Detection, Isolation, and Estimation of Cyber-Attacks in Presence of Faults in Cyber-Physical Systems

Reza Bahrevar Fetideh

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Electrical and Computer Engineering) at

Concordia University

Montreal, Quebec, Canada

March 2024

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Reza Bahrevar Fetideh**

Entitled: **Detection, Isolation, and Estimation of Cyber-Attacks in Presence of Faults in Cyber-Physical Systems**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Electrical and Computer Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Shahin Hashtrudi Zad

_____ Examiner
Dr. Shahin Hashtrudi Zad

_____ Examiner
Dr. Jun Yan

_____ Supervisor
Dr. Khashayar Khorasani

Approved by _____
Dr. Yousef Shayan, Chair
Department of Electrical and Computer Engineering

_____ 2024

_____
Dr. Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Detection, Isolation, and Estimation of Cyber-Attacks in Presence of Faults in Cyber-Physical Systems

Reza Bahrevar Fetideh

The security of Cyber-Physical Systems (CPS) has been the center of attention in the past decades. Developing methodologies to detect or estimate cyber-attacks on sensor measurements and actuator inputs is essential for ensuring the safe and reliable performance of these interconnected systems of systems. Considering the stealthy nature of cyber-attacks, combined with potential faults, additional challenges emerge, which this thesis addresses through the lens of control theory.

In control theory, several methodologies have addressed the decoupling of unknown inputs, such as faults and disturbances. However, the simultaneous presence of faults and cyber-attacks presents challenges that are not fully developed in the context of CPS. The first part of this thesis proposes a methodology, consisting of the construction of two plant-side monitoring filters to detect and isolate faults and cyber-attacks including covert and zero dynamic attacks. The findings are supported through analytical and simulation studies.

As the second challenge, a multi-rate approach is employed to estimate actuator cyber-attacks in the presence of sensors and actuators faults in CPS. A plant-side fault monitoring filter is augmented with the physical system, and its residual, along with the plant's outputs, is sent to C&C with a specified mechanism. Cyber-attacks are isolated from faults through the received information in the C&C and a secondary observer. Consequently, a delayed Unknown Input Observer (UIO) is constructed to estimate the actuator cyber-attack. The effectiveness of the proposed methodology is evaluated through numerical case studies.

Acknowledgments

Firstly, I want to thank Professor Khorasani for teaching me to have a critical view and guiding me toward a fundamental learning and research structure.

Secondly, I want to thank all my friends and loved ones who supported me through my studies.

Last but not least, I want to express my gratitude to Dr. Amir Baniamerian, who guided me and to whom I owe most of my theoretical and mathematical understandings in geometric and algebraic analysis of control systems.

Table of Contents

List of Figures

List of Tables

List of Acronyms

CPS Cyber-Physical Systems

UAV Unmanned Aerial Vehicle

PSF Plant Side Filter

CASF Cyber-Attack Sensitive Filter

CSF Control Side Filter

PTC Positive Train Control

KFF Kalman Fault Filter

ASF Attack Sensitive Filter

AKFF Auxiliary based Kalman Fault Filter

FDI Fault Detection and Isolation

PFTC Passive Fault Tolerant Control

AFTC Active Fault Tolerant Control

UIO Unknown Input Observer

ZDA Zero Dynamics Attacks

FP False Positive

FN False Negative

TP True Positive

TN True Negative

DoS Denial of Service

DDoS Distributed Denial of Service

Chapter 1

Introduction and Literature Review

Advancement and breakthroughs in the three realms of science, including system theory, communication networks, and processing systems have enabled the manipulation of machinery and information. Developments in information technology and digital systems accompanied by the advent of distributed embedded sensing, processing, and control made a revolutionary impact on various domains such as science, transportation, energy, and medical systems [32, 50]. The distributed embedded systems established the framework for the integration of control, communication, and computing, and the capacity for more efficient and reliable systems like intelligent power distribution [101] and autonomous vehicles have increased. These types of network-reliant systems are referred to as Cyber-Physical Systems (CPS).

As the field of CPS encompasses the diverse realms of science, it also absorbs all the obstacles that pertained to them. CPS are subject to new threats by adversaries with financial, harmful, and malicious motivations that would endanger the reliability and safety of these systems.

1.1 Security Issues in CPS

The reliability of CPS is entangled with the confidentiality, safety, and security of both physical and cyber parts. To detect and elevate the abnormality, it is beneficial to understand some of the problems causing security breaches in cyber-physical systems. The data in the CPS network

(a) (b)

Figure 1.1: Examples of CPS: Left an Unmanned Aerial vehicle [84] (a) and right side a Positive Train Control (PTC) over a communication network [21] (b).

is transferred through packets of information with a specified destination, which is exposed to tremendous dangers. The term packet sniffer is used to describe software and hardware that are the cause of such threats. Wireshark, TCPDUMP, and soft perfect network analyzer are a number of tools available for spying in the communication networks [86]. While non-encrypted communications are an easy target for packet sniffers, through the use of deciphering methods even encrypted communications are not immune from leakages of data.

Deciphering encrypted communication between embedded systems is not a simple job, and it is neither inconceivable. One of the common ways to access these encrypted information is side-channel attacks, such as retrieving intelligence about the cryptographic operation, which can be used for calculating the encryption key [7]. After gathering enough information about the system, an adversary can perform a form of cyber-attack called the man-in-the-middle attack.

Security breaches such as packet sniffers, manipulation of physical components, denial of service, man-in-the-middle attacks, side-channel attacks, and malicious code injection are causing serious concerns in maintaining the safety of CPS. From the control point of view, these security

breaches directly or indirectly result in disruption or manipulation of the sensor measurements or control outputs. For example, the adversary can utilize packet sniffers to gather information about the system and design an offline cyber-attack [14, 32]. Adversaries with malicious or financial motives are persistently looking for unprecedented ways to breach the CPS systems, which signifies the importance of devising robust and reliable schemes to counter cyber-attacks. The ultimate solution leading to reliable system performance is plausible when an intelligent system is conceived that can distinguish malicious abnormalities from the system's faults.

As a real-world example, cyber-attacks on Maroochy Shire Council's sewage control system in Queensland [91] can be mentioned, where they caused a lot of confusion and false alarms, such as flooding nearby hotel, park, and the river. Overall, through cyber-attacks, the adversary aims to utilize his knowledge and access (or sometimes both) of the physical system, control, and communication network to target the vulnerabilities in the CPS, and the objective of this thesis is to propose methodologies that can counter the threats posed by them. In this thesis, tackling these problems from the viewpoint of control theory is of main interest.

Two of the preliminary term in this topic are cyber-attacks and adversary. Here, an adversary is someone who wants to endanger the availability, reliability, and integrity of the control system. A cyber-attack is malicious data through which the attacker compromises the integrity of the outputs or command inputs data. The attacker intends to target the vulnerabilities to deviate the states of the system, create misleading data, and evade detection by the monitoring system.

Geneally a the defender must design a methodology:

- Protecting integrity and confidentiality
  - ✓ Detection and isolation of adversary attacks.
  - ✓ Minimizing adversaries' access (or revealing him/her) by cryptographic techniques.
- Ensuring robustness
  - ✓ Ensuring the detectability of adversary attacks by introducing practical constraints on the communication structure of cyber-physical systems.

- ✓ Being able to recover the compromised state of the system (resilient estimation).

- Increasing safety and reliability

  - ✓ By designing protective measures that ensure detection and isolation of the attack w.r.t faults in the system.

1.2 Motivation

Cyber-physical techniques in real-world examples come in a variety of examples such as Positive Train Control (PTC) and Unmanned Aerial Vehicle (UAV). When dealing with positive train control, the movement of the train must be monitored through each step of the operation. and this supervision can take place through the C&C by consideration of traffic and all the operating trains or on the plant side. In dealing with PTC, how can the faults in the C&C be monitored while it is ensured that no stealthy cyber-attack can compromise the safety of the plan-side operation?

Convectional fault detection and isolation methods are unable to differentiate between faults and cyber-attacks [28]. Observers developed based on decoupling methods such as Unknown input observers [25], parity-based observers [79], or other observers such as Kalman filter are unable to decouple fault and cyber-attacks with their conventional design, since the residual generated through Kalman filter or decoupling methods such as Unknown Input Observers (UIO) can generate identical symptoms.

In this work, the first objective is to develop a methodology for which the fault detection and isolation methods can answer to the new demands for the safety of cyber-physical systems and second objective is to suggest cyber-attack estimation methodology that can eventually lead to faults and cyber-attacks tolerant control systems.

1.3 Type of Attacks

Cyber-attacks on control systems are classified into two branches: deception and Denial of

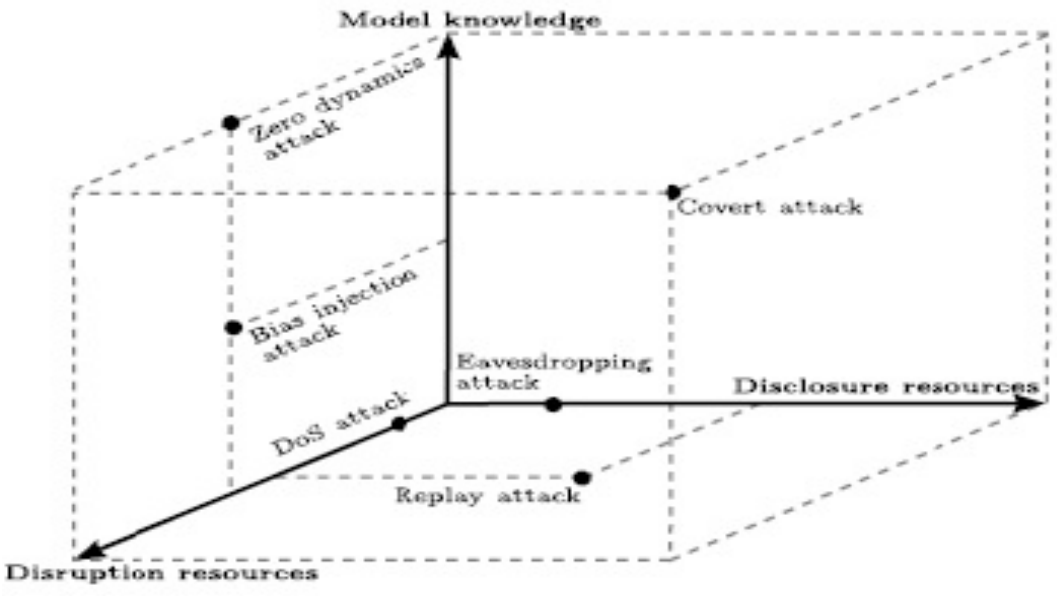

Figure 1.2: Different types of cyber-attacks against control systems are categorized based on their disruption power, need for internal information about the system, and the difficulty of their detection [98].

Service (DoS) attacks. Denial of service cyber-attacks is an example of disruption in the communication networks, where the adversary jam the communication link to interfere with transferred information to/from C&C [2]. On the other hand, the intent of the attacker from performing deception attacks is to compromise the integrity of control signals or sensors. Through deception attacks, the adversary manipulates the transferred data and causes unwanted behaviour while remaining undetected. Unlike DoS attacks, deception attacks are harder to detect, and due to their stealthiness, they are harder to implement. Therefore, the focus of this thesis is directed toward deception attacks. Replay, false data injection, covert, and its subclass zero dynamics attacks are commonly studied classes of deception attacks, where the focus of this study is mainly toward zero dynamics, covert, and false data injection attacks. In the following, some of the most recognized attacks based on their intelligent design, stealthiness, and potential threats that exist in the literature are mentioned:

1.3.1 Replay Attack

Incidents such as Stuxnet worm [54] are an indicator of the importance of finding ways to deal with replay attacks. In the case of a replay attack, the attackers hijack sets of sensors for a while, and then they replay the recorded signals to the command and control side, meanwhile the attackers send malicious control commands toward actuators to diverge the system from its desired path.

Physical watermarking is one of the first proposed methods against replay attacks [69]. Based on this method, the defender adds a zero-mean authenticating watermark inputs with a certain covariance to the main command inputs while assuming that the attacker does not know about the watermark. Therefore, in case of a replay attack, the adversary is not aware of the change in the distribution of the control inputs, and by replaying previous measurements, a detector such as the Kalman filter can identify it in the system. By using the watermark, we are essentially imposing a trade-off such that the defender is perturbing the command inputs and compromising the efficiency and safety of the system. In [43], instead of watermarking the inputs, a coding identifier on the output is considered as the main contribution for the detection of the replay attacks. An alternate control design strategy is proposed in [114], which provides a resilient control solution based on limiting the number of consecutive outputs that the attacker can repeat. Furthermore, a recently published paper discusses the simultaneous analysis of replay attack and physical fault in the system where an adaptive estimator is designed to discriminate the replay attack from faults [111].

1.3.2 False Data Injection Attack

False data injection is another class of deception attacks, and several studies in the literature considered these attacks regarding the situation where the attacker has partial information about the system[106], and a common form of these attacks are called bias injection attacks. This cyber-attack can target the inputs, outputs, or both of these channels [19, 27, 98]. This cyber-attack aims to successfully insert a data injection sequence under which the system is affected by maximum perturbation that causes the states of the system to diverge from desired values but simultaneously hides its effect from the detector. Coding the outputs sequence [66] is one of the suggested ways

of encountering these cyber-attacks. Some of the recent works have established robust and resilient ways for encountering these cyber-attacks [27]. Also, from a different perspective, different methods have been established that investigate the physical and economic impact of these attacks on the applications, such as the smart grid [52, 58].

1.3.3 Perfectly Undetectable Cyber-Attacks: Covert and Zero Dynamics Attacks

A cyber-attack is called a perfectly undetectable attack where the effect of a cyber-attack is completely removed from the measurements. The perfectly undetectable attack can be performed through zero dynamics attacks and covert attacks [8, 10].

The covert attack is the main and more general case of integrity attacks against the control systems, which usually assumes the attacker has the knowledge of system dynamics. The covert attacks can compromise the inputs in the control system and hide its effect by measuring and subtracting the inputs contribution from the outputs. There are studies on the design strategies for covert attacks, which do not require the perfect knowledge of the system [39]. According to [39], covert attacks can be accomplished by eavesdropping on the sensor and command inputs before implementing the attack. Most of the work in the literature addressed the issue by taking away a degree of information about the system from the attacker and designing a strategy that can reveal them [105].

Another important perfectly undetectable attack is called the zero dynamics attack. The attacker's objective is to design an output zeroing input. Since the effect of the attack on the output is zero, the detector would be unable to recognize any trace of that in the residuals of the system [20]. The difference between zero dynamics attacks and covert attacks is the access to the output channel by the adversary in the latter case.

Introducing the auxiliary system is one of the suggested ways in the literature to deal with the covert and zero dynamics attacks in networked control systems[30, 88, 104]. This method explores the idea of adding multiple switching auxiliaries on the site of the main process. The auxiliary

systems are authenticating virtual processes that possess specific properties. They are coupled with the main system's inputs and outputs and together form an augmented system. Designing an observer for the augmented dynamic while assuming that the attacker does not hold the perfect knowledge of the auxiliary (a certain point in time due to the switching auxiliaries) is the suggested way of this thesis to encounter the covert attacks. Implementation of this method requires to fulfil the assumption of synchronization between the system side and the C&C (it is the same as the observer side).

The study in [37] suggested an inputs modulation as the detection strategy against covert and zero dynamics attacks in the networked control systems and considers two scenarios: a fixed modulation matrix[1] or a varying modulation matrix. In the second case, the main advantage is the increased level of difficulty for the attacker to estimate a varying modulation, and as a consequence, the method is difficult to implement, which is due to the need for synchronization so that the observer and control law is updated based on the latest modulation matrix that multiplies the command inputs on the system side. Furthermore, in the first case, if the defender chooses to use static modulation, it will be easier for the attacker to break the defender's plan, which can jeopardize the security of the system.

The studies in [97, 111] are amongst the few works that addresses faults and stealthy attacks in the view of detection and isolation, where a decoupling methodology is developed that can identify which sensor or actuator is under fault or cyber-attack.

1.4 Faults and Their Importance in Dealing with Cyber-Attacks

In [45], faults are defined as "unpermitted derivation of one of the characteristics of the systems ". Then what is the difference between faults and attacks, as the attacks are also a form of unpermitted deviation? The main difference is the stealthiness and intelligent design of attacks. A fault does not try to hide from the observer, and chances that a fault signal can be described by a

[1]Modulation: An inputs modulation is a matrix multiplication of the inputs of the system for which it results in the change of mapping direction of inputs to states

signal based on the invariant zeros of the system are almost non-existence. In the following, first, the fault-related problems and their importance for studying cyber-attacks are explained.

1.4.1 Introduction to Faults

Detection, isolation, reconfiguration, and recovery are some of the important problems of control theory, and it is essential to examine them in order to tackle the problems concerning cyber-attacks in CPS. Detection and isolation of faults in control theory (Fault Detection and Isolation (FDI)) are explored by finding the abnormalities through symptoms and behaviours of the system in response to different inputs [45], and for this purpose model-based or knowledge-based methodologies can be applied.

Model-Based Fault Diagnosis and Fault-Tolerant Problem

One method to generate the symptoms for fault detection and isolation is through dynamical observers such as Leuenberger [44] and Kalman filter [12], by comparing the sensor measurements with the estimated outputs of these observers. The next step after residual generation is residual evaluation, where a decision is made according to created symptoms, corresponding to the presence of faults in the system, where algorithms such as threshold checking, maximum likelihood, and $\mathcal{X}^2$ are implemented for the processing the residual. If the evaluated residual leads to a unique feature that corresponds to sensor faults, actuator faults, or system faults, the isolation of faults are achieved. Methodologies such as parity approach [63], unknown input observer [25], and eigenstructure [4] are popular in the view of decoupling-based approaches, and there are H_∞ based approaches that consider disturbance mitigation as a solution in detection and isolation faults [87].

Decoupling-based methodologies follow certain criteria that are strongly related to fundamental concepts, such as strong observability and strong detectability [34], and the goal is to determine whether with the knowledge of systems structure is possible to uniquely recover the inputs information from outputs. On the other hand, by disturbance mitigation, the intent is to look for ways to design the fault detection and isolation filter such that sensibility of the design magnified in fault's

effect on residuals, in comparison to disturbance effects.

Ideally, it is preferred to maintain the physical safety of the faulty plant by decreasing the consequence of the faults and improving the reliability, safety, and performance of the system, and that's where the fault-tolerant control surfaces. The recovery after fault is often achieved either with robust passive methodologies (Passive Fault Tolerant Control (PFTC)), where recovery plan is predetermined, or algorithms that focus on control reconfiguration, which are also called Active Fault Tolerant Control (AFTC), such as adaptive or switching control laws [44, 112].

Knowledge-Based Fault Diagnosis

In fault detection and diagnosis, some methodologies are not reliant on the dynamical observer, but information processing through the inputs and outputs. These methodologies are often combined with expert knowledge of the system to generate assessable symptoms based on recorded abnormal events [45]. Generated symptoms can be considered raw data, feature extraction based on analysis of the raw data or frequency analysis of data, or a hybrid combination of available symptoms combined with expert knowledge of the system. Generating symptoms is considered as part of data preprocessing, while data processing usually includes implementation of an algorithm such as logic-based methods [94], machine learning [55], neural network [102] to make sense of the symptoms. Most of the methodologies around the concept of knowledge-based fault diagnosis follow categorizing faults based on symptoms and training an AI processing technique for learning and predicting future similar events, while learning can be through two categories of online or offline training[45]. Knowledge-based methodologies have received criticism over the designer's bias and algorithm bias that can cause over sensitiveness and inaccurate results.

Among knowledge-based fault diagnosis for control systems, the most practical methods involve a hybrid combination of expert knowledge and data mining, or methodologies, where distinguishable frequency characteristics can be extracted from the data. In this regard, fault diagnosis of bearing in rotary machinery is one of the most discussed case studies in the last decades, and many methodologies by the implementation of wavelet and Fast Fourier analysis tried to address

the different of faults bearing based on the expert knowledge of model for which different faults in bearing accrues [60]. Methodologies such as amplitude modulation detection based on power spectrum analysis [41], neural network-based wavelet analysis [53], and machine learning-based methodologies are some of the most discussed topics in literature.

Cyber-attacks and faults signals can have similar characteristics. The attack signal can be designed by the adversary to imitate any type of fault, and consequently, they can impact the outputs of the system similarly to faults. Therefore, due to similarities between the faults and cyber-attacks impact on the output knowledge-based methodologies may not guarantee the isolation of cyber-attacks and faults.

1.4.2 Multi-Agent Systems and Faults

Many studies have been conducted in the fault domain around multi-agent systems by considering topics such as obstacle avoidance [23], distributed estimation [38], distributed control [65], and distributed control and estimation [61]. In each of the aforementioned topics, the presence of fault can obstruct the consensus or the formation of agents. Furthermore, strategies with the concept of fault tolerance are also relevant in the context of multi-agent systems [26], however, the rationale behind the recovery procedure is situational and more subjective to designers' ideas, and it can cause the loss of an agent. The recovery plan can differ based on the expectancies of the performance, objectives, control recovery, and mission robustness, where it is also dependent on the communication protocol for the direct or indirect graph. Studies have been successful in fault-tolerant control and estimation of the multi-agent system through passive or active strategies, however, the challenge in the context of cyber-attack and faults is yet to be addressed[2].

[2]This topic is out of scope of this thesis

1.5 Methodologies and Their Effectiveness to Counter Cyber-Attacks

Auxiliary-based methodologies are unique approaches in tackling cyber-attacks that are investigated through event-triggered switching-based [88], and time-varying mechanisms [104]. Auxiliary-based methodologies are proven to be effective against covert and zero dynamics attacks, while there has been no discussion on their effectiveness against replay attacks. Utilizing auxiliaries involves expanding plant side dynamic by a computer-generated secondary system such that designers are flexible in choosing the auxiliaries inputs, while limited in its number of outputs depending on network loads and plants limitation such as battery consumption. The potential for the isolation of cyber-attacks and faults through auxiliary systems in the C&C has yet to be explored.

Multi-rate consideration of inputs or outputs rates is another approach that deals with zero dynamics attacks, where the designer eliminates the sampling zeros of the system by considering a fast output rate for the sensors. Fast outputs multi-rate strategy results in a strongly detectable system [36]. While solely dealing with zero dynamics attacks, the multi-rate strategy has been used in attack estimation [36] and resilient control [77]. The fast rate sampling based attack estimation is also effective against bias data injection attacks, while it is not tested against replay or covert attacks. Another way of detecting zero dynamics attacks is discussed in [6], which explores an intermittent zero hold rate mechanism. These methodologies by manipulating the sampling rate of the system, are trying to change the mapping function between the inputs and outputs such that there exists no non-zero inputs that its evolution locates in the kernel space of the outputs. The multi-rate idea is already been explored in the fault diagnosis domain for many years, while various examples of Leuenberger-like observers [113] and Unknown Input Observers (UIO) [57] are explored for this specific type of sampling to help with the residual generation and isolate faults in the system, however, they are currently no studies that address the simultaneous presence of faults and cyber-attacks.

Furthermore, there are reachable space optimization-based methods that search for cyber-attacks

in the entire set of estimates [31]. Secure estimation under adversarial attack is an example of these works [29], where an observer is constructed that searches the entire space of data through a highly complex optimization problem to correctly estimate the entire state of the system at all times, while a Leuenberger observer converges to the estimated value after a period of time. While adding to the complexity, these methods do not have significant protection over traditional observers.

Set membership-based approaches are another type of unique work in the literature in dealing with the problem of attacks. In [59, 73], an ellipsoidal set membership-based estimate to "detect the attacks on the outputs or through the control inputs channel" [73], where it is defined in the framework of the networked control system and provides an estimate of the system based on the set-theoretic estimation [22], by considering noise distribution in certain sets that are Unknown But Bounded (UBB) [3, 24, 73]. Overall set membership approaches are extremely situational, and while computationally expensive, have yet to bring a noticeable advantage compared to traditional observers in countering faults and attacks.

Knowledge-based methodologies against attacks are limited to machine learning-based and neural network-based cyber-attack detection [80]. There is also limited work concerning faults and attacks in the cyber-physical system using data-based analysis, while it is not formally proven how these methodologies can validate their approach to effectively separate the space of faults and attacks [99]. Overall, as of now there is no concrete evidence of isolation of faults and attacks in the feature space.

1.6 Four Aspects in Dealing with Simultanous Faults and Cyber-Attacks Problems

1.6.1 Achieving Security Through Zero Analysis

Analyzing the system through its invariant zeros in algebraic view or equivalently, the nulling space of the inputs of the system in the geometric point of view, is one of the main tools that is

utilized in this thesis for dealing with attacks in CPS.

Why is the analysis of the security through invariant zeros important? Throughout the years, many papers have been trying to address a type of attack that is called the perfectly stealthy attack [98]. As an example, a perfectly stealthy attack called zero dynamics attacks can be designed with the invariant zeros of the system, which denotes their effect is completely decoupled from the outputs of the system. By analyzing zero dynamics attacks or covert attacks, the type of loopholes that the attacker can use to remain completely stealthy from the potential detector are investigated. This viewpoint is addressed in several papers through the concept of the security index [9, 35, 67]. Security index implies in the perspective of a defender, which actuator or sensor should be protected to prevent a perfectly stealthy attack [67]. Analyzing the security index is beneficial to prevent worst-case attacks. Researchers constructed other types of attacks, such as the one mentioned in [19], while not perfectly stealthy, that are still considered a significant threat. In [19], cyber-attacks are designed with a running optimization problem that the attacker has to continuously optimize so that its effect on the residue is below the alarm threshold of the system.

This analogy has been used in multiple works to take away the perfect knowledge of the attacker and reveal stealthy attacks such as zero dynamics attacks. Usage of axillary systems [104], or inputs modulation [37] is a defensive action that researchers proposed for changing the mapping function from inputs to outputs.

Recently, researchers have paid attention to the multi-rate sampled data system and how changing the speed of data sampling in outputs or received inputs would affect the attackers' ability to find attack sequences that can remain stealthy [77]. Furthermore, the study of fault diagnosis through a multi-rate system has already been investigated in several works [57, 90], and it has been extensively studied for designing a stable control [18, 49, 108, 109]. However, since this application has the ability to change the location of zeros in the sampled data control system, recently, it has been utilized for increasing the security in cyber-physical systems through fast-rate sampling. This idea has also been used for the estimation of the zero dynamics attacks [36]. The real value of this work is that every possible combination of attack will affect the outputs response, and now the

adversary has to increase its design effort to find a sequence for which the attack will not raise an alarm.

1.6.2 Simultanous Analysis of Faults and Cyber-Attacks

Besides the zero analysis, methodologies beyond the existing literature that can isolate the residual effect of a fault with respect to an attack need to be established. Conventional fault isolation methodologies are proven to be effective to isolate and locate actuator faults and sensor faults. In this regard, one of the important theories called as detectability theory of the fault is established by Nikookhah [78], which compares image space of the fault with respect to a secondary unknown inputs such as disturbance. The detectability theory in [78] is the basis of many decoupling-based observers, such as UIO and parity-based observers, where one can design an observer that is not too sensitive to a specific unknown inputs. When analyzing faults and attacks, one might notice the intersection of the image space of these two unknown inputs, therefore, finding a definite answer to whether the nature of the presented anomaly is related to faults or attacks is not possible. Fault analysis commonly starts with fault detection [45], then fault isolation [45], and finally the mitigation problem and recovery [44, 112]. However, how can the same problem through the lens of faults and attacks be defined and analyzed?

1.6.3 Dealing with Cyber-Attacks in Single-Agent or Multi-Agent CPS

Depending on dealing with single-agent systems or multi-agent systems, the perspective on the fault and attack isolation problem will be different. Given single-agent systems, a network control system such as UAV, where there are certain established inputs or outputs for the agents can be considered. However, in the case of multi-agent, there can exist numerous agents with different channels for sent or received information, different communication topologies in form of directed or undirected graphs, and a variety of control law or observer protocols [110]. Therefore, a methodology developed for a single-agent system might not necessarily be suitable in a multi-agent case and vice versa.

Several work toward attack detection in multi-agent systems have been established in recent years such as [96] and [68]. In [27], a detection and mitigation strategy for mitigating biasing attacks in sensor networks is presented. In [96] , analysis toward understanding and detecting attacks in multi-agent systems is provided. In [28], a fault and attack detection strategy for the multi-agent system is provided by using a Markovian approach , however, a methodology that decouples fault and cyber-attacks has yet to be established.

1.6.4 Practicality of the Methodology

Depending on the problem in terms of single-agent system or multi-agent system, the rate of data transfer in the overall sampled data system, energy consumption, and computational complexity should be considered. One of the major improvements in attack detection can be the utilization of the event-triggered schemes to make the methodology more suitable for real-world scenarios [96].

Another question is that what is the next step after the detection and diagnosis when dealing with faults and attacks? In case of faults, one of the main concerns is the resilient estimation of the states of the system, which is one of the first steps and methodologies that can be used towards fault tolerance in the system. Resilient estimation is based on estimating the true state of the system [34], which denotes estimating them without knowledge of the unknown inputs.

The literature encompasses a variety of approaches toward resilient estimation. One of these approaches considers strong detectability as the dominant condition for resilient estimation, meaning if there is a way to achieve strong detectability, estimation of the true state would be possible. Alternatively, when dealing with any type of unknown inputs, by utilizing restrictive assumptions on the model of unknown inputs, one can also achieve its recovery. For example, [27] suggests a restrictive model assumption on the attacks for the recovery of true state in multi-agent systems, however, it neglects the presence of faults. The aforementioned problem by [27] is not solved for a situation dealing with both faults and attacks, therefore, the assumption of the knowledge about the attack model still can be improved.

1.7 Thesis Contributions

1. In Chapter 3, a methodology is presented that enhances the security of Cyber-Physical Systems (CPS) against covert and zero dynamics attacks. It will be formally proven that under certain conditions, the presented methodology is immune to zero dynamics and covert attacks. Furthermore, in this chapter, cyber-attacks and faults will be isolated from each other from the plant side perspective.

2. In Chapter 4, a multi-rate methodology for the estimation of cyber-attacks and faults is presented. This chapter, for the first time, presents a fault and attack problem in a multi-rate framework. This chapter also establishes a cost-effective strategy for the estimation of cyber-attack, including zero dynamics and bias injection attacks with the presence of faults in the multi-rate framework.

1.8 Thesis Layout

1. Chapter 1 provides a literature review of cyber-attacks (Section 1.3) and faults (Section 1.4) and appropriate actions against them in the context of model-based and knowledge-based methodologies (Section 1.6, 1.4, and 1.5) and some of the other aspects in dealing with simultaneous cyber-attacks and faults.

2. Chapter 2 provides background information on sampled data systems (section 2.1) and cyber-physical systems. CPS can be considered as a type of networked sampled data systems. Furthermore, this chapter reviews some of the definitions regarding faults and cyber-attacks in control systems 2.3. Lastly, some of the concepts regarding the cyber-attack and detectors are reviewed 2.5.

