Observer-based Anomaly Diagnosis and Mitigation for Cyber-Physical Systems

by

Zheng Wang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mechanical Engineering)
in the University of Michigan
2018

Doctoral Committee:

Professor Dawn Tilbury, co-chair
Dr. James Moyne, co-chair
Dr. Dhananjay Anand, National Institute of Standards and Technology
Professor Kira Barton
Professor Necmiye Ozay

Zheng Wang

zhengwa@umich.com

ORCID iD: 0000-0003-0306-8636

DEDICATION

For my parents Zhanling Wang and Xiuqin Liang who are always supportive.

For those who will never fade away in my memory.

ACKNOWLEDGEMENTS

I was told that writing acknowledgements is the most enjoyable part because it means I am almost done with my dissertation. Meanwhile, this is also one of the most important parts of my dissertation. My dissertation could not be done without the help and support from other people and funding from the National Institute of Standards and Technology. I would like to express my gratitude to who have helped me in achieving my doctoral degree.

My advisors Professor Dawn Tilbury and Dr. James Moyne gave me tremendous help. I can feel that I have progressed a lot during my Ph.D. study. I used to do what parents and teachers in school told me to do so that I would get compliments from my parents and teachers. Throughout the Ph.D. study, I understand that being a leader and actively thinking are more fun and those experiences help me discover a different world. During the dinner after my final defense, my advisors told me two things I think I shall never forget: 1) I should be a leader instead of a follower; 2) As a female engineer, I should contribute to changing this male-dominated society.

I am also grateful for the help I received from my committee members Professor Kira Barton, Professor Necmiye Ozay, and Dr. Dhananjay Anand. Especially, Dr. Dhananjay Anand and I have been working together throughout my whole Ph.D. program. He is very intelligent and motivated. Additionally, I would like to thank my other collaborators Dr. Farshad Harirchi and CheeYee Tang. Both of them provided me valuable suggestions for my work. My life in my office would not be so much fun without labmates from the Tilbury and Barton research groups. They helped me not only with my research, but also we joked around during the day. The Sweetland

Writing Center staff also helped me a lot with editing my paper.

My family is always my strongest support in every important decision I made so far. They supported me both financially and emotionally. My father wrote me a traditional Chinese poem after my final defense.

励志征程攀天梯，异域他国探真知。
贰拾贰载艰辛苦，终成正果工博士。

I would also like to express my gratitude to my friends. They are like my families in the United States. Besides those in real life, I would like to thank Mo Xu, a virtual character in the game Love & Producer, who brings iridescence to my virtual world.

The last yet the most important thing is to thank myself. Thanks to my perseverance, I was able to go through the hardship during my life. Thanks to my resilience, I am able to see the beauty of the world. Thanks to my curiosity, I will continue to discover the wonders around me. I wrote a traditional Chinese poem for myself before my final defense as an epilogue of my Ph.D. journey.

一入象牙廿贰载，出塔已过豆蔻年。
答辩前夕再回首，一把辛酸不言中。
漂洋过海求学路，渐把他乡作故乡。
寂寞安村寂寞雪，寂寞科研寂寞人。
可观可控可求导，连续离散非线性。
日夜颠倒证稳定，黑白不分求收敛。
学术高楼起平地，如今终临毕业期。
一路坎坷见成长，磐石淬炼成璞玉。
幸得友人慰吾心，同哭同笑同戚戚。
逝去随风化叹息，未来人生尚可期。

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

APPENDIX

ABSTRACT

Cyber-Physical Systems (CPS) seamlessly integrate computational devices, communication networks, and physical processes. The performance and functionality of many critical infrastructures such as power, traffic, and health-care networks and smart cities rely on advances in CPS. However, higher connectivity increases the vulnerability of CPS because it exposes them to threats from both the cyber domain and the physical domain. An attack or a fault within the cyber or physical domain can subsequently affect the cyber domain, the physical domain, or both, resulting in anomalies. An attack or a fault on CPS can have serious or even lethal consequences. Traditional anomaly diagnosis techniques mainly focus on cyber-to-cyber or physical-to-physical interactions. However, in practice they can often be subverted in the face of cross-domain attacks or faults. In summary, the safety and reliability of CPS become more and more crucial every day and existing techniques to diagnose or mitigate CPS attacks and faults are not sufficient to eliminate vulnerability.

The motivation of this dissertation is to enhance anomaly diagnosis and mitigation for CPS, covering physical-to-physical and cyber-to-physical attacks or faults. With the advantage of dealing with system uncertainties and providing system state estimation, observer-based anomaly diagnosis is of great interest. The first task is to design a multiple observers framework to diagnose sensor anomalies for continuous systems. Since CPS contain both continuous and discrete variables, CPS are modeled as hybrid systems. Utilizing the relationship between the continuous and discrete variables, a conflict-driven hybrid observer-based anomaly detection method is proposed, which checks for conflicts between the continuous and discrete variables to detect anomalies. Lastly, the observer design for hybrid systems is improved to enable observer-based anomaly diagnosis for a wider class of hybrid systems.

The novel observer-based anomaly diagnosis and mitigation approaches introduced in this dissertation can not only diagnose anomalies caused by traditional faults, but also anomalies caused by sophisticated attacks. This research work can benefit the overall security of critical infrastructures, preventing disastrous consequences and reducing economic loss. The effectiveness of the proposed approaches is demonstrated mathematically and illustrated through applications to various simulated systems, including a suspension system, the Positive Train Control system and a microgrid system.

CHAPTER I

Introduction

Through new measurement science, advanced Internet of Things (IoT) technologies, increasing computational power, and communication network speed, the cyber world and the physical world are integrated together in a scalable way to form new type of systems, Cyber-Physical Systems (CPS). CPS combine physical processes, computational resources, and communication capabilities in a unified design effort [12]. CPS are ubiquitous in critical infrastructures, such as transportation systems, power systems, and industrial control processes [12]. CPS security research faces unique challenges due to the complex ways cyber components interact with physical systems. To improve the security of CPS, the contribution of this dissertation is to extend the ability of anomaly diagnosis and mitigation to address more types of anomalies, which are caused by not only faults but also attacks.

I.1 Security of Cyber-Physical Systems

Faults or attacks on CPS can cause damage to public safety as well as economic losses. With the integration of the cyber and physical components, the security of CPS requires a three component perspective: cyber, physical and cyber-physical interaction. Traditional security techniques mainly focus on either cyber or physical components. Security of cyber components is usually associated with mechanisms such as cryptography, intrusion detection, and many other solutions commonly used in IT systems. Security of physical components is usually associated with model-based or non-model based fault diagnosis and fault tolerant control traditionally used in industrial control systems. Although the traditional cyber and physical security mechanisms can improve the security of CPS, in practice, they can be subverted by cross-domain attacks or faults because the cyber-physical interaction is not taken into consideration [12,40]. The failure to detect attacks or faults is especially undesirable as these attacks or faults can result in safety concerns. Much more needs to be done to secure CPS.

An example physical-to-cyber attack is side channel attack [27]. If an audio recorder is placed closed to a 3D printer, the attacker was able to regenerate the G code of the 3D printer by analyzing the recorded sound. Some attacks from the cyber domain can impact the physical assets, especially the critical infrastructure. In 2000, some hackers attacked Maroochy Shire Council's sewage control systems in Queensland, Australia, and caused flooding with millions of liters of sewage [81]. In 2010, Stuxnet was found to sabotage Iran's nuclear facilities, causing the nuclear centrifuges to spin out of control [45]. In 2015, a YouTube video demonstrated how cyber-attacks allow hackers to remotely gain control of a vehicle through the 3G network while a driver is driving it on a highway [60]. These real-life examples illustrate the importance and urgency of developing new approaches to expand the capabilities of traditional techniques regarding cross-domain attacks or faults.

Three security goals are required for a CPS: integrity, availability, and confidentiality [13]. Integrity refers to the trustworthiness of data or resources, which include the data or resources sent or received by the sensors, actuators or controllers. Availability refers to the ability of a system to be accessible and usable upon demand. Confidentiality refers to the ability to keep information secret from unauthorized users. Both faults and attacks can cause anomalous behaviors in systems, impacting the ability to achieve at least one of these three goals.

We give the formal definitions of anomaly, fault and attack [57].

Definition 1. *An anomaly is an occurrence that is different from what is standard, normal, or expected.*

Definition 2. *A fault is an anomaly that is related to an unwanted situation and may be associated* *with failure, malfunction, or quality degradation.*

Definition 3. *An attack is a purposeful action by an element external to the system that breaches the security goals of the system.*

In this dissertation, we are concerned with the anomalies caused by both faults and attacks.

Recognizing that the national and economic security depends on the reliable functioning of critical infrastructure, National Institute of Standards and Technology (NIST) has developed a voluntary framework for reducing cyber risks to critical infrastructure, which consists of standards, guidelines, and best practices [21]. The framework core includes five high-level functions: identify, protect, detect, respond, and recover. These five functions are not only applicable to cybersecurity risk management but also to risk management at large. The work of this dissertation falls into the detect and respond functions. We expand the capability of traditional anomaly diagnosis and mitigation by considering the cyber-physical interaction components. Diagnosis consists of detection and isolation[1], which are defined as [87]

Definition 4. *Detection makes a binary decision on whether an anomaly has occurred or not.*

Definition 5. *Isolation determines the location, and assesses the extent of the anomaly.*

Additionally, mitigation is defined as [25]

Definition 6. *Mitigation reduces the effect of the anomaly.*

The most popular anomaly diagnosis techniques can be classified into two main categories: model-based diagnosis and non-model based diagnosis [89]. Model-based anomaly diagnosis requires a process model running in parallel with a physical process and diagnoses an anomaly by comparing the estimates and measured process [42]. However, it assumes a system model is available. The system model is either built based on expert knowledge or learned based on a set of data or a combination of the two. Non-model based diagnosis checks the symptoms of a set of data, such as mean values or trends, to diagnose an anomaly [70]. This method has the advantage

[1] In the NIST security framework, detection process identifies the occurrence of a security event [21], which is different from the definition of detection in this dissertation. The definition of detection process in NIST security framework corresponds to detection and isolation in this dissertation.

when a system model is unavailable. However, non-model based methods are limited to a process in the steady state and not applicable to a process with a wide operation range. Additionally, due to the lack of system knowledge, non-model based methods have to rely on data and data history to determine acceptable operating region and therefore can be subject to false positives or false negatives.

As a starting point, we begin with model-based anomaly diagnosis. Among various model-based anomaly diagnosis techniques, the observer-based anomaly diagnosis technique is one of the central schemes, and it has the advantage of reducing the impact of system uncertainties [80]. Gaps still exist in addressing anomalies caused by cross-domain faults or attacks for CPS using observer-based anomaly diagnosis framework. In the next section, we discuss the gaps of observer-based anomaly diagnosis in detail and the specific contributions this dissertation makes towards filling the gaps.

