ell)}, \tag{5.12}$$

where $\mathcal{I}_\mathbf{x} \triangleq \{\ell \in \mathcal{I} : \mathbb{1}_{X_\ell}(\mathbf{x}) = 1\}$, and $V_{|X_\ell}$ is the restriction of V on X_ℓ. The MDP defined by $\tilde{H}$ is the target MDP to be learned by the training algorithm, and

can be in turn used to approximate the original multiple POMDPs with $\eta(\ell)$ as a weight in case ℓ.

5.3.2 Training the Detector

With above formulation, we obtain a standard RL problem with X as state space and U as control space. Such a problem can be solved by many different RL methods and we use Q-learning [131] as an example. The pseudocode is given in Algorithm 2. To address the exploration and exploitation trade-off, a distribution γ for initial states of SSP is specified. In addition, we denote by $i \sim \gamma(I)$ a sample from distribution γ defined on I, and similar notation is used for $\ell \sim \eta(\mathcal{I})$. We denote by $j \sim \text{SSP}(i, u)$ the sampled next state of SSP given current state and control pair (i, u), and $o \sim \ell\text{-POMDP}(i, u)$ the sampled observation of ℓ-POMDP given current state and control pair (i, u). With a slight abuse of notation, we denote by $u \sim \min_v Q_\varepsilon(\mathbf{x}, v)$ the sampled control from a greedy exploration ε policy given the current $Q(\,\cdot\,,\,\cdot\,)$, the current state $\mathbf{x}$ and exploration rate ε. The complete implementation process is shown in Figure 5.4. We will provide more details in Section 5.4.

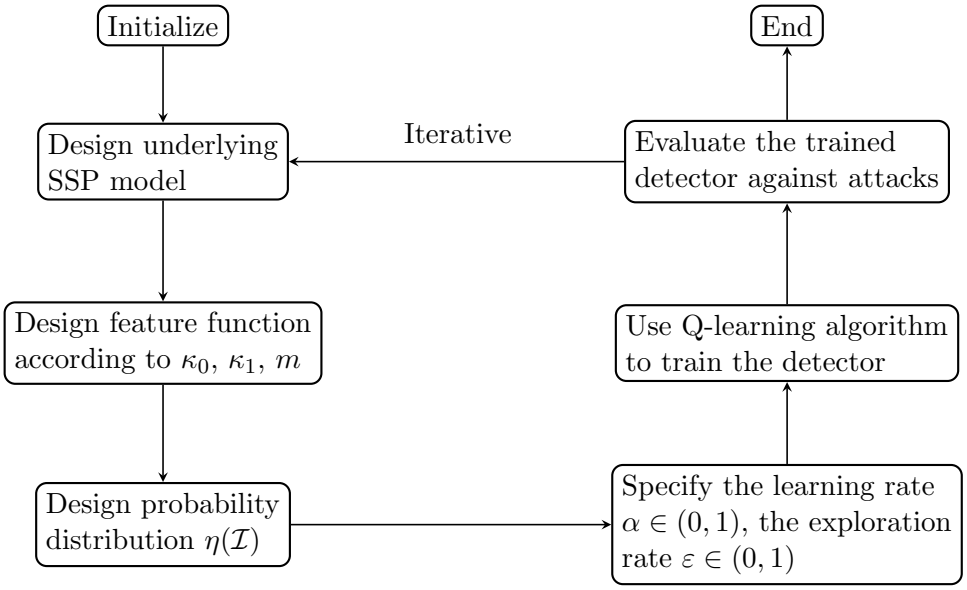

FIGURE 5.4: The implementation process.

Algorithm 2 Q-learning for detector training

Require: Problem data, the number of episodes N, and the learning rate $\alpha \in (0, 1]$, the exploration rate $\varepsilon \in (0, 1)$, initial state distribution γ.
Ensure: The optimal state-action value function $Q(\mathbf{x}, u)$

Initialize $Q(\mathbf{x}, u)$, $\forall \mathbf{x} \in X$, $u \in U$. $i \sim \gamma(I)$, $u \leftarrow u_c$.
for each $n \in N$ **do**
  $\ell \sim \eta(\mathcal{I})$, $\mathbf{o} \sim \ell$-POMDP(i, u).
  while $i \neq t$ **do**
    $\mathbf{x} \leftarrow \phi(\mathbf{o})$, $u \sim \min_v Q_\varepsilon(\mathbf{x}, v)$.
    if $u = u_s$ **then**
      $c \leftarrow g(i, u, t)$, $Q(\mathbf{x}, u) \leftarrow (1 - \alpha)Q(\mathbf{x}, u) + \alpha c$
    else
      $i' \sim \text{SSP}(i, u)$, $\mathbf{o}' \sim \ell$-POMDP(i', u).
      $\mathbf{x}' \leftarrow \phi(\mathbf{o}')$, $c \leftarrow g(i, u, i')$, $u' \in \min_u Q(\mathbf{x}', u)$,
      $Q(\mathbf{x}, u) \leftarrow (1 - \alpha)Q(\mathbf{x}, u) + \alpha\,[c + Q(\mathbf{x}', u')]$,
      $i \leftarrow i'$, $\mathbf{o} \leftarrow \mathbf{o}'$.
    end if
  end while
end for

5.4 Simulation

In this section, the numerical evaluation of the proposed method is presented. For some given sets of probability measures κ_θ, $\theta = 0, 1$, the parameters of SSP used for modeling the attack, the fixed-length window, the feature functions, and parameters used in Q-learning are given. Via tuning the cost defined in the SSP model, the obtained detector exhibits that there is a trade-off between detecting attack early and giving false alarm. In addition, the detector, without knowing the binary state θ, has a comparable performance with the classical QCD algorithm that equips the true value θ.

5.4.1 Simulation Setup and Modelling Parameters

We consider a flip attack detection problem with probability measures induced by the sensor measurements governed by normal distributions. The density functions under normal operation is given by $\mathcal{N}(\nu_\theta, \sigma_\theta^2)$ where $\theta = 0, 1$, $\nu_0 = -\nu_1$, and $\sigma_0 = \sigma_1 = 1$. Various values of ν_θ have been tested. The remote state is measured by $m = 5$ sensors with index set $\mathcal{S} = \{1, \ldots, 5\}$. Due to Assumption 5.2, it is assumed that there can be at most $l = 2$ sensors under attack. Figure 5.5 illustrates

the measurements of all the sensors in one trail where $\theta = 0$, $\nu_0 = 0.7$, and the attack occurred at $k = 50$ on sensor 1.

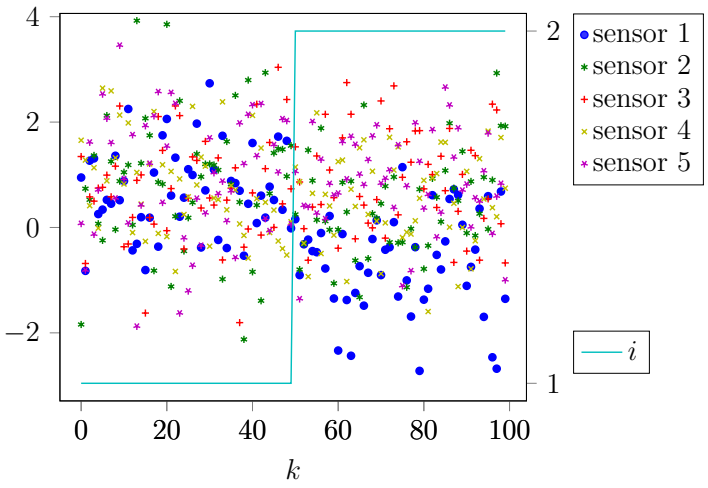

FIGURE 5.5: Sensor measurements in one trail where $\theta = 0$, $\nu_0 = 0.7$, and the attack occurs at $k = 50$ on sensor 1.

The aforementioned problem is modelled by SSP, as introduced in Section 5.2.1, with transition probability given by (5.4) where

$$p_{11}(u_c) = 0.95,\ p_{12}(u_c) = 0.05.$$

The transition probability from normal operation to under attack is assumed and kept fixed throughout the training of the detector. It will be shown later in the evaluation results that the trained detector is robust to this assumed transition probability. The cost per stage is given by (5.6), with

$$g(1, u_s, t) = 1,\ g(1, u_c, 2) = g(2, u_c, 2) \in (0, 1).$$

Recall that it is required to have $g(1, u_c, 2) < g(1, u_s, t)$ in order to make μ_d, the desired policy, the only policy whose cost function is the fixed point of the Bellman equation (5.7) and at the same time an optimal policy. The specific value of $g(1, u_c, 2)$ serves as the tuning parameter of the model and $g(2, u_c, 2)$ is set to be the same value in order to reduce the number of tuning parameter.

The size of the fixed-length window involves a trade-off between encoding more information and demanding more memory and consequently causing bigger dimension of state space. In this example, window size between $w = 2$ to 5 are explored.

This is by no means guaranteed to be optimal and it may be tuned to get better results.

Designing the feature function of the observation is typically challenging and requires domain specific knowledge. Here we use arithmetic means as features. With a better crafted feature functions, a performance improvement may be expected. The feature used here at time k is defined as $\mathbf{x}_k \triangleq [x_1(k)\ x_2(k) \cdots\ x_m(k)]^\intercal$, where for $s \in \mathcal{S}$, and $\mathcal{W} = \{0, \ldots, w-1\}$,

$$x_s(k) = \sum_{\ell \in \mathcal{R}} \left(\ell \times \mathbb{1}_{R_\ell} \left(\frac{\sum_{j \in \mathcal{W}} y_s(k-j)}{w} \right) \right), \tag{5.13}$$

with $\{R_\ell\}_{\ell \in \mathcal{R}}$ as a finite partition of the $\mathbb{R}$ whose index set is $\mathcal{R}$. The partition serves as a tuning parameter.

Apart from the above setting, a distribution $\eta(\mathcal{I})$ is specified, which weighs the chance of different sensors getting attacked. Here, $\eta(\mathcal{I})$ is defined such that the chance of one sensor under attack is 80% and of two as 20% while the chances of θ being 0 and 1 are equal. In addition, the initial state i of SSP for each episode is given by a Bernoulli distribution, with probability 0.3 that $i = 1$ at the beginning of each episode and $i = 2$ otherwise.

5.4.2 Training Setup and Performance Criteria

The model introduced above forms a standard RL problem, which in principle can be solved by many RL methods, and we apply Q-learning as an example. Q-learning algorithm is applied to obtain a MDP in form of (5.12) that approximates the behavior of POMDP introduced in Section 5.2. The learning rate α is set to be constant and different values of α have also been explored. The learning rates that fulfill the Robbins-Monro conditions, which is required to have the convergent behaviors, have also been tested. It results in no clear improvement and therefore is not presented here. The number of training episodes N varies between 300 thousands to 1 million depending on the size of the state space. Once the training process is complete, a table of Q values with data size less than 0.5 MB is obtained.

To test the obtained detector, Monte Carlo simulations with 20 thousands trials are used, half of which are always in normal operation, while the other half always

under attack. They corresponds to cases where the transition probability $p_{12}(u_c)$ of SSP is set to be 0 and 1 respectively. Those transition probabilities are significantly different from the model used to train the detector, in order to test its sensitivity to the assumed attack probability. To evaluate the detector performance, the false alarm rate (FAR) and the average detection delay (ADD) are used as criteria, where FAR and ADD are computed as follows:

$$\begin{aligned} \text{FAR} &= \frac{\sum_{j=1}^{10^4} \mathbb{1}_{\mathcal{H}_j}(u_s)}{10^4}, && i = 1, \\ \text{ADD} &= \frac{\sum_{j=1}^{10^4} \min\{\ell : u(\ell) \in \mathcal{U}_j \wedge u(\ell) = u_s\}}{10^4}, && i = 2, \end{aligned}$$

where $\mathcal{H}_j = \{i, u, g, \cdots\}$ is the history of j-th evaluation episode where $i = 1$ and $\mathcal{U}_j = \{u(0), u(1), \cdots\}$ is the history of control in j-th evaluation episode where $i = 2$.

5.4.3 Evaluation Results

The evaluation results are summarized here. The detector has been tested with various ν_θ and some of the evaluation results are summarized here. Table 5.1 lists the performance under different costs with mean $\nu_0 = 0.7$. The number of training episodes N is set to be 1 million. It is shown that as the cost increases, the corresponding FAR increases. This is expected since the increase of the cost $g(1, u_c, 2)$ signals a bigger emphasis on minimizing detection delay during the training phase. Therefore, the FAR and ADD increases and decreases, respectively. Table 5.2 shows the effect of different window size w when $\nu_0 = 1$. The number of training episodes N is set to be 300 thousands. It shows that as the window size increases, the corresponding FAR decreases. It is reasonable because with w increasing, the detector will be closer to be optimal.

In addition to the trade-off behavior between FAR and ADD discussed above, the performance also varies with the number of sensors under attack. Let us see these two tables again. The FAR of 1 sensor attacked is almost similar to the one of 2 sensors attacked. The ADD of 1 sensor attacked is larger than the one of 2 sensors attacked. In other words, the detector spends longer time to judge whether there

TABLE 5.1: Performance under different costs with window size $w = 6$ and learning rate $\alpha = 0.005$.

Type / Cost	1 sensor attacked		2 sensors attacked	
	FAR(%)	ADD	FAR(%)	ADD
0.001	4.95	15.5038	4.98	7.8317
0.002	6.98	11.3086	6.75	7.2053
0.005	9.03	9.8902	9.71	6.8299
0.010	18.01	9.1583	18.16	6.4030

TABLE 5.2: Performance under different window sizes with learning rate $\alpha = 0.005$ and cost $g(1, u_c, 2) = 0.001$.

Type / w	1 sensor attacked		2 sensor attacked	
	FAR(%)	ADD	FAR(%)	ADD
2	14.83	14.3542	15.33	5.4918
3	4.51	8.8770	4.63	5.4022
4	1.25	7.8350	1.43	5.4208
5	0.27	8.5059	0.32	6.3201

is an attack. It is easy to understand since it is more difficult to distinguish for the scenario of 1 sensor attacked than the one of 2 sensors attacked.