3. Chapter 3 begins with an introductory (Section 3.1) of CPS and related works regarding the detection of cyber-attacks in CPS, faults, and the importance of their isolation from the cyber-attacks. This chapter continues with a problem statement (Section 3.2) regarding the

isolation of faults and cyber-attacks in CPS. In this chapter a theory is developed for the isolation of faults and cyber-attacks in CPS, which is backed by numerical example (Section 3.5), comparative study (Section 3.7), and simulation of the comparative study (Section 3.6).

4. Chapter 4 studies the estimation of actuator cyber-attacks in CPS with the presence of sensor and actuator faults through a multi-rate-based methodology. First (Section 4.1), this chapter includes an introduction to the multi-rate-based methodology and how it can be used to detect and estimate cyber-attacks. Next, the problem in the framework of an auxiliary-based approach for achieving the cyber-attack estimation is formalized in Section 4.2. In the Section 4.3 of this chapter, we provide the solution to this problem. Finally, we provide a numerical simulation demonstrating the estimation of actuator cyber-attack in the presence of faults.

5. Chapter 5 is the conclusions and future works.

Chapter 2

Background Information

2.1 Sampled Data Systems [17]

Sampled data systems are a combination of both continuous time and discrete time signals. "Sampled data systems while operating in continuous time, while some of their continuous time signals are sampled at certain time instants (usually periodically) [17]."

The topic sampled data systems in control theory is a subdomain of digital systems, where the focus of the problem is not the quantization effects nor the issues regarding the real-time software [17].

Where G is the generalized system consisting of physical components such as the main dynamical system, sensors, and actuators. r, d, u, and y are continuous time signals, for which r is an exogenous input including reference command, sensor noise, and disturbances, and d desired signal to be controlled. ϕ and ψ are digital signals, μ is a microprocessor or processing unit of digital computer.

In general a single sampled data control system can be shown according to Fig. 2.1 that includes

1. G is the generalized system consist of physical components such as the main dynamical system, sensors, and actuators.

2. r, d, u, and y are continuous time signals, for which r is an exogenous input including refer-

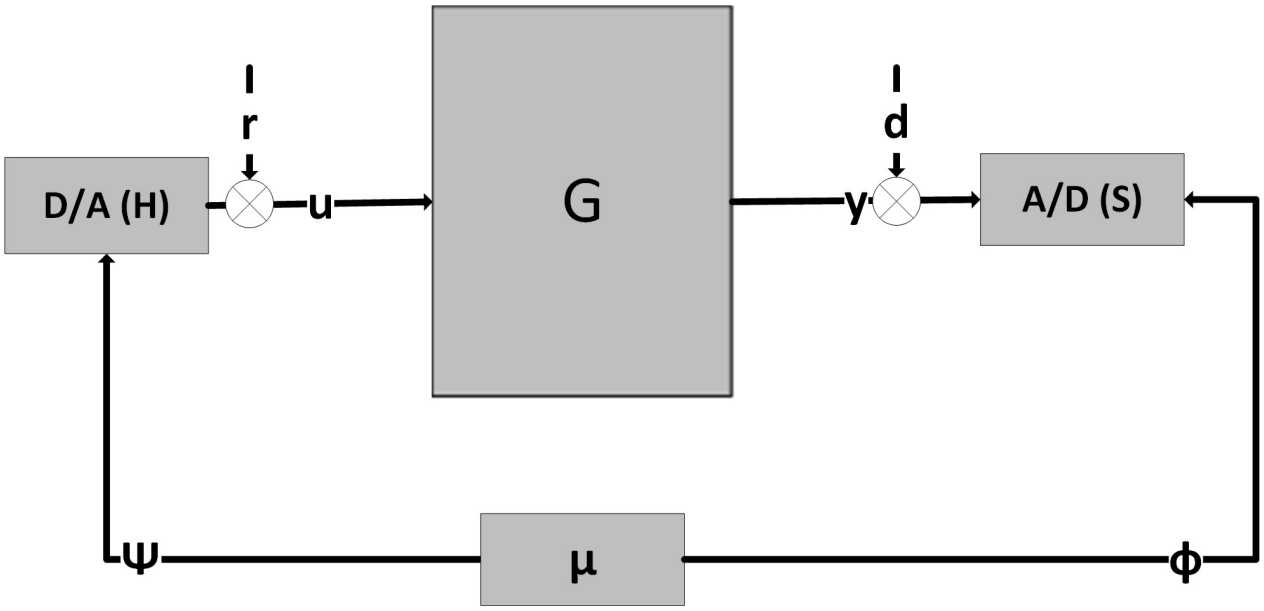

Figure 2.1: A simple representation of sampled data control system.

ence command, sensor noise, and disturbances, and d desired signal to be controlled.

3. ϕ and ψ are digital signals.

4. μ is a microprocessor or processing unit of a digital computer with a control algorithm K_d.

5. In general, for multi-inputs and multi-outputs signals, $A\backslash D$ is called analog to digital converter.

6. In general, for multi inputs and multi-outputs signal, D/A is called digital to analog converter.

2.1.1 Single-Rate Systems [17]

For the single rate system sampling operator S and hold operator H can be considered in one of the following situations:

- S is the ideal sampler with a sampling period of T and it will periodically samples the outputs for which $\phi(k) = y(kT)$, and $k = 0, 1, ...$

- H is the hold operator, which is tasked with the conversion of discrete time signal ψ to continuous time.

For the single rate system control law consists of three components $K = SK_dH$, and the discrete LTI plant for the period of T will be defined as $G_d = SGH$.

2.1.2 Multi-Rate Systems [17]

Multi-rate sampling and hold are applied in digital control and digital signal processing when it is necessary to sample the outputs or update the inputs with different rates [11]. For system G in Fig. 2.1, if sampling operator S and hold operator H consists of multiple rates such that $S = \{S_{T_1} \dots S_{T_i}\}$ and $H = \{H_{T'_1} \dots H_{T'_j}\}$, where T_i and T'_j are a finite number of periodic sampling and hold rates, then the system $G_m = SGH$ is called as multi rate system. The application of this topic in CPS security will be formally addressed in Chapter 4.

2.2 Cyber-Physical Systems

Cyber-physical systems are sampled data systems and generally can be defined as concatenated systems of systems, where by the combination of computing, communication, and control, they perform their desired task. In the following, we explain this concept over real-world examples:

1. **Swarm of UAV:** In a swarm of UAV the communication can take place distributively between different UAV, where collectively UAVs can accomplish distributed observation or distributed control for performing tasks such as environmental mapping and coordinated military operations.

2. **Communication-based Train Control:** Systematic railway safety has also been subjected to major upgrades and communication-based supervision. In this example, the entirety of trains is considered the element of CPS, where each train communicates its location to the C&C,

and a decision for train movement with respect to its communicated sensor measurements and placement of others in the network will be made.

2.3 Faults and Cyber-Attacks

2.3.1 Actuator Faults

Actuators are devices with mechatronic features that are responsible for controlling a mechanism [85]. They typically convert a form of input energy, such as electrical energy, into a mechanical energy. A malfunction of this device causing an unpermitted deviation from the desired energy in their outputs is called as an actuator fault [45].

2.3.2 Sensor Faults

"A sensor is a transducer that measures and monitors the status of the system without influencing it [85]". Any deviation between the actual measurements and the monitored one is called a fault [45].

2.4 Cyber-Attacks

2.4.1 Preliminaries

In this part, the preliminary information of cyber-attacks is provided, and moving forward, the investigated type of cyber-attacks in this thesis is explained.

1. **Basic definitions:**

    - Adversary: An adversary is someone who wants to endanger the availability, reliability, and integrity of the control system.
    - Attack: An attack is a malicious data through which the attacker compromises the integrity of the outputs or command inputs data.

- Defender: A defender is someone that comes up with a methodology to counter the attack.

2. **An attacker's objective:**

   - The attacker intends to target the vulnerabilities to deviate the states of the system, create misleading data and evade detection by the monitoring system.

3. **A defender's objective:**

   - Protecting integrity: Integrity denotes the trustworthiness of data and resources, a lack of integrity results in deception [62].
   - Protecting confidentiality: It denotes protecting information from getting accessed by unauthorized parties.
   - Ensuring availability: It means the information must remain accessible for authorized users.
   - Increasing safety: Design of protective measures that ensure the safety of the cyber-physical systems in the presence of attacks and faults.

2.4.2 Type of Cyber-Attacks

Cyber-attacks, in general, can affect the networked communicated inputs or outputs of the cyber-physical plant in different ways. The categories of cyber-attacks include disruption attacks such as DoS or Distributed Denial of Service (DDoS) attacks, or deception attacks [70], where it includes cyber-attacks such as replay-attacks, false data injection attacks, zero dynamics attacks, and covert attacks. The deception attack is introduced as the following:

- Replay attacks: In replay attacks, the attacker hijacks a set of sensors for a period of time, transmits the recorded measurements, and tries to manipulate the control system [71].
- False data injection attacks: It refers to data injection attacks through inputs, outputs, or both channels w.r.t. In part of the literature, this cyber-attack is categorized based on the situations

where the attacker has partial information about the system [52, 106]. However, some studies in literature expand this category to different types of cyber-attack, where the attacker has access to complete knowledge of system model, such as covert and zero dynamics attacks [30].

- Covert attacks: In covert attacks, the attacker is assumed to possess knowledge of system dynamics, or it can obtain it through eavesdropping. In this type of attack, adversary can compromise the inputs and hide its effect by measuring and subtracting the inputs contribution from the outputs [39].
- Zero dynamics attack: An output zeroing input [83].
- Denial of Service (DoS) and DDoS attacks: A denial-of-service attack (DoS attack) is a cyber-attack that targets availability of the communicated data by huge amounts of spam requests, in DDoS attacks, these requests originates from many sources [46].

2.4.3 Investigated Cyber-Attacks

Consider the following discrete linear time-invariant system as shown in Fig. 2.2, with the assumption that its inputs and outputs can be subjected to attack:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + B^{a^u} a_k^u \\ y'_k &= Cx_k \\ y_k &= Cx_k + D^{a^y} a_k^y = y'_k + D^{a^y} a_k^y \end{aligned} \tag{2.1}$$

where $x_k \in \mathbb{R}^n$ is the system state, $u_k \in \mathbb{R}^m$ is the command control signal, $a_k^u \in \mathbb{R}^{m_{a^u}}$ denotes the injected cyber-attacks through command input channel, $a_k^y \in \mathbb{R}^{m_{a^y}}$ denotes the injected cyber-attacks though the communicated measurements channel, $y_k \in \mathbb{R}^p$ is the measurements received on the C&C, $y'_k \in \mathbb{R}^p$ is the measurements obtained on the plant-side.

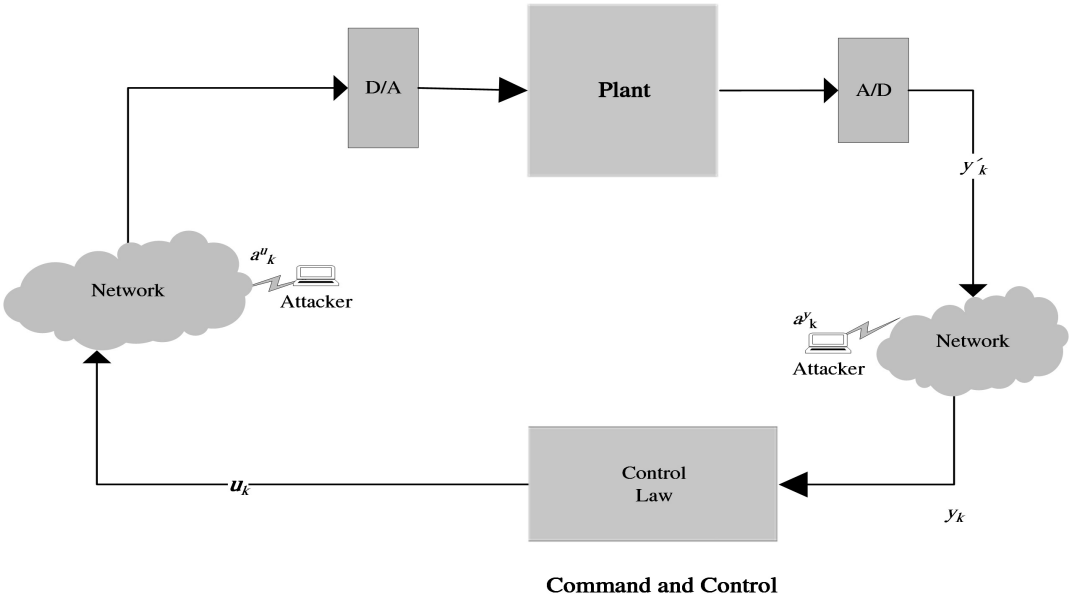

Figure 2.2: A CPS under cyber-attack.

Definition 2.1 (Actuator Cyber-Attack) *: In the considered representation of cyber-physical systems (2.1), with states defined as $x \in \mathscr{X}$, an actuator cyber-attack $a^u \in \mathscr{A}_u$ is associated with maps B^{a^u}, shown in equation (2.1), such that $B^{a^u} : \mathscr{A}_u \longrightarrow \mathscr{X}$, and $Im(B^{a^u}) \subseteq Im(B)$.*

Definition 2.2 (Sensor Cyber-Attack) *: In the considered representation of cyber-physical systems (2.1), with outputs defined as $y \in \mathscr{Y}$, a sensor cyber-attack $a^y \in \mathscr{A}_y$ is associated with maps D^{a^y}, shown in equation (2.1), such that $D^{a^y} : \mathscr{A}_y \longrightarrow \mathscr{Y}$.*

The investigated cyber-attack in the thesis include zero-dynamics and covert attacks as well as bias data injection attacks. The aforementioned cyber-attacks are a type of data injection attacks into the inputs or outputs of the systems and are similar in corruption of communicated data while different in terms of the access to the information about the physical plant and its monitoring system. In the following, description, objectives, and resource access of these cyber-attacks are provided:

1. Bias Injection Attacks: Several types of false data injection attacks exist, which can be injected in the outputs, inputs or both channels[98][19]. In this thesis, bias injection attacks are investigated that are a type of non-stealthy attack. Bias injection attacks are constant additive

functions which their effect on the outputs of the system can be similar to the additive actuator or sensor faults. For injecting bias injection cyber-attacks, a constant signal a_k^u will be added to the command inputs u_k.

In order to apply more sophisticated attacks, such as zero dynamics and covert attacks, we navigate some of the definitions of linear system theory through an adversary model used by the attacker and analyze the adversary attacks that are used in this thesis. Consider the adversary applies cyber-attacks with knowledge of networked linear time-invariant sample data system with the rate of T and utilizes the following model:

$$\begin{aligned} x_{k+1}^a &= Ax_k^a + B^{a^u} a_k^u \\ y_k^a &= a_k^y = Cx_k^a \end{aligned} \tag{2.2}$$

$x^a \in \mathcal{X}^a$ demonstrates the evolution of state subject to adversary inputs $a^u \in \mathcal{A}^{a^u}$, $a^y = y^a \in \mathcal{A}^{a^y}$ demonstrates the evolution of outputs under the adversary inputs.

2. Zero Dynamics Attacks

By definition a zero dynamics attack exist if the system under attack (2.2) is not strongly observable. Consider the following pencil matrix

$$\mathcal{P}(z) = \begin{pmatrix} A - z_0 I & B^{a^u} \\ C & 0 \end{pmatrix} \tag{2.3}$$

For a system to be strongly observable the following condition must hold:

$$rank(\begin{pmatrix} A - z_0 I & B^{a^u} \\ C & 0 \end{pmatrix}) = n + m \tag{2.4}$$

Definition 2.3 (Strong observability) *[100] System (2.2) is called strongly observable if for*

all initial condition $x_0^a \in \mathcal{X}^a$ and for every input function the following holds [100]:

$$\forall k \geq 0,\ y^{a^u}(k, x_0) = 0\ implies\ x_0 = 0. \tag{2.5}$$

Definition 2.4 (Zero Dynamics Attack for plant) *[10, Definition 2] For $z_0 \in \mathbb{C}$, if $rank\big(\mathcal{P}(z_0)\big) \leq n+m$, where $n+m$ is equal to normrank of $\mathcal{P}(z)$, then there exist an attack vector $a_k = (a_k^{uT}\ 0_{1\times p}^T)^T$ called zero dynamics attack for plant, where a_k is equal to $a_0^u z_0^k$, s.t z_0, a_0, and x_0 satisfy the following:*

$$\mathcal{P}(z_0)\begin{pmatrix} x_0^T & a_0^{uT} & 0_{1\times p}^T \end{pmatrix}^T = 0 \tag{2.6}$$

Definition 2.5 (Weakly unobservable point) *[100] For the system (2.2), a point $x_0 \in \mathcal{X}$ is called weakly unobservable if $\exists a^u \in \mathcal{U} \subset R^m$ $s.t\ \forall k \geq 0$, $y^a(k, x_0) = 0$.*

Definition 2.6 (Weakly unobservable subspace) *[100] Set of all weakly unobservable point named $\mathcal{V}$ is called the weakly unobservable subspace.*

Thus, the attacker's objective is to design an output zeroing input. Since the effect of the attack on the output is zero, the detector would be unable to recognize any trace of it in the residual of the system. This attack can be designed as the following:

Step 1: Find all the transmission zeros denoted by z_0:

$$det(\begin{pmatrix} A - z_0 I & B^a \\ C & 0 \end{pmatrix}) = 0 \tag{2.7}$$

Step 2: Find a_0^u corresponds to the transmission zero.

Step 3: Set $a_k^u = a_0^u z_0^k$

Step 4: Add the attack inputs to the known inputs: $u_k + a_k^u$

3. Covert Attacks: Consider the following pencil matrix for equation (2.2):

$$P(z) = \begin{pmatrix} A - zI & B^{a^u} & 0 \\ C & 0 & D^{a^y} \end{pmatrix} \tag{2.8}$$

Definition 2.7 (Covert Attack) *For* $\forall z \in \mathbb{C}$ *and* $\forall x_0 \in \mathbb{R}^n$, *the attack vector* $a_{\mathscr{Z}} = (a_k^{uT} a_{\mathscr{Z}}^{y\ T})^T$ *called covert attack for the plant s.t* z, $a_{\mathscr{Z}}$, *and* $x_{\mathscr{Z}}$ *satisfy the following:*

$$P(z) \begin{pmatrix} x_{\mathscr{Z}}^T & a_{\mathscr{Z}}^{u\ T} & a^{yT}_{\mathscr{Z}} \end{pmatrix}^T = 0 \tag{2.9}$$

For designing the covert attack the following steps are taken:

Step 1 : Add the attack inputs to the known inputs: $u_k + a_k^u$;

Step 2 : Compute a^y corresponding to the a^u according to equation (2.2)

Step 3 : Apply a^y as a sensor cyber-attack

2.5 Model-Based Detector Design

Model-based detector design can be divided into two fields of active and passive [30].

1. Passive: Normally, model-based detector design includes using residual generation alongside a dynamic filter such as Kalman or Luenberger filter combined with detection methodology such as $\mathcal{X}^2$ [30].

   Passive detection methodologies are unable to deal with sophisticated attacks such as zero dynamics attacks. For instance in order to design an observer for detection problems, it is important to know if the system is observable:

   Theorem 2.1 (Chapter8, Theorem 9) *[13] The system in equation* (2.1) *is observable if and only if* $\forall z \in \sigma(A) \subset \mathbb{C}$ *the following holds:*

$$rank(\begin{pmatrix} A - zI \\ C \end{pmatrix}) = n \tag{2.10}$$

for which $\sigma(A)$ *denotes eigenvalues of* $A \in \mathbb{R}^{n \times n}$.

2. Active: active detection can take place by adding extra information to inputs and outputs such that it can help to reveal the cyber-attacks [30].

   Cyber-attacks are often designed to remain stealthy for an extended time window. Cyber-attacks, such as zero dynamics and covert attacks, will continue to stay hidden and require active detection methodologies to be revealed. Watermarking [72], auxiliary-based methodologies [88], and coding-scheme solutions [66] are all part of active detection-based methodologies that add an extra layer of information to help detect covert attacks.

Methodologies for detection, whether passive or active, should be modified in such a way that they can distinguish cyber-attacks from faults [97]. For this purpose, the next section provides a review of an important concept in control theory, dealing with the unknown inputs of systems. Although cyber-attacks are intelligent, they are still considered an unknown input to the system and must adhere to the fundamental laws established in control theory for unknown inputs in order to be distinguishable from other types of unknown inputs in the system.

2.5.1 Detectability of Unknown Inputs

The topic of detectability for unknown inputs dates back in the literature to the work of Nikoukhah [78] and Nyberg [79]. In the following sections, we first present a linear control system under fault and disturbance to illustrate two important theorems discussed in Nyberg's work [79], which aligns with the conclusion drawn by Nikoukhah in [78]:

Consider a discrete-time representation of linear-time invariant systems with fault and distur-

bance.

$$x_{k+1} = Ax_k + B^u u_k + B^d d_k + B^f f_k$$
$$y_k = Cx_k + D^u u_k + D^d d_k + D^f f_k \tag{2.11}$$

where $(A, B^u, B^d, B^f, C, D^u, D^d$, and $D^f)$ are known system matrices of appropriate dimensions, x indicates the states, u is the known inputs, d indicates the disturbance, and f indicates the fault.

Theorem 2.2 *[79, Theorem 3] Faults are detectable in a system iff the following condition holds:*

$$Im(\begin{pmatrix} B^f \\ D^f \end{pmatrix}) \nsubseteq Im(\begin{pmatrix} zI - A & B^d \\ C & D^d \end{pmatrix}) \tag{2.12}$$

The above theorem denotes the direction of space for which the fault causes an effect on outputs will continue to do so if we decide to decouple the direction related to the disturbance. This result is developed and proved in [79] for the case of any residual generator.

Another interesting aspect of the Nyberg's work is presented in the following:

take:

$$M(z) = \begin{pmatrix} zI - A & B^d \\ C & D^d \end{pmatrix} \tag{2.13}$$

and $\mathcal{N}_L\{\{M(z)\})$ is the left null space of $M(z)$.

Theorem 2.3 *[79, Theorem 4]*

A fault is strongly detectable iff the following condition holds:

$$N_{M(0)} \begin{pmatrix} B^f \\ D^f \end{pmatrix} \neq 0 \tag{2.14}$$

where the rows of $N_{M(z)}$ are a basis for $\mathcal{N}_L\{\{M(z)\})$.

Nyberg elaborates on the residual generator and shows that the effect of weakly detectable faults can appear as a short pulse, which highlights the importance of establishing strong detectability for faults. Consequently, when we are dealing with an attack, we have to find a strategy that helps us to separate the space of the faults and attack such that the residual effect of the attack does not collide with the space of faults.

In view of faults and cyber-attacks, the problem of fault detectability in the presence of disturbance can be treated as the detectability of cyber-attacks from faults. Meaning, the equation (2.11) can be reformulated in the following manner:

$$\begin{aligned} x_{k+1} &= Ax_k + B^u u_k + B^{a^u} a^u_k + B^f f_k \\ y_k &= Cx_k + D^u u_k + D^{a^y} a^y{}_k + D^f f_k \end{aligned} \tag{2.15}$$

for which a^u and a^y are cyber-attacks that are respectively directed at the inputs channel or the outputs channel. B^{a^u} and B^{a^y} are the direction matrices for which these two vectors are applied on the inputs or outputs. Consider the vector a_k defined as $\begin{pmatrix} a^u_k & a^y_k \end{pmatrix}^T$. The equation (2.15) can be reformulated as the following:

$$\begin{aligned} x_{k+1} &= Ax_k + B^u u_k + \bar{B}^a a_k + B^f f_k \\ y_k &= Cx_k + D^u u_k + \bar{D}^a a_k + D^f f_k \end{aligned} \tag{2.16}$$

for which $\bar{D}^a = \begin{pmatrix} 0 & D^{a^y} \end{pmatrix}$ and $\bar{B}^a = \begin{pmatrix} B^{a^u} & 0 \end{pmatrix}$.

Furthermore, the equation (2.12) can be reformulated for faults and cyber-attacks:

$$Im(\begin{pmatrix} \bar{B}^a \\ \bar{D}^a \end{pmatrix}) \nsubseteq Im(\begin{pmatrix} zI - A & B^f \\ C & D^f \end{pmatrix}) \tag{2.17}$$

It signifies that cyber-attacks are detectable from faults if the above condition holds. Therefore, the evolution space of cyber-attacks in the outputs space must not be a subset of the evolution space of faults in the outputs space.

2.5.2 Model-Based Observers for Unknown Inputs

Fault-tolerant control often employs methodologies capable of obtaining the system's true state. This means that, even when faults are present, the expected estimate of the system's states does not differ from the actual states. One approach to achieve this is through Unknown Input Observers (UIOs) [40]. These model-based observers are designed in such a way that they can estimate the true state and also determine the value of the system's unknown input. However, UIOs require the system to be strongly detectable. As reported in [95], for the types of unknown inputs defined in (2.11) and assuming $d = 0$, the following condition is necessary and sufficient for the construction of UIOs to estimate the faults:

Theorem 2.4 (Theorem 6) *[95] A UIO for a discrete-time system is applicable if and only if the system, in the form shown in (2.11), is strongly detectable:*

$$\text{rank}(\begin{pmatrix} zI - A & B^f \\ C & D^f \end{pmatrix}) = n + m^f, \forall |z| \geq 1, z \in \mathbb{C} \tag{2.18}$$

where m^f denotes the dimension of the unknown input f.

2.6 Conclusion

The presence of both cyber-attacks and faults poses a significant challenge for fault and cyber-attack diagnosis. As discussed in Section 2.4.3, stealthy cyber-attacks are intelligently designed to evade the monitoring capabilities of traditional observers. Control theory provides various strategies for handling unknown inputs. However, when dealing with network-based unknown inputs, it is necessary to leverage existing knowledge in CPS security, such as multi-rate systems, active detection, and theories related to the isolation and estimation of unknown inputs, to devise solutions that can isolate and estimate cyber-attacks in the presence of faults. For instance, through equation (2.17), we discuss how cyber-attacks can be isolated from faults by identifying separate contributions of cyber-attacks to an output space that does not overlap with the output space of faults.

Chapter 3

Fault and Attack Isolation in Single Agent Cyber-Physical Systems

3.1 Introduction

Cyber-attacks on the inputs and outputs of cyber-physical systems (CPS) have introduced a different class of unknown inputs aside from recognized classes of machine-induced faults or disturbances. Since the traditional filters fail to separate their effect from the machine-induced faults, this has posed a significant threat to CPS security.

In [45], faults are defined as "unpermitted derivation of one of the characteristics of the systems." Attacks are also a form of unpermitted deviation, while the main difference between faults and attacks is the stealthiness and intelligent design of attacks considering a fault does not try to hide from the monitoring system. Zero dynamics attacks [8, 82] can be considered as one of the examples that show the differences between machine-induced faults and cyber-attacks, where these attacks are specifically designed to perform based on the output zeroing modes.

Fault detection and diagnosis consist of well-developed literature, including methodologies that can efficiently decouple these anomalies from the disturbance of the system. Numerous studies on the topic of fault detection and isolation through diagnostic methods such as Kalman filter [64],

Unknown Input Observer (UIO) [25, 74], and parity-based methods [63] are conducted. The mentioned fault diagnosis methodologies intend to generate a residual and relate the residual deviation to the malfunction of the actuators, sensors, or components of the system.

Different approaches for analyzing attacks in CPS can be divided into two categories: in the first one, the study addresses a single type of attack such as replay attacks [43, 71, 81] or false data injection attack [1, 52], and the second category involves investigating several types of attacks. Studies on security index [9, 15] or the development of sophisticated adversary attacks known as zero dynamics or covert attacks [51, 105] can be categorized as the second group, since the methodologies developed for revealing covert or zero dynamics attacks [10, 77, 88] usually cause other adversarial unknown inputs such as false data injection attacks to have a trace on the outputs of the system.

Most of the developed methodologies concerned with adversarial attacks do not address situations where actuator or sensor faults are present in the system. For example, from the viewpoint of the C&C, the range space of possible faults entries of actuators in the outputs can be a subset of the range space of the possible attacks entries, therefore in a situation where the system is under simultaneous faults and cyber-attacks, the conventional filters are not able to distinguish between them. [111] is one of few works in literature that address simultaneous faults and cyber-attack detection and isolation, and it is limited to replay attacks.

This work addresses the detection and isolation of cyber-attacks and faults in CPS in the plant side perspective through a specific structure shown in Fig.3.2. For the presented structure in Fig.3.2 it will be formally proven that CPS is secure against covert and zero dynamics attacks. The investigated CPS infrastructure includes a control loop responsible for tracking control and generating the command inputs of a physical plant.

In order to accomplish the detection and isolation of anomalies, two filters are constructed, one sensitive to faults and another one sensitive to cyber-attacks. The proposed methodology is provided under the assumption that the C&C control law is established through a specific dynamic of Kalman filter-based feedback control for which the attack-sensitive filter can isolate the zero

dynamics and covert attacks. Lastly, the proposed methodology is compared to an auxiliary-based attack detection method that is established in [88]. Upon detection of fault or cyber-attacks at the plant side, a decision can be made that can include self mitigatory plan or communication with the C&C, which is out of the scope of this paper.

The remainder of the chapter is as follows. Section. 3.2 provides the problem formulation. Section 3.3 explains proposed approach. Section 3.4 includes the analysis of the stealthy attacks for proposed methodology. Section 3.5 is the simulation results of the proposed methodology. Section 3.6 is the comparative study. Section 3.7 is the simulation results corresponding to the comparative study. Finally, Section 3.8 is the conclusion.

3.2 Problem Statement and Formulation

3.2.1 Considered System

The considered CPS in this chapter includes control law as depicted in Fig. 3.1. The control law performs tracking control of the plant. The C&C receives the outputs y_k and produces the command inputs u_k.