I.2 Contributions to the Research Areas

To improve the security of CPS, this dissertation makes contributions to the following three research areas.

We first focus on anomalies in sensors because sensors play a vital role in CPS estimation and control and they are also the most vulnerable part of CPS [82]. We model the CPS as continuous systems and work on anomaly diagnosis for continuous systems. Under the observer-based framework, we propose new detection, isolation, and mitigation methods to improve the overall performance of sensor anomaly diagnosis.

Then, we extend our work to hybrid systems which consist of both continuous dynamics and discrete behavior. For CPS, considering a continuous system alone is not adequate, because CPS contain both continuous and discrete variables. Additionally, the anomaly type is not limited to sensor anomalies. Assuming that the discrete behavior of the hybrid systems is current-state observable[2], we propose a new method that utilizes the relationship between the continuous and discrete variables to identify anomalous behaviors.

Finally, we consider a wider class of hybrid systems, including hybrid systems with unobservable discrete events. For some hybrid systems, the discrete system is not current-state observable. One of the reasons is that the discrete system contains some unobservable discrete events. To extend our work to hybrid systems with unobservable discrete events, we propose another method which can determine the current discrete state of the system by estimating the current continuous dynamics of the system.

We further describe our contributions in the three main research areas in the following subsections.

I.2.1 Sensor Anomaly Diagnosis and Mitigation for Continuous Systems

Although observer-based anomaly diagnosis has been developed for decades for continuous systems, existing methods mainly focus on anomalies that occur during steady state operation. However, an anomaly caused by an attack can happen any time, including during an observer's transient state. In addition, an attack can be designed to bypass a closed-loop observer-based anomaly detection method. For example, an attack targets critical sensors which are essential for system observability and the attack signal gradually changes the sensor value. Moreover, because sensor anomaly diagnosis takes time, a mitigation method is needed to potentially reduce the impact of a sensor anomaly during the time required for anomaly diagnosis.

Consider a train system as an example. Suppose there are two noisy sensors measuring the train position and velocity, respectively, and an observer estimating the train position and velocity. Since the initial state of the train is not precisely known, it takes some time for the observer to converge. During the observer's transient state, the estimation error is large and thus the residual

[2] A discrete system is current-state observable if the discrete state of the system can be uniquely determined after a finite number of discrete events. The formal definition is introduced in Chapter IV.

signal is large. A large residual signal may trigger an alarm even though the system is under normal operation. Traditionally, anomaly diagnosis is disabled during observer's transient state to reduce false alarms. But anomalies occurring during the observer's transient state are not detected, which is undesired for critical infrastructure. In addition, the position sensor is a critical sensor because it is indispensable for state estimation. If the measurement given by the position sensor is gradually drifting, the residual signal may not exceed the threshold and a residual-based method using closed-loop observer fails to detect it. As shown in Fig.I.1, to diagnose more sensor anomalies and mitigate sensor anomaly during anomaly diagnosis, our contributions are to propose three new methods that respectively:

1. enable sensor anomaly detection and reduce false alarms during the observers' transient state;
2. detect anomalies on critical sensors; and
3. potentially mitigate the impact of the anomalous sensor during the anomaly diagnosis process.

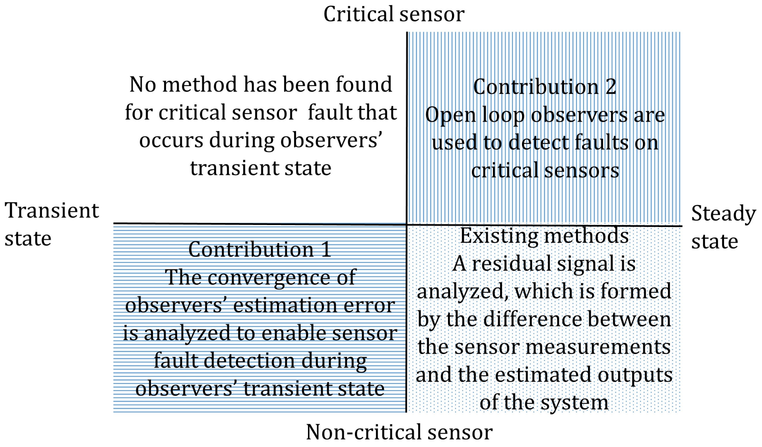

Critical sensor
No method has been found for critical sensor fault that occurs during observers' transient state
Contribution 2
Open loop observers are used to detect faults on critical sensors
Transient state
Steady state
Contribution 1
The convergence of observers' estimation error is analyzed to enable sensor fault detection during observers' transient state
Existing methods
A residual signal is analyzed, which is formed by the difference between the sensor measurements and the estimated outputs of the system
Non-critical sensor

Figure I.1: The two contributions to sensor anomaly diagnosis positioned in the space of critical vs. non-critical sensors, and observers' transient vs. steady state

I.2.2 Anomaly Detection for Hybrid Systems With Current-State Observable Discrete Dynamics

The improved sensor anomaly diagnosis framework for continuous systems is not enough for systems that have both continuous variables and discrete variables, i.e., hybrid systems. In this section, we first introduce the limitation of the sensor anomaly diagnosis framework for hybrid systems. Then we describe the contributions to the area of anomaly detection for hybrid systems.

I.2.2.1 Limitation of the Sensor Anomaly Diagnosis Framework for Hybrid Systems

Hybrid systems contain both continuous dynamics and discrete behavior. The improved sensor anomaly diagnosis framework only considers the continuous dynamics to diagnose anomalies, which has limitations when used for hybrid systems. As an example, a train system can have different operation modes, i.e., discrete states, under different scenarios. When the train is running freely on the track, a speed controller is regulating the train speed. When the train is approaching the next scheduled station, a position controller makes sure that the train stops at the designated position. Since the train system considered in Section I.2.1 is a continuous system, we consider a hybrid system example to study anomaly detection for hybrid systems with current-state observable discrete dynamics [69].

I.2.2.2 Anomaly Detection for Hybrid Systems Utilizing the Relation Between the Discrete and Continuous Components

In order to detect more types of anomalies including anomalies in continuous variables mentioned in Section I.2.1, we expand our work to hybrid systems which include both continuous and discrete states. Here, we consider that each discrete state has an invariant which describes the set of allowable continuous states and there is no discontinuity in continuous states when discrete transition occurs. The proposed anomaly detection method benefits from a hybrid observer, which consists of a Finite State Machine (FSM) and a Set-Valued Observer (SVO). Based on all possible anomalies for hybrid systems, three types of conflicts are defined. A conflict is a contradiction between the continuous and discrete variables. The conflict-driven anomaly detection method takes advantage of the knowledge from these continuous-discrete interactions to identify anomalous behaviors. In the work presented in Chapter IV, the contributions are four-fold:

1. We propose a conflict-driven method with three conflict types defined based on the relation between the discrete and the continuous variables of the hybrid systems. In addition to anomalies that can be detected by traditional observer-based and residual-based methods, the conflict-driven anomaly detection approach is capable of providing guarantees on the detection of some types of attacks and faults that are undetectable using the traditional methods.

2. We define a classification taxonomy for anomalies in hybrid systems. An anomaly in a hybrid system may affect the continuous variables or the discrete variables or both. Some anomalies are undetectable by only considering the continuous component of the system because the anomalous system may have a consistent input-output data with the system model under normal operation. Some anomalies are undiagnosable by only considering the discrete component of the system because the observed discrete event sequence of the anomalous system is the same as the system under normal operation. In this dissertation, we classify the anomalies into eight different types based on the variables that are affected, input-output data consistency, and diagnosability of the anomaly.

3. We develop a new hybrid observer for anomaly detection. We use a Set-Valued Observer (SVO) as the continuous state observer of the hybrid observer. With the SVO, we can apply the conflict-driven method to hybrid systems with unobservable continuous components.

4. We provide a mapping between conflict types and anomaly types. Based on the occurrence of the conflict types, we can identify if the anomaly is related to the continuous component of the system, the discrete component or both.

Additionally, we illustrate the effectiveness of the conflict-driven method in detecting different types of anomalies in a realistic system, namely a simulated Positive Train Control (PTC) system that is used as the illustrative example [69].

I.2.2.3 Positive Train Control System

The PTC system is designed to ensure safe and collision-free operation as well as high throughput of trains in a safety-critical environment [69]. The PTC system is a hybrid system that consists of a train and a Radio Block Controller (RBC). The train is modeled as a continuous system, and the RBC is modeled as a discrete system. Faults or attacks can occur in either the train system or the RBC system or both.

I.2.3 Anomaly Detection for Hybrid Systems With Unobservable Discrete Events

The conflict-driven method described in Section I.2.2 is proposed for hybrid systems with current-state observable discrete components. However, some hybrid systems contain unobservable discrete events such that the discrete components are not current-state observable. In this section, we first discuss the limitation of the conflict-driven method for hybrid systems with unobservable discrete events. Then, we describe the contributions to the area of anomaly detection for hybrid systems with unobservable discrete events.

I.2.3.1 Limitation of the Conflict-driven Method for Hybrid Systems With Unobservable Discrete Events

The conflict-driven method described in Section I.2.2 assumes that 1) the discrete behavior of the hybrid systems under normal operation is observable, 2) the invariant of each discrete state under normal operation is known, and 3) no discontinuity exists in continuous variables when a discrete transition occurs. However, the above three assumptions may not be true for some hybrid systems. As an example, in the PTC system, consider a scenario where a train *a* is running to a railway junction with a sensor measuring the distance from train *a* to any front object. There is another train *b* running on one of the tracks and a human operator doing some work on the other track. Suppose the default position of the railroad switch indicates that the front object of train *a* is train *b*. Suddenly an unexpected (unobservable) fault occurs, causing the railroad switch to direct train *a* to the other track. The front object of train *a* changes to the human operator. Thus there is a discontinuity in the variable representing the position of the front object. Then the hybrid observer used in the conflict-driven method cannot be used to detect the anomaly in this scenario. A new anomaly detection method is thus needed for hybrid systems.

I.2.3.2 State Estimation and Anomaly Detection for Hybrid Systems With Unobservable Discrete Events

To estimate state and detect anomalies for hybrid systems with unobservable discrete events, we propose a new observer framework which consists of two continuous state observers. The two continuous state observers use different sets of sensors and the same continuous system model of the current estimated discrete state (assuming that the initial discrete state of the system is given) to estimate the continuous state of the system. Based on the estimated continuous state trajectories, the Recursive Least Squares (RLS) method is used to help identify the current continuous dynamics of the system, thus knowing the current discrete state of the system. In the work presented in Chapter V, the contributions are as follows:

1. We propose a new observer framework to estimate both the discrete and the continuous variables for hybrid systems with unobservable discrete events;
2. We use the proposed observer framework to detect anomalies which can be modeled as unobservable discrete events; and
3. We apply the proposed anomaly detection method to a realistic microgrid system to validate its

effectiveness.