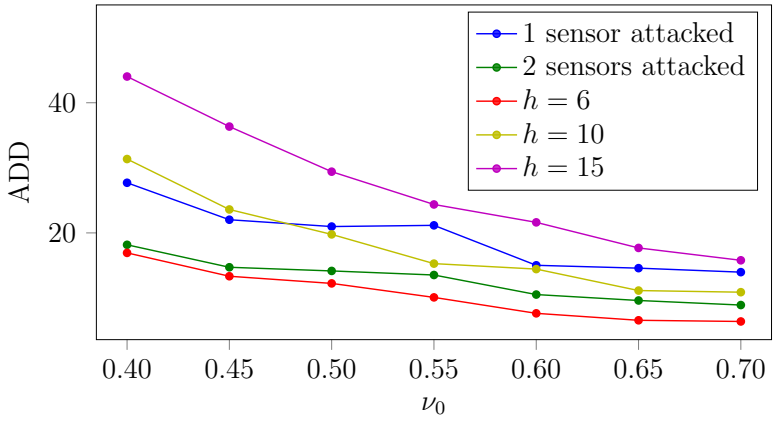

FIGURE 5.6: ADDs of the resulting detector and CUSUM algorithm. Parameter h refers to the tuning parameter used in CUSUM.

In Figure 5.6, the performance of our trained detector is compared with the one of classical quickest change detection algorithm (CUSUM) with different ν_0 [132] and we focus on comparing ADD, while all corresponding FAR are less than 5% in our methods. The blue and green lines denote ADD of 1 sensor attacked and 2 sensors attacked when choosing appropriate window size w, learning rate α and exploration parameters ε. The last three lines denote corresponding ADD when

the threshold h is set as 6, 10, 15. Note that the CUSUM requires the true value of θ, which is not needed in our detector. From this figure, one can see that the resulting detector has comparable performance with classical QCD algorithm.

5.5 Conclusion

In this chapter, we formulated a detection problem of flip attacks as a POMDP by assuming an attack probability and employed an MDP in the form of SSP to approximate the behavior of the POMDP by fixed-length window and state aggregation of observations. A standard Q-learning algorithm was then applied to derive the optimal solution of the approximated MDP. Numerical results were given to illustrate that the resulting detector can obtain a promising behavior.

Chapter 6

Rollout Approach to Sensor Scheduling for Remote State Estimation under Integrity Attack

Chapter 4 and 5 focus on the detection of cyber-attacks. However, it is difficult and unrealistic to prevent these attacks from happening or detect all potential attacks before they cause damage the system. Hence, how to mitigate the effects of cyber-attacks is of great importance for the security of CPS.

In this chapter, we explore the possibility of mitigating the effect of cyber-attacks. More specifically, the sensor scheduling problem for remote state estimation under integrity attacks is studied. We seek to optimize a trade-off between the accumulated energy consumption of communications and the state estimation error covariance when the acknowledgment (ACK) information, sent by the remote estimator to the local sensor, is compromised. The sensor scheduling problem is formulated as an infinite horizon discounted optimal control problem with infinite states. We first analyze the underlying Markov decision process (MDP) and show that the optimal schedule without ACK attack is of the threshold type. Thus, we can simplify the problem by replacing the original state space with a finite state space. For the simplified MDP, when ACK is under attack, the problem is modelled as a partially observable Markov decision process (POMDP). We analyze the induced MDP that uses a belief vector as its state for the POMDP. We investigate the properties of the exact optimal solution via contractive models and show that

the threshold type of solution for the POMDP cannot be readily established. A suboptimal solution is then obtained via a rollout approach which is a prominent class of reinforcement learning (RL) method that relies on an approximation in value space. We present two variants of rollout and provide performance bounds of those variants.

The rest of the chapter is organized as follows. Section 6.1 introduces the system model, smart sensor, remote state estimation as well as ACK feedback scheme. We also formulate the sensor scheduling problem under the ACK-based attack as an infinite horizon discounted problem in this section. The underlying MDP is studied and the optimal policy is shown to be of threshold type in Section 6.2. In the presence of attack, a decision making process of the sensor is modelled as a POMDP, which is analyzed in Section 6.3. In Section 6.4, some numerical examples are provided to demonstrate the effectiveness of the proposed strategy. Conclusions and future work are provided in Section 6.5.

6.1 Problem Formulation

6.1.1 System Model

An LTI system is considered in this chapter and it is the same as the one in Chapter 3. For the sake of readability, we rewrite the system model here again:

$$x_{k+1} = Ax_k + w_k,$$
$$y_k = Cx_k + v_k,$$

where $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}^m$ are the state vector and all sensor measurements at time k, respectively. $w_k \in \mathbb{R}^n$ denotes the process noise and $v_k \in \mathbb{R}^m$ is the measurement noise. $w_k \in \mathcal{N}(0, Q)$ and $v_k \in \mathcal{N}(0, R)$, where $Q \succeq 0$ and $R \succ 0$, respectively. It is assumed that $w_0, w_1, \ldots$ and $v_0, v_1, \ldots$ are mutually independent. Besides, we give the following assumption about the system matrices:

Assumption 6.1. The pair (A, C) is detectable and $(A, \sqrt{Q})$ is stabilizable.

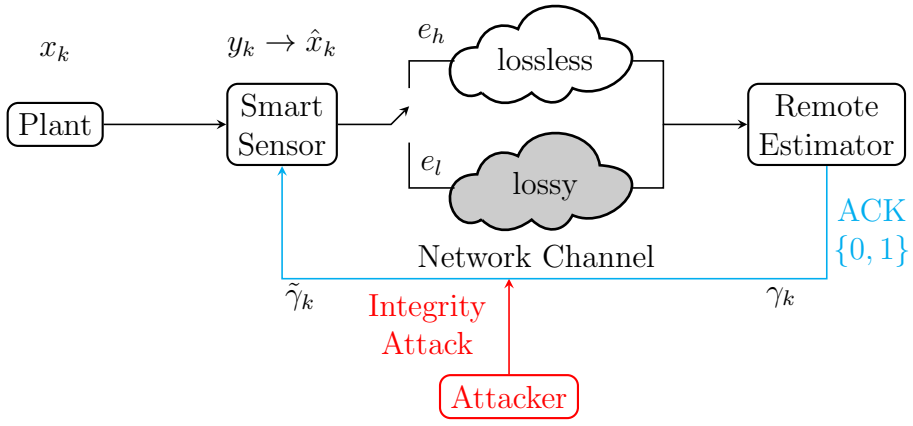

FIGURE 6.1: The system diagram

6.1.2 Smart Sensor

Similar to Chapter 3, we consider the smart sensor as described in [39], which first locally estimates the state x_k based on all the measurements it has collected up to time k and then transmits its local estimate to the remote estimator.

Denote $\hat{x}_k^s$ and $\hat{P}_k^s$ as the sensor's local minimum mean squared error (MMSE) estimate of the state x_k and the corresponding error covariance:

$$\hat{x}_k^s = \mathbb{E}\left[x_k \,|\, y_1, y_2, \cdots, y_k\right], \tag{6.1}$$

$$\hat{P}_k^s = \mathbb{E}\left[(x_k - \hat{x}_k^s)(x_k - \hat{x}_k^s)^\intercal \,|\, y_1, y_2, \cdots, y_k\right], \tag{6.2}$$

which can be calculated by a standard Kalman filter. Under Assumption 6.1, the estimation error covariance of the Kalman filter converges to a unique value from any initial condition. Without loss of generality, we assume that the Kalman filter at the sensor side has entered the steady state and simplify our subsequent discussion by setting:

$$\hat{P}_k^s = P,\ k \geq 1, \tag{6.3}$$

where P is the steady-state error covariance. For notational ease, we define the Lyapunov operator h: $\mathbb{S}_+^n \to \mathbb{S}_+^n$ as:

$$h(X) \triangleq AXA^\intercal + Q.$$

Then P is given by the unique positive semi-definite solution of the following equation:

$$X = h(X) - h(X)C^{\intercal}[Ch(X)C^{\intercal} + R]^{-1}Ch(X).$$

After obtaining $\hat{x}_k^s$, the sensor will transmit it as a data packet to the remote estimator. Due to fading and interference, random data drops may occur. We assume that the sensor has two choices: one is to send the local state estimate with high power, which will consume energy e_h ($e_h > 0$); the other is to send the estimate with low power, which will consume energy e_l ($0 < e_l < e_h$). We assume that for the first choice, the packet will always arrive at the remote estimator, while for the other one, the successful arrival rate is $\nu \in (0, 1)$. In Figure 6.1, we use two lines to denote these two choices. The upper line refers to the choice e_h, and the lower line denotes the choice e_l.

6.1.3 Remote State Estimation

The transmission of $\hat{x}_k^s$ between the sensor and the remote estimator can be characterized by a binary random variable γ_k, $k \in \mathbb{N}$:

$$\gamma_k = \begin{cases} 1, & \text{if } \hat{x}_k^s \text{ arrives at time } k, \\ 0, & \text{otherwise (regarded as dropout).} \end{cases}$$

Denote $\hat{x}_k$ and $\hat{P}_k$ as the remote estimator's own MMSE state estimate and the corresponding error covariance based on all the sensor data packets received up to time step k. The remote state estimate $\hat{x}_k$ obeys the recursion:

$$\hat{x}_k = \begin{cases} \hat{x}_k^s, & \text{if } \gamma_k = 1, \\ A\hat{x}_{k-1}, & \text{if } \gamma_k = 0. \end{cases} \tag{6.4}$$

The corresponding state estimation error covariance $\hat{P}_k$ satisfies:

$$\hat{P}_k = \begin{cases} P, & \text{if } \gamma_k = 1, \\ h(\hat{P}_{k-1}), & \text{if } \gamma_k = 0. \end{cases} \tag{6.5}$$

The objective of sensor scheduling for the smart sensor is to optimize a trade-off between the energy consumption of communications and the trace of the estimation error covariance of the remote estimator. We formally formulate the cost that the sensor aims to minimize as:

$$J_{obj} = \lim_{N\to\infty} \mathbb{E}\Big\{ \sum_{k=0}^{N} \alpha^k g_k \Big\},$$

where $\alpha \in (0,1)$ is the discount factor and g_k denotes the stage cost given by:

$$g_k = \beta \cdot (\text{energy consumption}) + (1-\beta) \cdot (\text{trace of error covariance}),$$

where $\beta \in (0,1)$ is the weighting coefficient, the term "energy consumption" is related with e_l and e_h and the term "error covariance" is related with Eq. (6.5). Here, we need to emphasize that the state estimation error covariance and the choice of sending local state estimate with high or low energy at every time depend on the current state which is the holding time since the most recent successful reception of the data from the sensor by the remote estimator and the definition of which will be detailed later. Hence, it is of great importance for the sensor to obtain an accurate state. Thus, the following acknowledgment scheme is employed.

6.1.4 ACK Feedback Scheme

To improve the estimation performance under an energy constraint, an online power schedule based on the ACK from the remote estimator was proposed in [133]. We consider the same setting and use akin scheme. The remote estimator generates a 1-bit ACK signal to indicate whether the data packet is delivered successfully or not, which is illustrated in Figure 6.1. Namely, $\gamma_k = 1$ means that information $\hat{x}_k^s$ has been delivered to the remote state estimator, and $\gamma_k = 0$ otherwise. In this way, the sensor can obtain real-time information from the remote estimator. However, ACK scheme faces the risk of being attacked. The next section shows a possible ACK-based attack model, which results in the degradation of the performance of sensor scheduling.

6.1.5 Attack Model

While the ACK feedback scheme is simple and easy to implement, the simple structure makes it a likely target of an attacker who can carry out integrity attacks or Denial of Service (DoS) attacks [134]. In this section, we propose a possible attack strategy for the attacker and investigate the corresponding mitigation of this kind of attack in the rest of the paper.

When the ACK channel is perfect, the smart sensor will receive real-time ACK. If there is an integrity attack launched by "Attacker" shown in Figure 6.1, the ACK may be modified, i.e., $\gamma_k = 0$ may be modified to 1 and $\gamma_k = 1$ to 0, according to certain probability defined by the attacker. In order to differentiate the ACK signal under the normal and attack scenarios, we use $\tilde{\gamma}_k$ to denote the ACK received by the smart sensor under possible attack for later discussion. We denote Γ as the observation space of $\tilde{\gamma}_k$ and $\Gamma = \{0, 1\}$. We will use $\tilde{\gamma}$ to refer to a value in Γ. Note that $\tilde{\gamma}_k$ may not be equal to γ_k. We denote κ_0 and κ_1 as the probability that $\gamma_k = 0$ and $\gamma_k = 1$ are flipped by the attacker, respectively, i.e.,

$$\kappa_0 = p(\tilde{\gamma}_k = 1 | \gamma_k = 0),$$
$$\kappa_1 = p(\tilde{\gamma}_k = 0 | \gamma_k = 1),$$

where $p(\cdot)$ denotes the probability. It is assumed that the attack probabilities do not vary during the attack process. Our goal is to design a power scheduling approach, knowing that the ACK information is under the above flip attack.

Remark 6.1. We give a further discussion about this attack model. If the smart sensor is only concerned with the error covariance of remote state estimation, the optimal attack strategy for an attacker is given as $(\kappa_0, \kappa_1) = (1, 0)$, while if the energy consumption of smart sensor is the only concern, corresponding attack probabilities should be given as $(\kappa_0, \kappa_1) = (0, 1)$. Since the smart sensor aims to optimize a trade-off between the remote state estimation accuracy and transmission energy, intuitively, the optimal attack probability should be between 0 and 1. This can be also observed in the example in Section 6.4.1.

Remark 6.2. Note that the flipping probability of the attack may be obtained by the smart sensor which transmits its state estimate with high power e_h over a period of time and computes the statistics of the ACK signals. In view of this, we

assume that the system has the knowledge of attack probabilities and focus on the mitigation of this kind of attack.

Remark 6.3. The related study on the fake acknowledgement attack can be found in [134]. The optimal DoS attack on the feedback channel against ACK-based sensor power schedule is studied in [34]. In [135], a game-theoretic approach to acknowledgement attack is proposed and the Nash equilibrium is studied. Note that the above works consider the DoS attack and emphasize the performance analysis of attacks on the feedback channel. Our work focuses on the mitigation of the integrity attack.