The governed system is represented by linear discrete-time invariant dynamics under the influence of actuator faults, sensors faults, cyber-attacks on the communication channels transmitting control command, and cyber-attacks on transmitted measurements as the following (Fig. 3.2):

$$\begin{aligned} x_{k+1} &= Ax_k + \underbrace{Bu_k + B^{a^u}}_{B\bar{u}_k} a_k^u + B^{f^u} f_k^u + w_k \\ y'_k &= Cx_k + D^{f^s} f_k^s + v_k^s \\ y_k &= Cx_k + D^{f^s} f_k^s + v_k^s + D^{a^y} a_k^y = y'_k + D^{a^y} a_k^y \end{aligned} \tag{3.1}$$

where $x_k \in \mathbb{R}^n$ is the system state, $u_k \in \mathbb{R}^m$ is the command control signal, $a_k^u \in \mathbb{R}^{m_{a^u}}$ denotes the injected cyber-attacks through command input channel, $a_k^y \in \mathbb{R}^{m_{a^y}}$ denotes the injected cyber-

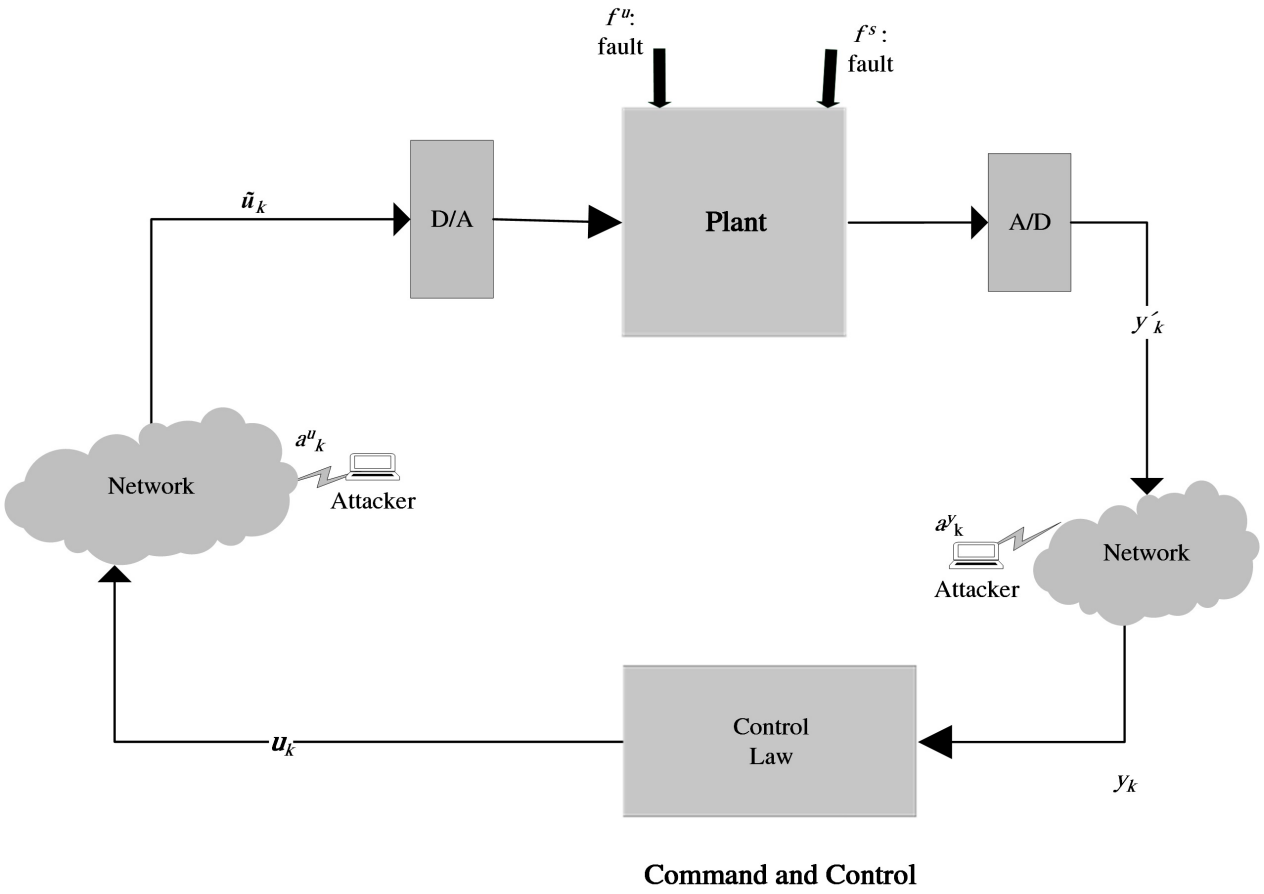

Figure 3.1: The CPS framework.

attacks though the communicated measurements channel, $f_k^u \in \mathbb{R}^{m_{f^u}}$ denotes the actuator faults, $f_k^s \in \mathbb{R}^{m_{f^s}}$ denotes the sensor faults, $y_k \in \mathbb{R}^p$ is the measurements received on the C&C, $y'_k \in \mathbb{R}^p$ is the measurements obtained on the plant-side, w_k, v_k^s are process noise and sensor noise with Gaussian distribution, $\tilde{u}_k = u_k + \Gamma a_k^u$, and $\Gamma = [\alpha_{i,j}] \in R^{m \times m_a}$ and it is a diagonal or rectangular diagonal matrix such that based on the attacked channels entries, $\alpha_{i,i}$ are either zero or one, and the rest of the entries are zero. Moreover, matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$,$B^{a^u} = B\Gamma \in \mathbb{R}^{n \times m_{a^u}}$, $B^{f^u} \in \mathbb{R}^{n \times m_{f^u}}$ $C \in \mathbb{R}^{p \times n}$, $D^{a^y} \in \mathbb{R}^{p \times m_{a^y}}$, $D^{f^s} \in \mathbb{R}^{p \times m_{f^s}}$ are known and are based on the model of the physical plant.

Remark 3.1 *In equation (3.1) A, B, C is representative of physical plant and local control system if the CPS needs a local control system such that it ensures stability on the plant side, the system matrices can be representative of the physical plant if all command inputs are received from C&C.*

Therefore, the proposed methodology is not restrictive of the existence of local controller.

In the following formal definitions of actuator faults and sensor faults are provided. The formal definition for actuator cyber-attack and sensor cyber-attack has been provided in Definitions 2.1 and 2.2.

Definition 3.1 (Actuator Fault) *: In the considered representation of cyber-physical systems denoted in (3.1), with states defined as $x \in \mathscr{X}$, an actuator fault signal $f^u \in \mathscr{F}_u$ is associated with a map B^{f^u}, such that $B^{f^u} : \mathscr{F}_u \longrightarrow \mathscr{X}$.*

Definition 3.2 (Sensor Fault) *: In the considered representation of cyber-physical systems (3.1), with outputs defined as $y \in \mathscr{Y}$, a sensor fault $f^s \in \mathscr{F}_s$ is associated with maps D^{f^s}, shown in equation (3.1), such that $D^{f^s} = [\lambda_{i,j}] : \mathscr{F}_s \longrightarrow \mathscr{Y}$. According to the faulty sensors, $\lambda_{i,i}$ entries are either zero or one, and the rest of the entries are zero.*

Moreover, let us consider Kalman filter based control law in the C&C that can be represented through the following equation:

$$\begin{aligned} x^{oc}_{k+1} &= A_{oc}x^{oc}_k + Bu_k + L^{oc}_k \overbrace{(y'_k + D^{a^y} a^y_k)}^{y_k} \\ y^{oc}_k &= Cx^{oc}_k \\ u_k &= -K^{oc}x^{oc}_k \end{aligned} \tag{3.2}$$

where $A_{oc}, K^{oc}, L^{oc}_k, D^{oc}, B$ is the matrix representation describing the dynamical control, x^{oc} represents its state, y^{oc} is the outputs of Kalman filter, u_k is the command inputs at C&C, with the following filter parameters including L^{oc}_k Kalman filter gain, error covariance matrix P^{oc}_k

$$\begin{aligned} A_{oc} &= (A - L^{oc}_k C), L^{oc} = L^{oc}_k \\ L^{oc}_k &= AP^{oc}_k C^T (CP^{oc}_k C^T + R_k)^{-1} \\ P^{oc}_{k+1} &= A(P^{oc}_k - P^{oc}_k C^T (CP^{oc}_k C^T + R_k)^{-1} CP^{oc}_k)A^T + Q_k \end{aligned} \tag{3.3}$$

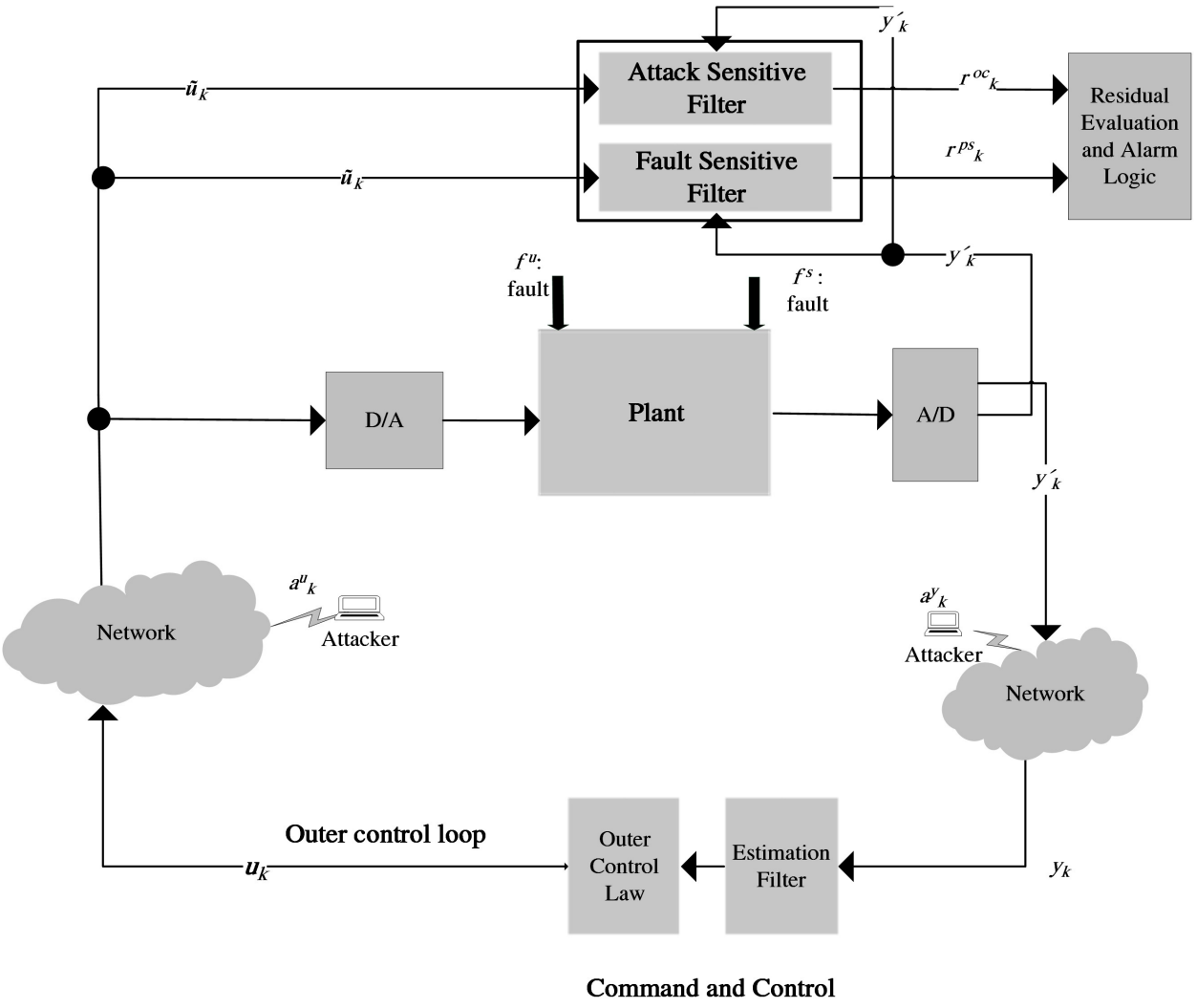

Figure 3.2: A CPS system under fault and attack.

in which Q^{oc} ad R^{oc} are constant covariance matrices of system noises.

Assumption 3.1 *There are no assumptions regarding the number of inputs and outputs.*

3.2.2 Cyber-Attacks and Faults Isolation Problem

In this work, the main objective is to isolate cyber-attacks and faults from each other on the plant side for a specific class of Kalman filter-based control law on the C&C. For this purpose, in Fig. 3.2, we present a schematic that makes the isolation possible while granting the CPS presented in (3.1) immunity to covert and zero dynamics attacks.

To achieve this, we deal with two main challenges:

1. Consider the the control law model (3.2) with the state defined as x^{oc} is available on C&C. The challenge is to design two filters on the plant side called attack sensitive and fault sensitive filters:

   - **Plant Side Attack Sensitive Filter:** The monitoring filter for attack detection and isolation is located on the plant side, which identifies the existence of malicious inputs in the received inputs $\tilde{u}_k$ and the transmitted measurements y_k.

     The task is to generated a residual r^{oc} by generating an error e^{oc} through the plant side generated value $\bar{x}^{oc}$ for the control model such that $e^{oc} = \bar{x}^{oc} - x^{oc}$. r^{oc} must become free of any fault such that it detects and isolates the attack, and its design is explained in Section 3.3.1. It will also be shown that such an attack filter will cancel the effect of the noise.

   - **Plant Side Fault Sensitive Filter:** The fault monitoring filter is a traditional Kalman filter located on the plant side. The fault filter utilizes plant-side received control command inputs and transmitted outputs to generate a residual r^{ps} that is attack-free and recognizes machine-induced behaviors that affect the transmitted measurements to the C&C, and it is explained in Section 3.3.2.

   One of the main challenges in dealing with faults and attacks is that the conventional monitoring system is unable to generate a residual that leads to their detection and isolation. As shown in Fig. 3.2, two residuals r_k^{ps} and r_k^{oc}, will be generated. For r_k^{ps} that belongs to the Kalman filter residual, a $\mathcal{X}^2$ test will be performed, for which a value $J_k^{ps} = r_k^T \Sigma_k^{-1} r_k$ will be

generated, where $\Sigma_k = CP_kC^T + R_k$:

$$\begin{cases} \text{if} & J_k^{ps} > T^{th}, |r_k^{oc}| <= \epsilon \\ \text{then} & f_k^s \; or \; f_k^u \neq 0 \text{ and } a_k^u \; \& \; a_k^u = 0 \\ \text{if} & J_k^{ps} <= T^{th}, r_k^{oc} > \epsilon \\ \text{then} & a_k^y \; or \; a_k^u \neq 0 \text{ and } f_k^s \; \& \; f_k^u = 0 \\ \text{if} & J_k^{ps} > T^{th}, r_k^{oc} > \epsilon \\ \text{then} & a_k^y \; or \; a_k^u \neq 0 \text{ and } f_k^s \text{ or } f_k^u \neq 0 \end{cases}$$

ϵ denotes small amount and the threshold for J^{ps} will be determined by setting F1 score [92] set in range of 92 to 93 %, where F1 score is defined as:

$$F1 = 100\% \frac{TP}{TP + \frac{1}{2}(FP + TN)} \tag{3.4}$$

where True Positive (TP), False Positive (FP), True Negative (TN) , and False Negative (FN) are shown. The false alarm rate and detection rate are also defined as follows:

$$false\ alarm = \frac{FP}{TN + FP} \tag{3.5}$$

$$detection\ rate = \frac{TP}{FN + TP} \tag{3.6}$$

2. The second challenge is to determine whether the attack-sensitive filter is able to reveal covert and zero dynamics cyber-attacks directed at the plant and how can they be avoided, for which an analysis will be provided in Section 3.4.

3.3 Attack Sensitive and Fault Sensitive Filters

In this section, the design of attack-sensitive and fault-sensitive filters are explained.

3.3.1 Representation of the Plant-side Attack Monitoring Filter

In this subsection, we propose our cyber-attack sensitive filter. This filter uses the plant side inputs for updating the state of its estimator:

$$\bar{x}^{oc}_{k+1} = (A_{oc} - M^{oc}K^{oc})\bar{x}^{oc}_{k} + \underbrace{B\tilde{u}_k}_{Bu_k + B^{a^u}a^u_k} + L^{oc}y'_k + M^{oc}(\underbrace{-K^{oc}x^{oc}_k + \Gamma a^u_k}_{\tilde{u}_k}) \tag{3.7}$$

$$\bar{\tilde{u}}_k = -K^{oc}\bar{x}^{oc} \tag{3.8}$$

$$r^{oc}_k = \bar{\tilde{u}}_k - \tilde{u}_k \tag{3.9}$$

$\tilde{u}_k$ is the plant side command inputs, $\bar{x}^{oc}_k$ is the state of attack monitoring filter, r^{oc}_k is the generated residual that will be utilized for attack isolation, M^{oc} is the design parameter, and $\bar{\tilde{u}}^{oc}_k$ is the plant-side estimated of the command inputs. In the following, a theorem is provided that proves the monitoring filter in (3.7) with residual (3.9) is able to isolate the cyber-attacks at the plant-side while omitting the effects of sensor noise and measurements noise.

Theorem 3.1 *The residual generated r^{oc}_k through the proposed filter in (3.9) is only sensitive to attack if there exists a M^{oc} s.t $A_{oc} - M^{oc}K^{oc}$ is Schur stable* [1].

Proof *The error dynamic $e^{oc}_{k+1} = \bar{x}^{oc}_{k+1} - x^{oc}_{k+1}$ between the proposed monitoring attack filter in (3.7) and the control law in (3.2) is derived according to the following. L^{oc}_k is considered constant*

[1] Schur stability: A square complex matrix is considered Schur stable if its eigenvalues are inside the unit circle in the complex plane.

upon convergence of Kalman filter and is represented by L^{oc}.

$$\begin{aligned} e^{oc}_{k+1} &= (A_{oc} - M^{oc}K^{oc})e^{oc}_k - L^{oc}D^{a^y}a^y_k \\ &+ (B^{a^u} + M^{oc}\Gamma)a^u_k \end{aligned} \tag{3.10}$$

For the proposed filter if M^{oc} is designed such that $A_{oc} - M^{oc}K^{oc}$ is Schur stable, then $e^{oc}_k \longrightarrow -L^{oc}D^{a^y}a^y_k + (B^{a^u} + M^{oc}\Gamma)a^u_k$ as $k \longrightarrow \infty$. The proposed residual in (3.9) is calculated by $r^{oc}_k = K^{oc}L^{oc}D^{a^y}a^y_k + (B^{a^u} + M^{oc}\Gamma)a^u_k) - \Gamma a^u_k$. That shows that the generated residual r^{oc}_k is only dependent on the attack.

□

3.3.2 Fault Sensitive Filter

This filter utilizes the traditional design of the Kalman filter method, and its objective is to generate a residual, based on the comparison of the measured and estimated outputs of the plant on the plant-side:

$$\begin{aligned} \hat{x}^{ps}_{k+1} &= (A - L^{ps}_k C)\hat{x}^{ps}_k + \underbrace{B\tilde{u}_k}_{Bu_k + B^{a^u}a^u_k} + L^{ps}_k y'_k \\ L^{ps}_k &= AP^{ps}_k C^T (CP^{ps}_k C^T + R_k)^{-1} \\ P^{ps}_{k+1} &= A(P^{ps}_k - P^{ps}_k C^T (CP^{ps}_k C^T + R_k)^{-1} CP^{ps}_k)A^T + Q_k \\ r^{ps}_k &= y'_k - C\hat{x}^{ps}_k \end{aligned} \tag{3.11}$$

In which r^{ps} is the plant side fault detector residual, $\hat{x}^{ps}$ is the estimated value of the plant, L^{ps} is the Kalman filter gain, and P^{ps} is the error covariance of the Kalman filter.

Lemma 3.1 *[103, Proposition 1] A Kalman filter of the form (3.2) (in which $u_k = 0$, $a^u_k = 0$, and $a^y_k = 0$) converges if the following conditions hold:*

(i) $L^{oc}QL^{oc\prime} + R > 0$

(ii) The pair (A, L^{oc}) *is detectable*

(iii) $\bar{U}'\bar{z} = 0$ *for any arbitrary vectors* $\bar{z}$ *implies* $L^{oc}\bar{z} = 0$*, where* $Q = \bar{U}\bar{U}'$ *and* $\bar{U} \in \mathbb{R}^{q \times rank(Q)}$

Theorem 3.2 *The residual* r_k^{ps} *generated by a plant-side Kalman filter according to (3.11) will converge to the only fault.*

Proof *The error dynamic between Kalman filter in (3.11) and system (3.1) is derived as the following:*

$$\begin{aligned} e_{k+1}^{ps} = x_{k+1} - \hat{x}_{k+1}^{ps} = (A - L^{ps}C)e_k^{ps} + B_k^{f^u} f_k^u \\ - L^{ps}D^{f^s} f_k^s + \varepsilon_k(w_k, v_k) \end{aligned} \tag{3.12}$$

In which $\varepsilon_k(w_k, v_k)$ *is the optimal effect of noise on the error dynamic due to Kalman filter. If the stability conditions in* $Lemma3.1$ *is satisfied for the proposed Kalman filter in (3.11), then* $A - L^{pc}C$ *will be Schur stable. The residual for such an observer will converges to* $r^{ps} \longrightarrow C(B_k^{f^u} f_k^u - L^{ps}D^{f^s} f_k^s + \varepsilon_k(w_k, v_k))$ *as* $k \longrightarrow \infty$*. Therefore the* r_k^{ps} *is dependent on the fault and free of any attack.* □

3.3.3 Attack and Fault Isolation Alarm Logic

Through the equations (3.9) and (3.11), it is demonstrated how to generate the residuals r_k^{ps} and r_k^{oc}. From these residuals, faults and attacks can be isolated by a simple threshold checking as the residues are respectively faults sensitive and attack sensitive per demonstration in Theorem 3.1 and Theorem 3.2. The isolation logic can be demonstrated as follows:

$$\begin{cases} \text{if} \quad J_k^{ps} > T^{th} \text{ and } |r_k^{oc}| <= \epsilon \longrightarrow \text{ Fault} \\ \text{if} \quad r_k^{oc} > \epsilon \text{ and } J_k^{ps} <= T^{th} \longrightarrow \text{Attack} \\ \text{if} \quad r_k^{oc} > \epsilon \text{ and } J_k^{ps} > T^{th} \longrightarrow \text{Fault and attack} \end{cases}$$

Threshold (T^{th}) is selected according to $\mathcal{X}^2$ detection methodology by consideration of $F1$ score in range of (92-93)% averaged over 100 simulation.

3.4 Analysis of the Defined Class of Attack Filters Against Zero Dynamics and Covert Attacks

In this section, the analysis for the zero dynamics and covert attacks is performed for the specified class of control law in (3.2), from the attacker's point of view. The analysis in this section seeks to understand whether covert or zero dynamics attacks that are completely stealthy for the control-side filter are also stealthy for the plant-side filter.

For this purpose, first, the error dynamic of the proposed filter is augmented with the open-loop structure of the plant (3.1) in the attacker's point of view, then covert and zero dynamics attacks based on the definition provided in [10] are redefined for the presented scenarios.

First, the error dynamic presented in equation (3.10) for zero analysis of our system in the view-point of the attacker with consideration of the residuals as the outputs is written as the following:

$$\begin{aligned} e^{oc}_{k+1} &= A^{cl}_{oc} e^{oc}_k - L^{oc} D^{a^y} a^y_k + (B^{a^u} + M^{oc}\Gamma) a^u_k \\ r^{oc}_k &= -K^{oc} \bar{e}^{oc}_k - K^{oc}(L^{oc} D^{a^y} a^y_k \\ &\quad + (B^{a^u} + M^{oc}\Gamma) a^u_k) + \Gamma a^u_k \end{aligned} \tag{3.13}$$

where A^{cl}_{oc} = $A_{oc} - M^{oc}K^{oc}$.

Next, the above equation (3.13) is augmented with equation (3.1) for performing the analysis:

consider the augmented error dynamic of the filter and the plant as the following:

$$\begin{aligned}
\begin{pmatrix} x_{k+1} \\ e^{oc}_{k+1} \end{pmatrix} &= \begin{pmatrix} A & 0 \\ 0 & A^{cl}_{oc} \end{pmatrix} \begin{pmatrix} x_k \\ e^{oc}_k \end{pmatrix} + \begin{pmatrix} B^{a^u} \\ B^{a^u} + M^{oc}\Gamma \end{pmatrix} a^u_k \\
&\quad + \begin{pmatrix} 0 \\ -L^{oc}D^{a^y} \end{pmatrix} a^y_k \\
\begin{pmatrix} y_k \\ r^{oc}_k \end{pmatrix} &= \begin{pmatrix} C & 0 \\ 0 & -K^{oc} \end{pmatrix} \begin{pmatrix} x_k \\ e^{oc}_k \end{pmatrix} \\
&\quad + \begin{pmatrix} 0 \\ -K^{oc}(B^{a^u} + M^{oc}\Gamma) + \Gamma \end{pmatrix} a^u_k \\
&\quad + \begin{pmatrix} D^{a^y} \\ -K^{oc}L^{oc}D^{a^y} \end{pmatrix} a^y_k
\end{aligned} \tag{3.14}$$

Subsequently, three pencil matrices for the plant, filter (3.13), and the augmented view are defined:

1. The pencil matrix of plant

$$\mathcal{P}_{plant}(z) = \begin{pmatrix} zI - A & B^{a^u} & 0 \\ C & 0 & D^{a^y} \end{pmatrix} \tag{3.15}$$

2. The pencil matrix of the filter

$$\mathcal{P}_{filter}(z) = \begin{pmatrix} zI - A^{cl}_{oc} & B^{a^u} + M^{oc}\Gamma & -L^{oc}D^{a^y} \\ -K^{oc} & -K^{oc}(B^{a^u} + M^{oc}\Gamma) + \Gamma & -K^{oc}L^{oc}D^{a^y} \end{pmatrix} \tag{3.16}$$

3. The pencil matrix of augmented view

$$\mathcal{P}(z) = \begin{pmatrix} zI - A & 0 & B^{a^u} & 0 \\ 0 & zI - A_{oc}^{cl} & B^{a^u} + M^{oc}\Gamma & -L^{oc}D^{a^y} \\ C & 0 & 0 & D^{a^y} \\ 0 & -K^{oc} & -K^{oc}(B^{a^u} + M^{oc}\Gamma) + \Gamma & -K^{oc}L^{oc}D^{a^y} \end{pmatrix} \tag{3.17}$$

Next, based on the definitions of covert and zero dynamics attack by [10], the stealthy attack scenarios for the system are defined:

Definition 3.3 (Zero Dynamics Attack for Plant) *[10, Definition 2] For $\lambda_0 \in \mathbb{C}$, if $rank\big(\mathcal{P}_{plant}(\lambda_0)\big) \leq r$, where r is equal to normrank of $\mathcal{P}_{plant}(z)$, then there exist an attack vector $a_k = (a_k^{uT}\ 0_{1\times p}^T)^T$ called zero dynamics attack for plant, where a_k is equal to $a_0^u\lambda_0^k$, s.t λ_0, a_0, and x_0 satisfy the following:*

$$\mathcal{P}_{plant}(\lambda_0)\begin{pmatrix} x_0^T & a_0^{uT} & 0_{1\times p}^T \end{pmatrix}^T = 0 \tag{3.18}$$

Definition 3.4 (Zero Dynamics Attack for Filter) *[10, Definition 2] For $\gamma_0 \in \mathbb{C}$, if $rank\big(\mathcal{P}_{filter}(\gamma_0)\big) \leq r$, where r is equal to normrank of $\mathcal{P}_{filter}(z)$, then there exist an attack vector $a_k = (a_k^{uT}\ 0_{1\times p}^T)^T$ called zero dynamics attack for the filter, where a_k is equal to $a_0^u\gamma_0^k$, s.t γ_0, a_0, and e_0^{oc} satisfy the following:*

$$\mathcal{P}_{filter}(\gamma_0)\begin{pmatrix} x_0^T & e_0^{ocT} & a_0^{uT} & 0_{1\times p}^T \end{pmatrix}^T = 0 \tag{3.19}$$

Definition 3.5 (Zero Dynamics Attack for Augmented System) *[10, Definition 2] For $z_0 \in \mathbb{C}$, if $rank\big(\mathcal{P}(z_0)\big) \leq r$, where r is equal to normrank of $\mathcal{P}(z)$, then there exist an attack vector $a_k = (a_k^{uT}\ 0_{1\times p}^T)^T$ called zero dynamics attack for plant and the filter, where a_k is equal to $a_0^u z_0^k$, s.t z_0, a_0, and x_0, e_0^{oc} satisfy the following:*

$$\mathcal{P}(z_0)\begin{pmatrix} x_0^T & e_0^{ocT} & a_0^{uT} & 0_{1\times p}^T \end{pmatrix}^T = 0 \tag{3.20}$$

Definition 3.6 (Covert Attack for Plant) *For $\forall z \in \mathbb{C}$ and $\forall x_0 \in \mathbb{R}^n$, the attack vector $a_{\mathscr{Z}} = ({a^u_{\mathscr{Z}}}^T \ {a^y_{\mathscr{Z}}}^T)^T$ called covert attack for the plant s.t z, $a_{\mathscr{Z}}$, and $x_{\mathscr{Z}}$ satisfy the following:*

$$\mathcal{P}_{plant}(z)\begin{pmatrix} x^T_{\mathscr{Z}} & {a^u_{\mathscr{Z}}}^T & {a^y}^T_{\mathscr{Z}} \end{pmatrix}^T = 0 \tag{3.21}$$

Definition 3.7 (Covert Attack for the Detector) *For $\forall z \in \mathbb{C}$ and $\forall e_0 \in \mathbb{R}^n$, the attack vector $a_{\mathscr{Z}} = ({a^u_{\mathscr{Z}}}^T \ {a^y_{\mathscr{Z}}}^T)^T$ called covert attack for the filter s.t z, $a_{\mathscr{Z}}$, and $e^{oc}_{\mathscr{Z}}$ satisfy the following:*

$$\mathcal{P}_{filter}(z)\begin{pmatrix} {e^{oc}_{\mathscr{Z}}}^T & {a^u_{\mathscr{Z}}}^T & {a^y}^T_{\mathscr{Z}} \end{pmatrix}^T = 0 \tag{3.22}$$

Definition 3.8 (Covert Attack for Augmented System) *For $\forall z \in \mathbb{C}$ and $\forall x_0, e_0 \in \mathbb{R}^n$, the attack vector $a_{\mathscr{Z}} = ({a^u_{\mathscr{Z}}}^T \ {a^y_{\mathscr{Z}}}^T)^T$ called covert attack for both plant and filter s.t z, $a_{\mathscr{Z}}$, $x_{\mathscr{Z}}$, and $e^{oc}_{\mathscr{Z}}$ satisfy the following:*

$$P(z)\begin{pmatrix} x^T_{\mathscr{Z}} & {e^{oc}_{\mathscr{Z}}}^T & {a^u_{\mathscr{Z}}}^T & {a^y}^T_{\mathscr{Z}} \end{pmatrix}^T = 0 \tag{3.23}$$

Lemma 3.2 *Consider $(\mathscr{L}_u, \mathscr{L}_y)$ as the pair of all existing arbitrarily z-transform functions that can be injected on the inputs or outputs of the plant (3.1). The set of all possible covert attacks for this plant includes $(\mathscr{A}_u, \mathscr{A}_y)$ such that $\mathscr{A}_u = \mathscr{L}_u$ and $\mathscr{A}_y \subsetneqq \mathscr{L}_y$.*

Proof *The transfer matrices between the attacks and outputs for the (3.1) can form the following:*

$$\begin{pmatrix} zI - A & B^{a^u} & 0 \\ C & 0 & D^{a^y} \end{pmatrix} \begin{pmatrix} a^u_{\mathscr{Z}}(z) \\ a^y_{\mathscr{Z}}(z) \end{pmatrix} = 0 \tag{3.24}$$

Based on the above for forming a covert attack by an arbitrary $a^u_{\mathscr{Z}}(z) \in \mathscr{L}_u$, there exists a function a^y such that $D^{a^y} a^y_{\mathscr{Z}}(z) = -C(zI-A)^{-1}B^{a^u} a^u_{\mathscr{Z}}(z)$, and based on this equality, it can be concluded $a^y_{\mathscr{Z}}(z) = F(a^u_{\mathscr{Z}}(z))$, where $F : \mathscr{A}_u \to \mathscr{A}_y$. Therefore, there exist a $a^y_{\mathscr{Z}}(z) \in \mathscr{A}_y$ corresponding directly to the function of $a^u_{\mathscr{Z}}(z)$ such that $\mathscr{A}_y = \{F(a^u_{\mathscr{Z}}(\cdot))\} \subsetneqq \mathscr{L}_y$. □

Theorem 3.3 *Consider $(\mathscr{L}_u, \mathscr{L}_y)$ as the pair of all arbitrarily z-transform functions creating a cyber-attack, and $(\mathscr{A}_u = \mathscr{L}_u, \mathscr{A}_y \subsetneqq \mathscr{L}_y)$ the set of all possible covert attacks that target the plant*

(3.1). *If a covert attack directed at plant takes place, in the augmented view of the plant side and C&C, the filter described in* (3.7) *will limit the possible attacks that are covert for the filter, to sets of* $(\mathscr{A}_u^W, \mathscr{A}_y^W)$ *such that* $\mathscr{A}_u^W \subsetneqq \mathscr{A}_u$ *and* $\mathscr{A}_y^W \subsetneqq \mathscr{A}_y$.