The reason we use microgrid system, instead of the PTC system, to validate the effectiveness of the new observer framework is that the PTC system is not complex enough for our purpose. For the PTC system, we can use the measured distance from train *a* to the front object to reset the estimated continuous state by the continuous state observer. The continuous state observer can still give a good state estimation because the continuous dynamics of the PTC system are simple and stay the same before and after the fault mentioned in Section I.2.3.1. Such an example does not provide enough complexity to show the need for a new observer design for hybrid systems with unobservable discrete events. Therefore, we look to a power microgrid, which is a more complex system, to show that a new observer is needed for state estimation and anomaly detection for hybrid systems with unobservable discrete events.

I.2.3.3 Microgrid System

A microgrid system is an electrical energy generation, consumption, and grid-interaction system, which consists of Distributed Energy Resources (DER) such as solar, wind, fuel cells, etc., loads and transmission lines, as shown in Fig. I.2 [72]. Depending on the status of the system, either grid-tied or islanded, the switch at the Point of Common Coupling (PCC) will connect the microgrid to the main grid or not. Knowing the status of the microgrid can help ensure worker safety and DER management. However, the transition from the grid-tied to islanded is an unobservable discrete event in the case of unplanned islanding [56]. Islanding Detection Methodology (IDM) is an algorithm that allows for the presence of an electrical island to be detected [4]. Traditional IDM can be classified into remote and local methods [48]. Remote methods are based at the grid level where the communication between the utility and the DER is monitored. Local methods are based at the inverter where the information at DER side is gathered to determine whether or not the DER is islanded. Traditional IDMs are developed for single DER instead of a system with multiple DERs. In addition, the traditional IDMs cannot provide state estimation before the unplanned islanding is detected. Without a good state estimation during the unplanned islanding detection, the controller of the system cannot provide a good control performance, and may even damage the system in severe cases. An extended IDM theory is needed to address unplanned islanding detection problems in microgrids with multiple DERs, as well as provide good state estimation during the islanding detection time period. Our proposed state estimation and anomaly detection method for hybrid system with unobservable discrete events can be used to detect the unplanned islanding of the microgrid system consisting of multiple DERs.

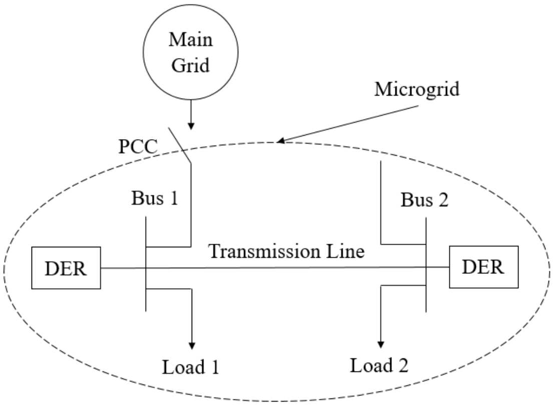


Figure I.2: Microgrid Architecture.

I.3 Expected Impact

This dissertation is expected to have the following impacts:

1. Enhance capabilities to diagnose anomalies and mitigate the impact of anomalies in CPS which are modeled as continuous systems;

2. Enhance capabilities to detect anomalies in hybrid systems with current-state observable discrete components; and

3. Enhance capabilities to detect anomalies in hybrid systems with unobservable discrete events.

With the proposed anomaly diagnosis and mitigation approaches, the overall security of CPS will be improved by expanding the types of anomalies that can be diagnosed.

I.4 Dissertation Outline

The rest of the dissertation is organized as follows. In Chapter 2, we review the relevant literature to identify the gaps of anomaly diagnosis and mitigation for continuous systems, discrete systems, and hybrid systems, respectively. In Chapter 3, we present our improved sensor fault diagnosis and mitigation framework. In Chapter 4, the proposed conflict-driven anomaly detection method for hybrid systems is introduced and validated using the PTC system. In Chapter 5, we describe the new observer framework and present the effectiveness in state estimation and anomaly detection for hybrid systems with unobservable discrete events. In Chapter 6, we conclude the work of this dissertation.

CHAPTER II

Background

A significant amount of research has been carried out for both observer-based anomaly diagnosis and mitigation. An observer is a system that provides an estimate of the internal state of a given system, using measurements of the input and output of the system. The estimated internal state of the system can give an insight into how the system is behaving internally, thus helping diagnose anomalies in the system. Different observers have been designed for different types of systems. For continuous systems, the continuous state observers can be classified into two major types. One gives a single estimated continuous state, such as the Kalman filter. The other is the Set-Valued Observer (SVO). In contrast to Kalman filter, the SVO takes a measurement history of some time horizon and gives a non-empty set of estimated continuous states [79]. For discrete systems, a discrete state observer is usually designed as a finite state automaton to estimate the discrete state of the system [14]. For hybrid systems, a hybrid observer is typically used as it is computationally efficient [5]. A hybrid observer consists of a continuous state observer and a discrete state observer, estimating the continuous state and the discrete state of the system, respectively.

In this chapter, we first summarize the methods that use observers to diagnose sensor anomalies and mitigate the impact of sensor anomalies for continuous systems. Then, we provide a brief review of observer-based anomaly detection in hybrid systems that have current-state observable discrete components. Finally, we review observer-based anomaly detection for hybrid systems with unobservable discrete events.

II.1 Sensor Anomaly Diagnosis and Mitigation for Continuous Systems

A significant amount of research has been carried out to diagnose sensor anomalies using observer-based methods due to their cost efficiency [41]. Observer-based approaches use a continuous state observer to estimate the continuous state of the system. As mentioned above, there are two major types of continuous state observers for continuous systems. If a continuous state observer, such as the Kalman filter, is utilized, we can estimate system output based on the estimated continuous state. Then a residual, which is the difference between the measured output and the estimated output, is analyzed to detect an anomaly [20,29,85]. If a SVO is utilized, an anomaly is detected when the estimated state set is empty [73,76]. Based on our literature review, we identified three research gaps in anomaly detection, isolation and mitigation, respectively.

The first research gap we identified is that no existing method detects a sensor anomaly during the observers' transient state. For sensor anomaly isolation, a system of multiple continuous state observers is usually used, which is called the Dedicated Observer Scheme (DOS). In [19], each continuous state observer in the DOS uses only one sensor for state estimation based on the assumption that the system is observable with any one of the sensors. Similarly, in [9], the authors design multiple robust sliding mode observers with different subsets of sensor measurements to generate residuals for sensor anomaly diagnosis. Each sliding mode observer is designed to exclude a particular sensor so that the residual generated by this observer is sensitive to an anomaly in this sensor, but insensitive to anomalies in other sensors. In addition to observers designed using different inputs and outputs of the physical system, some DOSs consist of unknown input observers. In [1], the authors combine multiple local unknown input observers which can decouple the unknown disturbances from the residual to achieve robust anomaly diagnosis for nonlinear systems. In [46], instead of isolating unknown disturbances, the authors consider a single additive anomaly as an unknown input, and attempt to reconstruct the anomaly with a bank of unknown input observers for each sensor. All of the methods mentioned above assume that the observers have reached their steady state, so that the effect of the uncertain initial condition on a residual has died out. Otherwise, the methods may miss alarms or generate false alarms. A method that can detect a sensor anomaly during the observers' transient state is needed.

The second research gap we identified is that no existing method can detect critical sensor anomalies without hardware redundancy. Although the above mentioned methods are developed to diagnose anomalies that occurs during the observers' steady state, some anomalies may not be detected by traditional anomaly diagnosis, such as anomalies caused by intelligent attacks. In [55], the authors propose a cyber attack that injects false data in the sensor measurements and show that a static residual-based anomaly detector cannot detect this attack. In [8], the authors propose to protect the subset of sensor measurements which are necessary to ensure the system observability with a static residual-based anomaly detector in order to detect sensor anomalies caused by the cyber attacks introduced in [55]. In [62], the authors propose another kind of cyber attack, which can bypass not only a static anomaly detector but also an anomaly detector utilizing the system dynamics, such as a χ^2 anomaly detector[1]. The failure to detect a sensor anomaly caused by the cyber attack occurs because the system is not detectable according to classical control theory[2] when removing the anomalous sensor and, as a result, the attacker could impose arbitrarily large errors between the anomalous sensor measurements and the actual system outputs. The anomalous sensors in [62] are a subset of the critical sensors that are indispensable for system observability. A method using open-loop observers instead of closed-loop observers is needed to detect critical sensor anomalies.

The third research gap is that no method can mitigate the impact of sensor anomalies during the diagnosis process [50]. Some anomalies may happen quickly in systems with fast dynamics. Although the diagnosis of an anomaly can lead to appropriate maintenance [17, 26], the physical system may be in jeopardy during the diagnosis process. A timely mitigation technique during the

[1] A χ^2 anomaly detector converts a Gaussian distributed residual to a χ^2 distributed signal and detects an anomaly by comparing the χ^2 distributed signal with a pre-defined threshold.

[2] In control theory, a system is detectable if all the unobservable states are stable.

diagnosis process may help maintain acceptable performance of the physical system. To the best of our knowledge, sensor anomaly mitigation techniques that can be applied during the diagnosis process have not yet been developed for sensor anomalies [50].

Based on our literature review of sensor anomaly diagnosis and mitigation for continuous systems, three research gaps are identified for sensor anomaly diagnosis and mitigation for continuous systems:

1. no existing method detects a sensor anomaly during the observers' transient state;
2. no existing method can detect critical sensor anomalies without hardware redundancy; and
3. no method can mitigate the impact of sensor anomalies during the diagnosis process.

II.2 Anomaly Detection for Hybrid Systems with Current-State Observable Discrete Component

Hybrid systems consist of both continuous dynamics and discrete behavior. We have reviewed anomaly detection techniques for continuous systems in the previous section. Even though the aforementioned methods are effective for systems with continuous dynamics, they are computationally demanding for hybrid systems with many different continuous dynamics in different discrete states because observers with different continuous models need to run in parallel [91].

Various discrete model-based anomaly detection methods have been proposed up to date. The fault diagnosis problem is closely related to anomaly detection. To diagnose a fault, fishbone diagramming and fault-tree analysis are popular approaches because they are easily understood. A fishbone diagram is a cause-effect diagram, which maps potential root causes to the problems of the system [32]. Fault-tree analysis is a structural logic diagram (fault-tree) representing a physical system, in which low-level (software failure, hardware failure or human errors) causes are combined with boolean logic leading to one specified top event (undesired system failure) [49].