Remark 6.4. It is worth noticing that our proposed attack model is different form the lossy acknowledgement channel model. The key difference is that the remote estimator sends a feedback at each time and the attacker modifies the feedback signal with certain probabilities.

6.1.6 Preliminary Analysis

In this section, we define the state, control, and observation of our problem. We apply the POMDP framework to model the process given that ACK is under possible attack. Note that due to the recursion of the dynamics in (6.5), the covariance P_k can only take value in the infinitely countable set $\{P, h(P), h^2(P), \cdots\}$. Denote $s_k \in \mathbb{Z}$ as the holding time since the most recent successful reception of the data from the sensor by the remote estimator:

$$s_k \stackrel{\triangle}{=} k - 1 - \max_{1 \le t \le k-1} \{t : \gamma_t = 1\}. \tag{6.6}$$

Therefore, $s_k = 0$ means the message sent at time $k-1$ has been successfully received. We denote by $\mathcal{S}$ the state space of s_k. The state space has countable elements $0, 1, \cdots$, standing for the holding time s_k, and we will use s_k to represent the unspecified state in $\mathcal{S}$. Denote u_k as the control option and $\mathcal{U}$ as the control space of u_k. The control options are to send the state estimate with high or low energy, denoted as 1 and 0, respectively, i.e., $\mathcal{U} = \{0, 1\}$, and we will use u_k to represent the unspecified control at time k. In particular, we use s and u to refer to a value in $\mathcal{S}$ and $\mathcal{U}$. We denote by $p_{ss'}(u)$ the transition probability from s to s' under control u and denote by $g(s, u, s')$ a function which returns the cost when such transition occurs.

It is clear from our assumption that the probabilities are given as

$$p_{ss'}(1) = \begin{cases} 1, & s' = 0; \\ 0, & \text{o.w.}; \end{cases} \quad p_{ss'}(0) = \begin{cases} \nu, & s' = 0; \\ 1-\nu, & s' = s+1; \\ 0, & \text{o.w.}. \end{cases} \tag{6.7}$$

For the power scheduling problem of our interest, the following functions are usually used as cost per stage

$$\begin{aligned} g(s, 1, s') &= \beta e_h + (1-\beta)\,\mathrm{tr}(P), \\ g(s, 0, s') &= \beta e_l + (1-\beta)\,\mathrm{tr}(h^{s'}(P)). \end{aligned} \tag{6.8}$$

Due to the presence of attack, the true state s_k can not be computed according to Eq. (6.6) since γ_k is potentially compromised. Therefore, the observation received at current time is conditioned on the current state s, denoted by $q_{\tilde{\gamma}}(s)$, and its conditional probability is given as

$$q_1(s) = \begin{cases} 1-\kappa_1, & s = 0; \\ \kappa_0, & \text{o.w.}; \end{cases} \quad q_0(s) = \begin{cases} \kappa_1, & s = 0; \\ 1-\kappa_0, & \text{o.w.}. \end{cases} \tag{6.9}$$

6.1.7 Problem of Interest

Given that we know the successful arriving rate ν of packet when transmitted with low energy, the presence of the attacker, the observation probabilities $q_{\tilde{\gamma}}(s)$, and the observations $\tilde{\gamma}_k$, we are interested in obtaining a scheduling rule that minimizes the following expectation of the infinite horizon accumulated cost of $g(s_k, u_k, s_{k+1})$ given by (6.8) with a given discount factor of $\alpha \in (0, 1)$:

$$J(s) = \lim_{N\to\infty} \mathbb{E}\Big\{ \sum_{k=0}^{N} \alpha^k g(s_k, u_k, s_{k+1}) \,|\, s_0 = s \Big\}.$$

The challenge of the problem is twofold. First, the problem is a POMDP, which is difficult in its own right. Secondly, the underlying MDP has a state space that is composed of infinite states. In what follows, we will first investigate the properties of the MDP. It is shown that the optimal policy, if the states are known, is of the

threshold type. Therefore, it is then sufficient to consider the finitely many state case, thus obtaining a MDP with a truncated state space. Then we will show for the POMDP induced by the truncated state space, contractive models apply and some performance guarantees can be established.

6.2 Simplification of the Underlying MDP

6.2.1 Analysis on the Properties of MDP

Recall that the Lyapunov operator is defined as:

$$h(X) \triangleq AXA^{\intercal} + Q.$$

For the error covariance iterated through the Lyapunov operator, the following results are given by [104].

Lemma 6.1 ([104])**.** *The following inequality holds:*

$$\mathrm{tr}(P) = \mathrm{tr}(h^0(P)) < \mathrm{tr}(h^1(P)) < \mathrm{tr}(h^2(P)) < \cdots < \mathrm{tr}(h^k(P)) < \cdots, \forall k \in \mathbb{N}.$$

From Lemma 6.1, one could obtain that the sequence $\{\mathrm{tr}(h^k(P))\}_{k=0}^{\infty}$ is monotone non-decreasing as k grows.

Lemma 6.2. *If $\rho(A) \geq 1$, $\mathrm{tr}(h^k(P))$ approaches infinity as k goes to infinity. Otherwise, the sequence $\{\mathrm{tr}(h^k(P))\}_{k=0}^{\infty}$ is bounded.*

Recall the definition of the cost g given by Eq. (6.8), we have $g \geq 0$ viz., the MDP we are investigating here is nonnegative. This is formalized as the following lemma.

Lemma 6.3. *The cost function g satisfies*

$$g(s, u, s') \geq 0, \forall (s, u, s') \in \mathcal{S} \times \mathcal{U} \times \mathcal{S}.$$

In this chapter, we will focus on the case where the stage cost is unbounded, which is formalized in the following assumption.

Assumption 6.2. The system A fulfills $\rho(A) \geq 1$.

Remark 6.5. When $\rho(A) < 1$, in view of Lemma 6.2, the stage cost is bounded for all states and controls, in which case a large portion of the following analysis still applies. However, it would be possible, depending on the tuning parameter β, to have low energy transmission always as the optimal policy, which would not be practically interesting to investigate.

For the nonnegative MDP with state $\mathcal{S}$, control $\mathcal{U}$, we denote $\bar{\mu} : \mathcal{S} \to \mathcal{U}$ as a fixed policy, and $\bar{\pi} = \{\bar{\mu}_0, \bar{\mu}_1, \dots\}$ as a sequence of $\bar{\mu}_k$. Thus, the cost function under $\bar{\pi}$ and $\{\bar{\mu}, \bar{\mu}, \dots\}$ are defined as

$$J_{\bar{\pi}}(s) = \lim_{N\to\infty} \mathbb{E}\Big\{ \sum_{k=0}^{N} \alpha^k g(s_k, \bar{\mu}_k(s_k), s_{k+1}) \,|\, s_0 = s \Big\},$$
$$J_{\bar{\mu}}(s) = \lim_{N\to\infty} \mathbb{E}\Big\{ \sum_{k=0}^{N} \alpha^k g(s_k, \bar{\mu}(s_k), s_{k+1}) \,|\, s_0 = s \Big\},$$

in view of g being nonnegative. For such problems, we are interested in obtaining the optimal cost function J^* and optimal stationary policy $\bar{\mu}^*$ given as

$$J^*(s) = \inf_{\bar{\pi}} J_{\bar{\pi}}(s), \;\; J_{\bar{\mu}^*}(s) = J^*(s), \;\; \forall s \in \mathcal{S}. \tag{6.10}$$

For the problem of interest, the optimal cost function can be achieved since the cost function is nonnegative (Lemma 6.3) and the control space is compact under an unbounded cost. This is formalized as the following lemma.

Lemma 6.4 (Chapter 6, [136])**.** *About the cost function, we have*

(a) The optimal cost function J^* of (6.10) satisfies Bellman's equation. Namely, for all $s \in \mathcal{S}$,

$$J^*(s) = \min_{u\in\mathcal{U}} \sum_{s'\in\mathcal{S}} p_{ss'}(u)\big\{g(s,u,s') + \alpha J^*(s')\big\}. \tag{6.11}$$

(b) Let $\bar{\pi} = \{\bar{\mu}, \bar{\mu}, \cdots\}$ be an admissible stationary policy. We have for all $s \in \mathcal{S}$,

$$J_{\bar{\mu}}(s) = \sum_{s'\in\mathcal{S}} p_{ss'}\big(\bar{\mu}(s)\big)\Big\{g\big(s, \bar{\mu}(s), s'\big) + \alpha J_{\bar{\mu}}(s')\Big\}.$$

(c) There is an optimal stationary policy $\bar{\mu}^*$ that satisfies (6.10).

(d) Starting with $J_0(s) \equiv 0$, the value iteration sequence $\{J_k\}_{k=0}^{\infty}$ generated by

$$J_{k+1}(s) = \min_{u \in \mathcal{U}} \sum_{s' \in \mathcal{S}} p_{ss'}(u)\big\{g(s,u,s') + \alpha J_k(s')\big\}, \tag{6.12}$$

for all $s \in \mathcal{S}$, converges pointwise to J^*.

In view of the above nice properties for the positive MDP problems, we are ready to show that the underlying MDP with infinite states can be simplified.

Theorem 6.1. *Let Assumptions 6.1 and 6.2 hold. The optimal policy $\bar{\mu}^*$ defined by (6.10) is of the threshold type, viz., there exists a constant $\epsilon^* \in \mathcal{S}$ such that*

$$\bar{\mu}^*(s) = \begin{cases} 0, & \text{if } s < \epsilon^*, \\ 1, & \text{if } s \geq \epsilon^*. \end{cases} \tag{6.13}$$

Moreover, the optimal cost is strictly increasing for all $s \leq \epsilon^*$, viz., $J^*(s-1) < J^*(s)$ when $s \leq \epsilon^*$, and remains constant for $s \geq \epsilon^*$.

Proof. For the simplicity of notations, we define:

$$\begin{aligned} g_h &\triangleq g(s,1,0) = \beta e_h + (1-\beta)\operatorname{tr}(P), \\ g_l &\triangleq g(s,0,0) = \beta e_l + (1-\beta)\operatorname{tr}(P), \end{aligned}$$

where $g_l < g_h$ since $e_l < e_h$ if $\beta \neq 0$.

By Lemma 6.4 (b) and (c), we have

$$J^*(s) = \min\Big\{g_h + \alpha J^*(0), \nu\left[g_l + \alpha J^*(0)\right] + (1-\nu)\left[g(s,0,s+1) + \alpha J^*(s+1)\right]\Big\}. \tag{6.14}$$

Since for a policy $\bar{\mu}$ such that $\bar{\mu}(0) = 1$,

$$J_{\bar{\mu}}(0) = \sum_{k=0}^{\infty} \alpha^k g_h = \frac{g_h}{1-\alpha} < \infty,$$

and $J^*(0) \leq J_{\bar{\mu}}(0)$ per definition, we have for all $s \in \mathcal{S}$,

$$J^*(s) \leq g_h + \alpha J^*(0) \leq g_h + \frac{g_h}{1-\alpha} = \frac{2-\alpha}{1-\alpha} g_h < \infty.$$

By applying value iteration with $J_0 \equiv 0$, we have

$$J_{k+1}(s) = \min\left\{g_h + \alpha J_k(0), \nu\left[g_l + \alpha J_k(0)\right] + (1-\nu)\left[g(s,0,s+1) + \alpha J_k(s+1)\right]\right\}.$$

We now claim that for all k, it holds that

$$J_k(s) \leq J_k(s'), \quad \forall s' \geq s.$$

We establish the result by induction. The above assertion holds for J_0. Assume now it holds for k, then for $k+1$, we have

$$\begin{aligned} J_{k+1}(s) &= \min\left\{g_h + \alpha J_k(0), \nu\left[g_l + \alpha J_k(0)\right] + (1-\nu)\left[g(s,0,s+1) + \alpha J_k(s+1)\right]\right\}, \\ J_{k+1}(s+1) &= \min\left\{g_h + \alpha J_k(0), \nu\left[g_l + \alpha J_k(0)\right] + (1-\nu)\left[g(s+1,0,s+2) + \alpha J_k(s+2)\right]\right\}. \end{aligned}$$

By induction assumption and Lemma 6.1, we have

$$g(s,0,s+1) + \alpha J_k(s+1) \leq g(s+1,0,s+2) + \alpha J_k(s+2).$$

Therefore, we have $J_{k+1}(s+1) \leq J_{k+1}(s+2)$, and by induction, we have $J_{k+1}(s) \leq J_{k+1}(s'), \forall s' \geq s$.

Since by Lemma 6.4 (d), VI converges pointwise to J^*, we thus have

$$J^*(s) = \lim_{k\to\infty} J_k(s) \leq \lim_{k\to\infty} J_k(s') = J^*(s').$$

Namely, $J^*(s)$ is monotonically increasing. In addition, the term $g(s,0,s+1)$ grows unbounded with s due to $\rho(A) \geq 1$. Thus, there exists some $\bar{s}$ such that $\bar{\mu}^*(\bar{s}) = 1$ and $J^*(\bar{s}) = g_h + \alpha J^*(0)$ and $J^*(s') \geq J^*(\bar{s}) = g_h + \alpha J^*(0)$, $\forall s' > \bar{s}$. On the other hand, as discussed above, for all s, we have $J^*(s) \leq g_h + \alpha J^*(0)$. Thus, $J^*(s') = g_h + \alpha J^*(0)$. Define $\epsilon^* \triangleq \arg\min\{\bar{s} \,|\, \bar{\mu}^*(\bar{s}) = 1\}$.

The above proof has shown that for all $s \geq \epsilon^*$, the optimal cost $J^*(s)$ is constant and equals $g_h + \alpha J^*(0)$. In view of definition of ϵ^*, we see that $J^*(\epsilon^* - 1) < J^*(\epsilon^*)$ [since otherwise, we would have $\bar{\mu}^*(\epsilon^* - 1) = 1$, contradicting the definition of ϵ^*]. For $s \leq \epsilon^* - 2$, if $J^*(s) = J^*(s+1)$, we have, by (6.14),

$$\begin{aligned} &(1-\nu)\left[g(s,0,s+1) + \alpha J^*(s+1)\right] \\ =&(1-\nu)\left[g(s+1,0,s+2) + \alpha J^*(s+2)\right]. \end{aligned}$$

However, this contradicts with $g(s, 0, s+1) < g(s+1, 0, s+2)$ and $J^*(s+1) \leq J^*(s+2)$. Thus, the assumption is false and the proof is complete. □

Remark 6.6. Although similar results are available in [107, 137], our problem setup is different. To the best of our knowledge, it is the first time to study the discounted infinite horizon optimal control problem with an unbounded stage cost in the context of sensor scheduling.