Proof *To analyze this, first the transfer matrix between the attack and outputs will be written and it will be shown that it is belongs to smaller subspace:*

$$\mathcal{P}(z) = \begin{pmatrix} zI - A & 0 & B^{a^u} & 0 \\ 0 & zI - A_{oc}^{cl} & B^{a^u} + M^{oc}\Gamma & (M^{oc}D^{oc} - L^{oc})D^{a^y} \\ C & 0 & 0 & D^{a^y} \\ 0 & -K^{oc} & -K^{oc}(B^{a^u} + M^{oc}\Gamma) + \Gamma & -K^{oc}L^{oc}D^{a^y} \end{pmatrix} \begin{pmatrix} x_{\mathscr{Z}}(z) \\ e_{\mathscr{Z}}(z) \\ a^u_{\mathscr{Z}}(z) \\ a^y_{\mathscr{Z}}(z) \end{pmatrix} = 0 \tag{3.25}$$

in which $x_{\mathscr{Z}}(z)$, $e_{\mathscr{Z}}(z)$, $a^u_{\mathscr{Z}}(z)$, $a^y_{\mathscr{Z}}(z)$ *are corresponding to states, error dynamics, attacks through the inputs, and attacks through the outputs in the frequency domain. First, the following will be extracted the following equations from* (3.25)*:*

$$(zI_{n\times n} - A)x_{\mathscr{Z}}(z) + B^{a^u}a^u_{\mathscr{Z}}(z) = 0 \tag{3.26}$$

$$D^{a^y}a^y_{\mathscr{Z}}(z) = -Cx_{\mathscr{Z}}(z) \tag{3.27}$$

$$(zI_{n\times n} - A^{cl}_{oc})e_{\mathscr{Z}}(z) + (B^{a^u} + M^{oc}\Gamma)a^u_{\mathscr{Z}}(z) - L^{oc}D^{a^y}a^y_{\mathscr{Z}}(z) = 0 \tag{3.28}$$

$$(-K^{oc}(B^{a^u} + M^{oc}\Gamma) + \Gamma)a^u_{\mathscr{Z}}(z) = K^{oc}e_{\mathscr{Z}}(z) + (K^{oc}L^{oc}D^{a^y})a^y_{\mathscr{Z}}(z) \tag{3.29}$$

By combining equations (3.27) and (3.29):

$$(-K^{oc}(B^{a^u}+M^{oc}\Gamma)+\Gamma)a^u_{\mathscr{Z}}(z)=K^{oc}e_{\mathscr{Z}}(z)$$
$$-K^{oc}L^{oc}Cx_{\mathscr{Z}}(z) \tag{3.30}$$

In the augmented structure, the aim is to prove that the $a^u_{\mathscr{Z}}(z)$ cannot be selected arbitrarily. Therefore, (3.26) and (3.28) are combined with (3.30) and the following equation is resulted from it:

$$\begin{aligned}(-K^{oc}(B^{a^u}+M^{oc}\Gamma)+\Gamma)a^u_{\mathscr{Z}}(z)=\\ K^{oc}(zI_{n\times n}-A^{cl}_{oc})^{-1}\Big((B^{a^u}+M^{oc}\Gamma)a^u_{\mathscr{Z}}(z)\\ +L^{oc}D^{a^y}C(zI_{n\times n}-A)^{-1}B^{a^u}a^u_{\mathscr{Z}}(z)\Big)\\ -K^{oc}L^{oc}C(zI_{n\times n}-A)^{-1}B^{a^u}a^u_{\mathscr{Z}}(z)\end{aligned} \tag{3.31}$$

With simplifying the above equation:

$$\begin{aligned}&\Bigg((-K^{oc}(B^{a^u}+M^{oc}\Gamma)+\Gamma)\dots\\ &-K^{oc}(zI_{n\times n}-A^{cl}_{oc})^{-1}\Big((B^{a^u}+M^{oc}\Gamma)\\ &+L^{oc}D^{a^y}C(zI_{n\times n}-A)^{-1}B^{a^u}\Big)\\ &-K^{oc}L^{oc}C(zI_{n\times n}-A)^{-1}B^{a^u}\Bigg)a^u_{\mathscr{Z}}(z)\\ &=W(z)a^u_{\mathscr{Z}}(z)=0\end{aligned} \tag{3.32}$$

Let us break down the above equation into three scenarios that are considered for every possible $a^u_{\mathscr{Z}}(z)\in\mathscr{A}^W_u$:

1. *$W(z)$ is not full column rank for all z.*

   (a) *In the general case, there exists a direction set of z-transforms $a^u_{\mathscr{Z}}(z)\in ker(W(z))$ as the specific answer of this solution, and $D^{a^y}a^y_{\mathscr{Z}}(z)=-C(zI-A)^{-1}B^{a^u}a^u_{\mathscr{Z}}(z)$.*

(b) $W(z)$ *is not full column rank and also rank deficient for some* z_0*:*

In this case there exist a direction $a^u_{\mathscr{Z}}(z_0) \in ker(W(z_0))$ *for which the* (3.32) *is equal to zero, and* $D^{a^y} a^y_{\mathscr{Z}}(z_0) = -C(z_0 I - A)^{-1} B^{a^u} a^u_{\mathscr{Z}}(z_0)$.

2. $W(z)$ *is full column rank for almost all* z.

   (a) In the general case, if $W(z)$ *is full column rank for almost all* z, $a^u_{\mathscr{Z}}(z) = 0$ *and since* $D^{a^y} a^y_{\mathscr{Z}}(z) = -C(zI - A)^{-1} B^{a^u} a^u_{\mathscr{Z}}(z) = 0$, *therefore* $a^y_{\mathscr{Z}}(z) = 0$.

   (b) $W(z)$ *is full column rank for almost all* z, *and* $W(z_0)$ *is rank deficient for some* z_0*:*

   If $W(z_0)$ *is rank deficient for some* z_0, *in this case there exist a direction* $a^u_{\mathscr{Z}}(z_0) \in ker(W(z_0))$ *for which the* (3.32) *is equal to zero, and* $D^{a^y} a^y_{\mathscr{Z}}(z_0) = -\big(C(z_0 I - A)^{-1} B^{a^u} a^u_{\mathscr{Z}}(z_0)\big)$.

3. $W(z)$ *is full column rank for all* z. *In this case since* $W(z)$ *is full column rank for all* z *therefore,* $a^u_{\mathscr{Z}}(z)$ *must be zero and accordingly* $a^y_{\mathscr{Z}}(z) = 0$.

Therefore according to the above it can be concluded $\mathscr{A}^W_u \subsetneqq \mathscr{L}_u = \mathscr{A}_u$ *as it is a set including all the above solutions and* $\mathscr{A}^W_y = \{F(a^u_{\mathscr{Z}}(\cdot))\}$ *has a specific answer corresponding to the* $a^u_{\mathscr{Z}}(z)$ *such that* $D^{a^y} a^y_{\mathscr{Z}}(z) = -C(zI_{n\times n} - A)^{-1} B^{a^u} a^u_{\mathscr{Z}}(z)$, *and since there are less possible choices for* $a^u_{\mathscr{Z}}(z)$ *in augmented case compared to what is shown in Lemma* 3.2 *for system without the novel filter, then* $\mathscr{A}^W_y \subsetneqq \mathscr{A}_y$. □

Remark 3.2 *By comparing Theorem* 3.3 *and Lemma* 3.2, *significant improvement can be concluded by employing the proposed isolation methodology. Theorem* 3.3 *reaches a strict system structural dependent condition by introducing* $W(z)$ *that can limits the possible number of stealthy attacks to a finite set or even zero, which is shown through the numerical example provided in Section* 3.5.

3.5 Numerical Example: Attack Sensitive Filter and Kalman Fault Filter

In order to demonstrate our proposed approach, certain simulation results are provided in this section. The simulations belong to a quadruple tank process provided in [48], where it is assumed C&C is in charge of all commands inputs, and therefore, the model in (3.1) will be representative of the physical plant and not the combined control and physical system. We assume that the C&C uses Kalman filter-based control, and according to simulation the maximum convergence time S for its filter gains L^{oc} is equal to $89.93s$. We also use the outputs of Kalman filter y_k^{oc} to produce an attack or fault detection residual $\bar{r}_k^{oc} = y_k^{oc} - y_k$ in $C\&C$ called as control side filter Control Side Filter (CSF) residual. Furthermore, in the following figures, the Kalman Fault Filter (KFF) at the plant side, and the novel plant-side Attack Sensitive Filter (ASF) are presented according to their appropriate abbreviation. All the attacks or faults are injected at the time of $200s$. The thresholds for KFF are selected based on $F1$ score of 92 % to 93 % over 100 simulation (Fig. 3.3). The threshold for the CSF follows is considered the same as the KFF. ASF is not sensitive to noise, and the threshold is set to the small amount of $|r^{oc}| < 0.05$ for indication of no attack. The structure of the example is defined as follows:

$$A = \begin{pmatrix} 0.9843 & 0 & 0.0251 & 0 \\ 0 & 0.9892 & 0 & 0.0175 \\ 0 & 0 & 0.9747 & 0 \\ 0 & 0 & 0 & 0.9823 \end{pmatrix},$$

$$B = \begin{pmatrix} 0.0478 & 0.0010 \\ 0.0005 & 0.0348 \\ 0 & 0.0765 \\ 0.0554 & 0 \end{pmatrix}, C = \begin{pmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \end{pmatrix} \tag{3.33}$$

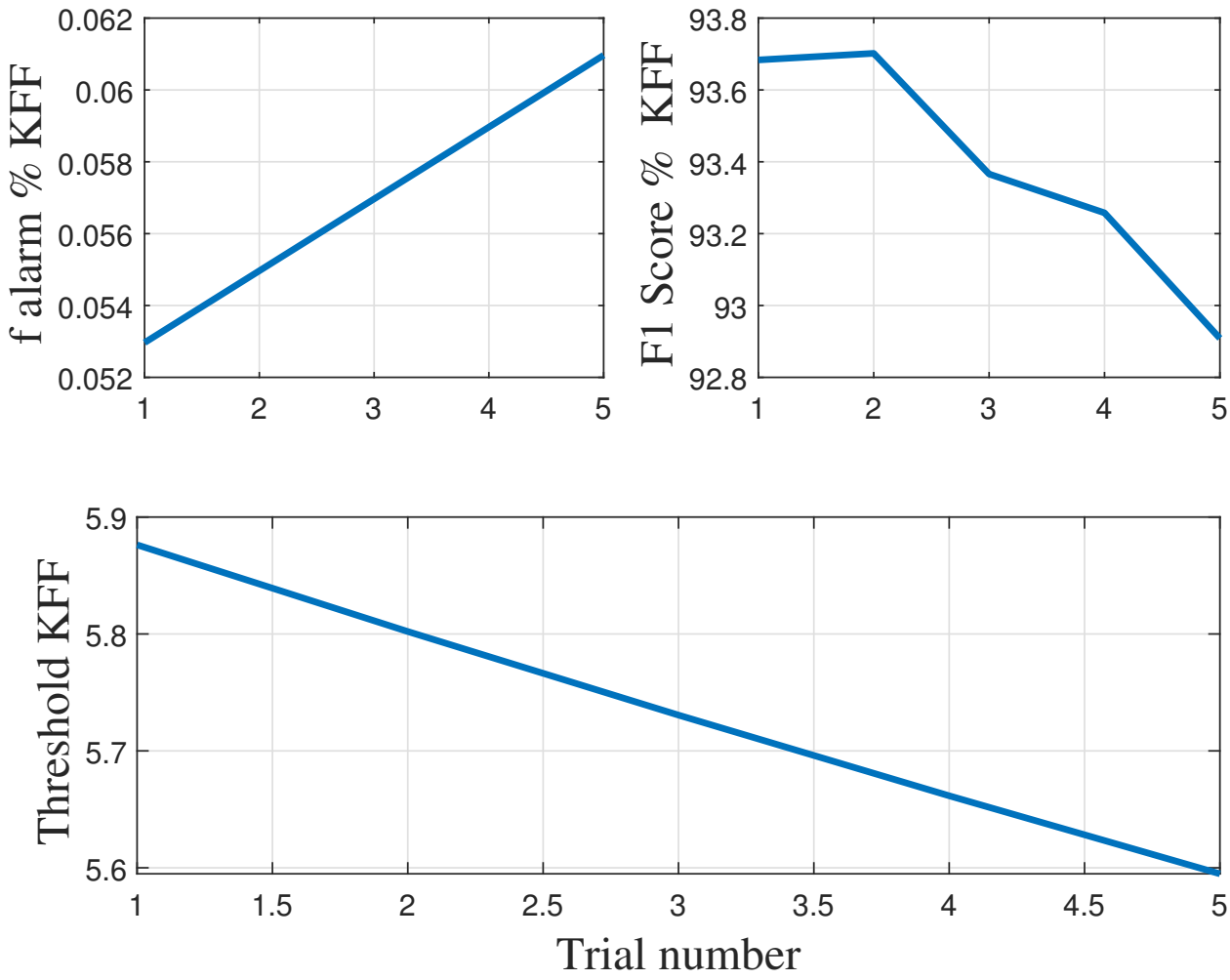

Figure 3.3: Trial for achieving the F1 score between 92 % and 93 %. Each trial includes 100 simulation.

$$
\begin{aligned}
L^{oc} &= \begin{pmatrix} 0.2013 & 0 & 0.0712 & 0 \\ 0 & 0.2049 & 0 & 0.0752 \end{pmatrix} \\
K^{oc} &= \begin{pmatrix} 1.3130 & 1.1251 & -0.3967 & 2.4002 \\ 0.9876 & 0.8680 & 2.4766 & -0.7160 \end{pmatrix}, \\
M^{oc} &= \begin{pmatrix} 0.0484 & 0.0239 & 0.0150 & 0.1623 \\ 0.0251 & 0.0383 & 0.1604 & -0.0096 \end{pmatrix}
\end{aligned} \tag{3.34}
$$

The scenarios include sensor fault without attack (Fig. 3.4), actuator fault without attack (Fig. 3.5), covert attack directed at plant without fault (Fig. 3.8), and zero dynamics attack (Fig. 3.6). In case of covert attack and actuator fault for a discrete system with the period of $1s$ the following inputs are added to the command inputs ($u + a^u(orf^u)$):

$$
a^u(k)\ orf^u(k) = \begin{cases} [0 \quad 0]^T, & if\ 0 \leq k < 200 \\ -[k/8 \quad k/8]^T, & if\ 200 \leq k < 300 \\ [0 \quad 0]^T, & \text{if } k \geq 300 \end{cases} \tag{3.35}
$$

for zero dynamics attacks the inputs is equal to the following:

$$
a^u(k) = \begin{cases} [0 \quad 0]^T, & if\ 0 \leq k < 200 \\ -1.0128^{k-200} \begin{pmatrix} -0.3398 \\ 0.3157 \end{pmatrix}, & if\ 200 \leq k < 300 \\ [0 \quad 0]^T, & \text{if } k \geq 300 \end{cases} \tag{3.36}
$$

For the covert attack the following outputs is subtracted from the sent measurements ($y - y^a$):

$$\begin{cases} y_k^a = [0 \quad 0]^T, & if\ 0 \leq k < 200 \\ x_{k+1}^a = Ax_k^a + Ba_k^u & \\ y_k^a = Cx_k^a, & if\ 200 \leq k < 300 \\ y_k^a = [0 \quad 0]^T, & \text{if } k \geq 300 \end{cases} \tag{3.37}$$

And for the sensor fault a bias of 10 has been added to sensor measurements.

Fig. 3.4 and Fig. 3.5 show the isolation of faults from attacks through KFF residual while insensitivity of ASF residual. Fig. 3.6 and Fig. 3.8 respectively show the isolation of zero dynamics and covert attacks from faults through ASF and KFF residual, while these attacks are shown to be stealthy for the CSF.

In this example, by setting $B^{a^u} = B$ and $D^{a^u} = I_{2\times 2}$, it can be concluded that $W(z)$ is full rank for almost all z. However, a finite number of undetectable attacks can be applied contrary to an infinite case without the novel filter, as the function $W(z)$ has 41 unstable zeros and 111 stables zeros. The calculation of $W(z)$ can be done using Matlab and is omitted due to spacing.

3.6 Comparative Study

To study the effectiveness of the proposed methodology, it will be compared to an auxiliary-based attack detection methodology presented in [88]. The original methodology presented by [88], consists of a switching strategy between auxiliary modes to reveal the stealthy attack. Based on [88], attackers' incomplete information about the auxiliaries will lead to detection of the attack.

First, it will be investigated how the auxiliary-based methodology will respond to the simultaneous presence of faults and attacks. Afterward, the simulation and comparison of the proposed filter to the auxiliary-based methodology will be provided.

In the example presented in this section, two modes are utilized, wherein mode 1 the attacker

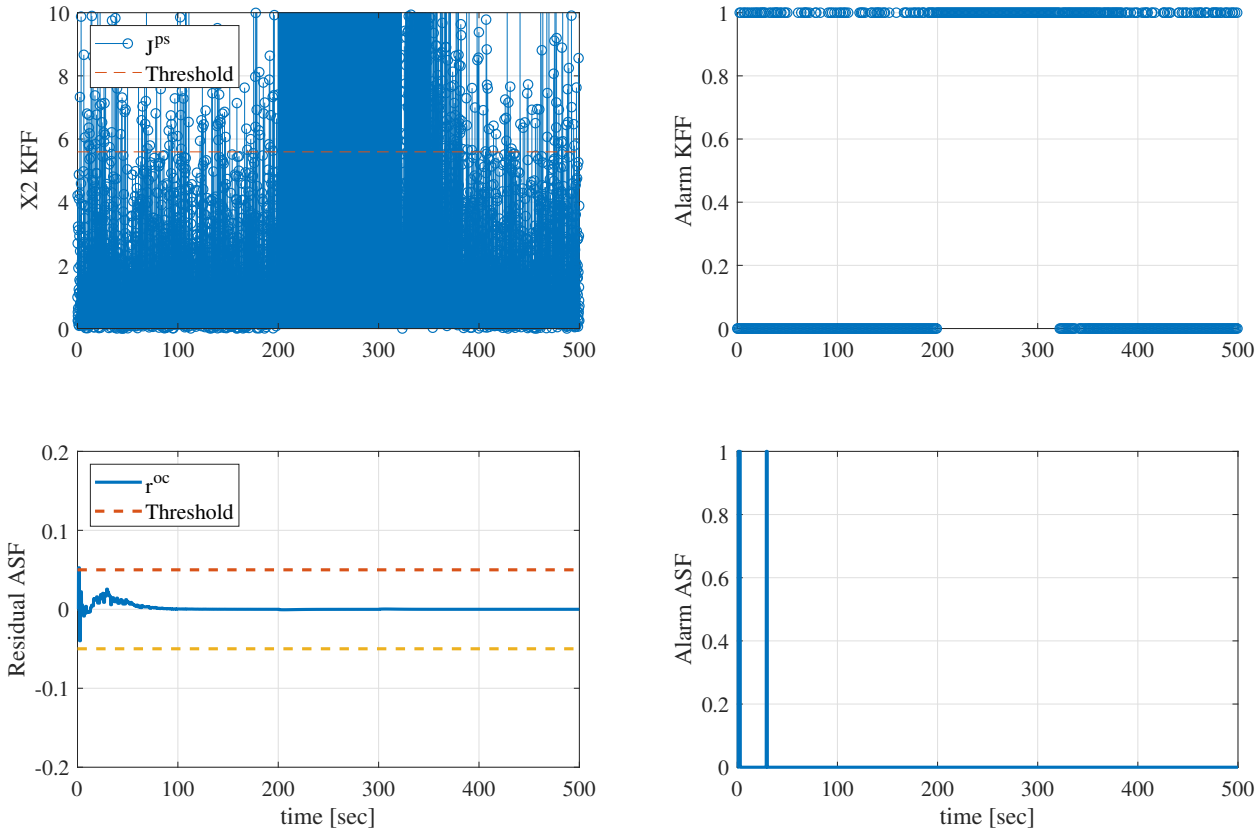

Figure 3.4: The response of the plant side filters to sensor faults on both sensors starting at $200s$.

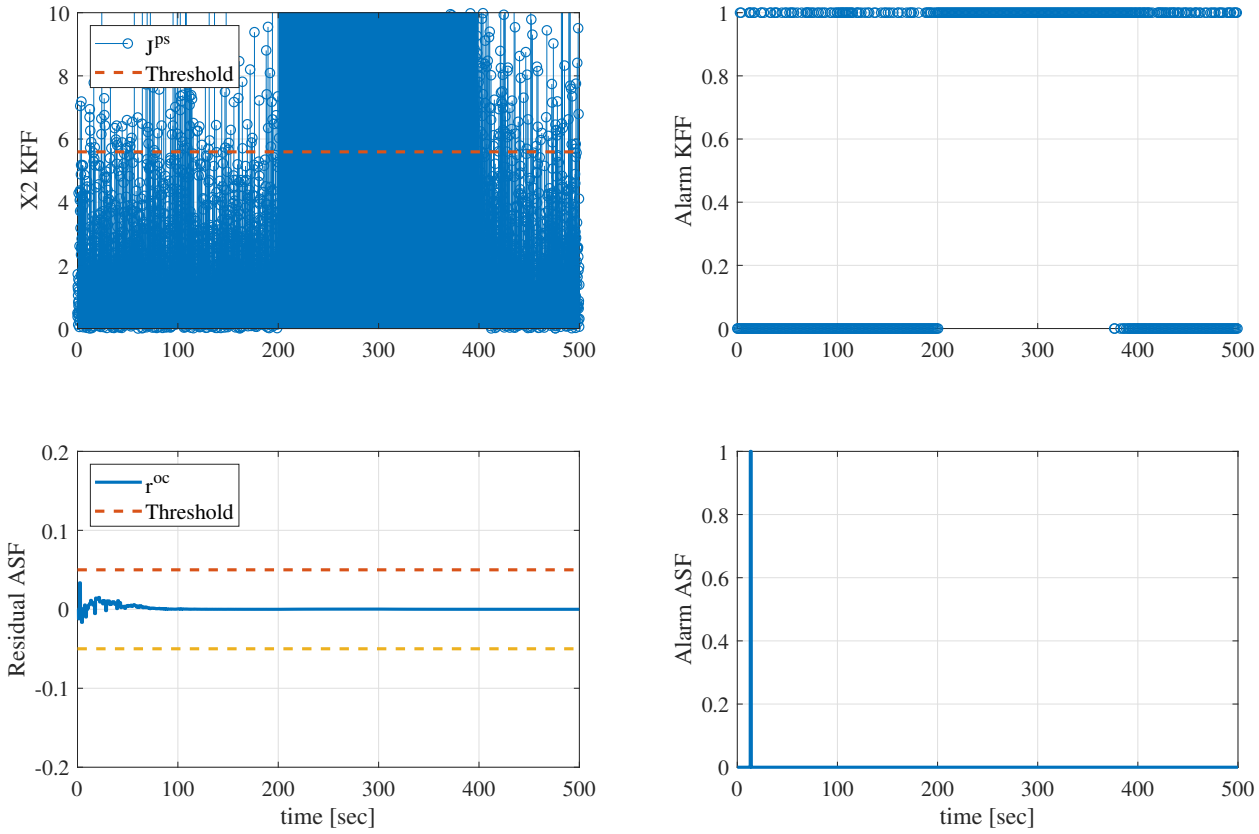

Figure 3.5: The response of the plant side filters to actuator faults on both actuators starting at $200s$.

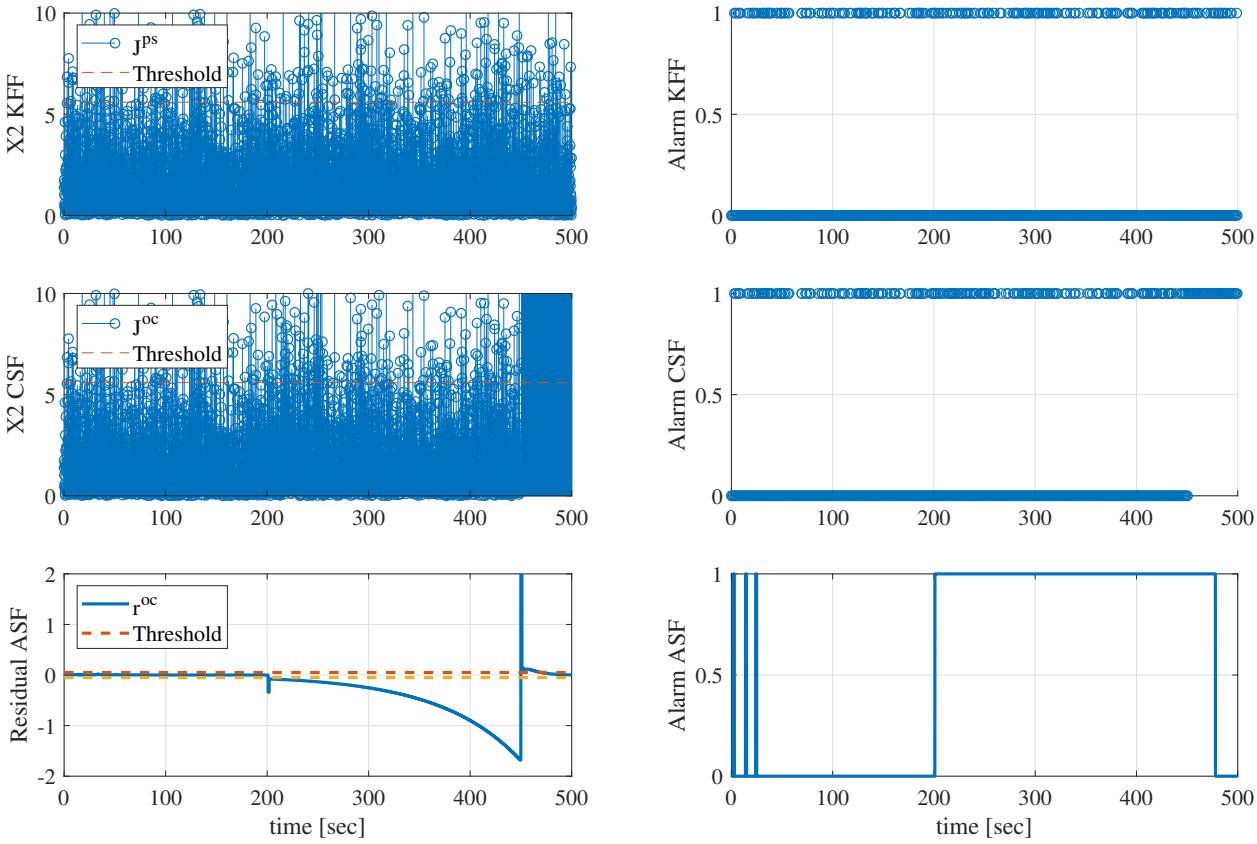

Figure 3.6: The response of the plant side and C&C filters to zero dynamics attacks starting at $200s$.

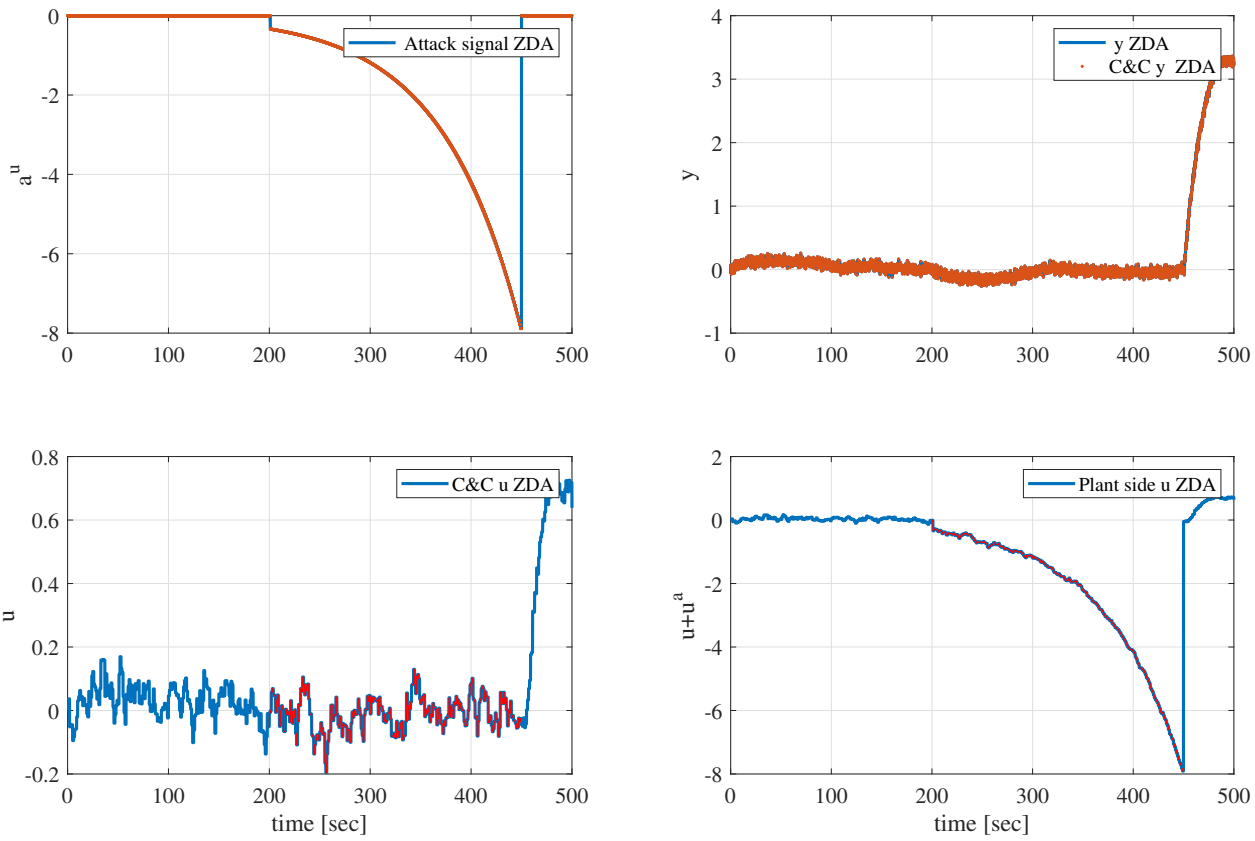

Figure 3.7: The response of the plant to ZDA starting at $200s$.