However, both fishbone diagramming and fault-tree analysis have certain limitations particularly in incorporating the ordering of events [90], which can be captured by finite-state automata. Two types of anomaly detection methods based on finite-state automata have been proposed: state-based and event-based approaches [14,54]. A state-based approach determines if the current discrete state is nominal or anomalous. An event-based approach determines whether an (unobservable) anomalous event has occurred or not, based on observed events. Even though discrete model-based methods are computationally efficient [78], they cannot provide sufficient resolution of continuous degradations for hybrid systems [31]. Discrete model-based methods can only detect drastic anomalies, such as a valve stuck closed. Other types of anomalies such as small changes in sensors or actuators cannot be addressed because they cannot be efficiently modeled in a discrete system framework [90].

As many CPS consist of both continuous dynamics and discrete behavior, hybrid model-based approaches are promising in anomaly detection. Hybrid model-based methods include set membership-based methods [36] and observer-based methods [39]. Given a data trajectory generated by the system, set membership-based methods check whether or not it is possible that the trajectory is generated by the model of the system. In [35], the concept of T-detectability is defined as a time horizon length that is enough to provide detection guarantees when given nominal and anomalous system models. Although these methods provide necessary and sufficient conditions in some cases for anomaly detection, they are computationally demanding as they require costly set calculations or mixed integer programming. Set membership-based methods are also utilized in active anomaly detection, where the goal is to design a minimal excitation that guarantees the detection of anomalous behavior [11,34,65]. In observer-based methods, a hybrid observer is used as it is computationally efficient [39]. A hybrid observer consists of two components: a discrete state observer identifying the current discrete state and a continuous state observer estimating the continuous state [5,39]. With the estimated discrete state given by the discrete state observer, the continuous state observer uses the corresponding continuous model and provides an estimate of the continuous state of the system. Usually a continuous state observer that gives a single estimated state is used for anomaly detection. With the hybrid observer framework, traditional residual-based methods can be applied for hybrid systems, including different residual generation methods, such as the unknown input observer scheme [15], and some residual evaluation methods, such as fuzzy decision logic approaches [2].

Even though residual-based methods with a hybrid observer are efficient, intuitive and easy to implement, they can easily be circumvented by a smart attacker or by sensor faults [61, 62]. This is because the relationship between the continuous component and discrete component is not considered in the residual-based methods. The research gap we identified is that no existing method utilizes the relationship between the continuous component and discrete component to detect anomalies in hybrid systems.

II.3 Anomaly Detection for Hybrid Systems with Unobservable Discrete Events

In the previous section, we mentioned that a hybrid observer is typically used in anomaly detection for hybrid systems. In order to use a hybrid observer, the discrete component of the system should be current-state observable such that the discrete state observer, which is designed as a finite state automaton, can give a unique estimated discrete state after a finite number of discrete transitions [5]. However, for some hybrid systems, some of the discrete events may be unobservable, thus the discrete component is not current-state observable. For these systems, the type of hybrid observer proposed in [5] cannot be used for anomaly detection.

Traditionally, if the discrete events are unobservable, a bank of continuous state observers is designed, each corresponding to the continuous system in one discrete state [6, 28]. Based on the estimated continuous state provided by each continuous state observer, we can calculate a residual, which is the the difference between the measured output and the estimated output. By analyzing the residuals, the discrete state can be uniquely determined if the continuous dynamics of different discrete states are distinguishable. However, this method is computationally complex for hybrid systems with a large number of different continuous dynamics [91]. Multiple-model estimation algorithms have been presented in [33,51,52,92] which track the most likely discrete state set, but the set is still very large for online estimation.

More methods have been proposed for hybrid systems with autonomous discrete transitions triggered by continuous dynamics, where the real-time discrete events are unobservable but the discrete transitions (guard condition[3], reset function[4], and linear-time properties[5]) are known *a priori*. In [44], a particle filter-based estimation algorithm is proposed, assuming that the guard conditions are known *a priori*. A method based on qualitative reasoning mechanisms using the discrete knowledge is proposed in [63, 64] assuming that *a priori* and *a posteriori* state vector values corresponding to the discrete transitions are known.

However, these above approaches cannot be used to distinguish discrete states of hybrid systems when only the continuous variables are measurable, the discrete events are unobservable, and the guard conditions and the reset functions are unknown *a priori*. Additionally, most practical anomalies occur with unknown conditions and times, and cannot be represented as observable discrete events. These anomalies should be described as unobservable discrete events which transition the system from nominal discrete states to anomalous discrete states. Additionally, the guard conditions and the reset functions corresponding to the anomaly discrete transitions are unknown *a priori*. In [84], a robust hybrid observer is proposed for hybrid systems. The robust hybrid observer consists of a continuous state observer and a discrete state observer. The continuous state observer estimates the continuous state and monitors the discrete transitions by comparing the residual with a threshold. When the residual exceeds the threshold, a discrete transition is detected and the discrete state observer is activated to identify the new discrete state. The discrete state observer consists of a bank of mode isolators. A mode isolator is an algorithm which checks

[3] Guard condition indicates when the discrete transition occurs.

[4] Reset function resets the value of the continuous state of the hybrid system when the corresponding discrete transition occurs.

[5] Linear-time properties specify the traces that a discrete system should exhibit [3].

whether or not the input-output signal from the system is consistent with one specific continuous model of the system. The mode isolators are designed as unknown input extended Kalman filters in [84]. The number of mode isolators is the same as the number of continuous models of the system. However, method of using mode isolators to track the emergence of unforeseen discrete states with continuous dynamics is computationally complex when identifying the discrete state if the hybrid system contains a large number of different continuous dynamics. Additionally, this method assumes that there is no discontinuity in continuous states. To estimate state and detect anomalies for hybrid systems with discontinuities in continuous variables, a predictor–corrector set-membership method is proposed in [71]. For the prediction step, a forward reachable set is calculated based on a union of zonotopes constructed by the current possible continuous states for each discrete state. For the correction step, the reachable set is filtered using the image projected by the measurements. The emptiness of the filtered reachable set indicates the infeasibility of the corresponding discrete state. With the prediction and correction steps, the inconsistent discrete states and inconsistent continuous state vectors in each consistent discrete state are discarded at each time step. However, this predictor-corrector set-membership method has the drawback of being computationally demanding because it requires intense set computation for several discrete states at each time step.

Based on this literature review, existing state estimation and anomaly detection methods for hybrid systems with unobservable discrete events at least have one of the following drawbacks:

1. *a priori* knowledge of discrete transitions is needed;
2. no discontinuity exists in continuous variables; or
3. they are computationally demanding.

The research gap we identified is that no existing observer design can effectively estimate states and detect anomalies for hybrid systems with unobservable discrete events, with discontinuity in continuous variables, and without *a priori* knowledge of discrete transitions.

II.4 Summary

In this chapter, we provided a literature review of sensor anomaly diagnosis and mitigation for continuous systems, anomaly detection for hybrid systems with current-state observable discrete transitions, and anomaly detection for hybrid systems with unobservable discrete events. Based on the detailed literature review, we identified several research gaps that need to be addressed in order to enhance the capability of observer-based anomaly diagnosis and mitigation to improve the overall security of CPS.

The research gaps identified for sensor anomaly diagnosis and mitigation for continuous systems are addressed in Chapter III. The research gaps for anomaly detection for hybrid systems with current-state observable discrete components are addressed in Chapter IV. The research gaps found for anomaly detection for hybrid systems with unobservable discrete events are addressed in Chapter V.

CHAPTER III

Improved Sensor Anomaly Diagnosis and Mitigation Using Multiple Observers Approach

III.1 Introduction

As introduced in Chapter I, sensors are considered to be the weak link in Cyber-Physical Systems (CPS) [13, 22]. A sensor anomaly that may be caused by a fault or an attack can be a major problem that may degrade the performance of the CPS, and even put the CPS in jeopardy in severe cases. Anomaly diagnosis and mitigation mechanisms are crucial for protecting a system that is susceptible to sensor faults or attacks. As defined in Chapter I, diagnosis consists of detection and isolation. Sensor anomaly detection determines the occurrence of a sensor anomaly. Sensor anomaly isolation identifies the anomalous sensor and estimates the anomaly signal. Sensor anomaly mitigation reduces the impact of the sensor anomaly [87]. Traditional anomaly diagnosis mechanisms mainly focus on anomalies caused by sensor faults. However, CPS are also subject to cross-domain attacks, i.e., attacks from the cyber domain that can cause anomalies in the physical domain. Some cross-domain sensor attacks can bypass the existing sensor anomaly diagnosis mechanisms.

As a starting point for developing a sensor anomaly diagnosis and mitigation mechanism for CPS, we model a Cyber-Physical System as a continuous system and assume that only one sensor is anomalous at a time. Based on our literature review in Chapter II, we identified the following three research gaps:

1. No existing method detects a sensor anomaly during the observers' transient state;

2. No existing method can detect critical sensor anomalies without hardware redundancy; and
3. No method can mitigate the impact of sensor anomalies during the diagnosis process.

We will fill these research gaps in this chapter.

In reviewing the existing methods in Section II.1 we note that one of the central schemes for sensor anomaly diagnosis is the Dedicated Observer Scheme (DOS), which consists of multiple observers [19]. Different observer-based approaches have been proposed under the DOS framework [1, 9, 46]. Most of these methods analyze residuals, which are the difference between the measured output and the estimated output, to diagnose anomalies. However, these residual-based methods assume that the observers have reached their steady state, so that the effect of the uncertain initial condition on a residual has died out. Otherwise, these methods may generate false alarms during the observers' transient state. An anomaly caused by an attack can happen any time, including during the observers' transient state. If we disable anomaly detection during the observers' transient state, then we can have missed alarms, resulting in severe consequences. Therefore, the first research gap we identified is that no existing method detects a sensor anomaly during the observers' transient state.

Based on the assumption of only one anomalous sensor, the sensors can be divided into two sets: *critical sensor* set and *non-critical sensor* set. Critical sensors are indispensable for system observability; if any critical sensor is removed, the system is unobservable. Non-critical sensors are redundant; the system is still observable if any one of the non-critical sensors is removed. An anomaly in a non-critical sensor can be diagnosed using a closed-loop observer which is designed excluding the anomalous sensor. Some anomalies in the critical sensors may not be detected by existing anomaly diagnosis approaches, such as anomalies caused by False Data Injection Attack [55, 61, 62]. Therefore, the second research gap we identified is that no existing method can detect critical sensor anomalies without hardware redundancy.

Some anomalies may happen quickly in systems with fast dynamics. Although the diagnosis of an anomaly can lead to appropriate maintenance, the physical system may be in jeopardy during the diagnosis process. A timely mitigation technique during the diagnosis process may help maintain acceptable performance of the physical system [50]. Therefore, the third research gap we identified is that no method can mitigate the impact of sensor anomalies during the diagnosis process.

In this chapter, we propose three new methods to improve the performance of the traditional sensor anomaly diagnosis and mitigation by filling the three research gaps. With respect to the research gaps, the contribution of this chapter is to propose three new methods that respectively

1. enable anomaly detection for some sensor anomalies during the observers' transient state;
2. detect some anomalies on critical sensors; and
3. potentially mitigate the impact of the anomalous sensor during the diagnosis process.