6.2.2 Computational Approach

Due to the fact that the optimal policy is of the threshold nature, we can thus consider a truncated state space $\mathcal{S}_t = \{0, 1, \ldots, \ell\}$ with $\ell > \epsilon^*$. However, ϵ^* is unknown. For the truncated problem, with a slight abuse of notion, we will still use $p_{ss'}(u)$ to denote the transition probability, with the modification that $p_{\ell\ell}(0) = \nu$ and $p_{\ell 0}(0) = 1-\nu$, while the stage costs remain the same as for the original problem. The following theorem shows that the optimal control and cost of the truncated problem are the same as those of the original problem that has an infinite state space provided that $\ell > \epsilon^*$. In order to establish the desired relation, we recall one additional lemma, which shows better properties for finite state problems.

Lemma 6.5 (Props. 4.3.1, 4.3.2 [111])**.** *For a MDP problem with finite state space $\mathcal{S}_t$ and finite control space $\mathcal{U}$, the solution of (6.11) is unique. Moreover, the VI algorithm (6.12) starting from any $J_0 : \mathcal{S}_t \to \mathbb{R}$ converges pointwise to the solution of (6.11).*

Theorem 6.2. *For the truncated problem with state space $\mathcal{S}_t = \{0, 1, \ldots, \ell\}$, the optimal cost function and optimal policy fulfill the following conditions:*

(a) If $\ell \geq \epsilon^*$, the optimal control of the truncated problem is also given by (6.13), while the optimal cost function is $J^*_{|\mathcal{S}_t}$, the restriction of optimal cost of original MDP to $\mathcal{S}_t$.

(b) If $\ell < \epsilon^*$, the optimal control of the truncated problem is 0 for all $s \in \mathcal{S}_t$ while the optimal cost is upper bounded by $J^*_{|\mathcal{S}_t}$.

Proof. In order to differentiate the optimal cost function for the truncated and the original problems, we temporally denote $\bar{J}^*(s)$ as the optimal cost for the truncated problem.

For the truncated problem, Lemmas 6.4 and 6.5 hold. By Lemma 6.4 (b) and (c), for $s \in \mathcal{S}_t, s \neq \ell$, we have

$$\bar{J}^*(s) = \min\Big\{g_h + \alpha\bar{J}^*(0), \nu\left[g_l + \alpha\bar{J}^*(0)\right] + (1-\nu)\left[g(s,0,s+1) + \alpha\bar{J}^*(s+1)\right]\Big\}, \tag{6.15}$$

and for $s = \ell$:

$$\bar{J}^*(\ell) = \min\Big\{g_h + \alpha\bar{J}^*(0), \nu\left[g_l + \alpha\bar{J}^*(0)\right] + (1-\nu)\left[g(\ell,0,\ell+1) + \alpha\bar{J}^*(\ell)\right]\Big\}. \tag{6.16}$$

(a) Now we look at the optimal cost of the original problem J^*. We will show that restriction $J^*_{|\mathcal{S}_t}$ is a fixed point of the Bellman equations (6.15) and (6.16) of the truncated problem, and then by the uniqueness property asserted in Lemma 6.5, we have $\bar{J}^* = J^*_{|\mathcal{S}_t}$. For $s \in \mathcal{S}_t$ and $s \neq \ell$, it is clear that the Bellman equations of the original problem and the truncated problem are the same so J^* is the solution of Eq. (6.15). For $s = \ell$, J^* fulfills the condition

$$J^*(\ell) = \min\Big\{g_h + \alpha\bar{J}^*(0), \nu\left[g_l + \alpha\bar{J}^*(0)\right] + (1-\nu)\left[g(\ell,0,\ell+1) + \alpha J^*(\ell+1)\right]\Big\}.$$

Since $\ell \geq \epsilon^*$, by Theorem 6.1, we have $J^*(\ell+1) = J^*(\ell) = g_h + \alpha J^*(0)$. Thus, the restriction $J^*_{|\mathcal{S}_t}$ is a fixed point of the Bellman equations (6.15) and (6.16) of the truncated problem. By uniqueness of the solution, we have $\bar{J}^* = J^*_{|\mathcal{S}_t}$.

(b) Apply the same VI arguments as used in Theorem 6.1, we can show that $\bar{J}^*$ is monotonically increasing. Thus, for the truncated problem, if there exists some $\bar{\epsilon}^*$ such that $\bar{J}^*(\bar{\epsilon}^*) = g_h + \alpha\bar{J}^*(0)$, then $\bar{J}^*(\bar{\epsilon}^*) = g_h + \alpha\bar{J}^*(0)$ for $s = \bar{\epsilon}^*, \bar{\epsilon}^*+1, \ldots, \ell$. We will prove there exists no such $\bar{\epsilon}^*$ by contradiction.

Assume there exists such $\bar{\epsilon}^*$, such that

$$\bar{J}^*(s) = \nu\left[g_l + \alpha\bar{J}^*(0)\right] + (1-\nu)\left[g(s,0,s+1) + \alpha J^*(s+1)\right]$$

when $s < \bar{\epsilon}^*$ and $\bar{J}^*(s) = g_h + \alpha\bar{J}^*(0)$, $s \geq \bar{\epsilon}^*$. Then by setting $\bar{J}_0 = J^*_{|\mathcal{S}_t}$, and in view of monotonicity of $\bar{J}^*_{|\mathcal{S}_t}$, we can see that the sequence $\{\bar{J}_k\}$ generated by VI is monotonically decreasing and converges to $\bar{J}^*$. Thus, $\bar{J}^*$ is upper bounded by $J^*_{|\mathcal{S}_t}$. In addition, we have $\bar{J}_1(\ell) < \bar{J}_0(\ell) = J^*_{|\mathcal{S}_t}(\ell)$. Now we define

a new policy for the original MDP as

$$\bar{\mu}(s) = \begin{cases} 0, & s < \bar{\epsilon}^*, \\ 1, & s \geq \bar{\epsilon}^*, \end{cases}$$

Then we can see that $J_{\bar{\mu}} \geq J^*$ by optimality. However, it also holds that $J_{\bar{\mu}|\mathcal{S}_t} = \bar{J}^* \leq J^*_{|\mathcal{S}_t}$ and $\bar{J}^*(\ell) < J^*_{|\mathcal{S}_t}(\ell)$. Thus $\bar{J}^*(\ell) < J^*_{|\mathcal{S}_t}(\ell)$ and $\bar{J}^* \geq J^*_{|\mathcal{S}_t}$ hold simultaneously, which cannot be true.

□

In view of Theorem 6.2 and Lemma 6.4 (d), given some ℓ, we can use VI to get the corresponding optimal policy for the truncated problem. If it appears to be the threshold type, it corresponds to Theorem 6.2 (a) and the optimal cost and policy obtained are exactly the same as the underlying MDP the one with the truncated state space. If it is not the threshold type, it corresponds to Theorem 6.2 (b) and we need to increase ℓ to get the threshold ϵ^*. From now on, we refer to as the underlying MDP the one with the truncated state space provided that $\ell > \epsilon^*$, and with a slight abuse of notion, we will still use $p_{ss'}(u)$ to denote the transition probability, with the modification that $p_{\ell\ell}(0) = \nu$ and $p_{\ell 0}(0) = 1 - \nu$, while the transition costs for all $s \in \mathcal{S}_t$ remain unchanged, as is indicated in the proof of Theorem 6.2. Also, we will still use J^* to denote the optimal cost function.

6.3 Scheduling under Integrity Attack

When the acknowledgement information is under attack, we have a POMDP problem where the underlying state $\mathcal{S}$ is infinite dimensional, with known probability of observation. Based on the study of the previous section, the original state space can be replaced by a truncated version $\mathcal{S}_t$ with no impact of control selection. Thus, the POMDP of our concern is one with state space $\mathcal{S}_t$, control $\mathcal{U}$, observation Γ, transition probability given by Eq. (6.7), stage cost given by Eq. (6.8), observation probability given by Eq. (6.9), with the exception that $p_{\ell\ell}(0) = \nu$ and $p_{\ell 0}(0) = 1 - \nu$, while the transition costs for all $s \in \mathcal{S}_t$ remain unchanged.

6.3.1 Properties of the Exact Optimal Solution

For the POMDP introduced above, we analyze the induced MDP that uses a state as a belief vector. It is well-known that those two problems are equivalent, albeit the problem with belief state is infinite-dimensional [138]. For this study, we apply the contractive model detailed in [128]. To this end, we consider as states the functions $b : \mathcal{S}_t \to \mathbb{R}$ such that $\sum_{s\in\mathcal{S}_t} b(s) = 1$, and $b(s) \geq 0$, $\forall s$. It is easy to see that b is a vector and $b \in \mathbb{R}^{\ell+1}$. We denote as $\mathcal{B}$ the set of all such belief states and $\mathcal{B} \subset \mathbb{R}^{\ell+1}$. Denote $V : \mathcal{B} \to \mathbb{R}$ as a function defined on $\mathcal{B}$, and $\mathcal{V}$ as the set of functions that contains all V where $\|V\|_\infty < \infty$. Given a belief state b and a control u, the distribution of $\tilde{\gamma}$ can be computed as

$$\hat{p}(\tilde{\gamma} \,|\, b, u) = \sum_{s=0}^{\ell} b(s) \sum_{s'=0}^{\ell} p_{ss'}(u) q_{\tilde{\gamma}}(s').$$

The dynamics of b is governed by a Bayesian estimator denoted as $\Phi : \mathcal{B}\times\mathcal{U}\times\Gamma \to \mathcal{B}$, and given by

$$\Phi(b,0,1) = \begin{bmatrix} \dfrac{\nu(1-\kappa)}{\nu(1-\kappa)+(1-\nu)\kappa} \\ \dfrac{(1-\nu)\kappa b(0)}{\nu(1-\kappa)+(1-\nu)\kappa} \\ \vdots \\ \dfrac{(1-\nu)\kappa(b(\ell)+b(\ell-1))}{\nu(1-\kappa)+(1-\nu)\kappa} \end{bmatrix}, \tag{6.17}$$

$$\Phi(b,0,0) = \begin{bmatrix} \dfrac{\nu\kappa}{\nu\kappa+(1-\nu)(1-\kappa)} \\ \dfrac{(1-\nu)(1-\kappa)b(0)}{\nu\kappa+(1-\nu)(1-\kappa)} \\ \vdots \\ \dfrac{(1-\nu)(1-\kappa)(b(\ell)+b(\ell-1))}{\nu\kappa+(1-\nu)(1-\kappa)} \end{bmatrix}, \tag{6.18}$$

$$\Phi(b,1,\tilde{\gamma}) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^\intercal, \ \forall \tilde{\gamma} \in \Gamma. \tag{6.19}$$

For the induced infinite-dimensional MDP, the stage cost $\hat{g} : \mathcal{B} \times \mathcal{U} \to \mathbb{R}$ is given as

$$\hat{g}(b,u) = \sum_{s=0}^{\ell} b(s) \sum_{s'=0}^{\ell} p_{ss'}(u) g(s,u,s').$$

Based on the above definitions, we introduce an abstract operator that defines the induced MDP: $H : \mathcal{B} \times \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ given as

$$H(b, u, V) = \hat{g}(b, u) + \alpha \sum_{\tilde{\gamma} \in \Gamma} \hat{p}(\tilde{\gamma} \,|\, b, u) V\big(\Phi(b, u, \tilde{\gamma})\big). \tag{6.20}$$

We denote μ as a function $\mu : \mathcal{B} \to \mathcal{U}$ and all such functions form the set $\mathcal{M}$. In addition, we denote π as a sequence of admissible policies $\{\mu_k\}_{k=0}^{\infty}$ and Π as the set of all π. Then we introduce two additional Bellman operators $T_\mu, T : \mathcal{V} \to \mathcal{V}$ given by

$$(T_\mu V)(b) = H(b, \mu(b), V), \ (TV)(b) = \inf_{\mu \in \mathcal{M}} (T_\mu V)(b). \tag{6.21}$$

Theorem 6.3 (Fixed Point Properties)**.** *Let Assumptions 6.1 and 6.2 hold. For every T_μ and T, we have the following hold:*

(a) There exists unique V_μ, $V^* \in \mathcal{V}$ such that

$$V_\mu = T_\mu V_\mu, \ V^* = TV^*. \tag{6.22}$$

(b) V^* given in Eq. (6.22) is the optimal cost function, viz., $V^* = \inf_{\pi \in \Pi} V_\pi$, where V_π is given as

$$V_\pi = \lim_{N \to \infty} \sum_{k=0}^{N} T_{\mu_N}\Big(T_{\mu_{N-1}}\big(\cdots (T_{\mu_0} V_0) \cdots\big)\Big),$$

with $V_0 \equiv 0$. In particular, $V^* = \inf_{\mu \in \mathcal{M}} V_\mu$.

(c) For a policy $\mu^* \in \mathcal{M}$, $V_{\mu^*} = V^*$ if and only if $T_{\mu^*} V^* = TV^*$.

Proof. Given that the space $\mathcal{V}$ is complete, we only need to verify that the operators defined in Eq. (6.21) has the following properties:

(1) For all V, $V' \in \mathcal{V}$ such that $V \leq V'$, it holds that

$$T_\mu V \leq T_\mu V', \ \mu \in \mathcal{M}, \ TV \leq TV'.$$

(2) For all V, $V' \in \mathcal{V}$,

$$\|T_\mu V - T_\mu V'\|_\infty \leq \alpha \|V - V'\|_\infty, \ \forall \mu \in \mathcal{M}.$$

Then the results stated in (a), (b) and (c) follow from Prop. 2.1.1 and Prop. 2.1.2 in [128].