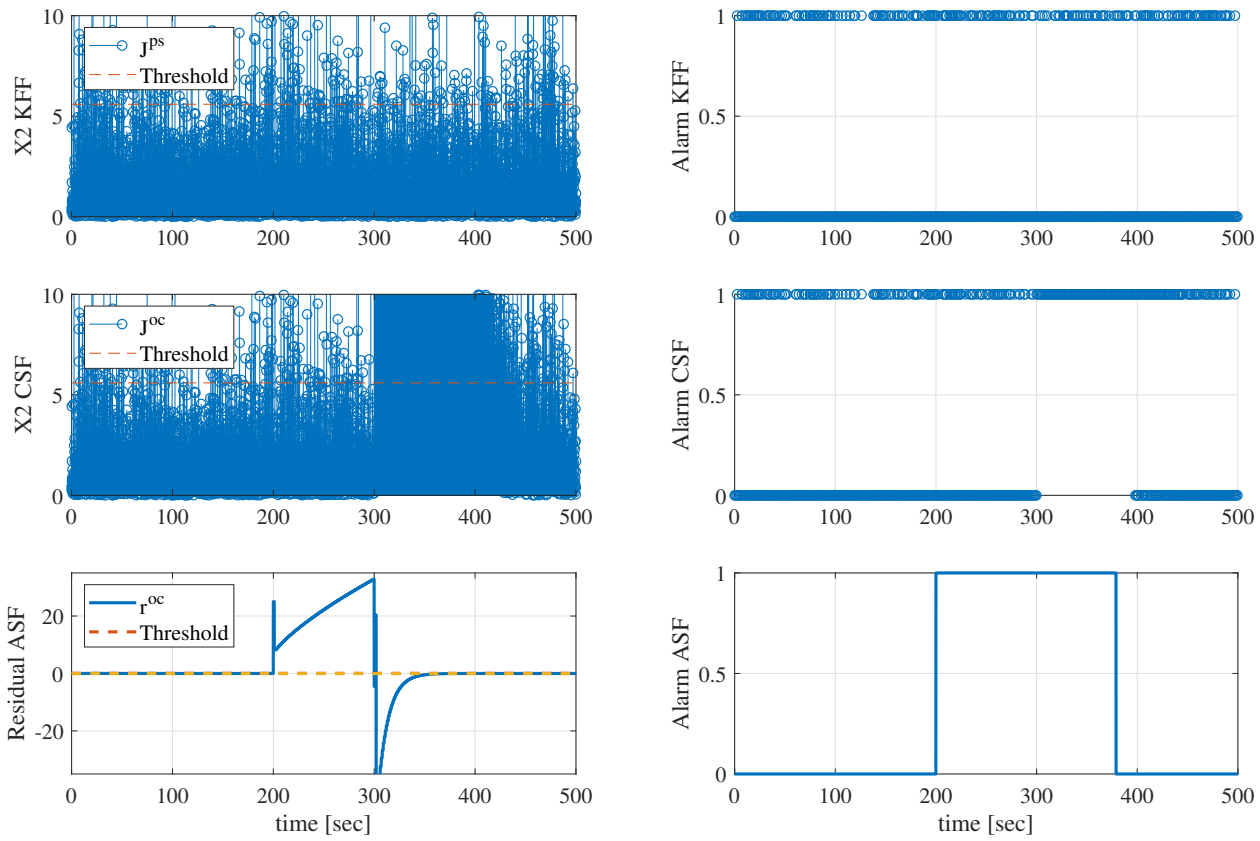

Figure 3.8: The response of the plant side and C&C filters to covert attacks starting at $200s$.

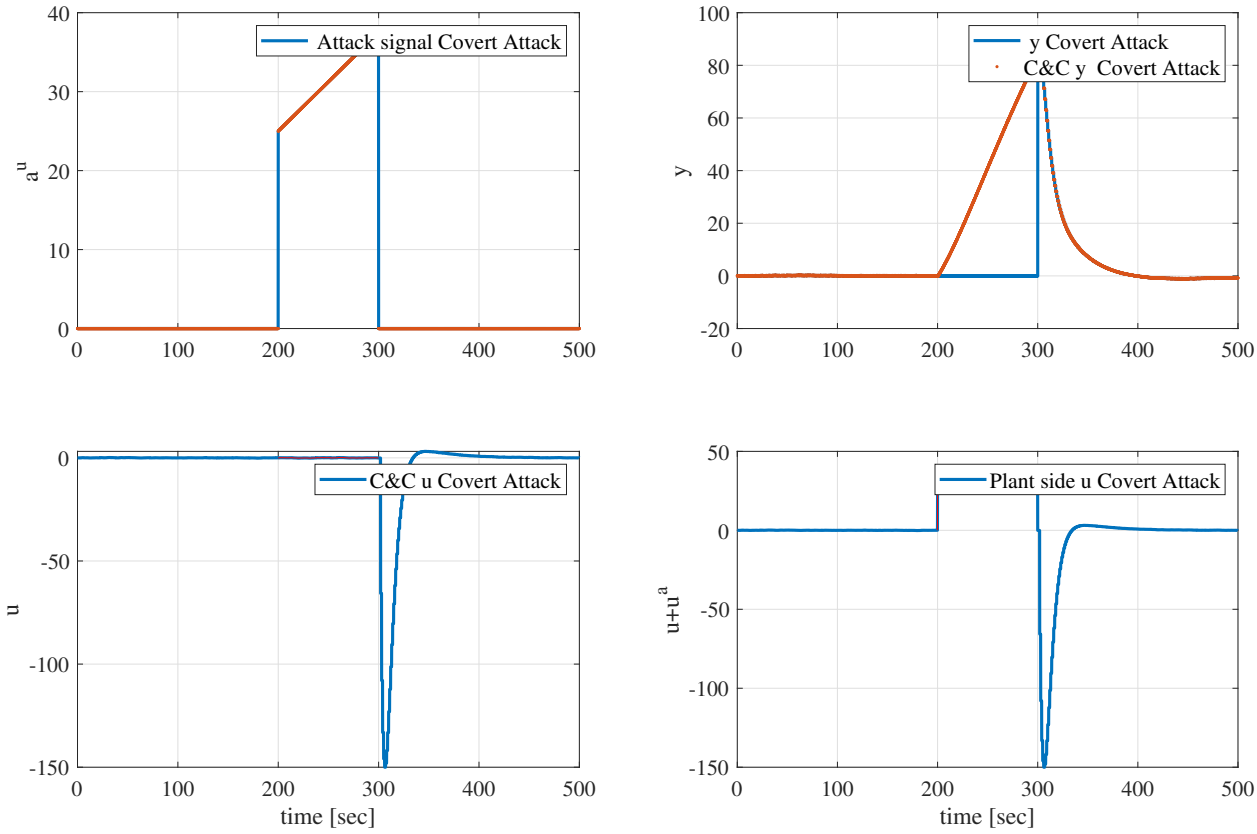

Figure 3.9: The response of the plant to covert attack starting at $200s$.

has full knowledge of the system, and in the mode 2, due to the switching, the attacker has partial knowledge, which according to [88] leads to detection of covert and zero dynamics attacks.

3.6.1 Original Work by [88]

In the methodology presented by [88], they consider a computer-generated auxiliary system on the plant-side, its output information along with the system's output measurements are transferred to the C&C, however, the effectiveness of this methodology is to detect zero dynamics and covert attacks at the C&C, and in this section, it will be compared to the proposed method and justify why such method is not suitable for distinguishing faults and attacks:

- The considered representation in [88]:

$$
\begin{aligned}
\eta_{k+1} = \begin{pmatrix} x_{k+1} \\ \tilde{x}_k \end{pmatrix} &= \underbrace{\begin{pmatrix} A & 0 \\ A_{coup} & A_{aux} \end{pmatrix}}_{A_{comb}} \underbrace{\begin{pmatrix} x_k \\ \tilde{x}_k \end{pmatrix}}_{\eta_k} \\
&+ \underbrace{\begin{pmatrix} B & 0 \\ B_{coup} & B_{aux} \end{pmatrix}}_{B_{comb}} \underbrace{\begin{pmatrix} u_k \\ \tilde{u}_k \end{pmatrix}}_{\delta_k} \\
&+ \underbrace{\begin{pmatrix} B^a & 0 \\ B^a_{coup} & B^a_{aux} \end{pmatrix}}_{B^a_{comb}} \underbrace{\begin{pmatrix} a^u_k \\ \tilde{a}^u_k \end{pmatrix}}_{\delta^{au}_k} + \underbrace{\begin{pmatrix} w_k \\ w^{aux}_k \end{pmatrix}}_{\tilde{w}} \\
\zeta_k = \begin{pmatrix} y_k \\ \tilde{y}_k \end{pmatrix} &= \underbrace{\begin{pmatrix} C & 0 \\ C_{coup} & C_{aux} \end{pmatrix}}_{C_{Comb}} \underbrace{\begin{pmatrix} x_k \\ \tilde{x}_k \end{pmatrix}}_{\eta_k} - \underbrace{\begin{pmatrix} y^a_k \\ \tilde{y}^a_k \end{pmatrix}}_{\zeta^a_k} + \underbrace{\begin{pmatrix} v_k \\ v^{aux}_k \end{pmatrix}}_{\tilde{v}_k}
\end{aligned}
\tag{3.38}
$$

In plant-side view:

$$\zeta'_k = \begin{pmatrix} y'_k \\ \tilde{y}'_k \end{pmatrix} = C_{comb}\eta_k + \tilde{v}_k \tag{3.39}$$

$$\begin{aligned} &A_{Coup} = Q_{Exp1,y}Q_{Red,y}C, B_{Coup} = Q_{Exp,u}Q_{Red,u}, \\ &C_{Coup} = Q_{Exp2,y}Q_{Red,y}C \end{aligned} \tag{3.40}$$

In the above equation, $\tilde{x}_k$ is the state of the auxiliary system, $\tilde{y}_k$ is the outputs of the auxiliary system, $\tilde{a}_k^u$ cyber-attack directed at the auxiliary system, $\tilde{y}_k^a$ is the attack on the outputs of the auxiliary, w_k^{aux} and v^{aux_k} computer based generated Gaussian process and measurement noise for auxiliary system with covariance matrices of Q_k^{aux} and R_k^{aux}, η_k are is the augmented state of the system plus the auxiliary, ζ_k is the augmented outputs of the system and auxiliary, δ is the augmented inputs of the system and auxiliary, $\delta_k^{a^u}$ is the augmented attack vectors of the system plus auxiliary, $\tilde{w}_k$ is the augmented vector of process noise, and $\tilde{v}_k$ is the augmented vector of measurements noise.

The design information for the auxiliary system is a procedure that is introduced by [88]. The design information includes matrices $Q_{Exp1,y}, Q_{Red,y}, Q_{Exp,u}, Q_{Red,u}, Q_{Exp2,y}$,$Q_{Red,y}$,C_{aux}, A_{aux}, and B_{aux} that are part of the novelty of the work by [88], and further information can be found in the referenced paper. Red subscript denotes reduction, and Exp subscript denotes expansion. $Q_{Exp1,y}$ denotes expansion at the direction of y, and $Q_{Red,y}$ denotes reducing dimension at the direction of y. Selecting the mentioned matrices is based on the designer's choices such that the multiplication of the maps as provided in (3.40) results in an auxiliary system with a reduced dimension. Important information to consider is that in the above representation, $A_{coup}x_k$ is $Q_{Exp1,y}Q_{Red,y}y'_k$ in which the y'_k represents the plant side outputs of the system.

- Attackers model-based on [88]:

$$\eta^a_{k+1} = \begin{pmatrix} x^a_{k+1} \\ \tilde{x}^a_{k+1} \end{pmatrix} = \begin{pmatrix} A & 0 \\ A_{coup} & A_{aux} \end{pmatrix} \underbrace{\begin{pmatrix} x^a_k \\ \tilde{x}^a_k \end{pmatrix}}_{\eta^a_k} + B^a_{comb} a^\delta_k$$

$$\zeta^a_k = \begin{pmatrix} y^a_k \\ \tilde{y}^a_k \end{pmatrix} = C_{comb} \eta^a_k \tag{3.41}$$

and the defender utilizes a Kalman filter designed as the following:

$$\hat{\eta}_{k+1} = A_{comb}\hat{\eta}_k + B_{comb}\delta_k + L^{comb}_k(\zeta_k - \hat{\zeta}_k)$$

$$\hat{\zeta}_k = C_{comb}\hat{\eta}_k \tag{3.42}$$

$$r_k = V(\zeta_k - \hat{\zeta}_k) \tag{3.43}$$

$$L^{comb}_k = A_{comb} P^{comb}_k C^T (C_{comb} P^{comb}_k C^T + R_k)^{-1}$$

$$P^{comb}_{k+1} = A_{comb}\big(P^{comb}_k - P^{comb}_k C^T_{comb}(C_{comb} P^{comb}_k C^T + R^{comb}_k)^{-1} C_{comb} P^{comb}_k\big) A^T + Q^{comb}_k \tag{3.44}$$

For which $\hat{\eta}$ is the estimated augmented state, $\hat{\zeta}$ is the estimated augmented outputs, L^{comb}_k is the Kalman filter gain, $Q^{comb}_k = \begin{pmatrix} Q_k & 0 \\ 0 & Q^{aux}_k \end{pmatrix}$, $R^{comb}_k = \begin{pmatrix} R_k & 0 \\ 0 & R^{aux}_k \end{pmatrix}$ and V is an arbitrary constant gain.

Remark 3.3 *$A_{coup}x_k$ results in no physical connection, but A_{coup} is designed such that its calculation is replaced form by an amount derived from the outputs of the system.*

3.6.2 Extended Representation

The representation in [88] is extended so that it includes faults. This representation helps us demonstrate why the methodology proposed in [88] is unable to differentiate faults and attacks :

$$\begin{aligned}\eta_{k+1} &= A_{comb}\eta_k + B_{comb}\delta_k + B^a_{comb}\delta^{a^u}_k + \underbrace{\begin{pmatrix} B^{f^u} & 0 \\ 0 & 0 \end{pmatrix}}_{B^f_{comb}} \underbrace{\begin{pmatrix} f^u_k \\ 0 \end{pmatrix}}_{\delta^{f^u}_k} + \tilde{w}_k \\ \zeta_k &= C_{comb}\eta_k - \zeta^a_k + \underbrace{\begin{pmatrix} D^{f^s} & 0 \\ 0 & 0 \end{pmatrix}}_{D^f_{comb}} \underbrace{\begin{pmatrix} f^s_k \\ 0 \end{pmatrix}}_{\delta^{f^s}_k} + \tilde{v}_k \end{aligned} \tag{3.45}$$

Remark 3.4 *Residual r_k generated for system (3.45) through (3.43) is unable to differentiate between the faults and attacks at C&C.*

Proof *The error dynamic between (3.42) and equation (3.45) can be written as the following:*

$$\begin{aligned} e^\eta_{k+1} &= \eta_{k+1} - \hat{\eta}_{k+1} = A_{comb}\eta_k + B_{comb}\delta_k + B^a_{comb}\delta^{a^u}_k \\ &+ B^{f^u}_{comb}\delta^{f^u} + \tilde{w}_k - A_{comb}\hat{\eta}_k + L^{comb}_k(C_{comb}\eta_k + \tilde{v}_k - \zeta^a_k \dots \\ &+ D^{f^s}\delta^{f^s}_k - C_{comb}\hat{\eta}_k) = (A_{comb} + L^{comb}_k C_{comb})e^\eta_k \\ &+ B_{comb}\delta_k + B^a_{comb}\delta^{a^u}_k + B^{f^u}_{comb}\delta^{f^u} \\ &+ L^{comb}_k(-\zeta^a_k + D^{f^s}\delta^{f^s}_k) + \epsilon_k(\tilde{v}_k, \tilde{w}_k) \\ r &= VC_{comb}e^\eta_k \end{aligned} \tag{3.46}$$

In which $\epsilon_k(\tilde{w}_k, \tilde{w}_k)$ is the optimal effect of noise on the error dynamic with respect to Kalman filter. In the above equation, although according to [88], the effect of attack can appear in $C_{comb}e^\eta_k$, due to the term C_{coup}, the residual effect of faults is coupled with the residual of attack. Therefore, a definite comment on the separation of their impact on the outputs cannot be made. □

3.7 Comparative Simulation

In the following, the capability of the methodology in [88] for isolation of actuator faults and actuator cyber-attacks is studied. The mentioned paper investigates the same quadruple tank process, and two of its auxiliary modes (mode 1, mode 2) that respectively include A^1_{aux}, A^2_{aux} are provided as follows:

$$
\begin{aligned}
A^1_{aux} &= \begin{pmatrix} 0.98267 & 3.06e10^{-5} \\ 3.06e10^{-5} & 0.98262 \end{pmatrix}, \\
A^2_{aux} &= \begin{pmatrix} 0.98273 & 1.93e10^{-5} \\ 1.93e10^{-5} & 0.98265 \end{pmatrix}, B_{aux} = \begin{pmatrix} 0.0282 \\ 0.0271 \end{pmatrix}, \\
C_{aux} &= \begin{pmatrix} 0.0615 & 0.1036 \\ 0.1756 & 0.0970 \end{pmatrix} \\
Q_{Red,y} &= \begin{pmatrix} 0.0407 & 0.1487 \end{pmatrix}, Q_{Red,u} = \begin{pmatrix} 0.1609 & 0.0150 \end{pmatrix} \\
Q_{Exp1,y} &= \begin{pmatrix} 0.1582 \\ 0.1547 \end{pmatrix}, Q_{Exp2,y} = \begin{pmatrix} 0.1670 \\ 0.0961 \end{pmatrix} \\
, Q_{Exp,u} &= \begin{pmatrix} 0.1396 \\ 0.1291 \end{pmatrix}
\end{aligned} \tag{3.47}
$$

Based on [88], at each mode one of auxiliary structures in equation (3.47) will be augmented with matrices in equation (3.33). C&C through a Kalman filter (equation (3.43)) will produce a residual r_k . It is assumed that the attacker will target mode 1 with the incorrect assumption that the auxiliary has not switched from mode 1 to mode 2, and this will lead to the detection of cyber-attacks through mode 2. As can be seen in $Fig.$3.10 and Fig. 3.11, both attack and fault trigger the alarm of the anomaly detector Auxiliary based Kalman Fault Filter (AKFF), and therefore no comment can be made on the origin of the anomaly. Furthermore, a comparison between the proposed filter and methodology presented in [88] is also provided through Table 3.1.

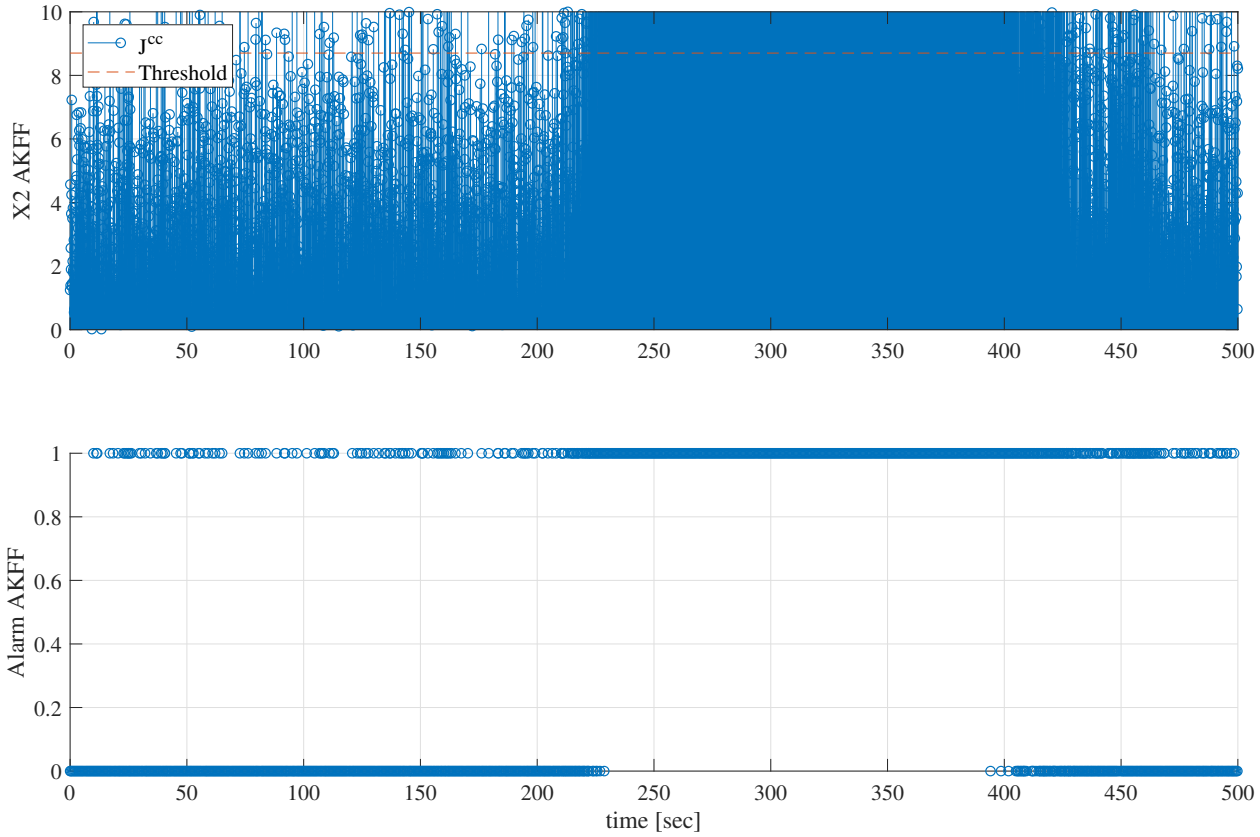

Figure 3.10: Covert attack in CPS with the starting time of $200s$.

The average $F1$ Score, d_{rate}, and f_{alarm} in Table 3.1 are computed with respect to scenarios where the CPS is under only cyber-attacks or under faults, as the mathematical analysis in (3.46) demonstrates that [88] is unable to separate faults and cyber-attacks. Therefore, anomaly detection comparison has been performed for only faults or only cyber-attacks. In this table, d_t represents the detection time that is computed for one of the simulations. d_t for faults and cyber-attacks is defined as the time past the 30th sample of 0.1 seconds for which there exist either thirty consecutive attack alarms or the current detection rate for 100 consecutive samples is bigger than 16%.

3.8 Conclusion

In this work, two filters are proposed on the plant-side for simultaneous isolation of faults and cyber-attacks. The first filter is only sensitive to inputs or output cyber-attacks, and it is shown that it can reveal covert and zero dynamics attacks directed at the plant. The second filter is also

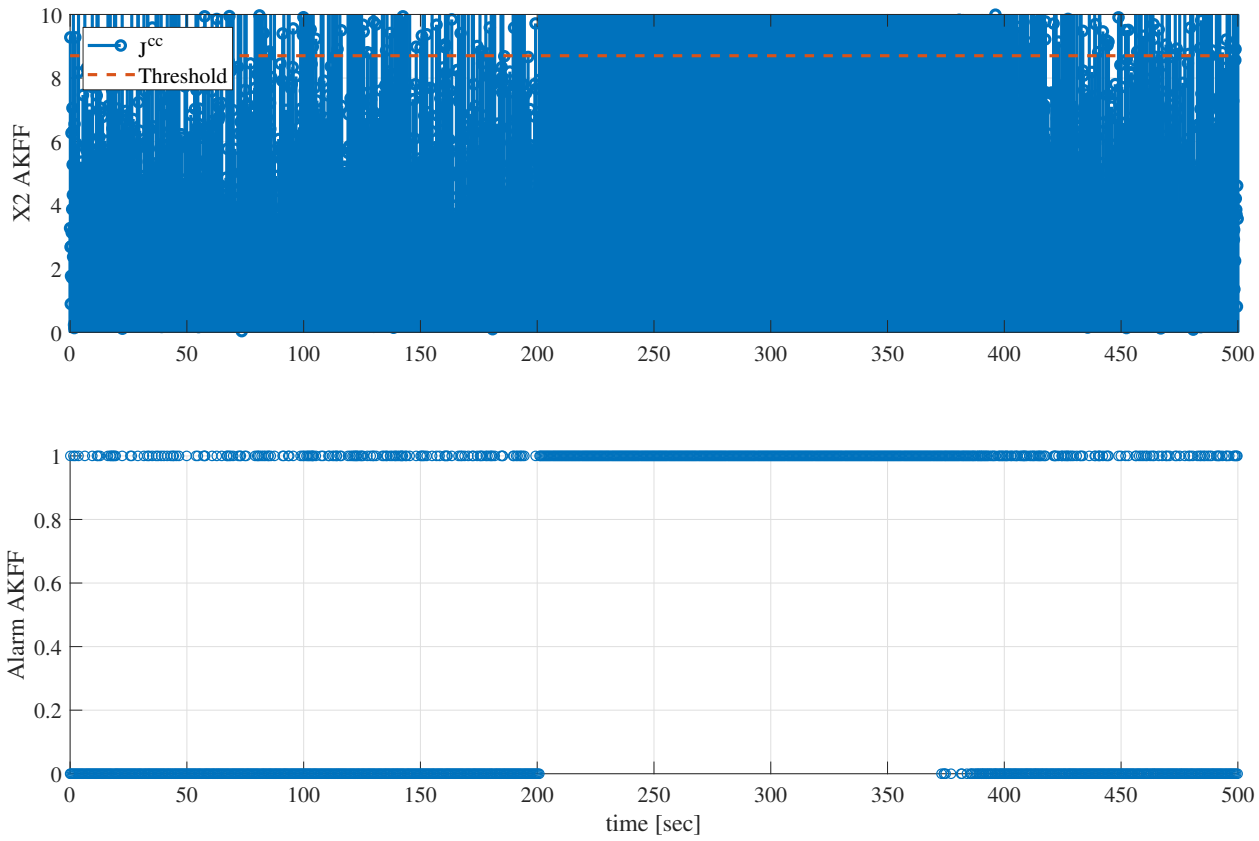

Figure 3.11: Fault on actuators with starting time of $200s$.

	Proposed method					**[88]**				
Anomaly Types	f_{alarm}	d_{rate}	F_1 Score	d_t	AFI	f_{alarm}	d_{rate}	F_1 Score	$d_t(sec)$	AFI
Covert attack	0.75	100	99.26	3.1	✓	11.4840	79.8	82.61	12	✗
Zero Dynamics Attack	1.72	99.60	99.11	3.7	✓	7.10	58.5	71.26	13.8	✗
Actuator Fault	6.82	99.05	93.14	3.3	✓	7.02	99.08	92.98	3.3	✗
Sensor Fault	7.01	100	93.44	3	✓	7.29	100	93.21	3.1	✗

Table 3.1: This table represents some of the advantages of proposed methodology with respect to an auxiliary-based control side filter presented by [88]. AFI denotes attack and fault isolation.

constructed on the plant side and is only sensitive to faults. In future work, a more sophisticated design may be developed to address the isolation of the cyber-attacks on sensors and actuators. The results are derived for a specific model of control law in the C&C, while it may also be possible to accomplish a similar result by considering other types of models for control law, however, the type of monitoring filter and the analysis provided in Section 3.4. We also provided detailed comparison of the proposed methodology with an auxiliary based methodology for which Table 3.1 provide a summary of this comparison.

Chapter 4

A Dual-Rate Auxiliary-Based Approach for Actuators Cyber-Attack Estimation in Presence of Faults

4.1 Introduction

Safety of control systems used to be particularly viewed through the faults as the main unknown inputs of the system, and an active path toward control reconfiguration is considered as unknown inputs estimation. Estimation of faults has been addressed in the literature through a variety of methodologies such as Kalman filter [42], UIO [16], adaptive [47], data-driven methodologies [75] for single-rate sampled-data systems as well as multi-rate systems. In this regard, multi-rate systems have been beneficial in terms of fault diagnosis and estimation applications while giving freedom to relocate non-minimum phase sampling zeros of systems and providing a framework for solving problems with uniform and non-uniform data sampling [5, 113]. However, with the advent of cyber-attacks in the era of cyber-physical technologies, these new unknown inputs should be considered in the study and analysis of CPS systems to enhance their security.

In the last decade, there have been numerous studies that address cyber-attacks as the stand-

alone threat against cyber-physical systems. One of the fundamental approaches to studying security against cyber-attacks is through the notion of invariant zeros and investigating zero dynamics attacks. Introducing the auxiliary system is one of the suggested techniques in the literature to deal with the zero dynamics attacks [8, 88], which can be considered as manipulating the invariant zeros of the overall augmented system of the plant plus auxiliary. In [88], auxiliary systems are defined as virtual processes that are coupled with the plant's inputs and output and together form an augmented system. Designing an observer for the augmented system while assuming that the attacker does not hold the perfect knowledge of auxiliary is the suggested way of [88] to encounter zero dynamics attacks. Utilizing the command input modulation matrix in the plant-side [37] is another defensive action that researchers proposed for changing the function that maps command inputs (entering from the network) to the output. Based on the idea that the attacker is not aware of the mentioned changes, these methodologies offer a change in mapping function from inputs passing through the network to the output, in which the attacker cannot exploit and insert a successful zero dynamics attack.

In [77], based on increasing the sampling rate and forming a dual-rate CPS, they prove all its sampling zeros are stable, and the CPS is immune to zero dynamics attacks. Dual-rate, which is defined as the special case of multi-rate systems where the inputs hold rate and output sampling rates are different, but fixed [77]. Based on [77], through the viewpoint of the (fast output-rate) dual-rate system, one could design a filter that can detect zero dynamics attacks targeting the unstable modes of the system in a situation where the adversary is not aware of the fast sampling rate, therefore, the attacker will not be able to have a successful zero dynamics attack due to a change in the mapping function between network command inputs and outputs, and the defender makes it impossible for the attacker to find an input function that belongs to the null space of fast rate output.