These three methods are then systematically integrated with a previously developed residual-based method to create an improved sensor anomaly diagnosis and mitigation framework. The first two contributions are outlined in Fig. I.1.

The rest of this chapter is organized as follows. In Section III.2, an overview of problem statement and solution is provided. In Section III.3, the mathematical description of the system is given. In section III.4, we introduce three new methods to address the research gaps, and the proposed methods are integrated with a previously developed method. In Section III.5, an illustrative example validates the proposed methods. The summary of this chapter is given in Section III.6.

III.2 Problem / Solution Overview

As mentioned in Section III.1, we model the CPS as a linear time-invariant discrete-time system. The problem addressed in this Chapter can be formulated as: given a linear time-invariant discrete-time system with multiple sensors, multiple observers, a state feedback controller, a residual-based anomaly detector, and the following assumption

Assumption 1. *Only one sensor is anomalous at a time.*

The specific goals of this chapter are to

- propose a non-residual based method for sensor anomaly detection during the observers' transient state (Contribution 1);
- propose a method for critical sensor anomaly diagnosis (Contribution 2);
- propose a method to potentially mitigate the impact of the anomalous sensor during the diagnosis process (Contribution 3); and
- systematically integrate the three new methods with a previously developed residual-based method for diagnosis and mitigation.

As described in Section III.1, the sensors can be divided into two sets: *critical* sensor set and *non-critical* sensor set. To diagnose anomalies in these two different sets of sensors, we use both closed-loop and open-loop observers. A closed-loop observer estimated the system internal state with a feedback from the sensor measurements. An open-loop observer is running in parallel with the physical system, reproducing the behavior of the system.

To detect anomalies in non-critical sensors, we design one closed-loop observer with all of the sensor measurements, and multiple closed-loop observers each with one non-critical sensor excluded. Each observer is compared with all other observers, and the difference of estimated states between two observers is decoupled to calculate the estimation errors of these two observers. Thus, each observer has multiple calculated estimation errors. These calculated estimation errors are combined to determine the overall estimation error of the observer. The convergence ratio of the estimation error of an observer should be related to the designed state matrix of the observer, and not affected by the uncertain initial condition. But a sensor anomaly or a disturbance can change the convergence ratio of the estimation error. Note that a disturbance is also a kind of anomaly. In this dissertation, we model the disturbance as the system disturbance, i.e., a disturbance added to the state equation. A disturbance should be distinguished from a sensor anomaly. To detect a sensor anomaly or a disturbance, we propose the Convergence Ratio (CR) method which can reduce the false alarms during the observers' transient state. To distinguish a sensor anomaly from a disturbance, bias analysis based on the calculated estimation errors is developed. In the ideal case, the biases calculated based on the estimation errors of all observers should be the same when the system is under disturbance, but should be different under sensor anomaly. With bounded system noise, the bound of the difference between the calculated bias and the actual disturbance signal can be determined. Therefore, a threshold can be selected and compared with the difference between any two calculated biases. The threshold is specific for each pair of biases. If at least one pair of them exceeds their threshold, the system is under sensor anomaly. If none of them exceeds their threshold, the system is under disturbance.

To diagnose anomalies in critical sensors, we design Multiple Open-Loop Observers (MOLO), and analyze the residuals formed based on the difference between the measured outputs of the system and the estimated outputs. This method is only applicable to an open-loop stable or marginally stable system. If the system is open-loop unstable, the estimation error of an open-loop observer could diverge exponentially. To increase the estimation accuracy, we periodically update the states of multiple open-loop observers with the state estimated by the closed-loop observer using all of the sensor measurements when no sensor anomaly is detected. There is a trade-off between estimation performance and the ability to detect a sensor anomaly. Therefore, we divide the multiple open-loop observers into several groups. The observers within the same group are updated with the same update frequency. To mitigate the impact of noise, the update time steps of the observers in the same group are distributed evenly within one update period, and the residuals generated by the observers within the same group are averaged. The averaged residual is compared with a threshold, which is related to the known upper bound of noise and the update frequency. If the residual is larger than the threshold, then an alarm is triggered and the states of the open-loop observers of that group are not updated with the estimated state of the closed-loop observer until the alarm is cleared. Logic is provided to determine whether or not the system is under sensor anomaly based on which groups of open-loop observers trigger alarms. Then the residuals of the groups that trigger alarms are analyzed to determine which sensor is anomalous.

For sensor anomaly mitigation, we also need to consider two cases: anomalies in critical sensors and anomalies in non-critical sensors. For anomalies in non-critical sensors, a closed-loop observer without the anomalous sensor provides a better state estimation, based on which a state feedback controller can give the control input closest to the ideal control input. Pinpointing this closed-loop observer during the diagnosis process is the key for sensor anomaly mitigation. Therefore, we propose the Calculated Control Input (CCI) method to switch among different observers, and potentially mitigate the impact of the anomaly in a non-critical sensor during the diagnosis process. For anomalies in critical sensors, none of the closed-loop observers can provide a good state estimation. If the system is open-loop stable, we can use an open-loop observer for state estimation to mitigate the impact of the sensor anomaly. If the system is marginally stable, the only way to mitigate the impact of sensor anomaly is to repair the anomalous sensor.

We also need a residual-based method based on closed-loop observers for non-critical sensor anomaly isolation. In this dissertation, we use a method adopted from [9], and call it the Calculated Outputs (CO) method. The method in [9] consists of several sliding mode observers, each excluding a particular sensor or actuator. The sliding mode observer without the anomalous sensor generates a significant residual signal. In contrast, we use a bank of Luenberger observers (or Kalman filters) [1] for the CO method. In this case, the observers with the anomalous sensor generate significant residuals. However, the CO method is not robust to disturbance in the system.

Table III.1 shows the abilities of the CO, CR, MOLO, and CCI methods. The CO method can detect and isolate non-critical sensor anomalies that occur during the observers' steady state. The CR method can detect non-critical sensor anomalies that occur during both the observers' transient and steady state. The MOLO method can detect and isolate critical sensor anomalies that occur during the observers' steady state. The CCI method can mitigate the impact of non-critical sensor anomalies during the sensor anomaly diagnosis process. Fig. III.1a shows when to use those four

[1] The reason we use the Luenberger observers (or Kalman filters) instead of sliding mode observers is that we can decouple observers' estimation errors for the CR method.

Table III.1: Abilities of the CO, CR, MOLO and CCI methods

	Anomaly Detection		Anomaly Isolation	Anomaly Mitigation
	Observers Transient State	Observers Steady State		
S_{nc}	CR	CR, CO	CO	CCI
S_c		MOLO	MOLO	

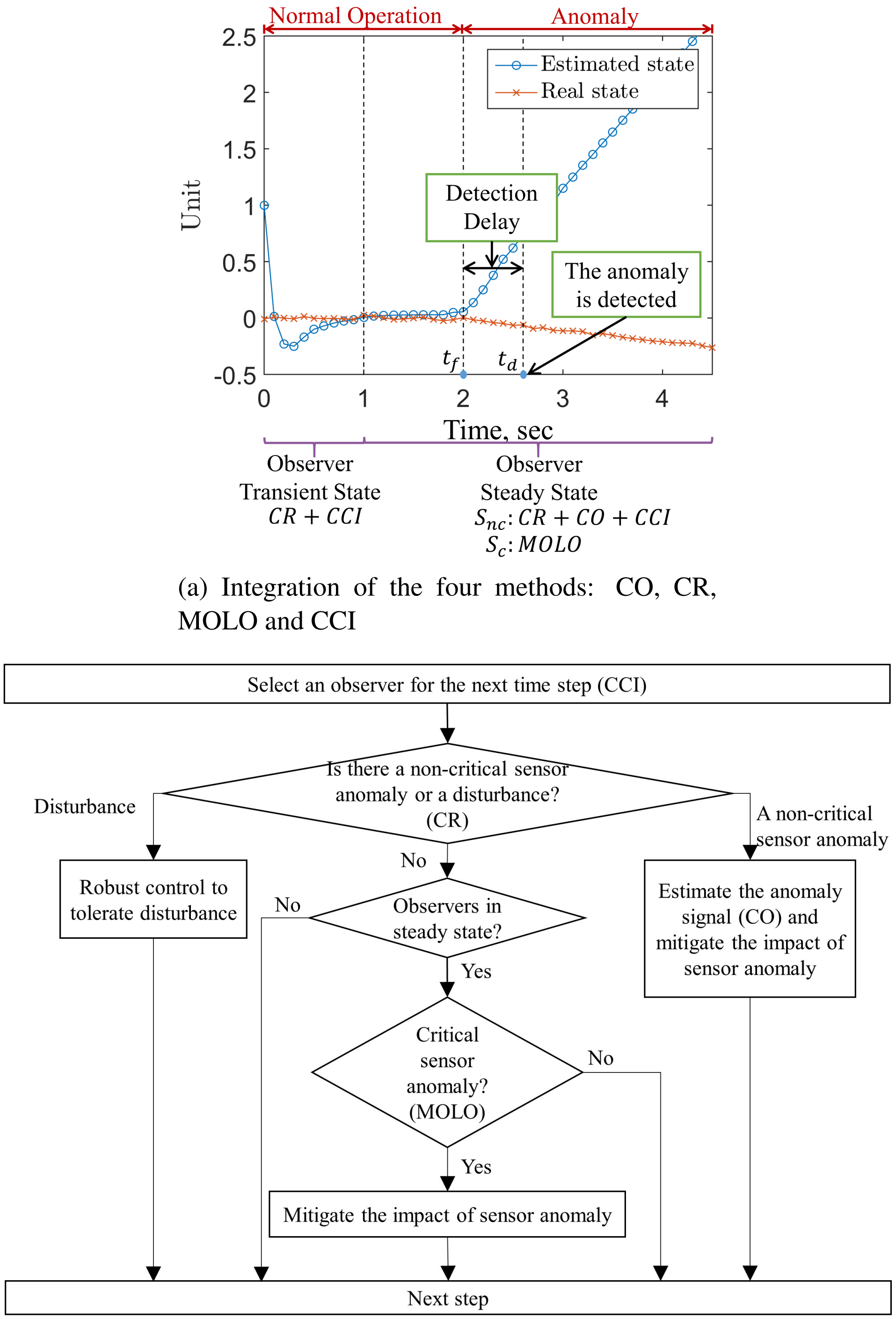
methods based on their abilities. We systematically integrate them as shown in Fig. III.1b. During the observers' transient state, we use the CR method for non-critical sensor anomaly detection. If a sensor anomaly is detected and the observers have already reached their steady state, then we use the CO method for sensor anomaly isolation. The CCI method is used for non-critical sensor anomaly mitigation during both the observers' transient state and steady state. Suppose an anomaly on a non-critical sensor starts at t_f, and it is detected and isolated at t_d. During the detection delay $t_d - t_f$, the CCI method may have already switched to the observer without the anomalous sensor, providing estimated state to the controller. The MOLO method is running in parallel with the CR, CO, and CCI methods to diagnose a critical sensor anomaly.