(1) For all V, $V' \in \mathcal{V}$,

$$\begin{aligned}
&(T_\mu V)(b) = H(b, \mu(b), V) \\
=&\hat{g}(b, \mu(b)) + \alpha \sum_{\tilde{\gamma} \in \Gamma} \hat{p}(\tilde{\gamma} \,|\, b, \mu(b)) V\big(\Phi(b, \mu(b), \tilde{\gamma})\big) \\
\leq&\hat{g}(b, \mu(b)) + \alpha \sum_{\tilde{\gamma} \in \Gamma} \hat{p}(\tilde{\gamma} \,|\, b, \mu(b)) V'\big(\Phi(b, \mu(b), \tilde{\gamma})\big). \\
=&H(b, \mu(b), V') = (T_\mu V')(b),
\end{aligned}$$

$$(TV)(b) = \inf_{\mu \in \mathcal{M}} (T_\mu V)(b) \leq \inf_{\mu \in \mathcal{M}} (T_\mu V')(b) = (TV')(b).$$

(2) For all $\mu \in \mathcal{M}$, we have

$$\begin{aligned}
&(T_\mu V)(b) - (T_\mu V')(b) \\
=&\alpha \sum_{\tilde{\gamma} \in \Gamma} \hat{p}(\tilde{\gamma} \,|\, b, u) V\big(\Phi(b, u, \tilde{\gamma})\big) - \alpha \sum_{\tilde{\gamma} \in \Gamma} \hat{p}(\tilde{\gamma} \,|\, b, u) V'\big(\Phi(b, u, \tilde{\gamma})\big) \\
=&\alpha \sum_{\tilde{\gamma} \in \Gamma} \hat{p}(\tilde{\gamma} \,|\, b, u) \big[V\big(\Phi(b, u, \tilde{\gamma})\big) - V'\big(\Phi(b, u, \tilde{\gamma})\big)\big] \\
\leq&\alpha \|V - V'\|_\infty,
\end{aligned}$$

which holds for all $b \in \mathcal{B}$. Reverse the order of V and V', which implies

$$|V(b) - V'(b)| \leq \alpha \|V - V'\|_\infty, \, \forall b \in \mathcal{B}.$$

Take supremum over $\mathcal{B}$ and we get the desired result.

□

As is shown in the proof of Theorem 6.1, the optimal cost function is monotonically increasing. However, for the induced MDP where now the state b is an element in the simplex of $\mathbb{R}^{\ell+1}$, additional conditions are required to obtain structural results like the one in Theorem 6.1. One simple and useful approach is through Monotone

likelihood ratio (MLR) ordering [109]. For two belief states b and b', we call b dominates b' in the MLR sense if

$$b(s)b'(s') \leq b'(s)b(s'),\ s < s',\ s,\ s' \in \mathcal{S}_t.$$

Under suitable conditions, computing MLR ordering between two belief states b and b' is sufficient to conclude whether $V^*(b) \geq V^*(b')$. In particular, if b dominating b' in the MLR sense implies $V^*(b) \leq V^*(b')$, then we say the POMDP is MLR decreasing in b. For the result to hold true, one key property required for the underlying MDP is that the related matrix is totally positive of order 2 (TP2). A stochastic matrix (could be rectangle, [109, Definition 10.2.1]) is called TP2 if all its second-order minors are nonnegative. The relevant stochastic matrices of our concern under control u are the transition matrix $P(u)$ and observation matrix $M(u)$, given as

$$P(u) = \begin{bmatrix} p_{00}(u) & p_{01}(u) & \cdots & p_{0\ell}(u) \\ p_{10}(u) & p_{11}(u) & \cdots & p_{1\ell}(u) \\ \vdots & \vdots & \ddots & \vdots \\ p_{\ell 0}(u) & 0 & \cdots & p_{\ell\ell}(u) \end{bmatrix},$$

$$M(u) = \begin{bmatrix} q_0(0) & q_0(1) & \cdots & q_0(\ell) \\ q_1(0) & q_1(1) & \cdots & q_1(\ell) \end{bmatrix}^{\intercal}.$$

The following lemma outlines the sufficient conditions under which the POMDP is MLR decreasing in b.

Lemma 6.6 (Theorem 11.2.1 [109])**.** *For the POMDP with state space $\mathcal{S}_t$, control $\mathcal{U}$, observation Γ, and stage cost $g : \mathcal{S}_t \times \mathcal{U} \times \mathcal{S}_t \to \mathbb{R}$, if the following three conditions hold:*

(1) For each control $u \in \mathcal{U}$, the expected stage cost defined as $\sum_{s' \in \mathcal{S}_t} p_{ss'}(u) g(s, u, s')$ is decreasing in s;

(2) The transition matrix $P(u)$ is TP2 for all u;

(3) The observation matrix $M(u)$ is TP2 for all u;

then the POMDP is MLR decreasing in b.

Remark 6.7. The stage cost in [109] is defined as a function of current state and current action $\mathcal{S}_t \times \mathcal{U}$, while here our stage cost $g(s,u,s')$ also depends on the next state. Thus, the stage cost in [109] is the expected stage cost $\sum_{s' \in \mathcal{S}_t} p_{ss'}(u) g(s,u,s')$ in our paper.

In the following theorem, we will show that the problem of interest here does not fulfill the TP2 property, thus MLR decreasing in b cannot be established through the above conditions. For the following discussion, we introduce a new class of functions. We call $\sigma : \mathcal{S}_t \to \mathcal{S}_t$ a permutation if it is a bijection. For some bijection, we obtain a new POMDP corresponding to σ, with state space $\mathcal{S}_t$, control $\mathcal{U}$, observation Γ, transition probability given as $p^{\sigma}_{ss'}(u) = p_{\sigma^{-1}(s)\sigma^{-1}(s')}(u)$, transition cost given as $g^{\sigma}(s,u,s') = g\big(\sigma^{-1}(s), u, \sigma^{-1}(s')\big)$, observation probability $q^{\sigma}_{\tilde{\gamma}}(s) = q_{\tilde{\gamma}}\big(\sigma^{-1}(s)\big)$, transition matrix $P^{\sigma}(u)$ and observation matrix $M^{\sigma}(u)$ defined accordingly. Now we are ready to proceed to state the following result, which essentially means that for the problem of interest, the MLR decreasing relation cannot be established through Lemma 6.6.

Theorem 6.4. *Let Assumptions 6.1 and 6.2 hold. There exists no permutation σ such that the corresponding POMDP fulfills the conditions (1), (2), and (3) given in Lemma 6.6.*

Proof. In view of Lemma 6.1, stage cost $g(x,u,s')$ of the original POMDP is increasing in s. Thus, condition (1) in Lemma 6.6 is violated. To have condition (1) hold, the only valid permutation is σ given as

$$\sigma(s) = \ell - s,$$

Under σ, the stage cost $g^{\sigma}(s,u,s')$ is decreasing in s. However, we can see the transition matrix $P^{\sigma}(0)$ is not TP2. Indeed, the last second order minor of $P^{\sigma}(0)$ is given as,

$$\begin{aligned}\begin{vmatrix} p^{\sigma}_{(\ell-1)(\ell-1)}(0) & p^{\sigma}_{(\ell-1)(\ell)}(0) \\ p^{\sigma}_{(\ell)(\ell-1)}(0) & p^{\sigma}_{(\ell)(\ell)}(0) \end{vmatrix} &= \begin{vmatrix} p_{11}(0) & p_{10}(0) \\ p_{01}(0) & p_{00}(0) \end{vmatrix} = \begin{vmatrix} 0 & \nu \\ 1-\nu & \nu \end{vmatrix} \\ &= \nu(\nu-1) < 0,\end{aligned}$$

which means $P^{\sigma}(0)$ is not TP2. Thus, there exists no permutation σ such that its corresponding POMDP fulfills the conditions (1), (2), and (3) in Lemma 6.6. □

Theorem 6.4 shows that the structural result for the induced MDP is not readily available. Therefore, we move to explore a suboptimal solution based on approximation in value space, implemented through the use of rollout [111] which is a common approach to compute the exact solution.

6.3.2 Approximate Solution through Rollout

Given that the exact solution cannot be established through the MLR ordering for the belief states, we seek an approximate solution. Here, we use the rollout approach, which is a simulation-based approach. To this end, we fix certain base policy $\mu_b \in \mathcal{M}$. Given current belief state as b, we aim to obtain a control option as the following minimizer

$$\tilde{\mu}(b) \in \arg\min_{u\in\mathcal{U}} \hat{g}(b,u) + \alpha \sum_{\tilde{\gamma}\in\Gamma} \hat{p}(\tilde{\gamma}\,|\,b,u)\tilde{V}\big(\Phi(b,u,\tilde{\gamma})\big), \tag{6.23}$$

where we examine two options for $\tilde{V}$, which we denote as $\tilde{V}^r$ and $\tilde{V}^{(\lambda)}$. They refer to fixed steps and geometrically distributed steps, which are common in rollout approach. The first option $\tilde{V}^r$ is to evaluate base policy μ_b for some fixed $r \in \mathbb{N}$ steps, with final cost $\bar{V}(b) = \sum_{s\in\mathcal{S}_t} b(s)J^*(s)$, thus

$$\tilde{V}^r(b) = (T^r_{\mu_b}\bar{V})(b). \tag{6.24}$$

The second option $\tilde{V}^{(\lambda)}$ is to evaluate the base policy μ_b with geometrically distributed steps with final cost $\bar{V}$, thus $\tilde{V}$ is defined as

$$\tilde{V}^{(\lambda)}(b) = (T^{(\lambda)}_{\mu_b}\bar{V})(b),$$

where $\lambda \in (0,1)$ is some design parameter, and $T^{(\lambda)}_{\mu_b} : \mathcal{V} \to \mathcal{V}$ is defined as

$$(T^{(\lambda)}_{\mu_b}V)(b) = (1-\lambda)\sum_{\ell=1}^{\infty} \lambda^{\ell-1}\big(T^{\ell}_{\mu_b}V\big)(b). \tag{6.25}$$

Remark 6.8. Note that the principal aim of rollout is policy improvement. Two conditions that the base policy needs to satisfy to ensure the cost improvement property are sequential consistency and sequential improvement. More details

about these conditions could be found in [111]. It seems that the choice of the base policy is important to the performance of rollout. However, experiments evidence has shown that surprisingly good rollout performance may be attained even with a relatively poor base heuristic [111]. We also verify this in Section 6.4.3.

Note that Eq. (6.25) is a well-defined, infinite-dimensional operator with the same fixed point as T_{μ_b}, and for all $V, V' \in \mathcal{V}$, it holds that

$$\|T_{\mu_b}^{(\lambda)}V - T_{\mu_b}^{(\lambda)}V'\|_\infty \leq \alpha_\lambda \|V - V'\|_\infty,$$

where $\alpha_\lambda = \frac{\alpha(1-\lambda)}{1-\lambda\alpha}$. Refer to [139] for details of the above results.

When rollout method defined in Eq. (6.23) is implemented exactly, with $\tilde{V} = \tilde{V}^r$, the performance of $\tilde{\mu}$ with respect to the optimal cost V^* can be characterized by the following lemma, which is given as Proposition 2.2.1 [128].

Lemma 6.7 (Proposition 2.2.1, [128]). *Denote as $V_{\tilde{\mu}}$ the cost function of the policy $\tilde{\mu}$ that is given by Eq. (6.23), and $\tilde{V} = \tilde{V}^r$. Then the suboptimality of $V_{\tilde{\mu}}$ with respect to V^* is given by*

$$\|V_{\tilde{\mu}} - V^*\|_\infty \leq \frac{2\alpha}{1-\alpha}\|T_{\mu_b}^r \bar{V} - V^*\|_\infty. \tag{6.26}$$

Compared with above result, when $\tilde{V} = \tilde{V}^{(\lambda)}$, we have the following performance bound.

Theorem 6.5 (Performance bound). *Denote as $V_{\tilde{\mu}}$ the cost function of the policy $\tilde{\mu}$ that is given by Eq. (6.23), and $\tilde{V} = \tilde{V}^{(\lambda)}$. Then the suboptimality of $V_{\tilde{\mu}}$ with respect to V^* is given by*

$$\|V_{\tilde{\mu}} - V^*\|_\infty \leq \frac{2\alpha}{1-\alpha}\|T_{\mu_b}^{(\lambda)} \bar{V} - V^*\|_\infty. \tag{6.27}$$

Proof. We view $T_{\mu_b}^{(\lambda)}V_0$ as the final cost and the above suboptimality is a direct application of Proposition 2.2.1, [128]. □

Remark 6.9. The scalars in (6.26) and (6.27) are usually unknown, so the resulting analysis will have a mostly qualitative character. However, Lemma 6.7 and Theorem 6.5 provide some insight on the performance of the rollout approach. By decreasing the discount factor α and increasing the number of lookahead step, a better performance bound could be expected.

6.3.3 Rollout Implementation via Monte-Carlo Sampling

Here we exemplify a rollout implementation via Monte-Carlo simulation method detailed in [111]. The algorithm is summarized in Algorithm 3, where $\tilde{V} = \tilde{V}^r$. The algorithm starts by initializing a belief state b. Then we run N_s Monte Carlo simulations to decide the control input, where the function "SIMULATOR" is used to take the Monte Carlo simulations with parameter action u, belief state b and truncated steps r. Then the obtained control is applied and the observation is collected. Finally, the new belief state is updated and proceed to next stage. In total, we obtain a suboptimal action sequence via rollout policy. It is worth noting that if we use the second variant where $\tilde{V} = \tilde{V}^{(\lambda)}$, the truncated step r is not anymore fixed, but rather a random variable drawn from geometric distribution with parameter λ.