Furthermore, in [36], a methodology based on the aforementioned dual-rate approach of [77] is presented that designs an unknown input estimator based on the properties of the dual-rate system for an attack only scenario. Besides [36], there are a few studies in the literature that address the problem of attack estimation through resilient-based methodologies, which limit the type of

cyber-attacks to false data injection attacks [29, 56].

This chapter investigates the estimation of actuator cyber-attacks in the presence of actuator and sensor faults in the C&C from the viewpoint of fast output dual-rate systems while assuming that the output channels are safe from cyber-attack. In order to deal with this, a fault filter is employed at the plant side that securely transmits its residual to the C&C with a special mechanism that will be formally defined in Section II.

Employing fast output dual-rate sampling for the plant causes strongly detectability, which is a necessary and sufficient condition for designing delayed Unknown Input Observer (UIO) [95, Theorem 6]. Delayed UIO has been used for inversion-based tracking of the unknown inputs in works such as [74, 95], and has the decoupling capability, resilient estimation, and is not computationally expensive, unlike convex estimation methods such as [33]. Consequently, a UIO is considered to solve the actuator cyber-attack estimation problem in presence of faults.

The organization of this chapter is as the following. Section 4.2 is the problem statement and problem formulation. Section 4.3 is the actuator cyber-attacks and fault estimator design. Section 4.4 is the simulation and Section 4.6 is the conclusion.

4.2 Problem Statement and Formulation

4.2.1 Considered System

The considered CPS in this chapter includes an outer control law and inner or local control law, as depicted in Fig. 4.1. The inner control is responsible for the overall stability of the plant, while the outer control law performs tracking control of the combined plant as well as the inner control. The inner control consideration is not restrictive and does not affect the overall conclusion of the chapter, where it can be extended to situations where all control commands for the plants are decided in the C&C.

Through the outer control loop view Fig. 4.1 the physical system plus controller operate under two different rates, meaning the A/D converter sends the transmitted output $y(k)$ with the rate $\frac{T}{N}$

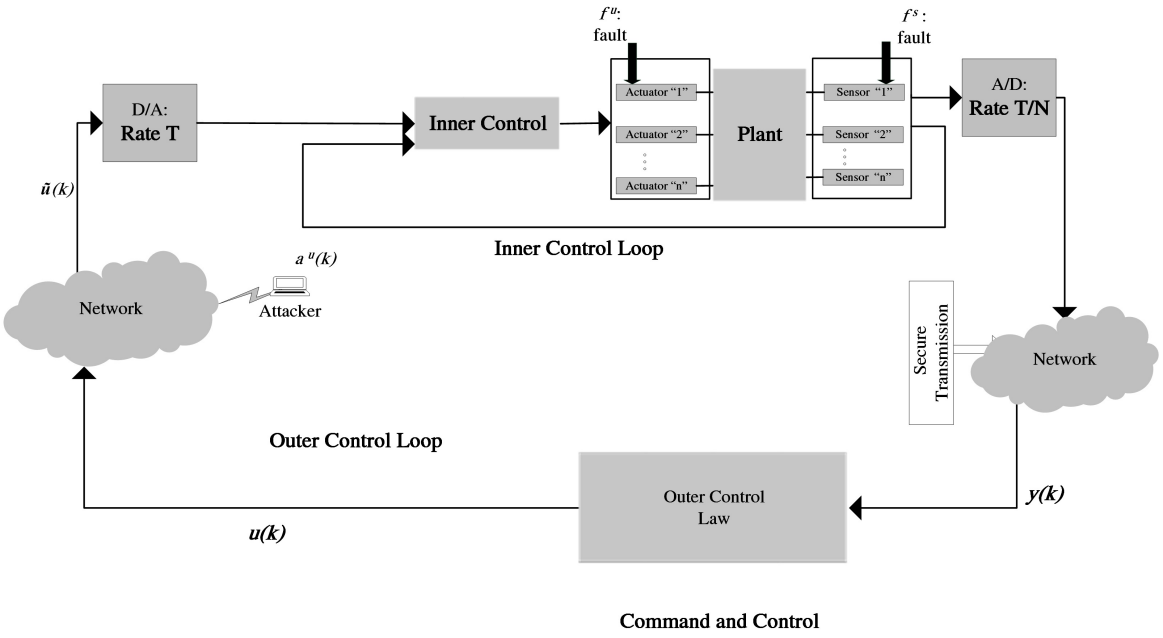

Figure 4.1: The CPS framework.

and the D/A converter holds the commands inputs $u(k)$ received from C&C for a period of T.

The governing dynamic for the CPS under actuator faults and actuator cyber-attacks can be defined in three steps:

Step 1: Continuous-Time Representation: Consider a continuous-time representation of cyber-physical systems. The following representation considers actuator faults as well as the cyber-attack through communication channels dedicated to the inputs:

$$\begin{aligned} \dot{x} &= A_c x + \underbrace{B_c u + B_c^{a^u} a^u}_{B\tilde{u}(k)} + B_c^{f^u} f^u \\ y &= C_c x + D^{f^s} f^s \end{aligned} \tag{4.1}$$

Definition 4.1 *Plant (4.1) is considered tall, meaning the number of outputs p is greater or equal to the number of unknown inputs.*

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ is the known inputs with dimension of m, $a^u \in \mathbb{R}^{m_{a^u}}$ denotes the actuator cyber-attacks with dimension of m_{a^u}, $f^u \in \mathbb{R}^{m_{f^u}}$ denotes the actuator faults with

dimension of m_{f^u}, $y \in \mathbb{R}^p$, $f^s(k) \in \mathbb{R}^{m_{fs}}$ denotes the sensor faults, $y \in \mathbb{R}^p$, $\tilde{u} = u + \Gamma a^u$, and $\Gamma = [\alpha_{i,j}] \in R^{m \times m_a}$. Γ is a diagonal or rectangular diagonal matrix such that based on the attacked channels entries $\alpha_{i,i}$ are either zero or one, and the rest of entries are zero. $A_c \in \mathbb{R}^{n \times n}$, $B_c \in \mathbb{R}^{n \times m}$,$B_c^{a^u} \in \mathbb{R}^{n \times m_{a^u}}$, $B_c^{f^u} \in \mathbb{R}^{n \times m_{f^u}}$, $C_c \in \mathbb{R}^{p \times n}$, and $D_c^{f^s} \in \mathbb{R}^{p \times m_{f^s}}$.

Remark 4.1 *In equation (4.1), A_c, B_c, and C_c is representative of a combined physical plant and control system if the CPS needs a local control system such that it ensures stability on the plant side. The system matrices can also be representative of the physical plant such that all command inputs are received from C&C. Therefore, the proposed methodology is not restrictive of the existence of a local controller.*

Step 2: Discrete-Time Representation: A discrete-time state-space representation of (4.1) with a selected period of $\frac{T}{N}$ can be achieved by the following transformation:

$$\begin{aligned} A = e^{A_c \frac{T}{N}}, B = \int_0^{\frac{T}{N}} e^{A_c \tau} B_c d\tau \\ B^{a^u} = \int_0^{\frac{T}{N}} e^{A_c \tau} B_c^{a^u} d\tau, B^{f^u} = \int_0^{\frac{T}{N}} e^{A_c \tau} B_c^{f^u} d\tau \\ D^{f^s} = D_c^{f^s}, C = C_c \end{aligned} \tag{4.2}$$

this representation is shown according to the following:

$$\begin{aligned} x(k+1) = Ax(k) + \underbrace{Bu(k) + B^{a^u} a^u(k)}_{B\tilde{u}(k)} + B^f f^u(k) \\ y(k) = Cx(k) + D^{f^s} f^s(k) \end{aligned} \tag{4.3}$$

where $k = 0, 1, \ldots$.

In the $\mathscr{Z}$ transform domain, the equation (4.3) can be described as the following:

$$\begin{aligned} y(z) &= (C(zI - A)^{-1}B)u(z) \\ &+ (C(zI - A)^{-1}B^{a})a^{u}(z) \\ &+ (C(zI - A)^{-1}B^{f^{u}})f^{u}(z) + D^{f^{s}}f^{s}(z) \end{aligned} \tag{4.4}$$

where z is a forward shift operator such that $zu(k) = u(k+1)$ or $zy(k) = y(k+1)$.

Step 3: **Dual-Rate Representation:** Next, we define a linear time-invariant representation for the system with operating rate T for command inputs and operating rate $\frac{T}{N}$ for outputs also known as blocked or lifted representation in multi-rate systems, which is a well-known methodology for transforming multi-rate representations to linear time-invariant representations of systems and is extensively explained throughout the literature [18, 108].

Consider the discrete-time system (4.3), with same sampling rate $\frac{T}{N}$ and changing of the hold rate to T. Since sampling to hold ratio of $y(k)$ and $\tilde{u}(k)$ is equivalent to $\frac{1}{N}$ in a uniform manner. For instances that $k = 0, N, 2N, \ldots$, we can define vectors $\mathcal{U}(k)$, $\mathcal{A}(k)$ as well as $\mathcal{F}^{s}(k)$ and $\mathcal{F}^{u}(k)$ in the following manner:

$$\begin{aligned} \mathcal{U}(k) &= \begin{pmatrix} u(k) & u(k) & \ldots & u(k) \end{pmatrix}^{T} \\ \mathcal{F}^{u}(k) &= \begin{pmatrix} f^{u}(k+1) & f^{u}(k+2) & \ldots & f^{u}(k+N-1) \end{pmatrix}^{T} \\ \mathcal{F}^{s}(k) &= \begin{pmatrix} f^{s}(k+1) & f^{s}(k+2) & \ldots & f^{s}(k+N-1) \end{pmatrix}^{T} \\ \mathcal{A}(k) &= \begin{pmatrix} a(k) & a(k) & \ldots & a(k) \end{pmatrix}^{T} \end{aligned} \tag{4.5}$$

Consequently, a shift operator Z is defined such that it satisfies $Z\mathcal{U}(k) = \mathcal{U}(k+N)$, $Z\mathcal{F}^{u}(k) = \mathcal{F}^{u}(k+N)$, $Z\mathcal{F}^{s}(k) = \mathcal{F}^{s}(k+N)$, $Z\mathcal{A}(k) = \mathcal{A}(k+N)$, $Zx(k) = x(k+N)$, $ZY(k) = Y(k+N)$.

With this modification, the blocked or lifted system will be given by:

$$\begin{aligned} x(k+N) &= \underline{\mathbf{A}}x(k) + \underline{\mathbf{B}}\mathcal{U}(k) + \underline{\mathbf{B}}^{f^u}\mathcal{F}(k) + \underline{\mathbf{B}}^{a^u}\mathcal{A}(k) \\ Y(k) &= \underline{\mathbf{C}}x(k) + \underline{\mathbf{D}}\mathcal{U}(k) + \underline{\mathbf{D}}^{f^u}\mathcal{F}^u(k) \\ &\quad + \underline{\mathbf{D}}^{f^s}\mathcal{F}^s(k) + \underline{\mathbf{D}}^{a^u}\mathcal{A}(k) \\ Y(k) &= \begin{pmatrix} y(k) & y(k+1) & \dots & y(k+N) \end{pmatrix}^T \end{aligned} \tag{4.6}$$

Where:

$$\begin{aligned} \underline{\mathbf{A}} &= A^N, \underline{\mathbf{B}}^{f^u} = \begin{pmatrix} A^{N-1}B^{f^u} & A^{N-2}B^f & \dots & B^{f^u} \end{pmatrix}, \\ \underline{\mathbf{B}}^a &= \begin{pmatrix} A^{N-1}B^a & A^{N-2}B^a & \dots & B^a \end{pmatrix}, \\ \underline{\mathbf{B}} &= \begin{pmatrix} A^{N-1}B & A^{N-2}B & \dots & B \end{pmatrix}, \\ \underline{\mathbf{D}} &= \begin{pmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-2}B & CA^{N-3}B & \cdots & D \end{pmatrix}, \underline{\mathbf{D}}^{a^u} = \begin{pmatrix} D^a & 0 & \cdots & 0 \\ CB^a & D^a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-2}B^a & CA^{N-3}B^a & \cdots & D^a \end{pmatrix} \\ \underline{\mathbf{D}}^{f^u} &= \begin{pmatrix} 0 & 0 & \cdots & 0 \\ CB^f & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-2}B^f & CA^{N-3}B^f & \cdots & 0 \end{pmatrix}, \underline{\mathbf{C}} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{N-1} \end{pmatrix} \\ \underline{\mathbf{D}}^{f^s} &= \begin{pmatrix} D^{f^s} & 0 & \cdots & 0 \\ 0 & D^{f^s} & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & D^{f^s} \end{pmatrix} \end{aligned} \tag{4.7}$$

and where $\underline{\mathbf{A}} \in \mathbb{R}^{n\times n}$, $\underline{\mathbf{B}} \in \mathbb{R}^{n\times Nm}$, $\underline{\mathbf{B}}^{a^u} \in \mathbb{R}^{n\times Nm_{a^u}}$, $\underline{\mathbf{B}}^{f^u} \in \mathbb{R}^{n\times Nm_{f^u}}$, $\underline{\mathbf{C}} \in \mathbb{R}^{Np\times n}$, $\underline{\mathbf{D}} \in$

$\mathbb{R}^{Np\times m}$,$\underline{\mathbf{D}}^{a^u} \in \mathbb{R}^{Np\times Nm_{a^u}}$,$\underline{\mathbf{D}}^{f^u} \in \mathbb{R}^{Np\times Nm_{f^u}}$,$\underline{\mathbf{D}}^{f^s} \in \mathbb{R}^{Np\times Nm_{f^s}}$ $Y(k) \in \mathbb{R}^{Np}$, Np is equal to N multiplied by p, Nm is equal to N multiplied by m, Nm_{f^s} is equal to N multiplied by m_{f^s}, and Nm_{a^u} is equal to N multiplied by m_{a^u}.

The above equation in the $\mathscr{Z}$ transform domain can be described as the following:

$$\begin{aligned} Y(z) &= (\underline{\mathbf{C}}(z^N I - \underline{\mathbf{A}})^{-1}\underline{\mathbf{B}} + \underline{\mathbf{D}})\mathcal{U}(z) \\ &+ (C(z^N I - \underline{\mathbf{A}})^{-1}\underline{\mathbf{B}}^a + \underline{\mathbf{D}}^a)\mathcal{A}(z) \\ &+ (\underline{\mathbf{C}}(z^N I - \underline{\mathbf{A}})^{-1}\underline{\mathbf{B}}^f + \underline{\mathbf{D}}^f)\mathcal{F}(z) \end{aligned} \tag{4.8}$$

Lastly, the equation (4.6) can be simplified as the following:

$$\begin{aligned} x(k+N) &= \tilde{A}x(k) + \tilde{B}u(k) + \tilde{B}^{a^u}a^u(k) + \tilde{B}^{f^u}\mathcal{F}^u(k) \\ Y(k) &= \tilde{C}x(k) + \tilde{D}u(k) + \tilde{D}^{a^u}a^u(k) \\ &+ \tilde{D}^{f^u}\mathcal{F}^u(k) + \tilde{D}^{f^s}\mathcal{F}^s(k) \end{aligned} \tag{4.9}$$

$$\begin{aligned} &\tilde{A} = \underline{\mathbf{A}}, \tilde{B} = \sum_{k=0}^{N-1} A^k B, \tilde{B}^{a^u} = \sum_{k=0}^{N-1} A^k B^{a^u}, \\ &\tilde{B}^{f^u} = \sum_{k=0}^{N-1} A^k B^{f^u}, \tilde{D} = \begin{pmatrix} 0 \\ CB \\ \vdots \\ C\sum_{k=0}^{N-2} A^k B \end{pmatrix} \\ &\tilde{D}^{a^u} = \begin{pmatrix} 0 \\ CB^{a^u} \\ \vdots \\ C\sum_{k=0}^{N-2} A^k B^{a^u} \end{pmatrix}, \tilde{D}^{f^u} = \begin{pmatrix} 0 \\ CB^{f^u} \\ \vdots \\ C\sum_{k=0}^{N-2} A^k B^{f^u} \end{pmatrix} \\ &\tilde{C} = \underline{\mathbf{C}}, \tilde{B}^{f^u} = \underline{\mathbf{B}}^{f^u}, \tilde{D}^{f^s} = \underline{\mathbf{D}}^{f^s} \end{aligned} \tag{4.10}$$

where $\tilde{B} \in \mathbb{R}^{n\times m}$, $\tilde{B}^{a^u} \in \mathbb{R}^{n\times m_{a^u}}$, $\tilde{D} \in \mathbb{R}^{Np\times m}$,$\tilde{D}^{a^u} \in \mathbb{R}^{Np\times m_{a^u}}$, and $\tilde{D}^{f^u} \in \mathbb{R}^{Np\times m_{f^u}}$.

4.2.2 Type of Cyber-Attacks

The type of considered attacks in this section is limited to zero dynamics and false data injection attacks.

1. Zero dynamics attacks: Consider the pencil matrix associated with the output entry of the system defined as the following:

$$\mathcal{P}_{plant}(z) = \begin{pmatrix} z^N I - \tilde{A} & \tilde{B}^{a^u} \\ C & 0 \end{pmatrix} \tag{4.11}$$

Definition 4.2 (Zero Dynamics Attack for Plant) *[10, Definition 2] For $z_0 \in \mathbb{C}$, if $rank\big(\mathcal{P}_{plant}(z_0)\big) \leq l$, where l is equal to normrank of $\mathcal{P}_{plant}(z)$, then there exist an attack vector $a_k = (a_k^{uT}\ 0_{1\times p}^T)^T$ called zero dynamics attack for plant, where a_k is equal to $a_0^u(z_0^N)^k$, s.t (z_0^N), a_0, and x_0 satisfy the following:*

$$\mathcal{P}_{plant}(z_0)\begin{pmatrix} x_0^T & a_0^{uT} & 0_{1\times p}^T \end{pmatrix}^T = 0 \tag{4.12}$$

2. False data injection attacks:

   Aside from zero dynamics attacks, in this work only non-stealthy false data injection attacks such as bias injection attacks.

4.2.3 Designing the Cyber-Attack Sensitive Filter

First, we consider two filters: one at the plant side (Plant Side Filter (PSF)) and one at the C&C (CSF) as shown in Fig. 4.2.

1. **Plant Side Filter (PSF):** Consider a fast rate fault filter with inputs and output rate of $\frac{T}{N}$ and residual output $\bar{r}_k$ that is constructed according to the following:

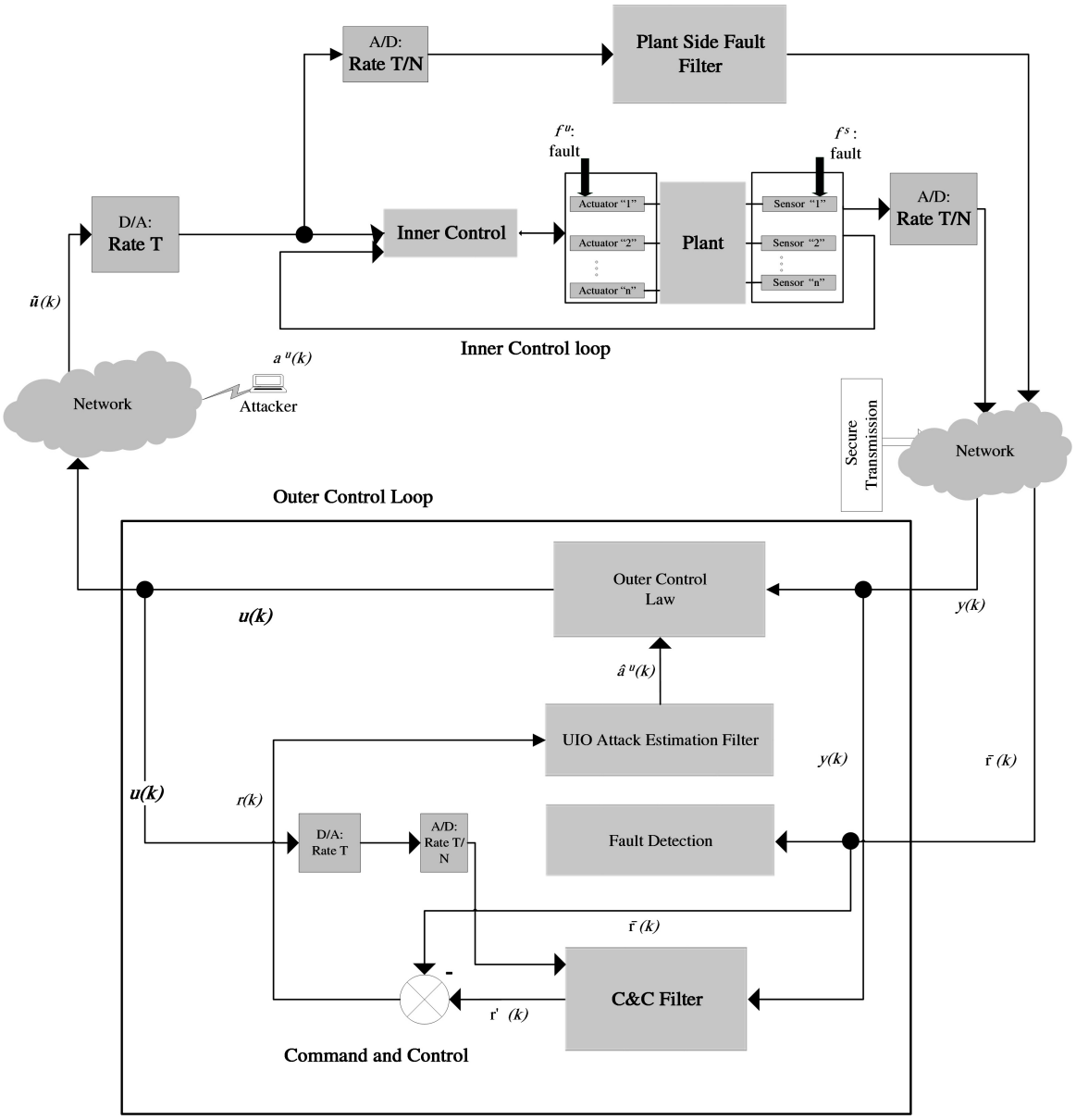

Figure 4.2: The schematic of our methodology. The system is assumed to be prone to simultaneous actuator cyber-attacks and actuator faults. Sensor channels are assumed to be secure and without any fault.

$$\bar{x}(k+1) = (A - LC)\bar{x}(k) + \underbrace{B\tilde{u}(k)}_{Bu(k)+B^{a^u}(k)a^u(k)} + Ly(k)$$
$$\bar{r}(k) = (C\bar{x}(k) - y(k)) \tag{4.13}$$

where L is a filter gain, $\bar{r}(k) \in \mathbb{R}^{p\times n}$ is the residual, and $\bar{x}(k) \in \mathbb{R}^{n\times n}$ is the state of the observer.

The error dynamic $\bar{e}(k+1) = x(k+1) - \bar{x}(k+1)$ of the above filter along with its residual $\bar{r}(k)$ can be defined as the following:

$$\bar{e}(k+1) = (A - LC)\bar{e}(k) + B^{f^u} f^u(k)$$
$$\bar{r}(k) = C\bar{e}(k) + D^{f^s} f^s(k) \tag{4.14}$$

The following mechanism based on a threshold level for each vector of $\bar{r}$ is considered for sending information to the C&C according the following:

$$\begin{cases} if & abs(\bar{r}_i(k)) > 0 \; for \; i = 0, 1, \ldots p \\ & \\ & then \; \bar{r}_i(k) \text{ transmitted to } C\&C \\ otherwise & \bar{r}_i(k) \text{ will be set to zero by C\&C} \end{cases} \tag{4.15}$$

2. **C&C Filter (CCF):**Next we construct a filter in a C&C in a similar manner to (4.13):

$$\begin{aligned} \acute{x}(k+1) &= (A - LC)\acute{x}(k) + Bu(k) + Ly(k) \\ \acute{r}(k) &= (C\acute{x}(k) - y(k)) \end{aligned} \tag{4.16}$$

where the error dynamic $\acute{e}(k+1) = \acute{x}(k+1) - x(k+1)$ is defined by the following equation:

$$\begin{aligned} \acute{e}(k+1) &= (A - LC)\acute{e}(k) + B^{f^u} f^u(k) + B^{a^u} a^u(k) \\ \acute{r}(k) &= C\acute{e}(k) + D^{f^s} f^s(k) \end{aligned} \tag{4.17}$$

3. **Cyber-Attack Sensitive Filter (CASF):**We define a attack model-based on the PSF and CCF with fast-rate error $e(k+1) = \acute{e}(k+1) - \bar{e}(k+1)$ and residual r defined as the following:

$$\begin{aligned} e(k+1) &= (A - LC)e(k) + B^{a^u} a^u(k) \\ &= A_e e(k) + B^{a^u} a^u(k) \\ r(k) &= Ce(k) \end{aligned} \tag{4.18}$$

and dual-rate representation defined as the following:

$$\begin{aligned} e(k+N) &= \tilde{A}_e e(k) + \tilde{B}_e^{a^u} a^u(k) \\ R(k) &= \tilde{C}_e e(k) + \tilde{D}_e^{a^u} a^u(k) \\ R(k) &= \begin{pmatrix} r(k) & r(k+1) & \dots & r(k+N) \end{pmatrix}^T \end{aligned} \tag{4.19}$$

in which by replacing $A_e \to A$ in (4.10), respectively, $\tilde{A}_e$, $\tilde{B}_e^{a^u}$, $\tilde{C}_e$, and $\tilde{D}_e^{a^u}$ can be realized in the same manner as $\tilde{A}$, $\tilde{B}^{a^u}$, $\tilde{C}$, and $\tilde{D}^{a^u}$.

4.2.4 UIO Attack Estimator

Next, an attack estimator constructed corresponding to (4.19), that has the following structure:

$$\begin{aligned}\hat{e}(k-L+N) &= E\hat{e}(k-L) + F\mathcal{R}(k-L:k)\\ \mathcal{R}(k-L:k) &= \mathcal{O}_L e(k-L) + \mathcal{J}_L a^u(k-L:k)\end{aligned} \tag{4.20}$$

where

$$\begin{aligned}\mathcal{R}(k-L:k) = (&R(k-L+N) \quad R(k-L+2N)\\ &\dots \quad R(k))^T\end{aligned} \tag{4.21}$$

$$\begin{aligned}a^u(k-L:k) = (&a^u(k-L+N) \quad a^u(k-L+2N)\\ &\dots \quad a^u(k))^T\end{aligned}$$

$$\mathcal{O}_L = \begin{pmatrix} \tilde{C}_e \\ \mathcal{O}_{L-N} A_e \end{pmatrix} = \begin{pmatrix} \mathcal{O}_{L-N} \\ \tilde{C}_e \tilde{A}_e^L \end{pmatrix} \tag{4.22}$$

$$\mathcal{J}_L = \begin{pmatrix} \tilde{D}_e^{a^u} & 0 \\ \mathcal{O}_{L-N}\tilde{B}_e^{a^u} & \mathcal{J}_{L-N} \end{pmatrix} = \begin{pmatrix} \mathcal{J}_{L-N} & 0 \\ \tilde{C}_e \mathcal{C}_{L-N} & \tilde{D}_e^{a^u} \end{pmatrix} \tag{4.23}$$

where $\mathcal{C}_{L-N}$ is the observability matrix between the pair $(\tilde{A}_e, \tilde{B}^{a^u})$. L represents a fixed number of delays, for the computation of the estimation, it can be selected through satisfying left invertibility condition, where the minimum number of delays is between N to $n \times (N-1)$, and the exact number of delays depends on the system [107, Theorem 1], E, F are the design parameters for the observers.

The error dynamic $e_e(k-L+N) = e(k-L+N) - \hat{e}(k-L+N)$ for this estimator, which

is based on equations (4.20) and (4.19) is described in the following:

$$\begin{aligned} e_e(k-L+N) &= \tilde{A}_e e(k-L) + \tilde{B}_e^{a^u} a^u(k-L) \\ &\quad - E\hat{e}(k-L) - F\mathcal{R}(k-L:k) \\ &= Ee_e(k-L) + (\tilde{A}_e - E - F\mathcal{O}_L)e(k-L) \\ &\quad - F\mathcal{J}_L a^u(k-L:k) + \tilde{B}^{a^u} a^u(k-L) \end{aligned} \tag{4.24}$$

4.2.5 Assumption

The key assumptions are as the following:

Assumption 4.1 *$\bar{r}(k)$ and $\bar{y}(k)$ are securely transmitted to $C\&C$.*

Assumption 4.2 *It is assumed that B^{a^u} is full column ranks, and system (4.3) has no zero on unit circle.*

Assumption 4.3 *The plant in the view of the outer control loop is tall, meaning the number of output $y(k)$ is more or equal to the number of command inputs $u(k)$.*

Remark 4.2 *The amount of delay L in this ac UIO is dependent on the satisfaction of the left invertibility condition [93]. According to [95], for a strongly detectable linear time-invariant system such a delay always exists.*

4.2.6 Objective and Motivation: Actuator Cyber-Attacks Estimation in Presence of Sensor and Actuator Faults

In this work, the main objective is to present a methodology for (4.9) and (4.13) that estimates actuator cyber-attacks in the presence of sensor and actuator faults in the C&C. We also for the first time present simultaneous occurrence of fault and attack in a multi-rate framework. For achieving the objectives we aim to accomplish the followings:

1. **Design a C&C Attack Sensitive Filter in Presence of Faults and Actuator Cyber-Attack :** A delayed UIO will be designed on the C&C that provides an estimate of actuator cyber-attacks $\hat{a}^u(k)$ with the information that includes $r(k)$, which is obtained in the command and control in Section 4.2.

2. **Provide a C&C Cyber-Attacks and Faults Isolation Scheme** Based on the estimated actuator cyber-attack $\hat{a}^u(k)$ and the residual $\bar{r}(k)$ (the sent information by the plant-side filter), faults and attacks are isolatable from each other:

$$\begin{cases} if \quad abs(\bar{r}(k)) > 0,\ (abs(\hat{a}^u(k)) > 0 \\ \qquad or\ abs(\hat{a}^u(k)) = 0) \rightarrow Fault \\ if \quad abs(\bar{r}(k)) = 0,\ (abs(\hat{a}^u(k)) > 0\ or\ abs(\hat{a}^u(k)) = 0) \\ \qquad \rightarrow No\ Fault \\ if \quad abs(\hat{a}^u(k)) > 0,\ (abs(\bar{r}(k)) > 0\ or\ abs(\bar{r}(k)) = 0) \\ \qquad \rightarrow\ Attack \\ if \quad abs(\hat{a}^u(k)) = 0,\ (abs(\bar{r}(k)) > 0\ or\ abs(\bar{r}(k)) = 0) \\ \qquad \rightarrow\ No\ Attack \end{cases} \tag{4.25}$$

4.3 Observer Design

In this section, an unknown input observer (UIO) is constructed according to [95] for the error dynamic system (4.24) such that this observer will achieve the attack estimation in the presence of faults through a delayed left inversion.