III.3 Mathematical Formulation of the Problem

The analysis is carried out based on a linear time-invariant discrete-time system equipped with multiple observers, a state feedback controller and a residual-based anomaly detector.

III.3.1 Notation

Let $||\cdot||$ denote ∞-norm, $\tilde{}$ denote estimated variables by a closed-loop observer, $\hat{}$ denote estimated variables by an open-loop observer. In addition, $\mathbf{x} \in \mathbb{R}^{n_x}$ represents a vector, where its i^{th} element is indicated by $\mathbf{x}^{(i)}$. $\mathbf{A} \in \mathbb{R}^{m\times n}$ represents a matrix, where its element at the i^{th} row and the j^{th} column is indicated by $\mathbf{A}^{(i,j)}$. $\mathbf{x_e}$ is the estimation error between system real state and the estimated state by an observer. $\mathbf{x}_{\mathbf{e},\mu,\nu}$ is the difference of estimated states between two closed-loop


Normal Operation
Anomaly
2.5
2
1.5
1
0.5
0
-0.5
Unit
0
1
2
3
4
Time, sec
Estimated state
Real state
Detection Delay
The anomaly is detected
t_f
t_d
Observer Transient State
$CR + CCI$
Observer Steady State
$S_{nc}: CR + CO + CCI$
$S_c: MOLO$
Select an observer for the next time step (CCI)
Is there a non-critical sensor anomaly or a disturbance? (CR)
Disturbance
A non-critical sensor anomaly
No
Robust control to tolerate disturbance
Observers in steady state?
No
Yes
Estimate the anomaly signal (CO) and mitigate the impact of sensor anomaly
Critical sensor anomaly? (MOLO)
No
Yes
Mitigate the impact of sensor anomaly
Next step

(a) Integration of the four methods: CO, CR, MOLO and CCI

(b) Flow chart of the integration

Figure III.1: The improved sensor anomaly diagnosis and mitigation framework description

observers μ and ν. $\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}$ is the calculated estimation error of closed-loop observer μ, and the calculation is based on $\mathbf{x}_{\mathbf{e},\mu,\nu}$. Detailed notations are shown in Appendix A.

III.3.2 Physical System

We model the physical system as a linear time-invariant discrete-time system. It has the following form:

$$\begin{aligned}\mathbf{x}(t+1) =& \mathbf{A}\mathbf{x}(t)+\mathbf{B}\mathbf{u}(t)+\mathbf{D}\mathbf{d}(t)+\mathbf{w}(t),\\ \mathbf{y}(t) =& \mathbf{C}\mathbf{x}(t)+\mathbf{v}(t)+\mathbf{\Gamma}\boldsymbol{\gamma}(t),\end{aligned} \tag{III.1}$$

where $\mathbf{x}(t) \in \mathbb{R}^{n_x}$ is the system state, $\mathbf{y}(t) \in \mathbb{R}^{n_y}$ is the sensor measurement, $\mathbf{u}(t) \in \mathbb{R}^{n_u}$ is the control input, $\mathbf{d}(t) \in \mathbb{R}^{n_d}$ is the unknown disturbance, $\boldsymbol{\gamma}(t) \in \mathbb{R}$ is the sensor anomaly signal added to the sensor measurements, the process noise $\mathbf{w}(t) \in \mathbb{R}^{n_x}$ and the sensor noise $\mathbf{v}(t) \in \mathbb{R}^{n_y}$ are zero mean random vectors with bounds $||\mathbf{w}(t)|| \leq w$ and $||\mathbf{v}(t)|| \leq v$, respectively, $\mathbf{A} \in \mathbb{R}^{n_x \times n_x}$, $\mathbf{B} \in \mathbb{R}^{n_x \times n_u}$, $\mathbf{C} \in \mathbb{R}^{n_y \times n_x}$, $\mathbf{D} \in \mathbb{R}^{n_x \times n_d}$ are real constant matrices, and $\mathbf{\Gamma} = [0 \quad \ldots \quad 1_{i_f} \quad \ldots \quad 0]^\intercal \in \mathbb{R}^{n_y}$ is a sensor anomaly vector, with 0 corresponding to the anomaly-free sensor, and 1_{i_f} corresponding to the anomalous sensor, and i_f is the index for the anomalous sensor. Based on Assumption 1, $\mathbf{\Gamma}$ has at most one non-zero element.

III.3.3 Closed-Loop Observers and Open-Loop Observers

At each time step, all of the sensor measurements $\mathbf{y}(t)$ and the control inputs $\mathbf{u}(t)$ are gathered for state estimation. Two different types of observers can be utilized: closed-loop observers and open-loop observers.

III.3.3.1 Closed-Loop Observers

A closed-loop observer corrects the estimation with a feedback from the sensor measurements as shown in Fig. III.2. Based on Assumption 1, sensor measurements can be divided into

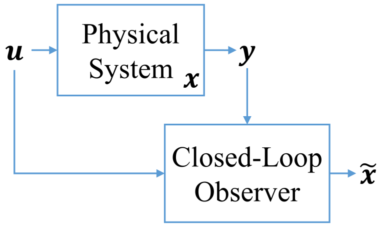


Figure III.2: Structure of a closed-loop observer

two sets: S_{nc} and S_c. S_{nc} contains m_{nc} non-critical sensors. S_c contains critical sensors. In order to design multiple closed-loop observers, we need the following assumption:

Assumption 2. *Set S_{nc} contains at least one non-critical sensor, i.e., $m_{nc} > 0$.*

We assume without loss of generality that the rows of the output matrix $\mathbf{C}$ are ordered such that the first m_{nc} sensors are non-critical sensors. Thus, $m_{nc}+1$ closed-loop observers can be designed. Observer 0 uses all of the sensor measurements. Observer i uses all but sensor i ($i = 1, 2, ..., m_{nc}$). For the closed-loop observers, we use Luenberger observers with the following form

$$\begin{aligned}\tilde{\mathbf{x}}_i(t+1) =&\mathbf{E}_i\tilde{\mathbf{x}}_i(t) + \mathbf{L}_i\mathbf{y}_i(t) + \mathbf{Bu}(t)\\ =&\mathbf{E}_i\tilde{\mathbf{x}}_i(t) + \mathbf{L}_i(\mathbf{C}_i\mathbf{x}(t) + \mathbf{v}_i(t) + \mathbf{\Gamma}_i\boldsymbol{\gamma}(t)) + \mathbf{Bu}(t),\end{aligned} \tag{III.2}$$

where $\tilde{\mathbf{x}}_i(t) \in \mathbb{R}^{n_x}$ is the state estimated by the closed-loop observer i ($i = 0, 1, 2, ..., m_{nc}$), $\mathbf{y}_i(t) \in \mathbb{R}^{n_y-1}$ is the sensor measurements used by observer i which does not contain the i^{th} element of $\mathbf{y}(t)$, $\mathbf{v}_i(t)$ does not contain the i^{th} sensor noise, $\mathbf{E}_i = \mathbf{A} - \mathbf{L}_i\mathbf{C}_i$, $\mathbf{L}_i \in \mathbb{R}^{n_x\times(n_y-1)}$ is the observer gain, placing the eigenvalues of $\mathbf{E}_i$ in the unit circle, $\mathbf{C}_i \in \mathbb{R}^{(n_y-1)\times n_x}$ is the output matrix for observer i and it does not contain the i^{th} row of $\mathbf{C}$, and $\mathbf{\Gamma}_i \in \mathbb{R}^{n_y-1}$ is the sensor anomaly vector of observer i which does not contain the i^{th} element of $\mathbf{\Gamma}$. If $i = i_f$, then $\mathbf{\Gamma}_i = \mathbf{0}^{n_y-1}$. This means that observer i_f does not use the anomalous sensor i_f for state estimation. The corresponding observer state matrix and observer gain that do not use the anomalous sensor are $\mathbf{E}_{i_f}$ and $\mathbf{L}_{i_f}$, respectively.

Remark 1. *Our assumption indicates that the system is detectable without one of the sensors in S_{nc}. If the system is detectable and the noise is truncated Gaussian, the time varying gain of a*

Kalman filter converges in a few steps. Therefore, for the closed-loop observers, we can also use Kalman filters with the steady-state Kalman gains [61].

III.3.3.2 Open-Loop Observers

An open-loop observer is running in parallel with the physical system, reproducing the behavior of the system as shown in Fig. III.3. Due to the lack of guaranteed estimation error convergence, the state of the open-loop observer is updated periodically by the closed-loop observer 0 which uses all of the sensor measurements. As mentioned in Section III.2, we design M groups of open-loop observers, each group with N observers. The observers in the same group have the same update period. Then, an open-loop observer has the following form after one update period

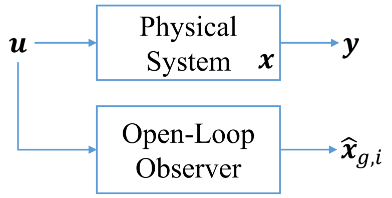


Figure III.3: Structure of an open-loop observer

$$\hat{\mathbf{x}}_{g,i}(t+\kappa_{f,g}) = \mathbf{A}^{\kappa_{f,g}}\tilde{\mathbf{x}}_0(t) + \Sigma_{j=0}^{\kappa_{f,g}-1}\mathbf{A}^j\mathbf{B}\mathbf{u}(t+\kappa_{f,g}-1-j), \tag{III.3}$$

where $\hat{\mathbf{x}}_{g,i}(t) \in \mathbb{R}^{n_x}$ is the state estimated by the open-loop observer i in group g ($i = 1,...,N$, $g = 1,...,M$), and $\kappa_{f,g}$ is the update period of group g, and $\tilde{\mathbf{x}}_0$ is the estimated state by closed-loop observer 0.

III.3.4 State Feedback Controller

A state feedback controller calculates a control command based on the system state, and applies it to the input of the system. The following assumption enables the utilization of a state feedback controller

Assumption 3. *The system is controllable.*

Since the real state of the system is unknown, the controller can only use the state estimated by a closed-loop observer with the following form [67]

$$\mathbf{u}(t) = \mathbf{F}\tilde{\mathbf{x}}_i(t), \tag{III.4}$$

where $\mathbf{F} \in \mathbb{R}^{n_u \times n_x}$ is the controller gain placing the eigenvalues of $\mathbf{A} + \mathbf{BF}$ in the unit circle. Notice that an open-loop observer cannot provide as good of an estimation of performance as a closed-loop observer due to system noise. Therefore, we use a closed-loop observer for the state feedback controller if the system is under normal operation or under non-critical sensor anomaly. If an open-loop stable system is under critical sensor anomaly, then we can switch to an open-loop observer to help mitigate the impact of the sensor anomaly.