Algorithm 3 Rollout policy for fixed truncated steps

Require: The iteration time steps N_i, the discount rate $\alpha \in (0, 1)$, the sample time N_s, the successful arrival rate under lower energy ν, the truncated steps r, and the base policy $\mu_b \in \mathcal{M}$

Ensure: A sensor schedule sequence.

function SIMULATOR(u, b, r)
  $v \leftarrow 0$
  Apply u and collect observation $\tilde{\gamma}$
  $v \leftarrow \hat{g}(b, u)$, $b \leftarrow \Phi(b, u, \tilde{\gamma})$
  for $l = 0 \rightarrow r - 1$ **do**
    $u \leftarrow \mu_b(b)$
    $v \leftarrow v + \alpha^l \hat{g}(b, u)$, $b \leftarrow \Phi(b, u, \tilde{\gamma})$
  end for
  $v \leftarrow v + \alpha^r \bar{V}(b)$
  return v
end function
Initialize the belief state b
for $j = 1 \rightarrow N_i$ **do**
  $v_0 \leftarrow 0$, $v_1 \leftarrow 0$
  for $k = 1 \rightarrow N_s$ **do**
    $v_0 \leftarrow v_0 +$ SIMULATOR$(0, b, r)$
    $v_1 \leftarrow v_1 +$ SIMULATOR$(1, b, r)$
    $\tilde{u} \in \arg\min_{u' \in \mathcal{U}} v_{u'}$
    Apply $\tilde{u}$ and collect observation $\tilde{\gamma}$
    $b \leftarrow \Phi(b, \tilde{u}, \tilde{\gamma})$
  end for
end for

6.3.4 Finite History Window Approach as a Baseline

In this subsection, we provide a finite history window approach which is widely employed in RL [140] for better comparison. Different from the above model-based and online method, it is a model-free and offline method. For this problem, the control space is the same as the problem in Section 3.1. We use o_k as an observation at time k, which includes a stored observation with length $m \geq 1$ and control input with length $n \geq 0$ given by

$$o_k \triangleq [\tilde{\gamma}_{k-m+1}, \cdots, \tilde{\gamma}_k, u_{k-n}, \cdots, u_{k-1}]^\intercal \in \mathbb{R}^{m+n}.$$

Here the control input also needs to be recorded as it also effects the decision-making of the sensor. Denote by $\mathcal{O}$ the set of all possible observations $o \in \mathbb{R}^{m+n}$, and $|\mathcal{O}| = |\Gamma|^m |\mathcal{U}|^n$. It is straightforward to see that the number of states grows exponentially with the length of the window.

Correspondingly, the cost per stage also depends on the random disturbance w_k and is denoted by $\bar{g}(o, u, o')$, which is given by:

$$\bar{g}(o, u, o') = \begin{cases} \beta e_h + (1-\beta)\,\mathrm{tr}(P), \\ \beta e_l + (1-\beta)\,\mathrm{tr}(h^{s'}(P)), \end{cases}$$

where s' is the hidden real state at next time.

With above formulation, we obtain a standard RL problem with $\mathcal{O}$ as state space and $\mathcal{U}$ as control space. Similarly, it can be solved by many different RL methods and we use Q-learning [131] as an example. The pseudocode of the algorithm is essentially same as Algorithm 2 in Chapter 5, so we omit it here.

Remark 6.10. Rollout is an online and simulation-based approach to obtain a sub-optimal control. Different from other offline approaches like Q-learning, it does not need to spend much time to train and not require to spend more space to store the lookup table. Here, we refer offline as the training process has been completed and the lookup table has been stored before online decision making. At the same time, rollout can avoid the curse of dimensionality. Finally, the performance of Q-learning will be affected by the choice of state, while the performance of rollout will not be limited by the base policy.

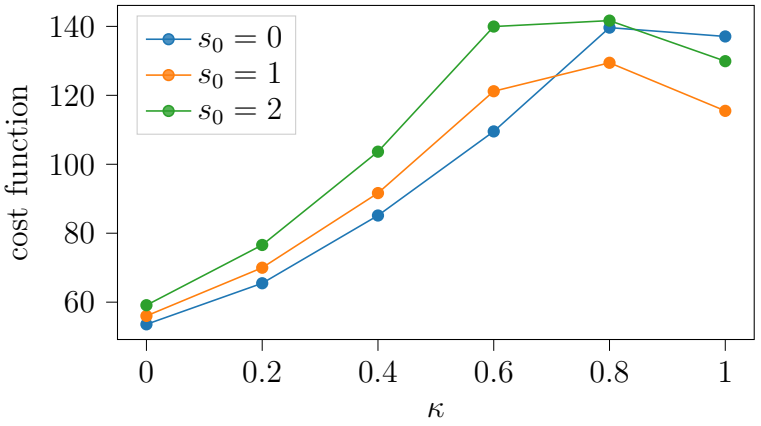

FIGURE 6.2: Cost functions of attacks with different probabilities and different initial states.

6.4 Simulation

6.4.1 A Simple Illustration of Attack Effect

An intuitive approach for the sensor scheduling would be to make the decision according to the most recent observation without remembering anything from the past. However, due to the existence of integrity attack, the sensor can not get the true state and it may not be optimal that the sensor decides whether to send the data with high or low power directly according to the current state. In order to give some insight to our proposed attack model, we provide a simple simulation to illustrate the effect of this kind of attack.

The system parameters are as follows:

$$A = \begin{bmatrix} 1.2 & 0.3 \\ 0.3 & 0.8 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1.7 \\ 0.3 & 1 \end{bmatrix}, \quad Q = R = I_2.$$

It is straightforward to obtain that the steady-state Kalman filtering error covariance $P = \begin{bmatrix} 1.7249 & -0.7250 \\ -0.7250 & 0.5144 \end{bmatrix}$. The energy consumptions of different levels are tuning parameters. We have tested different combinations and here we present one choice. They are set as $10 \operatorname{tr} P$ and $2 \operatorname{tr} P$. The successful arrival rate ν for lower energy is set as 0.4. The discount factor α is set to be 0.9.

When the above intuitive approach is employed for sensor scheduling, Figure 6.2 shows that the cost function values with different attack probabilities and initial

states. For the sake of simplicity, we set $\kappa_0 = \kappa_1 = \kappa$. From this figure, one can see that under the above parameter settings, the optimal attack is of flip probability between 0 and 1.

6.4.2 Threshold Policy of Underlying MDP

In this subsection, we provide a numerical example to verify the threshold policy with different weight parameters β's. Other related parameters are the same as the ones in the above subsection. Here, we use the value iteration to compute the optimal policy. The optimal policies under different weight parameter β are shown in Figure 6.3. From this figure, one can obtain that with the weight parameter β

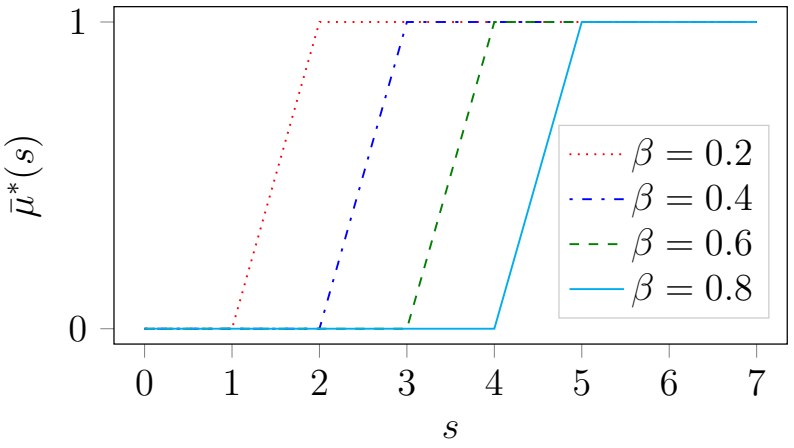

FIGURE 6.3: The optimal policy under different β.

increases, the optimal threshold value increases, which is expected due to higher weight on the energy consumption.

6.4.3 Comparison with Rollout Policy and Finite History Windows Approach

In this section, we consider the sensor scheduling under integrity attack. For simplicity of analysis, we truncated the state space and the state ℓ is set as 7. The attack probability κ is set as 0.5 and the weight parameter β is set as 0.6. Other related parameters are the same as the ones in the above subsection. From Figure 6.3, it is straightforward to get that the optimal threshold is $\epsilon^* = 4$ when $\beta = 0.6$.

Here, the base policy that we use is as follows:

$$\mu_b(b) = \begin{cases} 0, & \sum_{i=0}^{3} b(i) < \sum_{i=4}^{7} b(i), \\ 1, & \text{o.w..} \end{cases}$$

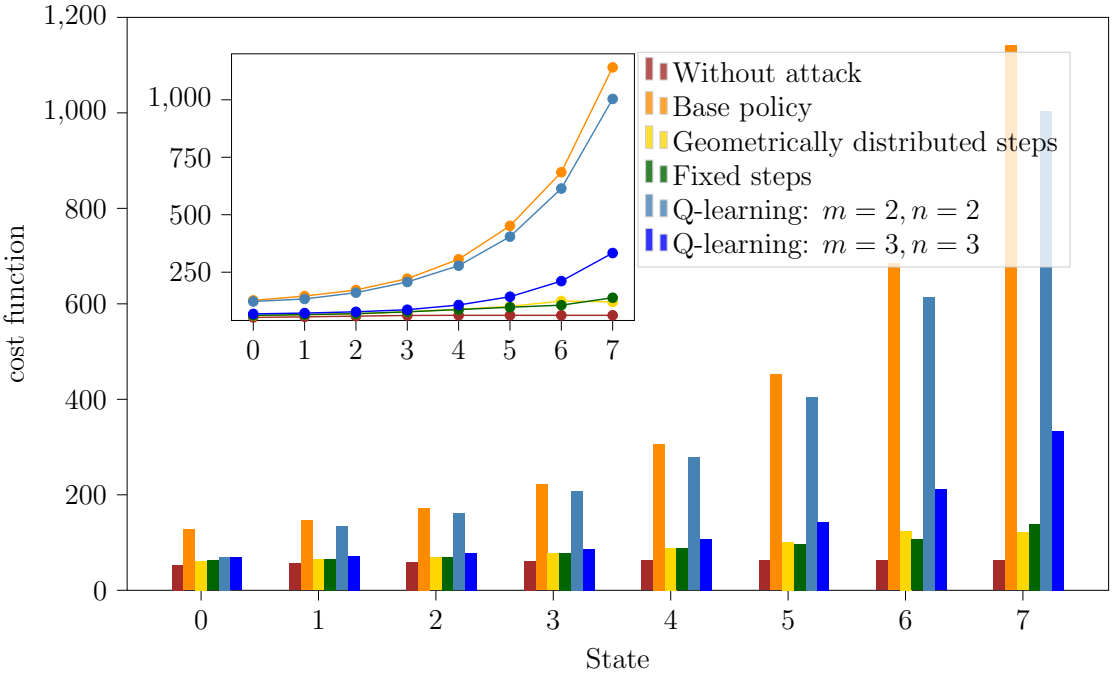

FIGURE 6.4: Performance under different approaches with $\kappa = 0.5$ and $\beta = 0.6$.

In Figure 6.4, the brown bar denotes the optimal value for each state, the dark orange, gold and dark green bars denote the approximate value functions using the base policy, rollout algorithm with geometrically distributed and fixed truncated steps. We can see that the rollout policy performs much better than the base policy. Also, the steel blue and blue bars denote the approximate value functions using Q-learning with different window sizes. It is shown that with the window size increases, the corresponding average cost decreases. This is expected as a better Q-value can be obtained with increased m and n. However, it is worth noticing that this approach is limited by the scale of the problem and system parameters. It shows a comparable performance with the rollout policy in the above figure. However, its performance is largely affected by the choice of state, the dimension of problem and other aspects. It is worth noticing that the performance of the rollout policy is not limited to the choice of base policy. The experiments show

that even the base policy is revised to the following extreme policy:

$$\mu_b(b) = \begin{cases} 1, & \sum_{i=0}^{3} b(i) < \sum_{i=4}^{7} b(i), \\ 0, & \text{o.w.}, \end{cases}$$

the rollout policy still can get a comparable results.

TABLE 6.1: The difference with the optimal value functions under different approaches.

Approach	Difference from optimal
Base policy	333.9849
Geometrically distributed steps	34.2081
Fixed steps	33.9491
Q-learning $m = 2, n = 2$	298.3401
Q-learning $m = 3, n = 3$	59.0104

By computing the stationary distribution of the problem, we can obtain that the $\phi = \{0.2, 0.2, 0.2, 0.2, 0.2\}$ under the designed parameters. We compute the two norm of the difference between optimal value function and other approximate value function from state 0 to state 4, which is shown in Table 6.1. From this table, we can see that the performance of the rollout policy is much better than the base policy.

6.5 Conclusion

This chapter studied the sensor transmission scheduling problem for remote state estimation under integrity attacks. It was proved that the underlying MDP has a threshold type optimal policy. Thus, we simplified the original problem by replacing a truncated state space. When integrity attack is present, the problem was formulated as a POMDP. The existence of optimal policy for the MDP with belief state induced by POMDP was studied and it was proved that the monotonicity of the value function cannot be established via MLR ordering. The main result of this chapter is a suboptimal, online and model-based approach based on the approximation in value space and implemented through rollout with fixed and geometrically distributed truncated steps, and corresponding performance guarantees were provided. Furthermore, numerical examples were provided to demonstrate

the effectiveness of the proposed approaches when compared with a finite history window approach.

Chapter 7

Conclusions and Future work

In this chapter, we summarize the main results of this thesis and outline some possible directions for future work.

7.1 Conclusions

We studied cyber-physical systems under potential cyber-attacks. More specifically, we investigated the performance of cyber-attacks, providing detection mechanisms, and developing feasible countermeasures against cyber-attacks.

In Chapter 3, we analyzed the performance for innovation-based remote state estimation under first-order false date injection attacks. The fundamental limits for innovation-based remote state estimation under linear attacks were characterized. The attacker was constrained to follow a linear attack type based on the past attack signal, the latest innovation and an additive random variable. The optimal attack strategies to achieve maximal performance degradation under a given stealthiness requirement were provided. Then we provided the maximal performance degradation ratio and the corresponding optimal attack strategy to achieve this maximum under strictly stealthy attacks for vector systems, which is a generalization of the previous work. For ϵ-stealthy attacks on scalar systems, the optimal attack strategy with an additive random noise was also presented. It was proven that the maximal performance degradation ration can be achieved without additive noise and the proposed strategy performs better than the existing linear attack strategies

in terms of performance degradation. Simulation results were presented to support the theoretical results.