The construction of such observers is dependent on satisfying the strong detectability condition [95, Theorem 6], which is proved to be necessary and sufficient [95, Theorem 6]. Before designing this estimator, first we investigate the properties of filters introduced in Section 4.2.

4.3.1 Filter Properties

First, we respectively prove the fault sensitivity and attack sensitivity for our proposed filters. The first filter (PSF) (4.13) is fault sensitive, while the CASF (4.18) is attack sensitive:

Theorem 4.1 *Residual $r(k)$ generated by CASF (4.18) is only attack sensitive and residual generated by plant side filter (PSF) (4.13) "$\bar{r}$" is only fault sensitive, if $A - LC$ is stable.*

Proof *The error dynamic (4.14) shows that by selecting a filter gain L for which $A - LC$ is stable, the residual of PSF $\bar{r} \to CB^{f^s} f^u(k) + D^{f^s} f^s(k)$ as $k \to \infty$. Furthermore, by satisfying the stability of $A-LC$, and investigating the error dynamic (4.18), it is concluded that the residual of CASF $r \to C(A - LC)e(k) + CB^{a^u} a^u(k)$ as $k \to \infty$.* □

Theorem 4.1 establishes that the filter (4.18) is only sensitive to attacks. Despite the attack sensitivity, without the dual-rate strategy, the single rate sampled data system in (4.18) is still prone to zero dynamics attacks.

Furthermore, A dual-rate filter is established in equation (4.19) to deal with zero dynamics attacks, and consequently, a UIO observer is designed for this filter. In the following, a Lemma is provided that helps us prove the immunity of dual-rate filter (equation (4.19)) to zero dynamics attacks.

Lemma 4.1 *[77, Lemma 10 and Proposition 12]: Consider a CPS with the following dual-rate representation defined through the following steps:*

1. *with the continuous representation*

$$\begin{aligned} \dot{v} &= \bar{A}_c v + \bar{B}_c l \\ z &= \bar{C}_c v \end{aligned} \tag{4.26}$$

in which v represents the state vector, z is the output vector, l is the unknown input. Moreover, $\bar{A}_c$, $\bar{B}_c$, and $\bar{C}_c$ are system matrices with appropriate dimension.

2. *The system has a fast rate of discretized representation with a rate of* $\frac{T}{N}$*:*

$$\begin{aligned} v(k+1) &= \bar{A}v(k) + \bar{B}l(k) \\ z(k) &= \bar{C}v(k) \end{aligned} \tag{4.27}$$

in which the discretization can take place according to (4.2) *by substituting* $\bar{A} \to A$*,* $\bar{B}^{\bar{a}^u} \to B$*, and* $\bar{C} \to C$*.*

3. *The dual-rate (inputs* T *and output* $\frac{T}{N}$*) representation is described according to the following:*

$$\begin{aligned} v(k+N) &= \tilde{\bar{A}}v(k) + \tilde{\bar{B}}l(k) \\ Z(k) &= \tilde{\bar{C}}v(k) + \tilde{\bar{D}}(k) \\ Z(k) &= \begin{pmatrix} z(k) & z(k+1) & \dots & z(k+N) \end{pmatrix}^T \end{aligned} \tag{4.28}$$

in which the system matrices can be driven by following the steps in (4.10) *by considering fast rate representation* (4.27) *and substituting* $\tilde{\bar{A}} \to \tilde{A}$*,* $\tilde{\bar{B}} \to \tilde{B}^{a^u}$*,* $\tilde{\bar{C}} \to \tilde{C}$*, and* $\tilde{\bar{D}} \to \tilde{D}^{a^u}$*.*

No invariant zeros with non-minimum phase zeros exist for the sampled data system (4.28) with inputs rate T*, and output rate* $\frac{T}{N}$ *if the following conditions hold:*

(i) $\bar{B}$ *is full column rank.*

(ii) The following matrix is full column rank (e.g. if system is observable):

$$\mathcal{O} = \begin{bmatrix} \bar{C} & \bar{A} & \dots & \bar{C}\bar{A}^{N-2} \end{bmatrix}^T \tag{4.29}$$

(iii) If the system (4.27) is tall and the respective sampled data system with the rate $\frac{T}{N}$ *has no zero at* $|z_0| = 1$*.*

Satisfying $Lemma4.1$ is equivalent to strong detectability for the dual-rate system, in which N is designed such that (4.29) holds (e.g. satisfying observability condition) [77]. Moreover, we have to show that investigated system in (4.19), has the same properties:

Lemma 4.2 *The investigated system* (4.19) *is strongly detectable if and only if the following condition holds:*

1. *The pair* $(A - LC, C)$ *is observable*
2. B^{a^u} *is full column rank*
3. $A - LC, B^{a^u}, C$ *has no zeros on the unit circle.*

Proof *We must prove the above conditions can hold in the same manner as the 3 conditions in* $Lemma4.1$.

First, the pair (A, C) *observable denotes* $rank(\begin{pmatrix} C \\ A - zI \end{pmatrix}) = n$, $\forall z \in \mathbb{C}$ *[100]. For the pair (A-LC, C) we can write the observability condition as* $rank(\begin{pmatrix} C \\ A - LC - zI \end{pmatrix}) = n$ *which is equivalent to* $rank(\begin{pmatrix} I & 0 \\ -L & I \end{pmatrix} \begin{pmatrix} C \\ A - zI \end{pmatrix}) = n.$

Condition (2) is the same as Assumption 4.2.

Condition (3) is also satisfied since the zeros of the system (4.19) *is equivalent to system* (4.3) *for which its zeros can be computed through pencil matrix* $\begin{pmatrix} zI - A & B^{a^u} \\ C & 0 \end{pmatrix}$. *Since* $\begin{pmatrix} zI - A + LC & B^{a^u} \\ C & 0 \end{pmatrix} = \begin{pmatrix} I & L \\ 0 & I \end{pmatrix} \begin{pmatrix} zI - A & B^{a^u} \\ C & 0 \end{pmatrix}$. *Additionally it is known that invariant zeros do not change under output feedback [89, Lemma 1].*

Therefore, the investigated system is also strongly detectable. □

The above conditions satisfy the strong detectability that is necessary and sufficient for designing a delayed UIO [95], which denotes we can design a delayed UIO according to (4.20) for the dual-rate system. Therefore, the following proposition addresses its design:

Proposition 4.1 *Observer (4.20) can be constructed if the following conditions holds:*

1. E *is stable*
2. $E = \tilde{A}_e - F\tilde{\mathcal{O}}_L$
3. $F\mathcal{J}_L = \begin{pmatrix} \tilde{B}_e^{a^u} & 0 & \cdots & 0 \end{pmatrix}$

Proof *These conditions for designing the delayed UIO are derived by setting* $e_e \longrightarrow 0$ *in* (4.24). *The proposed type for UIO has a well-known design procedure, and details of calculating* F *for this type of observer can be found in [95] (see Appendix). By satisfying above conditions, the UIO for auxiliary system will be constructed, for which* $e \longrightarrow 0$, *that denotes* $\hat{e}(k-L) - e(k-L) \longrightarrow 0$ *or* $\hat{e}(k-L) \longrightarrow e(k-L)$. □

Under the condition that $\begin{pmatrix} \tilde{B}_e^{a^u} & \tilde{D}_e^{a^u} \end{pmatrix}^T$ is full column rank there exist a matrix G such that the following holds:

$$G \begin{pmatrix} \tilde{B}_e^{a^u} & \tilde{D}_e^{a^u} \end{pmatrix}^T = I_{m_{a^u}} \tag{4.30}$$

After this step, we can calculate the unknown inputs of the system through a left inversion of the auxiliary system according to the following equation:

$$\hat{a}^u(k-L) = G \begin{pmatrix} \hat{e}(k-L+N) - \tilde{A}_e \hat{e}(k-L) \\ R(k-L) - \tilde{C}_e \hat{e}(k-L) \end{pmatrix} \tag{4.31}$$

Since $\hat{e}(k-L) - e(k-L) \longrightarrow 0$ then the $\hat{a}^u(k-L) \longrightarrow a^u(k-L)$ as $k \longrightarrow \infty$. The above estimation is a direct result of left invertiblity of the system.

4.4 Numerical Simulation Study

The performance of the methodology is tested by the estimation of unknown inputs through a framework of zero dynamics attacks developed by [76, 77]. We used the continuous-time representation denoted in their example, which is sampled with a rate of 0.5s. This would result in a sampling zero for the discretized system. Here, the transfer function for this continuous system is provided [76]:

$$G_c(s) = \frac{10}{\frac{1}{500}s^4 + \frac{67}{1000}s^3 + \frac{123}{200}s^2 + \frac{31}{20}s + 1} \tag{4.32}$$

The discrete-time minimal representation with a period of $\frac{T}{N} = 0.25s$ is computed as the following:

$$A = \begin{pmatrix} -0.1723 & -0.3513 & -0.0856 & -0.0250 \\ 0.2046 & 0.2560 & -0.2111 & -0.0656 \\ 0.2687 & 0.6648 & 0.9014 & -0.0310 \\ 0.0318 & 0.1001 & 0.2426 & 0.9976 \end{pmatrix}$$

$$B^{a^u} = B^{f^u} = \begin{pmatrix} 0.0511 \\ 0.1343 \\ 0.0636 \\ 0.0048 \end{pmatrix}, C = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 4.8828 \end{pmatrix}^T \tag{4.33}$$

And the lifted dual-rate representation can be seen in the following:

$$\tilde{A} = \begin{pmatrix} 0.0105 & 0.0246 & 0.0301 & 0.0112 \\ -0.0918 & -0.1817 & -0.1711 & -0.0367 \\ 0.1503 & 0.2687 & 0.1793 & -0.2298 \\ 0.2354 & 0.5115 & 0.6326 & 0.8918 \end{pmatrix}$$

$$, \tilde{B}^{a^u} = \tilde{B}^{f^u}_{sys} = \begin{pmatrix} 0.0001 \\ -0.0005 \\ 0.0012 \\ 0.0088 \end{pmatrix}$$

$$, \tilde{C} = \begin{pmatrix} 0 & 0 & 0 & 4.8828 \\ 0.1552 & 0.4889 & 1.1844 & 4.8713 \\ 0.5463 & 1.3458 & 2.1327 & 4.7871 \\ 0.9063 & 2.0498 & 2.7528 & 4.6077 \end{pmatrix}, \tilde{D}^{a^u} = \begin{pmatrix} 0 \\ 0.0236 \\ 0.1960 \\ 0.5634 \end{pmatrix} \tag{4.34}$$

The slow inputs rate of $0.5s$, based on system structure $\tilde{A}, \tilde{B}^{a^u}, C$, has unstable sampling zero at (-1.4196), for which the attacker can utilize and insert a zero dynamics attack. However, with the implementation of the proposed methodology, this attack can be estimated. In the following the design parameters for the UIO are shown:

$$E = \begin{pmatrix} -0.0048 & -0.0102 & 0.0184 & -0.0907 \\ 0.0519 & 0.0976 & -0.1840 & 0.9177 \\ 0.0705 & 0.1374 & -0.2559 & 1.2725 \\ 0.0110 & 0.0217 & -0.0402 & 0.1998 \end{pmatrix}$$

$$, F = \begin{pmatrix} -0.0674 & 0 & 0 & 0 & 0.0006 & -0.0152 & 0.0002 & 0.0006 \\ 0.3180 & 0 & 0 & 0 & -0.0020 & 0.0498 & -0.0006 & -0.0019 \\ 0.4639 & 0 & 0 & 0 & -0.0123 & 0.3121 & -0.0041 & -0.0116 \\ -0.1365 & 0 & 0 & 0 & -0.0058 & 0.1470 & -0.0019 & -0.0055 \end{pmatrix} \tag{4.35}$$

Mathematical description of simulated attacks and faults are provided in the following:

1. Zero dynamics actuator cyber-attack:

$$a^u(k) = \begin{cases} 0, & if\ 0 \leq k < 60 \\ -0.01 \times 1.4196^{k-60}, & if\ 60 \leq k \end{cases} \tag{4.36}$$

2. Sinusoidal actuator cyber-attack:

$$a^u(k) = \begin{cases} 0, & if\ 0 \leq k < 50 \\ 4sin(0.1k), & if\ 50 \leq k \end{cases} \tag{4.37}$$

3. Biasing actuator fault:

$$f^u(k) = \begin{cases} 0, & if\ 0 \leq k < 50 \\ 4, & if\ 50 \leq k \end{cases} \tag{4.38}$$

A few scenarios are considered by placing the UIO observer on the plant side as well as the command and control side. Fig. 4.3 demonstrates the sinusoidal actuator cyber-attack for which

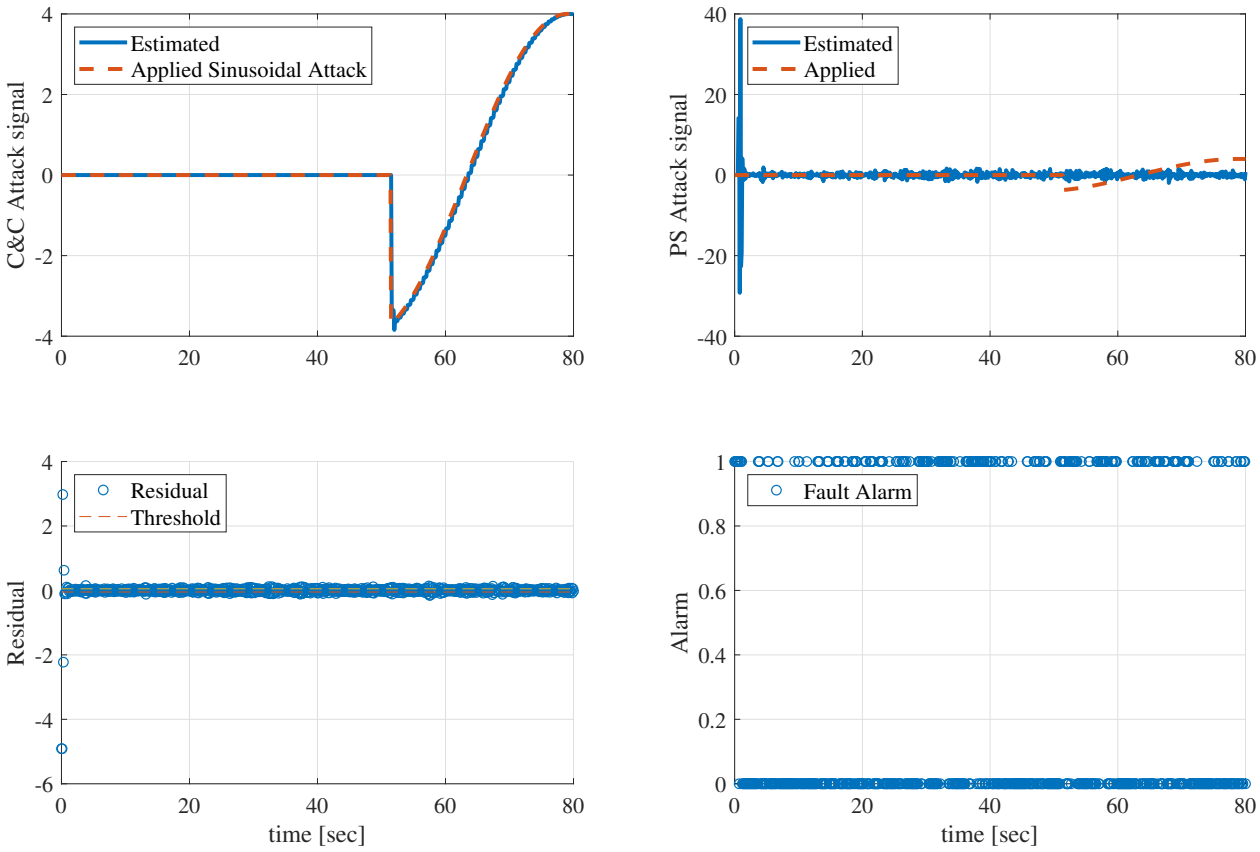

Figure 4.3: Sinusoidal attack estimation in the C&C, comparison of plant side and command and control side alongside with the fault alarm in this case.

the proposed methodology accurately estimates the cyber-attack in the C&C, while the same type of observer on the plant side is unable to estimate the attack. In addition, Fig. 4.3 shows no fault is detected through the residual generation for the sinusoidal actuator cyber-attack. Furthermore, the same result can be taken for zero dynamics actuator cyber-attacks through Fig. 4.4.

Furthermore, an additive actuator fault is investigated, for which Fig. 4.5 shows the sensitivity of the fault detector, and it also shows the insensitivity of the attack estimator in the C&C . In this case, the number of unknown inputs is equal to the number of the outputs of the system, which satisfies the minimal requirement for unknown input estimation, and it can be seen through Fig. 4.5 that the attack estimator can estimate this fault on the plant side. However, if sensor faults were also included, this estimator may not be able to estimate multiple unknown inputs.

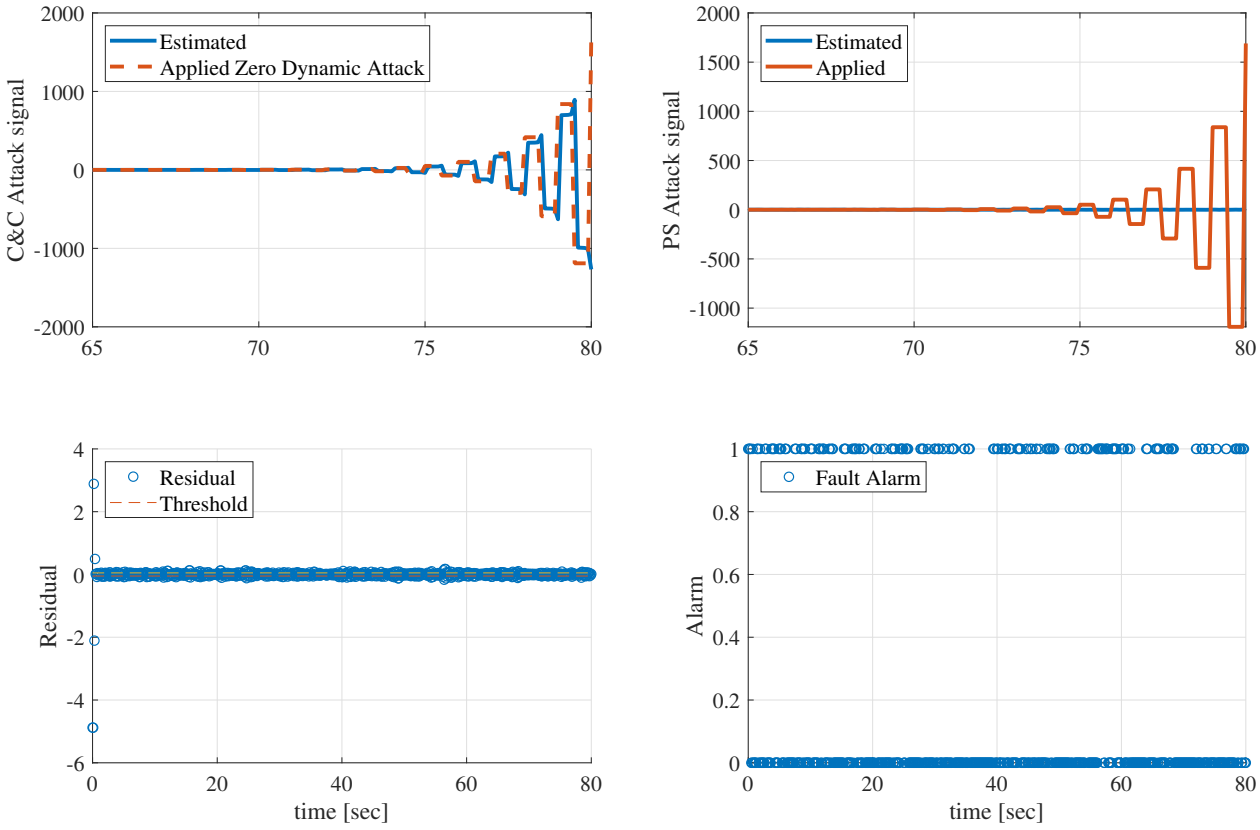

Figure 4.4: Zero dynamics attacks estimation, comparison of plant side and command and control side alongside with fault Detector with presence of zero dynamics attacks at C&C.

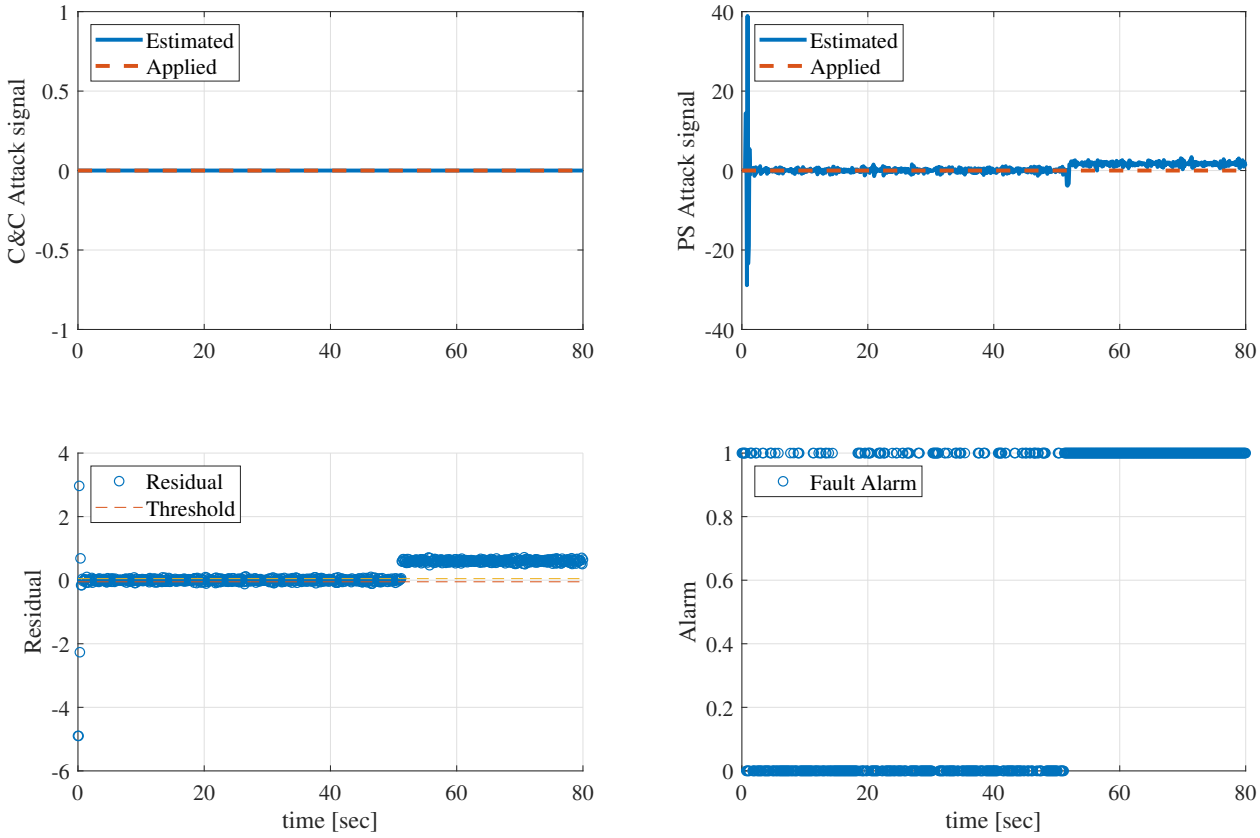

Figure 4.5: Bias fault scenario of magnitude 4 applied at time 50s, comparison of estimators placed in the C&C to Plant side estimator alongside with fault alarm at the C&C based on the received residual.

4.5 Comparison Study

The methodology is compared to the article [36] on the estimation of attack signal. As the novelty of our work remains in providing the fault and attack framework as well as estimating the attack in this framework, we also demonstrate the advantages or deficiency of the presented delayed UIO attack estimation methodology to methodology in [36].

By investigating the work in [36] and designing a UIO for (4.9) by consideration $\mathcal{F}^u, \mathcal{F}^s = 0$, we reach to the following estimator [36]:

$$\hat{x}(k+N) = E\hat{x} + LY(k) + \tilde{B}u(k) \tag{4.39}$$

for which E and L are design parameters. with error dynamic

$$\begin{aligned}
\hat{e}(k+N) &= \hat{x}(k+N) - x(k+N) \\
&= E\hat{x}(k) + LY(k) - \tilde{A}x(k) - \tilde{B}^{a^u}(k)a^u(k) \\
&= Ee(k) + (E - \tilde{A} + L\tilde{C})x(k) + (L\tilde{D}^{a^u} - \tilde{B}^{a^u})a^u(k) \\
s.t &: \\
E &\ is\ stable \\
L\tilde{D}^{a^u} &= \tilde{B}^{a^u} \\
E &= \tilde{A} - L\tilde{C}
\end{aligned} \tag{4.40}$$

In [36], the author mentions the use of "MATLAB Solver" for solving their constraint equations. We used an LMI solver to replicate their result for numerical example in equation (4.32). We provide a comparison for two scenarios in order to demonstrate the strength of our proposed methodology:

While performing the simulation, we applied small amount of noise on output and considered the following scenarios:

Scenario 1 :

For the Proposed Methodology, consider a threshold of $|T_r| < 0.15$ for attack and fault detection that is selected with respect to simulation noise and a situation such that plant side fault residual is communicated according to equation (4.15).

For [36], consider a threshold of $|T_r| < 0.15$ for attack detection.

Scenario 2 :

For the Proposed Methodology, reconsider a small threshold of $|T_r| < 0.05$ for attack detection and assume that plant side fault residual is continuously communicated to $C\&C$, which is also a special case of equation (4.15) when the fault is detected at plant side.

For [36], reconsider a small threshold of $|T_r| < 0.05$ for attack detection.

We have simulated the fast rate output system for our presented work and methodology in [36] in situations with $\frac{T}{N} = 0.25$ and $\frac{T}{N} = 0.5$ and considered the following factors for comparison purposes:

In the scenarios, the stability of E is noted to be stable (S) or unstable (NS) to compare the feasibility of delayed observer methodology and LMI-based methodology. Furthermore, d_{rate} for the given thresholds, $F1$ score which is explained in equation (3.4), and a average attack tracking error with (J_e^A) and without the attack (J_e^H) is provided:

$$J_e^A = \frac{\Sigma_{k_{START}}^{k_{END}} |a(k) - \hat{a}(k)|}{number\ of\ samples}, J_e^H = \frac{\Sigma_0^{k_{START}} |a(k) - \hat{a}(k)|}{number\ of\ samples} \tag{4.41}$$

In the above equation, k_{START} is associated with the sample for which the attack starts, and k_{END} is associated with the sample for which the attack finishes.

Discussion: Scenario 1

We found that overall the methodology in [36] can lead to a feasible solution, however, this is not the case for the provided example at $\frac{T}{N} = 0.5sec.$ In [36], their provided E matrix in the

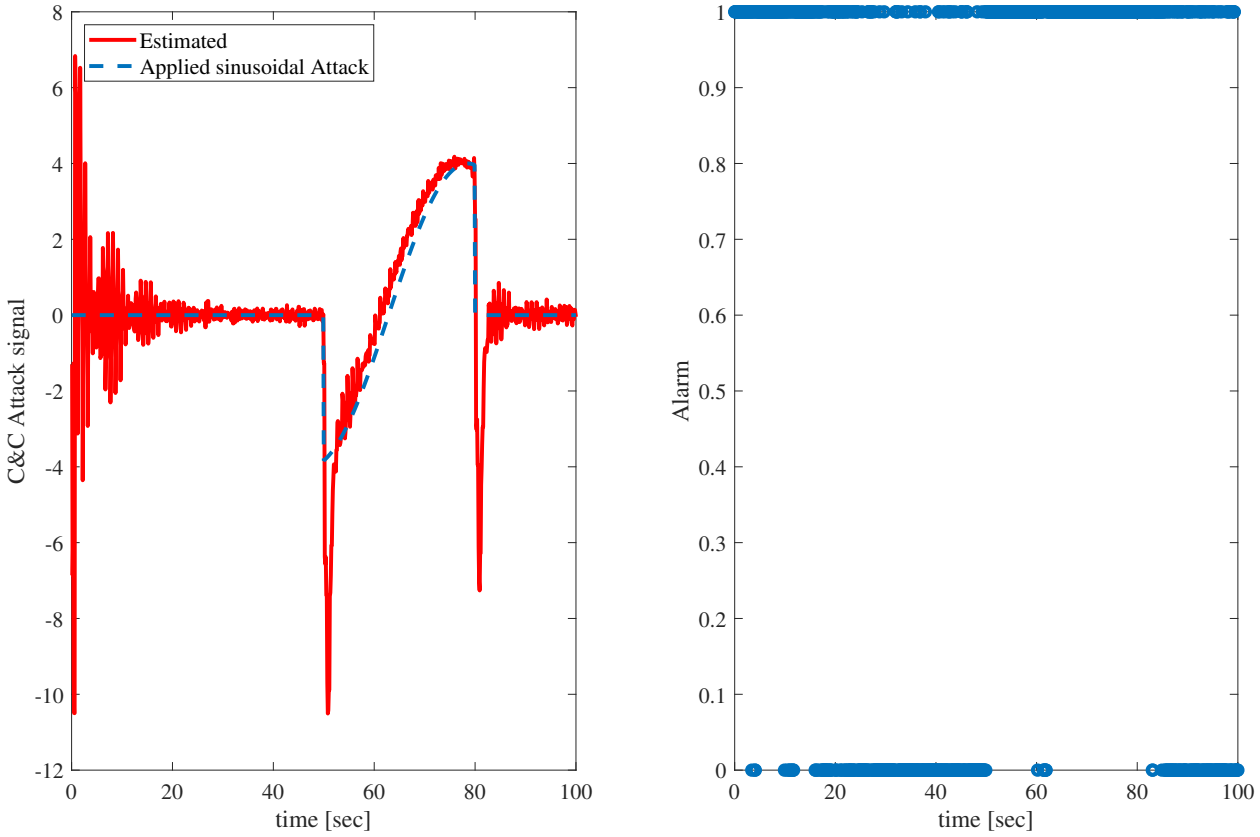

Figure 4.6: Sinusoidal attack scenario applied at time 50s, result based on methodology in [36].

example for the case of $\frac{T}{N} = 0.5sec$ is in fact unstable. Our simulation of their method for the case of $\frac{T}{N} = 0.5$ also reached an unstable E, which indicates that it is not suitable for the task of attack estimation. Through the provided Table 4.1 and Table 4.2, it is shown that the proposed methodology based on delayed UIO observer will demonstrate better tracking capability of attack for both cases while $\frac{T}{N} = 0.25$ and $\frac{T}{N} = 0.5$.