III.3.5 Residual-based Anomaly Detector

In this chapter, the residual-based anomaly detector uses the CO method, which is adopted from [9]. In contrast to the method in [9], the CO method consists of multiple Luenberger observers as shown in (III.2), and generates the residuals based on the subtraction between the sensor measurements $\mathbf{y}_i$ (without the i^{th} output) and the estimated outputs $\mathbf{C}_i\tilde{\mathbf{x}}_i$ as shown in (III.5)

$$\mathbf{r}_i(t) =(\mathbf{y}_i(t) - \mathbf{C}_i\tilde{\mathbf{x}}_i(t))^{\mathsf{T}}\mathbf{Q}_i(\mathbf{y}_i(t) - \mathbf{C}_i\tilde{\mathbf{x}}_i(t)), \tag{III.5}$$

where $\mathbf{Q}_i$ is a real constant weighting matrix for observer i[2], and $\mathbf{r}_i(t) \in \mathbb{R}$ is the residual generated based on observer i.

The residual generated based on observer 0 is compared with a selected threshold θ_{CO} to determine the occurrence of an anomaly. Note that the CO method cannot distinguish a disturbance from a sensor anomaly since Luenberger observer is not robust to disturbances. So the illustration of the CO method is under the assumption that the anomaly in the system is a sensor anomaly. When a sensor anomaly occurs, the closed-loop observer i_f, which does not use the anomalous sensor, is not affected by the sensor anomaly, and thus provides a better state estimation compared to other observers[3]. Then the residual generated by observer i_f is smaller than the residuals generated by other observers ($i \neq i_f$). Therefore, we can locate the anomalous sensor by finding the smallest residual among the observers from 1 to m_{nc}. After the anomalous sensor is located, the sensor anomaly vector $\mathbf{\Gamma}$ is known and the estimated sensor anomaly signal is given by

$$\tilde{\boldsymbol{\gamma}}(t) = \mathbf{\Gamma}^{\mathsf{T}}(\mathbf{y}(t) - \mathbf{C}\tilde{\mathbf{x}}_{i_f}(t)), \tag{III.6}$$

where $\tilde{\boldsymbol{\gamma}}(t) \in \mathbb{R}$ is the estimated sensor anomaly signal.

Algorithm 1 gives the procedure of the CO method. First, we calculate the residuals based on different observers. Then, we use the residual of observer 0 for anomaly detection, and compare the rest of the residuals for sensor anomaly isolation. The issue that the CO method cannot distinguish a disturbance from a sensor anomaly is addressed by complementing the CO method with the CR method introduced in Section III.4.3 which has the ability to distinguish a disturbance from a sensor anomaly.

[2] $\mathbf{Q}_i$ should be designed to make the element $\mathbf{y}_i^{(j)}(t) - \mathbf{C}_i^{(j,:)}\tilde{\mathbf{x}}_i(t)(j \in S_c)$, where j corresponds to the critical sensors, have larger weighting ratios than the element corresponding to non-critical sensors.

[3] The demonstration is shown in Appendix VI.2.

Algorithm 1: CO method for sensor anomaly diagnosis

```
function CO;
Input  : $\mathbf{y}(t), \tilde{\mathbf{x}}_i(t)$ $(i = 0, 1, ..., m_{nc})$
Output: $I_F, i_f$
//Residual generation for all observers;
for $i = 0$ to $m_{nc}$ do
    $\mathbf{r}_i(t) = (\mathbf{y}_i(t) - \mathbf{C}_i\tilde{\mathbf{x}}_i(t))^\mathsf{T}\mathbf{Q}_i(\mathbf{y}_i(t) - \mathbf{C}_i\tilde{\mathbf{x}}_i(t))$;
end
//Anomaly detection;
if $\mathbf{r}_0(t) \geq \theta_{CO}$ then
    $I_F = 1$;
    //Sensor anomaly isolation;
    $i_f = \min_i \mathbf{r}_i(t)$;
    $I_{FB} = i_f$;
    $\mathbf{\Gamma} = [0 \quad ... \quad 1_{i_f} \quad ... \quad 0]^\mathsf{T}$;
    $\tilde{\boldsymbol{\gamma}}(t) = \mathbf{\Gamma}^\mathsf{T}(\mathbf{y}(t) - \mathbf{C}\tilde{\mathbf{x}}_{i_f}(t))$;
else
    $I_{FB} = 0$;
end
```

III.4 Framework Components Description and Integration

Throughout this section, a simple system of a moving object is utilized as an illustration. First, we simulate sensor anomalies in the moving object system equipped with the CO method-based anomaly detector to understand its limitations. Then, three new methods are introduced and analyzed in the deterministic case (noise free). The impact of random system noise is discussed for each method thereafter. The simulation result shows the improvements of the proposed methods compared to the CO method. Finally, we provide an algorithm to integrate the CO method and the three new methods.

III.4.1 Moving Object System

The moving object system is a $1kg$ mass moving along a horizontal line. Two sensors are measuring the two outputs: the velocity y_v and the position y_p, respectively. A state feedback controller applies a horizontal force on the mass. The sampling time is $0.1s$. The system has an initial state $(0,0)$, the process noise with bound 0.001 (m/s or m), and the sensor noise with bound 0.01 (m/s or m). The initial states of the observers are chosen as $(1,0.5)$[4]. The state space representation of the moving object system is shown as

$$\begin{aligned} \mathbf{x}(t+1) =& \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{w}(t), \\ \mathbf{y}(t) =& \mathbf{C}\mathbf{x}(t+1) + \mathbf{v}(t), \end{aligned} \tag{III.7}$$

where $\mathbf{x} = \begin{bmatrix} x_v \\ x_p \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_v \\ y_p \end{bmatrix}$, $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0.1 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0.1 \\ 0.005 \end{bmatrix}$, and $\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

By checking the rank of observability matrix, the moving object system is observable with y_v and y_p or only y_p, but unobservable with only y_v. Therefore, $y_p \in S_c$, and $y_v \in S_{nc}$. Two observers can be designed with observer poles placed at [0.1 0.11]. Observer 0 uses both sensor measurements y_v and y_p. Observer 1 uses only y_p.

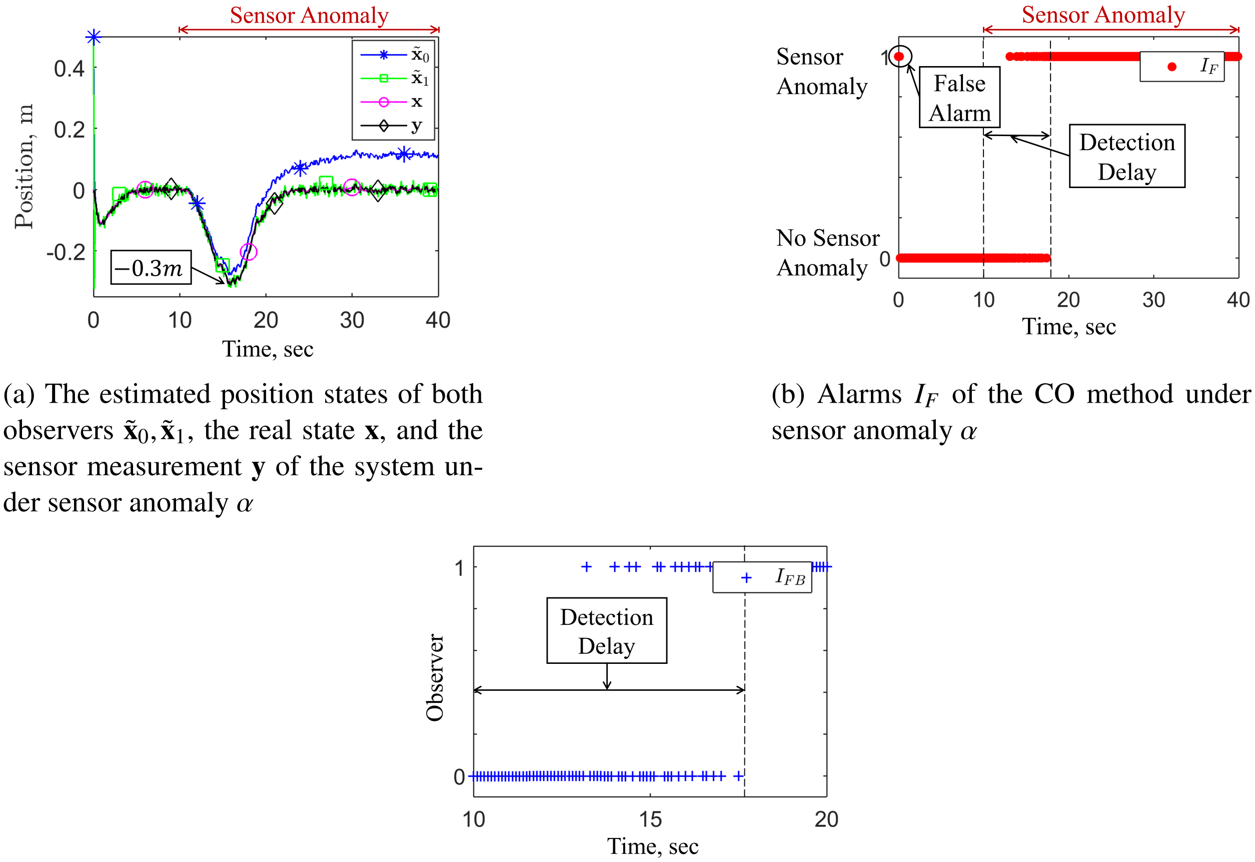
Two sensor anomaly scenarios are considered:

1. anomaly α: a ramp signal with slope $0.05m/s^2$ ($0.005m/s$ per time step) added to the velocity sensor y_v, saturating at $1m/s$; and

2. anomaly β: a ramp signal with slope $0.001m/s$ ($0.0001m$ per time step) added to the position sensor y_p, saturating at $1m$.

Both anomalies start at $10s$ and run until the end of the simulation. Here we consider ramp anomalies with small slopes because they are hard to detect compared to ramp anomalies with large slopes or step anomalies with large magnitudes.