In Chapter 4, we considered the problem of detection of the replay attacks under the assumption that the system parameters are unknown and need to be identified online. We proposed an algorithm that can simultaneously generate the watermarking signal and infer the system parameters. We proved that our algorithm converges to the optimal one and characterized an upper bound for the almost surely convergence rate.

In Chapter 5, we discussed the problem of detection of flip attacks. We formulated the detection problem as a POMDP by assuming an attack probability. Then, an MDP in the form of SSP was employed to approximate the behavior of the POMDP by fixed-length window and state aggregation of observations. Furthermore, a standard Q-learning algorithm was applied to derive the optimal solution of the approximated MDP. Numerical results were provided to verify the performance of the resulting detector.

Chapter 6 investigated the sensor scheduling problem for remote state estimation under integrity attacks. We proved that the underlying MDP has a threshold type optimal policy. Thus, we simplified the original problem by replacing a truncated state space. The problem was formulated as a POMDP. We explored the possibility of the existence of optimal policy for the MDP with belief state induced by the POMDP and proved that the monotonicity of the value function can not be established via MLR ordering. Furthermore, we proposed a suboptimal, online and model-based approach based on the approximation in value space and implemented through the rollout approach with fixed and geometrically distributed truncated steps, and corresponding performance guarantees were provided.

7.2 Future Work

There are many issues that deserve future research including:

- In Chapter 3, the maximal performance degradation that an adversary can induce by injecting any linear first-order false data injection attacks under ϵ-stealthy attacks for scalar systems was considered. What would be the

performance limits for a vector system and what would the corresponding attack strategy be to achieve this bound? Besides, we assume that the smart estimator transmits innovations based on a Kalman filter to the remote estimator and it is assumed that the attacker knows this. If the sensor decides to transmit either the innovation or the raw measurement at each time, the attacker will need to decide to use which attack strategy under some constraints including but not limited to energy, memory, and computation constraints. How to formulate this interaction between the attacker and the sensor is also worth to be explored.

- In Chapter 4, since we have only provided an upper bound for the convergence rate of the proposed online approach, an open question is how to quantify the exact convergence rate. It is of interest to study secure control in other cases, such as batch-operating process. Besides, the watermarking signal and detector are developed under the assumption that there is no attack during the learning phase. We are also interested in adversarial learning when the sensor data is compromised. Moreover, there are some other techniques that could be considered. For example, we can employ machine learning to estimate the system structure and parameters by training data. It is totally different from the method studied in this thesis.

- In Chapter 6, the research on the implementation of the rollout policy on a large sensor network is a potential research direction. Also, it is of great interest to explore alternatives to the rollout algorithms such as training some approximation architecture including a neural network. Besides, with the development of autonomous driving, how to secure autonomous driving systems with natural uncertainty is getting more and more attention. Furthermore, potential cyber-attacks on these systems increases the uncertainty. Hence, the research on how to deal with uncertainty is particularly important. As a useful tool to counter uncertainty, the rollout approach is worth being studied. The advance of rollout in solving POMDP problem is worth being explored.

List of Author's Publications[1]

Journal Articles

- Hanxiao Liu, Yilin Mo, Jiaqi Yan, Lihua Xie, and Karl Henrik Johansson, "An Online Approach to Physical Watermark Design," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3895-3902, 2020.
- Hanxiao Liu, Yuqing Ni, Lihua Xie, and Karl Henrik Johansson, "How Vulnerable is Innovation-based Remote State Estimation: Fundamental Limits under Linear Attacks," *Automatica*, accepted as Regular Paper.
- Hanxiao Liu, Yuchao Li, Karl Henrik Johansson, Jonas Mårtensson, and Lihua Xie, "Rollout Approach to Sensor Scheduling for Remote State Estimation under Integrity Attack," under review.

Conference Articles

- Hanxiao Liu, Yuchao Li*, Jonas Mårtensson, Lihua Xie, and Karl Henrik Johansson, "Reinforcement Learning Based Approach for Flip Attack Detection," in *Proceedings of the 59th IEEE Conference on Decision and Control*, pp. 3212-3217, 2020.
- Hanxiao Liu, Yuqing Ni, Lihua Xie, and Karl Henrik Johansson, "An Optimal Linear Attack Strategy on Remote State Estimation," in *Proceedings of IFAC World Congress*, 2020.
- Hanxiao Liu, Jiaqi Yan, Yilin Mo, and Karl Henrik Johansson, "An Online Design of Physical Watermarks," in *Proceedings of IEEE Conference on Decision and Control*, pp. 440-445, 2018.

[1]The superscript $*$ indicates joint first authors

Book Chapter

- Hanxiao Liu, Yilin Mo, and Karl Henrik Johansson, "Active Detection against Replay Attack: a Survey on Watermark Design for Cyber-physical Systems," *Safety, Security, and Privacy for Cyber-physical Systems*, *Springer*, 2020.

Bibliography

[1] G. W. Arnold, D. A. Wollman, G. J. FitzPatrick, D. Prochaska, D. G. Holmberg, D. H. Su, A. R. Hefner Jr, N. T. Golmie, T. L. Brewer, M. Bello *et al.*, "Nist framework and roadmap for smart grid interoperability standards, release 1.0," 2010. xxi, 2

[2] C. Lago and C. Trueman, "How singapore is driving the development of autonomous vehicles," 2019. [Online]. Available: https://www.cio.com/article/3294207/how-singapore-is-driving-the-development-of-autonomous-vehicles.html xxi, 3

[3] "What is the stuxnet virus," 2018. [Online]. Available: https://www.quora.com/What-is-the-Stuxnet-virus xxi, 3

[4] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *Proceedings of the 1st international conference on High Confidence Networked Systems*, 2012, pp. 55–64. xxi, 10

[5] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Worst-case stealthy innovation-based linear attack on remote state estimation," *Automatica*, vol. 89, pp. 117–124, 2018. xxi, 3, 17, 30, 38, 39, 40, 41, 43, 44, 45

[6] L. Monostori, B. Kádár, T. Bauernhansl, S. Kondoh, S. Kumara, G. Reinhart, O. Sauer, G. Schuh, W. Sihn, and K. Ueda, "Cyber-physical systems in manufacturing," *Cirp Annals*, vol. 65, no. 2, pp. 621–641, 2016. 1

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

[8] L. Wang, M. Törngren, and M. Onori, "Current status and advancement of cyber-physical systems in manufacturing," *Journal of Manufacturing Systems*, vol. 37, pp. 517–527, 2015. 1

[9] J. Wang, H. Abid, S. Lee, L. Shu, and F. Xia, "A secured health care application architecture for cyber-physical systems," *arXiv preprint arXiv:1201.0213*, 2011. 1

[10] N. Dey, A. S. Ashour, F. Shi, S. J. Fong, and J. M. R. Tavares, "Medical cyber-physical systems: A survey," *Journal of medical systems*, vol. 42, no. 4, pp. 1–13, 2018.

[11] I. Lee and O. Sokolsky, "Medical cyber physical systems," in *Design automation conference*. IEEE, 2010, pp. 743–748. 1

[12] S. Karnouskos, "Cyber-physical systems in the smartgrid," in *2011 9th IEEE international conference on industrial informatics*. IEEE, 2011, pp. 20–23. 1

[13] X. Yu and Y. Xue, "Smart grids: A cyber–physical systems perspective," *Proceedings of the IEEE*, vol. 104, no. 5, pp. 1058–1070, 2016.

[14] C.-C. Sun, C.-C. Liu, and J. Xie, "Cyber-physical system security of a power grid: State-of-the-art," *Electronics*, vol. 5, no. 3, p. 40, 2016. 1

[15] J. Kim, H. Kim, K. Lakshmanan, and R. Rajkumar, "Parallel scheduling for cyber-physical systems: Analysis and case study on a self-driving car," in *Proceedings of the ACM/IEEE 4th international conference on cyber-physical systems*, 2013, pp. 31–40. 1

[16] B. Chen, Z. Yang, S. Huang, X. Du, Z. Cui, J. Bhimani, X. Xie, and N. Mi, "Cyber-physical system enabled nearby traffic flow modelling for autonomous vehicles," in *2017 IEEE 36th International Performance Computing and Communications Conference (IPCCC)*. IEEE, 2017, pp. 1–6.

[17] C. Lv, X. Hu, A. Sangiovanni-Vincentelli, Y. Li, C. M. Martinez, and D. Cao, "Driving-style-based codesign optimization of an automated electric vehicle: A cyber-physical system approach," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 2965–2975, 2018. 1

[18] S. K. Khaitan and J. D. McCalley, "Design techniques and applications of cyberphysical systems: A survey," *IEEE Systems Journal*, vol. 9, no. 2, pp. 350–365, 2014. 1

[19] M. John *et al.*, "Israeli test on worm called crucial in iran nuclear delay." *The New York Times*, 2011. 2

[20] B. Kesler, "The vulnerability of nuclear facilities to cyber attack," *Strategic Insights*, vol. 10, no. 1, pp. 15–25, 2011. 2

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

[22] D. E. Whitehead, K. Owens, D. Gammel, and J. Smith, "Ukraine cyber-induced power outage: Analysis and practical mitigation strategies," in *2017 70th Annual Conference for Protective Relay Engineers (CPRE)*. IEEE, 2017, pp. 1–8. 2

[23] A. Di Pinto, Y. Dragoni, and A. Carcano, "Triton: The first ics cyber attack on safety instrument systems," in *Proc. Black Hat USA*, vol. 2018, 2018, pp. 1–26. 2

[24] J. Monlina Guzmán, "Origen de la falla eléctrica en venezuela," 2019. [Online]. Available: https://www.scribd.com/document/401835067/Origen-de-La-Falla-v2 2

[25] D. E. Sanger and N. Perlroth, "U.s. escalates online attacks on russia's power grid," *The New York Times*, 2020. 2

[26] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Optimal linear cyber-attack on remote state estimation," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 4–13, 2016. 3, 17, 29, 30, 37, 38

[27] Y.-G. Li and G.-H. Yang, "Optimal stealthy false data injection attacks in cyber-physical systems," *Information Sciences*, vol. 481, pp. 474–490, 2019. 3, 17, 38

[28] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *2009 47th annual Allerton conference on communication, control, and computing (Allerton)*. IEEE, 2009, pp. 911–918. 4, 14, 18, 19, 20, 60, 65

[29] M. N. Kurt, O. Ogundijo, C. Li, and X. Wang, "Online cyber-attack detection in smart grid: A reinforcement learning approach," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5174–5185, 2018. 5, 22

[30] Y. Chen, S. Huang, F. Liu, Z. Wang, and X. Sun, "Evaluation of reinforcement learning-based false data injection attack to automatic voltage control," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2158–2169, 2018. 5, 22

[31] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015. 5, 23

[32] V. V. Veeravalli and T. Banerjee, "Quickest change detection," in *Academic Press Library in Signal Processing*. Elsevier, 2014, vol. 3, pp. 209–255. 5, 23

[33] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic bayesian theory of quickest change detection," *Theory of Probability & Its Applications*, vol. 49, no. 3, pp. 458–497, 2005. 5, 23

[34] Z. Guo, J. Wang, and L. Shi, "Optimal denial-of-service attack on feedback channel against acknowledgment-based sensor power schedule for remote estimation," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 5997–6002. 5, 24, 113

[35] K. Ding, X. Ren, and L. Shi, "Deception-based sensor scheduling for remote estimation under dos attacks," *IFAC-PapersOnLine*, vol. 49, no. 22, pp. 169–174, 2016. 6, 24

[36] K. Ding, X. Ren, D. E. Quevedo, S. Dey, and L. Shi, "Defensive deception against reactive jamming attacks in remote state estimation," *Automatica*, vol. 113, p. 108680, 2020. 6, 24

[37] J. Andress, *The basics of information security: understanding the fundamentals of InfoSec in theory and practice*. Syngress, 2014. 9

[38] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015. 10

[39] F. L. Lewis, "Wireless sensor networks," *Smart environments: technologies, protocols, and applications*, pp. 11–46, 2004. 11, 109

[40] NCCIC, "Understanding denial-of-service attacks," 2013. [Online]. Available: https://www.us-cert.gov/ncas/tips/ST04-015 12

[41] V. D. Gligor, "A note on denial-of-service in operating systems," *IEEE Transactions on Software Engineering*, no. 3, pp. 320–324, 1984. 12

[42] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response," in *Proceedings of the 6th ACM symposium on information, computer and communications security*. ACM, 2011, pp. 355–366. 13

[43] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *International Workshop on Hybrid Systems: Computation and Control*. Springer, 2009, pp. 31–45.

[44] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 163–187, 2007. 13

[45] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on scada systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1396–1407, 2013. 14, 18, 19, 23, 65

[46] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 93–109, 2015. 14, 19, 20, 63

[47] Y. Mo and B. Sinopoli, "On the performance degradation of cyber-physical systems under stealthy integrity attacks," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2618–2624, 2015. 15

[48] C. Murguia, N. van de Wouw, and J. Ruths, "Reachable sets of hidden cps sensor attacks: Analysis and synthesis tools," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 2088–2094, 2017. 16

[49] C. Murguia and J. Ruths, "On reachable sets of hidden cps sensor attacks," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 178–184. 16

[50] H. Liu, B. Niu, and J. Qin, "Reachability analysis for linear discrete-time systems under stealthy cyber attacks," *IEEE Transactions on Automatic Control*, 2021. 16

[51] Y. Mo and B. Sinopoli, "False data injection attacks in control systems," in *Preprints of the 1st workshop on Secure Control Systems*, 2010, pp. 1–6. 16

[52] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *49th IEEE Conference on Decision and Control (CDC)*. IEEE, 2010, pp. 5967–5972.

[53] C. Kwon and I. Hwang, "Reachability analysis for safety assurance of cyber-physical systems against cyber attacks," *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 2272–2279, 2017.