Discussion: Scenario 2

Through Table 4.3 and Table 4.4, we provide a comparison of the proposed methodology to methodology in [36] for the second scenario. Our simulation of their method for the case of $\frac{T}{N} = 0.5$ has also reached an unstable E. However, through the provided Table 4.3 and Table 4.4, it is shown that the proposed methodology will demonstrate the better tracking and detection capability of attack for both cases, while the attack tracking ability of our methodology is improved in comparison to

	Proposed method (2 Sample Delay)					[36] (No Delay)				
$\frac{T}{N} = 0.5$	E	d_{rate}	F_1	Je_H, Je_A	AFI	E	d_{rate}	F_1	Je_H, Je_A	AFI
Bias Attacks	S	97.33	96.98	0.0998,0.06	✓	NS	✗	✗	✗	✗
Sinusoidal Attacks	S	94.19	95.35	0.099,0.23	✓	NS	✗	✗	✗	✗
Zero Dynamics Attacks	S	60.59	73.14	0.099,155.31	✓	NS	✗	✗	✗	✗
Actuator Faults	S	99	97.86	– , –	✓	✗	✗	✗	✗	✗
Sensor Fault	S	99.33	98.03	– , –	✓	✗	✗	✗	✗	✗

Table 4.1: For the Scenario 1, this table represents some of the advantages of the proposed methodology with respect to an attack signal estimation methodology proposed in [36]. AFI denotes attack and fault isolation. S denotes Stable, and NS denotes Not Stable. Each number represents the average result of a 50 simulation.

	Proposed method (2 Sample Delay)					[36] (No Delay)				
$\frac{T}{N} = 0.25$	E	d_{rate}	F_1	Je_H, Je_A	AFI	E	d_{rate}	F_1	Je_H, Je_A	AFI
Bias Attack	S	98.807	75.11	0.26,0.022	✓	S	100	73.29	0.46,0.41	✗
Sinusoidal Attack	S	96.22	74.819	0.259,0.111	✓	S	97.63	72.36	0.453,0.65	✗
Zero Dynamics Attack	S	76.2	52.8←6	0.24,48.49	✓	S	69.21	48.511	0.39,93.869	✗
Actuator Fault	S	99.053	98.297	– , –	✓	✗	✗	✗	✗	✗
Sensor Fault	S	95.93	96.68	– , –	✓	✗	✗	✗	✗	✗

Table 4.2: For Scenario 1, this table represents some of the advantages of the proposed methodology with respect to an attack signal estimation methodology proposed in [36]. AFI denotes attack and fault isolation. S denotes Stable, and NS denotes Not Stable. Each number represents the average result of a 50 simulation.

	Proposed method (2 Sample Delay)					[36] (No Delay)				
$\frac{T}{N} = 0.5$	E	d_{rate}	F_1	Je_H, Je_A	AFI	E	d_{rate}	F_1	Je_H, Je_A	AFI
Bias Attack	S	97.33	97.01	0.066,0.061	✓	NS	✗	✗	✗	✗
Sinusoidal Attack	S	97.33	97.01	0.066,0.225	✓	NS	✗	✗	✗	✗
Zero Dynamics Attack	S	68.5	78.96	0.055,155.307	✓	NS	✗	✗	✗	✗
Actuator Fault	S	99	97.44	– , –	✓	✗	✗	✗	✗	✗
Sensor Fault	S	99.35	97.67	– , –	✓	✗	✗	✗	✗	✗

Table 4.3: For the second scenario, this table represents some of the advantages of the proposed methodology with respect to an attack estimation methodology proposed in [36]. AFI denotes attack and fault isolation. S denotes Stable, and N denotes Not Stable. Each number represents the average result of a 50 simulation.

	Proposed method (2 Sample Delay)					[36] (No Delay)				
$\frac{T}{N} = 0.25$	E	d_{rate}	F_1	Je_H, Je_A	AFI	E	d_{rate}	F_1	Je_H, Je_A	AFI
Bias Attack	S	98	97.67	0.174,0.019	✓	S	100	61.205	0.45,0.417	✗
Sinusoidal Attack	S	96.67	96.99	0.17,0.105	✓	S	99.51	60.85	0.46,0.65	✗
Zero Dynamics Attack	S	72.5	82.153	0.145,48.44	✓	S	86.99	42.46	0.39,93.87	✗
Actuator Fault	S	99.17	96.16	– , –	✓	✗	✗	✗	✗	✗
Sensor Fault	S	96.77	94.87	– , –	✓	✗	✗	✗	✗	✗

Table 4.4: For the second scenario, this table represents some of the advantages of the proposed methodology with respect to an attack estimation methodology proposed in [36]. AFI denotes attack and fault isolation. S denotes Stable, and N denotes Not Stable. Each number represents the average result of a 50 simulation.

Scenario 1. By comparing Table 4.4 and Table 4.3 to Table 4.2 and Table 4.1, it is concluded that in the presence of added noise in the simulation, when the residual generated at the plant side is received at the $C\&C$ for all time, the proposed methodology performs better, even while selecting a small threshold. The proposed strategy provides better performance for a small amount of threshold since it inherently cancels the noise when C&C generated and plant-side generated residuals are both present. The proof can be realized through Theorem 4.1 and substituting for sensor fault f^s and actuator fault f^u such that $D^{f^s} f^s(k) \rightarrow D^{f^s} f^s(k) + w(k)$ and $D^{f^u} f^u(k) \rightarrow D^{f^u} f^u(k) + v(k)$. For which w demonstrates the process noise and v demonstrates the sensor noise. In other words, through the subtraction of the residual of two filters PSF and CSF, the proposed methodology cancels the noise contribution to the residual in the same manner that it deals with the faults.

4.6 Conclusion

In this chapter, estimation actuator cyber-attacks in the presence of faults are investigated. Based on a dual-rate strategy and generating fault-sensitive residual through a Plant side filter (PSF), we were able to design an unknown input observer that can estimate actuator cyber-attacks in the C&C in the presence of sensor and actuator faults. The delayed UIO also demonstrates a higher tracking ability of attack in comparison to the methodology presented in [36]. The presented work can be extended through a non-uniform fast rate periodic sampling to be able to detect covert cyber-attacks.

Chapter 5

Conclusion and Future Work

In this chapter we provide summary of the results of this work and provide potential future work directions:

5.1 Conclusion

- In Chapters 3, a CPS system is investigated that is simultaneously impacted by faults and cyber-attacks. On the plant side, to isolate the effects of cyber-attacks from faults, we try to formulate attack-sensitive ASF and fault-sensitive filters KFF. Through mathematical analysis of ASF filter (Section 3.4), we prove its capability to isolate stealthy cyber-attacks, including covert and zero dynamics attacks, at the plant side. In Section 3.3.1, it is also proven that KFF filter is only fault sensitive.

- In Chapter 3, by deriving equation (3.32) , an algebraic condition is developed to check for stealthy attacks in CPS. Figures 3.8, 3.6, 3.5,3.4, and 3.7 investigate different scenarios of covert and zero dynamics cyber-attacks as well as sensor and actuator faults to demonstrate the capability of the proposed filters. A comparative study is provided by considering an auxiliary-based attack detection methodology [88], and simulating it for the cyber-attack and fault scenarios demonstrated in Fig. 3.10, 3.11. Table 3.1 shows the performance comparison

of the established methodology with the methodology in [88]. Through this table, it is shown that the proposed methodology has a better performance in detecting cyber-attacks. There are no published paper that cover both sensor and actuator under faults and stealthy cyber-attack. In [111], replay attack isolation from sensor faults has been addressed, however it does not consider actuator faults or actuator cyber-attacks. The study in [28] also provide attack and fault isolation in view of multi-agent systems, however does not address stealthy cyber-attacks such as zero dynamics attacks.

- In Chapter 4, actuator cyber-attacks and faults in the multi-rate framework are investigated. In this part, we propose a strategy that isolates faults and cyber-attacks while estimating cyber-attacks on the command and control side. A plant-side auxiliary residual generator is established that sends its data to the plant side in events of fault occurrence (equation (4.25)). Unlike "cyber-attack only" detection auxiliary-based methodologies that are prevalent in the literature, such as [10, 88], the event-based (4.15) is designed such that it does not need constant communication to the C&C for isolating attacks from faults. Study in [36] is the only comparable research problem to the presented methodology, while it ignores the presence of faults and provides a cyber-attack estimation solution for dual-rate systems that is only subjected to cyber-attacks. In our proposed work, for the first time, the problem formulation of a dual-rate system is stated by including both faults and cyber-attacks. The proposed methodology isolates cyber-attacks from faults in the C&C with the help of the communicated residual (4.25) and is also able to estimate cyber-attacks while eliminating the effect of faults.

- In Chapter 4, the results of our numerical simulations are presented in Figures 4.4, 4.3, and 4.5. A comparative analysis, as shown in Tables 4.1, 4.2, 4.3, and 4.4, demonstrates that the Delayed-Observer-based methodology yields more reliable results compared to the LMI-based approach.

5.2 Future Work

1. Methodologies proposed in Chapters 3 and 4 can be investigated in a nonlinear framework.

   (a) The work in Chapter 3 can be extended to other linear or non-linear filter plus controllers in the $C\&C$. No matter the type of filter plus controller placed on the command and control side, such systems can be viewed as an auxiliary plant for which a plant-side filter needs to be established. The requirement is that the $C\&C$ filter plus controller needs to be observable through its output which is the communicated command and control inputs to the plant. In such a case, a geometric analogy can be established to prove the immunity to covert and zero dynamics attacks.

   (b) The work in Chapter 4 can also be extended to non-linear systems. With such modification, investigation can be done on the detection performance of control side and plant side filters for the dual-rate non-linear system.

2. The work in Chapter 3 can be investigated in the detection of replay attack, although its not formally proven in this thesis, it is probable that observability of matrix (A_{oc}, K) in equation (3.2) is sufficient for detection of reply attacks.

3. The work in Chapter 4 can be modified to detect covert attacks in CPS. By considering a constant rate for the inputs and a varying fast rate for the outputs it is possible to detect covert attacks. The defender has to make sure the knowledge of discrete-time representation of the CPS for the attacker remains uncertain by changing the output rate, therefore, making the attacker unable to perform a covert attack.

4. Mitigation strategies may be developed to deal with a reconfiguration of control law in the event of cyber-attack detection in Chapter 3 and Chapter 4.

5. A multi fast output rate and constant inputs rate strategy can be developed to deal with covert attacks in homogenous multi-agent systems for Chapter 4. It is particularly interesting to

how a varying output rate for a few agents can affect the collective zero dynamics of the multi-agent system.

6. In Chapter 3 and Chapter 4, we can also move forward and investigate fault and cyber-attack isolation in a way that is possible to determine which sensor or actuators are under cyber-attacks or fault.

7. Auxiliary consideration in 4 can be extended to multi-agents systems. Augmenting auxiliary systems in a multi-agent multi-rate frame work may prove effective in detecting stealthy cyber-attacks.

Appendix A

Computing Observer Gain

Here the methodology for computing the auxiliary's observer gain F^{aux} is explained, the observer gain F^{sys} can be computed:

Step for calculating F [95]:

1. Find L such that $rank(\mathcal{J}_L) - rank(J_{L-N}) = m_{a^u}$, that denotes satisfying left invertibility.
2. Find M such that $M\mathcal{J}_L = \begin{pmatrix} 0 & 0 \\ I_{m_{a^u}} & 0 \end{pmatrix}$.
3. Consider $F = \hat{F}M = \begin{pmatrix} \hat{F}_1 & \hat{F}_2 \end{pmatrix} M$ we want $\begin{pmatrix} \hat{F}_1 & \hat{F}_2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ I_{m_{a^u}} & 0 \end{pmatrix} = \begin{pmatrix} \tilde{B}^{a^u} & 0 \end{pmatrix}$ From this we get $\hat{F}_2 = \tilde{B}^{a^u}$.
4. According to proposition 4.1, we have $E = \tilde{A} - F\tilde{\mathcal{O}}_L = \tilde{A} - \begin{pmatrix} \hat{F}_1 & \tilde{B}^{a^u} \end{pmatrix} M\tilde{\mathcal{O}}_L$.
5. We set $M\tilde{\mathcal{O}}_L = \begin{pmatrix} S_1 & S_2 \end{pmatrix}^T$ and calculate S_1 and S_2.
6. Take $E = (\tilde{A} - \tilde{B}^{a^u} S_2) - \hat{F}_1 S_1$.
7. E has to be stable. Therefore, calculate $\hat{F}_1$ such that the pair $(\tilde{A} - \tilde{B}^{a^u} S_2, S_1)$ is detectable.

Bibliography

[1] M. Adeli, M. Hajatipour, M. J. Yazdanpanah, H. Hashemi-Dezaki, and M. Shafieirad, "Optimized cyber-attack detection method of power systems using sliding mode observer," *Electric Power Systems Research*, vol. 205, p. 107745, 2022.

[2] 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.

[3] M. Amozegar and K. Khorasani, "An ensemble of dynamic neural network identifiers for fault detection and isolation of gas turbine engines," *Neural Networks*, vol. 76, pp. 106–121, 2016.

[4] A. N. Andry, E. Y. Shapiro, and J. Chung, "Eigenstructure assignment for linear systems," *IEEE transactions on aerospace and electronic systems*, no. 5, pp. 711–729, 1983.

[5] S. E. Azam, E. Chatzi, and C. Papadimitriou, "A dual kalman filter approach for state estimation via output-only acceleration measurements," *Mechanical systems and signal processing*, vol. 60, pp. 866–886, 2015.

[6] J. Back, J. Kim, C. Lee, G. Park, and H. Shim, "Enhancement of security against zero dynamics attack via generalized hold," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 1350–1355.

[7] B. Badrignans, J. L. Danger, V. Fischer, G. Gogniat, and L. Torres, *Security trends for FP-*

GAS: From secured to secure reconfigurable systems. Springer Science & Business Media, 2011.

[8] A. Baniamerian, K. Khorasani, and N. Meskin, "A special class of zero dynamics cyber-attacks for siso time-delay systems," *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 4182–4187, 2021.

[9] A. Baniamerian and K. Khorasani, "Security index of linear cyber-physical systems: A geometric perspective," in *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE, 2019, pp. 391–396.

[10] A. Baniamerian, K. Khorasani, and N. Meskin, "Monitoring and detection of malicious adversarial zero dynamics attacks in cyber-physical systems," *2020 IEEE Conference on Control Technology and Applications (CCTA)*, pp. 726–731, 2020.

[11] S. Bittanti and P. Colaneri, *Periodic systems: filtering and control*. Springer Science & Business Media, 2009, vol. 5108985.

[12] B. Brumback and M. Srinath, "A chi-square test for fault-detection in kalman filters," *IEEE Transactions on Automatic Control*, vol. 32, no. 6, pp. 552–554, 1987.

[13] F. M. Callier and C. A. Desoer, *Linear system theory*. Springer Science & Business Media, 2012.

[14] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, S. Sastry *et al.*, "Challenges for securing cyber physical systems," in *Workshop on future directions in cyber-physical systems security*, vol. 5, 2009.

[15] C.-C. Chan, C.-Z. Yang, and C.-F. Fan, "Security verification for cyber-physical systems using model checking," *IEEE Access*, vol. 9, pp. 75 169–75 186, 2021.

[16] E. R. Chaves, F. d. A. André, and A. L. Maitelli, "Robust observer-based actuator and sensor

fault estimation for discrete-time systems," *Journal of Control, Automation and Electrical Systems*, vol. 30, no. 2, pp. 160–169, 2019.

[17] T. Chen and B. A. Francis, *Optimal sampled-data control systems*. Springer Science & Business Media, 2012.

[18] W. Chen, B. D. Anderson, M. Deistler, and A. Filler, "Properties of blocked linear systems," *Automatica*, vol. 48, no. 10, pp. 2520–2525, 2012.

[19] Y. Chen, S. Kar, and J. Moura, "Cyber-physical attacks with control objectives," *IEEE Transactions on Automatic Control*, vol. 63, no. 5, pp. 1418–1425, 2018.

[20] Y. Chen, S. Kar, and J. M. Moura, "Dynamic attack detection in cyber-physical systems with side initial state information," *ArXiv e-prints*, 2015.

[21] R. Collins. (2018) "ptc-ready" locomotives part of amtrak commitment to safety. [Online]. Available: https://media.amtrak.com/2017/09/amtrak-prepares-diesel-fleet-positive-train-control/

[22] P. L. Combettes, "The foundations of set theoretic estimation," *Proceedings of the IEEE*, vol. 81, no. 2, pp. 182–208, 1993.

[23] L. Dai, Q. Cao, Y. Xia, and Y. Gao, "Distributed mpc for formation of multi-agent systems with collision avoidance and obstacle avoidance," *Journal of the Franklin Institute*, vol. 354, no. 4, pp. 2068–2085, 2017.

[24] N. Daroogheh, N. Meskin, and K. Khorasani, "Ensemble kalman filters (enkf) for state estimation and prediction of two-time scale nonlinear systems with application to gas turbine engines," *arXiv preprint arXiv:1710.05244*, 2017.

[25] M. Darouach, M. Zasadzinski, and S. J. Xu, "Full-order observers for linear systems with unknown inputs," *IEEE transactions on automatic control*, vol. 39, no. 3, pp. 606–609, 1994.

[26] M. Davoodi, N. Meskin, and K. Khorasani, "Simultaneous fault detection and consensus control design for a network of multi-agent systems," *Automatica*, vol. 66, pp. 185–194, 2016.

[27] M. Deghat, V. Ugrinovskii, I. Shames, and C. Langbort, "Detection and mitigation of biasing attacks on distributed estimation networks," *Automatica*, vol. 99, pp. 369–381, 2019.

[28] A. Eslami, F. Abdollahi, and K. Khorasani, "Stochastic fault and cyber-attack detection and consensus control in multi-agent systems," *International Journal of Control*, pp. 1–19, 2021.

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

[30] M. Ghaderi, K. Gheitasi, and W. Lucia, "A blended active detection strategy for false data injection attacks in cyber-physical systems," *IEEE Transactions on Control of Network Systems*, vol. 8, no. 1, pp. 168–176, 2020.

[31] K. Gheitasi and W. Lucia, "A worst-case approach to safety and reference tracking for cyber-physical systems under network attacks," *IEEE Transactions on Automatic Control*, 2022.

[32] H. Giese, B. Rumpe, B. Schätz, and J. Sztipanovits, "Science and engineering of cyber-physical systems (dagstuhl seminar 11441)," in *Dagstuhl Reports*, vol. 1, no. 11. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2012.

[33] D. Han, Y. Mo, and L. Xie, "Convex optimization based state estimation against sparse integrity attacks," *IEEE Transactions on Automatic Control*, vol. 64, no. 6, pp. 2383–2395, 2019.

[34] M. L. Hautus, "Strong detectability and observers," *Linear Algebra and its applications*, vol. 50, pp. 353–368, 1983.

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

[36] N. H. Hirzallah, P. G. Voulgaris, and N. Hovakimyan, "On the estimation of signal attacks: A dual rate sd control framework," *2019 18th European Control Conference (ECC)*, pp. 4380–4385, 2019.

[37] A. Hoehn and P. Zhang, "Detection of covert attacks and zero dynamics attacks in cyber-physical systems," in *American Control Conference (ACC), 2016*. IEEE, 2016, pp. 302–307.

[38] Y. Hong, G. Chen, and L. Bushnell, "Distributed observers design for leader-following control of multi-agent networks," *Automatica*, vol. 44, no. 3, pp. 846–850, 2008.

[39] F. Hou and J. Sun, "Covert attacks against output tracking control of cyber-physical systems," *IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society*, pp. 5743–5748, 2017.

[40] M. Hou and P. Muller, "Disturbance decoupled observer design: A unified viewpoint," *IEEE Transactions on automatic control*, vol. 39, no. 6, pp. 1338–1341, 1994.

[41] T. Houtgast, "Frequency selectivity in amplitude-modulation detection," *The Journal of the Acoustical Society of America*, vol. 85, no. 4, pp. 1676–1680, 1989.

[42] C.-S. Hsieh, "Robust two-stage kalman filters for systems with unknown inputs," *IEEE Transactions on Automatic Control*, vol. 45, no. 12, pp. 2374–2378, 2000.

[43] T. Irita and T. Namerikawa, "Detection of replay attack on smart grid with code signal and bargaining game," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 2112–2117.

[44] R. Isermann, *Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems*. Springer Science & Business Media, 2011.

[45] R. Isermann and P. Ballé, "Trends in the application of model based fault detection and diagnosis of technical processes," *IFAC Proceedings Volumes*, vol. 29, no. 1, pp. 6325–6336, 1996.

[46] M. Jazzar and M. Hamad, "An analysis study of iot and dos attack perspective," *Proceedings of International Conference on Intelligent Cyber-Physical Systems*, pp. 127–142, 2022.

[47] B. Jiang and M. Staroswiecki, "Adaptive observer design for robust fault estimation," *International Journal of Systems Science*, vol. 33, no. 9, pp. 767–775, 2002.

[48] K. H. Johansson, "The quadruple-tank process: A multivariable laboratory process with an adjustable zero," *IEEE Transactions on control systems technology*, vol. 8, no. 3, pp. 456–465, 2000.

[49] P. Khargonekar, K. Poolla, and A. Tannenbaum, "Robust control of linear time-invariant plants using periodic compensation," *IEEE Transactions on Automatic Control*, vol. 30, no. 11, pp. 1088–1096, 1985.

[50] K.-D. Kim and P. R. Kumar, "Cyber–physical systems: A perspective at the centennial," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1287–1308, 2012.

[51] S. Kim, Y. Eun, and K.-J. Park, "Stealthy sensor attack detection and real-time performance recovery for resilient cps," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 11, pp. 7412–7422, 2021.

[52] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 645–658, 2011.

[53] A. Kumar, G. Vashishtha, C. Gandhi, Y. Zhou, A. Glowacz, and J. Xiang, "Novel convolutional neural network (ncnn) for the diagnosis of bearing defects in rotary machinery," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.

[54] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security & Privacy*, vol. 9, no. 3, pp. 49–51, 2011.

[55] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, p. 106587, 2020.

[56] L. Li, W. Wang, Q. Ma, K. Pan, X. Liu, L. Lin, and J. Li, "Cyber attack estimation and detection for cyber-physical power systems," *Applied Mathematics and Computation*, vol. 400, p. 126056, 2021.

[57] W. Li and S. Shah, "Fault detection and isolation in non-uniformly sampled systems," *IFAC Proceedings Volumes*, vol. 37, no. 9, pp. 59–64, 2004.

[58] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, 2016.

[59] L. Liu, L. Ma, J. Zhang, and Y. Bo, "Distributed non-fragile set-membership filtering for nonlinear systems under fading channels and bias injection attacks," *International Journal of Systems Science*, vol. 52, no. 6, pp. 1192–1205, 2021.

[60] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[61] M. G. Losada, F. R. Rubio, and S. D. Bencomo, *Asynchronous control for networked systems*. Springer, 2015.

[62] M. S. Mahmoud and Y. Xia, *Networked control systems: cloud control and secure control*. Butterworth-Heinemann, 2019.

[63] M.-A. Massoumnia and W. E. V. Velde, "Generating parity relations for detecting and identifying control system component failures," *Journal of Guidance, Control, and Dynamics*, vol. 11, no. 1, pp. 60–65, 1988.

[64] R. K. Mehra and J. Peschon, "An innovations approach to fault detection and diagnosis in dynamic systems," *Automatica*, vol. 7, no. 5, pp. 637–640, 1971.

[65] A. R. Mehrabian and K. Khorasani, "Constrained distributed cooperative synchronization and reconfigurable control of heterogeneous networked euler–lagrange multi-agent systems," *Information Sciences*, vol. 370, pp. 578–597, 2016.

[66] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding schemes for securing cyber-physical systems against stealthy data injection attacks," *arXiv preprint arXiv:1605.08962*, 2016.

[67] J. Milosevic, H. Sandberg, and K. H. Johansson, "A security index for actuators based on perfect undetectability: Properties and approximation," *arXiv preprint arXiv:1807.04069*, 2018.

[68] A. Mitra and S. Sundaram, "Secure distributed observers for a class of linear time invariant systems in the presence of byzantine adversaries," *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 2709–2714, 2016.

[69] Y. Mo and B. Sinopoli, "Secure control against replay attacks," *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pp. 911–918, 2009.

[70] 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, 2014.

[71] 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.

[72] ——, "Secure control against replay attacks," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*. IEEE, 2009, pp. 911–918.

[73] E. Mousavinejad, F. Yang, Q.-L. Han, and L. Vlacic, "A novel cyber attack detection method in networked control systems," *IEEE transactions on cybernetics*, no. 99, pp. 1–11, 2018.

[74] Naderi and Khorasani, "Inversion-based output tracking and unknown input reconstruction of square discrete-time linear systems," *Automatica*, vol. 95, pp. 44–53, 2018.

[75] E. Naderi and K. Khorasani, "Data-driven fault detection, isolation and estimation of aircraft gas turbine engine actuator and sensors," *Mechanical Systems and Signal Processing*, vol. 100, pp. 415–438, 2018.

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

[77] M. Naghnaeian, N. H. Hirzallah, and P. G. Voulgaris, "Security via multirate control in cyber–physical systems," *Systems & Control Letters*, vol. 124, pp. 12–18, 2019.

[78] R. Nikoukhah, "Innovations generation in the presence of unknown inputs: Application to robust failure detection," *Automatica*, vol. 30, no. 12, pp. 1851–1867, 1994.

[79] M. Nyberg, "Criterions for detectability and strong detectability of faults in linear systems," *International Journal of Control*, vol. 75, no. 7, pp. 490–501, 2002.

[80] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 8, pp. 1773–1786, 2015.

[81] M. Pajic, I. Lee, and G. J. Pappas, "Attack-resilient state estimation for noisy dynamical systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 82–92, 2016.

[82] G. Park, H. Shim, C. Lee, Y. Eun, and K. H. Johansson, "When adversary encounters uncertain cyber-physical systems: Robust zero-dynamics attack with disclosure resources," in *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 5085–5090.

[83] 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.

[84] H. Penney. (2020) Us state department must align uav export policy with american interests. [Online]. Available: https://www.defensenews.com/opinion/commentary/2020/06/11/us-state-department-must-align-uav-export-policy-with-american-interests/

[85] J. L. Pons, *Emerging actuator technologies: a micromechatronic approach*. John Wiley & Sons, 2005.

[86] M. A. Qadeer, A. Iqbal, M. Zahid, and M. R. Siddiqui, "Network traffic analysis and intrusion detection using packet sniffer," in *Communication Software and Networks, 2010. ICCSN'10. Second International Conference on*. IEEE, 2010, pp. 313–317.

[87] M. Saif and Y. Guan, "A new approach to robust fault detection and identification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 3, pp. 685–695, 1993.

[88] C. Schellenberger and P. Zhang, "Detection of covert attacks on cyber-physical systems by extending the system dynamics with an auxiliary system," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, Dec 2017, pp. 1374–1379.

[89] U. Shaked and N. Karcanias, "The use of zeros and zero-directions in model reduction," *international Journal of Control*, vol. 23, no. 1, pp. 113–135, 1976.

[90] M. A. Sid, "Sensor scheduling strategies for fault isolation in networked control system," *ISA transactions*, vol. 54, pp. 92–100, 2015.

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

[92] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation," pp. 1015–1021, 2006.

[93] E. Soroka and U. Shaked, "On the geometry of the inverse system," *IEEE transactions on automatic control*, vol. 31, no. 8, pp. 751–754, 1986.

[94] W. Sun, J. Chen, and J. Li, "Decision tree and pca-based fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 21, no. 3, pp. 1300–1317, 2007.

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

[96] Taheri, Khorasani, Shames, and Meskin, "Undetectable cyber attacks on communication links in multi-agent cyber-physical systems," *2020 59th IEEE Conference on Decision and Control (CDC)*, pp. 3764–3771, 2020.

[97] M. Taheri, K. Khorasani, I. Shames, and N. Meskin, "Cyber attack and machine induced fault detection and isolation methodologies for cyber-physical systems," *arXiv preprint arXiv:2009.06196*, 2020.

[98] 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.

[99] G. Tertytchny, N. Nicolaou, and M. K. Michael, "Classifying network abnormalities into faults and attacks in iot-based cyber physical systems using machine learning," *Microprocessors and Microsystems*, vol. 77, p. 103121, 2020.

[100] H. L. Trentelman, A. A. Stoorvogel, and M. Hautus, *Control theory for linear systems*. Springer Science & Business Media, 2012.

[101] G. K. Venayagamoorthy, "Dynamic, stochastic, computational, and scalable technologies for smart grids," *IEEE Computational Intelligence Magazine*, vol. 6, no. 3, pp. 22–35, 2011.

[102] V. Venkatasubramanian and K. Chan, "A neural network methodology for process fault diagnosis," *AIChE Journal*, vol. 35, no. 12, pp. 1993–2002, 1989.

[103] N. Watanabe, "Note on the kalman filter with estimated parameters," *Journal of Time Series Analysis*, vol. 6, no. 4, pp. 269–278, 1985.

[104] S. Weerakkody, O. Ozel, P. Griffioen, and B. Sinopoli, "Active detection for exposing intelligent attacks in control systems," in *Control Technology and Applications (CCTA), 2017 IEEE Conference on*. IEEE, 2017, pp. 1306–1312.

[105] 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.

[106] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. IEEE, 2010, pp. 226–231.

[107] T. YOSHlKAWA and T. Sugie, "Filtered inverse systems," *International Journal of Control*, vol. 43, no. 6, pp. 1661–1671, 1986.

[108] M. Zamani, B. D. Anderson, U. Helmke, and W. Chen, "On the zeros of blocked time-invariant systems," *Systems & Control Letters*, vol. 62, no. 7, pp. 597–603, 2013.

[109] M. Zamani, U. Helmke, and B. D. Anderson, "Zeros of networked systems with time-invariant interconnections," *Automatica*, vol. 61, pp. 97–105, 2015.

[110] H. Zhang, F. L. Lewis, and A. Das, "Optimal design for synchronization of cooperative systems: state feedback, observer and output feedback," *IEEE Transactions on Automatic Control*, vol. 56, no. 8, pp. 1948–1952, 2011.

[111] K. Zhang, C. Keliris, T. Parisini, and M. M. Polycarpou, "Identification of sensor replay attacks and physical faults for cyber-physical systems," *IEEE Control Systems Letters*, vol. 6, pp. 1178–1183, 2021.

[112] Y. Zhang and J. Jiang, "Bibliographical review on reconfigurable fault-tolerant control systems," *Annual reviews in control*, vol. 32, no. 2, pp. 229–252, 2008.

[113] M. Zhong, H. Ye, S. X. Ding, and G. Wang, "Observer-based fast rate fault detection for a class of multirate sampled-data systems," *IEEE Transactions on Automatic Control*, vol. 52, no. 3, pp. 520–525, 2007.

[114] M. Zhu and S. Martínez, "On distributed constrained formation control in operator–vehicle adversarial networks," *Automatica*, vol. 49, no. 12, pp. 3571–3582, 2013.