[54] Q. Zhang, K. Liu, Z. Pang, Y. Xia, and T. Liu, "Reachability analysis of cyber-physical systems under stealthy attacks," *IEEE Transactions on Cybernetics*, 2020. 16

[55] Y. Chen, S. Kar, and J. M. Moura, "Optimal attack strategies subject to detection constraints against cyber-physical systems," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1157–1168, 2017. 16

[56] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal denial-of-service attack scheduling with energy constraint," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 3023–3028, 2015. 16

[57] J. Qin, M. Li, L. Shi, and X. Yu, "Optimal denial-of-service attack scheduling with energy constraint over packet-dropping networks," *IEEE Transactions on Automatic Control*, vol. 63, no. 6, pp. 1648–1663, 2018. 16

[58] C.-Z. Bai and V. Gupta, "On Kalman filtering in the presence of a compromised sensor: Fundamental performance bounds," in *Proceedings of American Control Conference*. IEEE, 2014, pp. 3029–3034. 16

[59] C.-Z. Bai, F. Pasqualetti, and V. Gupta, "Security in stochastic control systems: Fundamental limitations and performance bounds," in *Proceedings of American Control Conference*. IEEE, 2015, pp. 195–200. 17, 31

[60] ——, "Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs," *Automatica*, vol. 82, pp. 251–260, 2017. 17

[61] E. Kung, S. Dey, and L. Shi, "The performance and limitations of *epsilon*-stealthy attacks on higher order systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 941–947, 2016. 17

[62] C.-Z. Bai, V. Gupta, and F. Pasqualetti, "On Kalman filtering with compromised sensors: Attack stealthiness and performance bounds," *IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6641–6648, 2017. 17, 31, 32, 33, 35

[63] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Worst-case innovation-based integrity attacks with side information on remote state estimation," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 1, pp. 48–59, March 2019. 17, 29, 30

[64] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Consequence analysis of innovation-based integrity attacks with side information on remote state estimation," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 8399–8404, 2017. 17, 30

[65] C. Zimmer, B. Bhat, F. Mueller, and S. Mohan, "Time-based intrusion detection in cyber-physical systems," in *Proceedings of the 1st ACM/IEEE International Conference on Cyber-Physical Systems*. ACM, 2010, pp. 109–118. 18

[66] R. Mitchell and R. Chen, "A hierarchical performance model for intrusion detection in cyber-physical systems," in *Wireless Communications and Networking Conference (WCNC), 2011 IEEE*. IEEE, 2011, pp. 2095–2100. 18

[67] R. Mitchell and I.-R. Chen, "A survey of intrusion detection techniques for cyber-physical systems," *ACM Computing Surveys (CSUR)*, vol. 46, no. 4, p. 55, 2014. 18

[68] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," in *American Control Conference (ACC), 2013*. IEEE, 2013, pp. 3344–3349. 18

[69] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013. 18

[70] S. Weerakkody and B. Sinopoli, "Detecting integrity attacks on control systems using a moving target approach," in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 5820–5826. 18

[71] ——, "A moving target approach for identifying malicious sensors in control systems," in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2016, pp. 1149–1156. 18

[72] J. Tian, R. Tan, X. Guan, Z. Xu, and T. Liu, "Moving target defense approach to detecting stuxnet-like attacks," *IEEE transactions on smart grid*, vol. 11, no. 1, pp. 291–300, 2019.

[73] P. Griffioen, S. Weerakkody, and B. Sinopoli, "An optimal design of a moving target defense for attack detection in control systems," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 4527–4534. 18

[74] R. Chabukswar, Y. Mo, and B. Sinopoli, "Detecting integrity attacks on scada systems," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 11 239–11 244, 2011. 19

[75] A. Khazraei, H. Kebriaei, and F. R. Salmasi, "A new watermarking approach for replay attack detection in lqg systems," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 5143–5148. 19

[76] ——, "Replay attack detection in a multi agent system using stability analysis and loss effective watermarking," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 4778–4783. 19

[77] S. Weerakkody, O. Ozel, and B. Sinopoli, "A bernoulli-gaussian physical watermark for detecting integrity attacks in control systems," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2017, pp. 966–973. 20

[78] B. Satchidanandan and P. R. Kumar, "Dynamic watermarking: Active defense of networked cyber–physical systems," *Proceedings of the IEEE*, vol. 105, no. 2, pp. 219–240, 2016. 20

[79] B. Satchidanandan and P. Kumar, "On the design of security-guaranteeing dynamic watermarks," *IEEE Control Systems Letters*, vol. 4, no. 2, pp. 307–312, 2019. 20

[80] J. Rubio-Hernán, L. De Cicco, and J. Garcia-Alfaro, "Revisiting a watermark-based detection scheme to handle cyber-physical attacks," in *2016 11th International Conference on Availability, Reliability and Security (ARES)*. IEEE, 2016, pp. 21–28. 20

[81] J. Rubio-Hernan, L. De Cicco, and J. Garcia-Alfaro, "On the use of watermark-based schemes to detect cyber-physical attacks," *EURASIP Journal on Information Security*, vol. 2017, no. 1, p. 8, 2017. 20

[82] ——, "Event-triggered watermarking control to handle cyber-physical integrity attacks," in *Nordic Conference on Secure IT Systems*. Springer, 2016, pp. 3–19. 20

[83] R. M. Ferrari and A. M. Teixeira, "Detection and isolation of replay attacks through sensor watermarking," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7363–7368, 2017. 20

[84] ——, "Detection and isolation of routing attacks through sensor watermarking," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 5436–5442. 20, 21

[85] A. M. Teixeira and R. M. Ferrari, "Detection of sensor data injection attacks with multiplicative watermarking," in *2018 European Control Conference (ECC)*. IEEE, 2018, pp. 338–343. 21

[86] R. C. B. Hink, J. M. Beaver, M. A. Buckner, T. Morris, U. Adhikari, and S. Pan, "Machine learning for power system disturbance and cyber-attack discrimination," in *2014 7th International symposium on resilient control systems (ISRCS)*. IEEE, 2014, pp. 1–8. 21

[87] J. Goh, S. Adepu, M. Tan, and Z. S. Lee, "Anomaly detection in cyber physical systems using recurrent neural networks," in *2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE)*. IEEE, 2017, pp. 140–145. 21

[88] Y. Luo, Y. Xiao, L. Cheng, G. Peng, and D. D. Yao, "Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities," *arXiv preprint arXiv:2003.13213*, 2020. 22

[89] X. Ren and Y. Mo, "Secure detection: Performance metric and sensor deployment strategy," *IEEE Transactions on Signal Processing*, vol. 66, no. 17, pp. 4450–4460, 2018. 22, 93

[90] J. Yan, X. Ren, and Y. Mo, "Sequential detection in adversarial environment," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 170–175.

[91] Z. Li, Y. Mo, and F. Hao, "Game theoretical approach to sequential hypothesis test with byzantine sensors," *arXiv preprint arXiv:1909.02909*, 2019. 22, 93

[92] M. Burmester, E. Magkos, and V. Chrissikopoulos, "Modeling security in cyber–physical systems," *International journal of critical infrastructure protection*, vol. 5, no. 3-4, pp. 118–126, 2012. 22

[93] B. Schneier, "Attack trees," *Dr. Dobb's journal*, vol. 24, no. 12, pp. 21–29, 1999. 22

[94] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, 2nd ed. MIT press, 2018. 22

[95] D. P. Bertsekas, *Reinforcement learning and optimal control*. Athena Scientific, 2019. 22

[96] ——, *Dynamic programming and optimal control*, 4th ed. Athena scientific Belmont, MA, 2017, vol. 1. 22

[97] Y. Mo and B. Sinopoli, "Secure estimation in the presence of integrity attacks," *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 1145–1151, 2014. 23

[98] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic control*, vol. 59, no. 6, pp. 1454–1467, 2014. 23

[99] ——, "Secure state-estimation for dynamical systems under active adversaries," in *2011 49th annual allerton conference on communication, control, and computing (allerton)*. IEEE, 2011, pp. 337–344. 24

[100] ——, "Security for control systems under sensor and actuator attacks," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 3412–3417. 24

[101] Y. Shoukry and P. Tabuada, "Event-triggered projected luenberger observer for linear systems under sparse sensor attacks," in *53rd IEEE Conference on Decision and Control*. IEEE, 2014, pp. 3548–3553. 24

[102] ——, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2079–2091, 2015. 24

[103] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach," *IEEE Transactions on Automatic Control*, vol. 62, no. 10, pp. 4917–4932, 2017. 24

[104] L. Shi, P. Cheng, and J. Chen, "Sensor data scheduling for optimal state estimation with communication energy constraint," *Automatica*, vol. 47, no. 8, pp. 1693–1698, 2011. 24, 115

[105] Y. Mo, B. Sinopoli, L. Shi, and E. Garone, "Infinite-horizon sensor scheduling for estimation over lossy networks," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 3317–3322. 24

[106] D. Han, P. Cheng, J. Chen, and L. Shi, "An online sensor power schedule for remote state estimation with communication energy constraint," *IEEE Transactions on Automatic Control*, vol. 59, no. 7, pp. 1942–1947, 2013. 24

[107] S. Wu, X. Ren, Q.-S. Jia, K. H. Johansson, and L. Shi, "Learning optimal scheduling policy for remote state estimation under uncertain channel condition," *IEEE Transactions on Control of Network Systems*, 2019. 25, 119

[108] A. S. Leong, A. Ramaswamy, D. E. Quevedo, H. Karl, and L. Shi, "Deep reinforcement learning for wireless sensor scheduling in cyber–physical systems," *Automatica*, vol. 113, p. 108759, 2020. 25

[109] V. Krishnamurthy, *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge University Press, 2016. 25, 125, 126

[110] G. Tesauro and G. R. Galperin, "On-line policy improvement using monte-carlo search," in *Advances in Neural Information Processing Systems*, 1997, pp. 1068–1074. 25

[111] D. P. Bertsekas, *Reinforcement learning and optimal control*. Athena Scientific Belmont, MA, 2019. 25, 119, 127, 128, 129

[112] D. P. Bertsekas and D. A. Castanon, "Rollout algorithms for stochastic scheduling problems," *Journal of Heuristics*, vol. 5, no. 1, pp. 89–108, 1999. 25

[113] D. P. Bertsekas, J. N. Tsitsiklis, and C. Wu, "Rollout algorithms for combinatorial optimization," *Journal of Heuristics*, vol. 3, no. 3, pp. 245–262, 1997.

[114] F. Tu and K. R. Pattipati, "Rollout strategies for sequential fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 33, no. 1, pp. 86–99, 2003.

[115] N. Secomandi, "A rollout policy for the vehicle routing problem with stochastic demands," *Operations Research*, vol. 49, no. 5, pp. 796–802, 2001.

[116] S. Bhattacharya, S. Badyal, T. Wheeler, S. Gil, and D. Bertsekas, "Reinforcement learning for pomdp: Partitioned rollout and policy iteration with application to autonomous sequential repair problems," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3967–3974, 2020. 25

[117] Y. Li, L. Shi, and T. Chen, "Detection against linear deception attacks on multi-sensor remote state estimation," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 846–856, 2017. 29

[118] A. Ribeiro, G. B. Giannakis, and S. I. Roumeliotis, "SOI-KF: Distributed Kalman filtering with low-cost communications using the sign of innovations," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4782–4795, 2006. 29

[119] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012. 32

[120] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951. 32

[121] J. H. Hubbard and B. B. Hubbard, *Vector calculus, linear algebra, and differential forms: a unified approach*. Matrix Editions, 2015. 41

[122] C.-T. Chen, *Linear system theory and design*. Oxford University Press, Inc., 1998. 59

[123] L. L. Scharf, *Statistical signal processing*. Addison-Wesley Reading, MA, 1991, vol. 98. 62

[124] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Computers & chemical engineering*, vol. 17, no. 3, pp. 245–255, 1993. 77

[125] N. L. Ricker, "Model predictive control of a continuous, nonlinear, two-phase reactor," *Journal of Process Control*, vol. 3, no. 2, pp. 109–123, 1993. 77

[126] Y. S. Chow, "On a Strong Law of Large Numbers for Martingales," *The Annals of Mathematical Statistics*, vol. 38, no. 2, pp. 610–610, apr 1967. 83

[127] D. R. Brillinger, "The analyticity of the roots of a polynomial as functions of the coefficients," *Mathematics Magazine*, vol. 39, no. 3, pp. 145–147, 1966. 86, 90

[128] D. P. Bertsekas, *Abstract dynamic programming*, 2nd ed. Athena Scientific, 2018. 95, 122, 124, 128

[129] D. P. Bertsekas and J. N. Tsitsiklis, "An analysis of stochastic shortest path problems," *Mathematics of Operations Research*, vol. 16, no. 3, pp. 580–595, 1991. 96

[130] S. Whitehead, "Reinforcement learning for the adaptive control of perception and action," Ph.D. dissertation, University of Rochester, 1992. 98

[131] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, King's College, Cambridge, 1989. 100, 130

[132] M. Basseville, I. V. Nikiforov *et al.*, *Detection of abrupt changes: theory and application*. prentice Hall Englewood Cliffs, 1993, vol. 104. 105

[133] Y. Li, D. E. Quevedo, V. Lau, and L. Shi, "Online sensor transmission power schedule for remote state estimation," in *52nd IEEE Conference on Decision and Control*. IEEE, 2013, pp. 4000–4005. 111

[134] Y. Li, D. E. Quevedo, S. Dey, and L. Shi, "Fake-acknowledgment attack on ack-based sensor power schedule for remote state estimation," in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 5795–5800. 112, 113

[135] ——, "A game-theoretic approach to fake-acknowledgment attack on cyber-physical systems," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 1, pp. 1–11, 2016. 113

[136] D. P. Bertsekas, *Dynamic Programming and Stochastic Control*. USA: Academic Press, Inc., 1976. 116

[137] A. S. Leong, S. Dey, and D. E. Quevedo, "On the optimality of threshold policies in event triggered estimation with packet drops," in *2015 European Control Conference (ECC)*. IEEE, 2015, pp. 927–933. 119

[138] K. J. Astrom, "Optimal control of markov processes with incomplete state information," *Journal of mathematical analysis and applications*, vol. 10, no. 1, pp. 174–205, 1965. 122

[139] Y. Li, K. H. Johansson, and J. Mårtensson, "Lambda-policy iteration with randomization for contractive models with infinite policies: Well posedness and convergence (extended version)," *arXiv preprint arXiv:1912.08504*, 2019. 128

[140] K. P. Murphy, "A survey of pomdp solution techniques," *environment*, vol. 2, p. X3, 2000. 130