[4] The initial estimation errors of the observers are large to help us understand the limitations of a residual-based method using closed-loop observers during the observers' transient state



(a) The estimated position states of both observers $\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_1$, the real state $\mathbf{x}$, and the sensor measurement $\mathbf{y}$ of the system under sensor anomaly α

(b) Alarms I_F of the CO method under sensor anomaly α

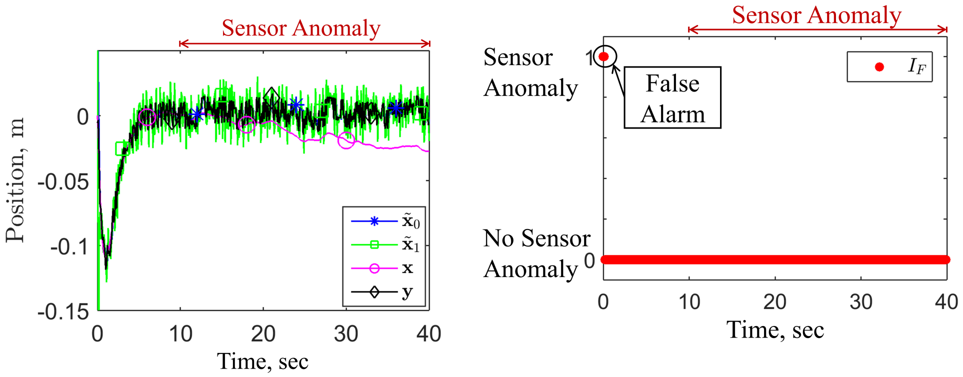
(c) Observer index I_{FB} selected for the state feedback controller under sensor anomaly α

Figure III.4: Moving object system under sensor anomaly α

III.4.2 The Impact of Sensor Anomalies

Two sensor anomaly scenarios are run on the moving object system equipped with the CO method-based anomaly detector to show the limitations of the CO method-based anomaly detector. Based on each limitation, a new method is discussed and proposed.

Fig. III.4-III.6 show the estimated position states of both observers $\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_1$, the real state $\mathbf{x}$, and the sensor measurement $\mathbf{y}$ of the system equipped with the CO method-based anomaly detector under sensor anomaly α, β and normal operation, respectively. In both Fig. III.4 and Fig. III.5, false alarms are generated by the CO method during the observers' transient state, which are about



(a) The estimated position states of both observers $\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_1$, the real state $\mathbf{x}$, and the sensor measurement $\mathbf{y}$ of the system under sensor anomaly β

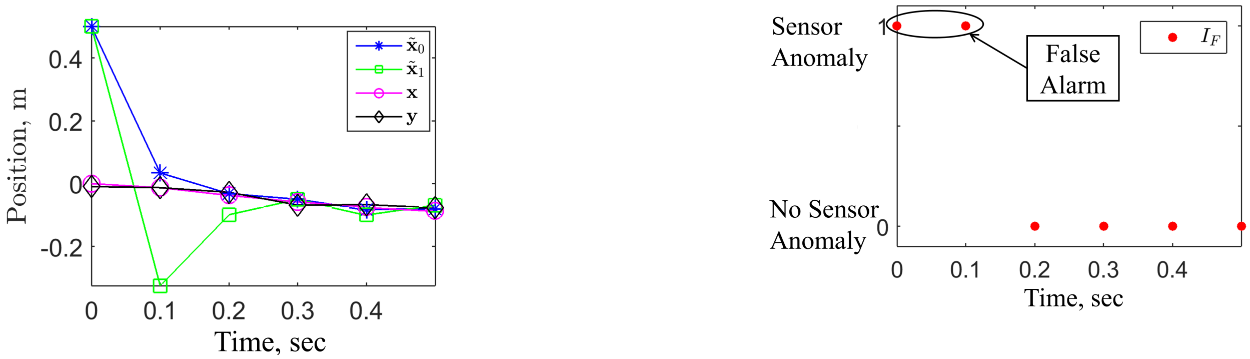
(b) Alarms I_F of the CO method under sensor anomaly β

Figure III.5: Moving object system under sensor anomaly β

0.2s, when the system is actually under normal operation. From Fig. III.6, it can be seen that the imperfect initial state of the observers causes the CO method to generate false alarms. According to (III.5), the residual $\mathbf{r}_i(t)$ of the CO method is a function of the observer's estimation error under normal operation. A large estimation error makes the residual exceed the threshold, causing false alarms during the observers' transient state. To enable sensor anomaly detection during observers' transient state, the CR method, described in Section III.4.3, which utilizes the convergence ratio of observers' estimation error, will be applied.

As shown in Fig. III.5b, when the system is under sensor anomaly β, no alarm is generated since the sensor anomaly is not detected by the CO method-based anomaly detector. The reason behind this behavior is that the system is not detectable when the position sensor y_p is removed, and the sensor anomaly signal is changing slightly at each time step to avoid significant change in the residuals. An open-loop observer (III.3) does not use any sensor for state estimation. Thus, this issue can be potentially addressed by the MOLO method introduced in Section III.4.4.

As shown in Fig. III.4b and Fig. III.4c, although the CO method successfully locates the anomalous sensor and then the system switches to observer 1 for state estimation after 18s, there is an 8s detection delay and the system switches between the two observers during 13s to 18s.



(a) The estimated position states of both observers $\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_1$, the real state $\mathbf{x}$, and the sensor measurement $\mathbf{y}$ of the system during the observers' transient state under normal operation

(b) Alarms I_F during the observers' transient state

Figure III.6: Moving object system under normal operation

This is caused by the relatively small sensor anomaly signal compared to the system noise and the threshold. Thus, the anomalous sensor cannot be located immediately. This detection delay makes the maximum absolute value of the position of the mass reach 30*cm* as shown in Fig. III.4a. The direct reason for this divergence is the discrepancy of the control input provided by the observer-based state feedback controller. To address this issue, we need to switch to the closed-loop observer without the anomalous sensor as soon as possible and continue using that observer during the sensor anomaly diagnosis process. Thus, we propose the CCI method to compare the control input calculated based on the state estimated by each closed-loop observer with an "ideal" control input calculated based on the state estimated by an open-loop observer, and to switch to the observer which gives the smallest difference between the calculated control input and the ideal control input. This method has the potential to mitigate the impact of a non-critical anomalous sensor during the sensor anomaly diagnosis process. The maximum absolute value of the position of the system under the CO method will be compared with that under the CCI method in Section III.4.5.

III.4.3 CR Method for Sensor Anomaly Detection during Transient and Steady State

This method is proposed to detect the occurrence of an anomaly based on the convergence of estimation error. It enables sensor anomaly detection during the observers' transient state. To achieve robust anomaly detection, a disturbance in the system is distinguished from a sensor anomaly by analyzing the bias of the estimation error. First, this method is introduced on an ideal control system. Then the impact of the process noise and the sensor noise are discussed.

III.4.3.1 Ideal System Case

Three different situations are considered for this method: normal operation, disturbance, and sensor anomaly. Estimation error $\mathbf{x}_{\mathbf{e},i}$ of closed-loop observer i, and the difference of estimated states $\mathbf{x}_{\mathbf{e},\mu,\nu}$ between two closed-loop observers μ and ν under three situations are shown in (III.8) through (III.13).

Under normal operation

$$\mathbf{x}_{\mathbf{e},i}(t+1) = \mathbf{x}(t) - \tilde{\mathbf{x}}_i(t) = \mathbf{E}_i\mathbf{x}_{\mathbf{e},i}(t), \tag{III.8}$$

$$\mathbf{x}_{\mathbf{e},\mu,\nu}(t+1) = \tilde{\mathbf{x}}_\mu(t+1) - \tilde{\mathbf{x}}_\nu(t+1) = \mathbf{E}_\nu\mathbf{x}_{\mathbf{e},\nu}(t) - \mathbf{E}_\mu\mathbf{x}_{\mathbf{e},\mu}(t). \tag{III.9}$$

Under disturbance

$$\mathbf{x}_{\mathbf{e},i}(t+1) = \mathbf{E}_i\mathbf{x}_{\mathbf{e},i}(t) + \mathbf{D}\mathbf{d}(t), \tag{III.10}$$

$$\mathbf{x}_{\mathbf{e},\mu,\nu}(t+1) = \mathbf{E}_\nu\mathbf{x}_{\mathbf{e},\nu}(t) - \mathbf{E}_\mu\mathbf{x}_{\mathbf{e},\mu}(t). \tag{III.11}$$

Notice that (III.9) and (III.11) are the same.

Under sensor anomaly

$$\mathbf{x}_{\mathbf{e},i}(t+1) = \mathbf{E}_i \mathbf{x}_{\mathbf{e},i}(t) - \mathbf{L}_i \mathbf{\Gamma}_i \boldsymbol{\gamma}(t), \tag{III.12}$$

$$\mathbf{x}_{\mathbf{e},\mu,\nu}(t+1) = \mathbf{E}_\nu \mathbf{x}_{\mathbf{e},\nu}(t) - \mathbf{E}_\mu \mathbf{x}_{\mathbf{e},\mu}(t) - (\mathbf{L}_\nu \mathbf{\Gamma}_\nu - \mathbf{L}_\mu \mathbf{\Gamma}_\mu)\boldsymbol{\gamma}(t). \tag{III.13}$$

The first step of the CR method is to calculate the estimation error of each closed-loop observer. The dynamics of $\mathbf{x}_{\mathbf{e},\mu,\nu}$ under both normal operation and disturbance are the evolution of the estimation errors of the two closed-loop observers $\mathbf{x}_{\mathbf{e},\mu}$ and $\mathbf{x}_{\mathbf{e},\nu}$. Therefore, the estimation errors of both observers can be decoupled over two time steps. However, the dynamics of $\mathbf{x}_{\mathbf{e},\mu,\nu}$ under sensor anomaly involves two unknown sensor anomaly vectors $\mathbf{\Gamma}_\mu$ and $\mathbf{\Gamma}_\nu$, and the unknown sensor anomaly signal $\boldsymbol{\gamma}(t)$. Thus, the estimation errors cannot be correctly decoupled under sensor anomaly. Lemma 1 gives the formulas for estimation error decoupling of any two different observers.

Lemma 1. *Given an ideal control system* (III.1) *with* $\mathbf{w}(t) = \mathbf{0}$ *and* $\mathbf{v}(t) = \mathbf{0}$*, the calculated estimation error* $\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}$ *and* $\tilde{\mathbf{x}}_{\mathbf{e},\nu(\mu)}$ *are derived based on* (III.14) *and* (III.15)*, respectively, with the following results:*

1. *When the system is under normal operation or under disturbance,* $\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)} = \mathbf{x}_{\mathbf{e},\mu}$ *and* $\tilde{\mathbf{x}}_{\mathbf{e},\nu(\mu)} = \mathbf{x}_{\mathbf{e},\nu}$*;*
2. *When the system is under sensor anomaly,* $\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)} \neq \mathbf{x}_{\mathbf{e},\mu}$ *and* $\tilde{\mathbf{x}}_{\mathbf{e},\nu(\mu)} \neq \mathbf{x}_{\mathbf{e},\nu}$ *if* $\mathbf{L}_\nu \mathbf{\Gamma}_\nu \neq \mathbf{L}_\mu \mathbf{\Gamma}_\mu$.

$$\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t) = (\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}(\mathbf{x}_{\mathbf{e},\mu,\nu}(t+1) - \mathbf{E}_\nu \mathbf{x}_{\mathbf{e},\mu,\nu}(t)), \tag{III.14}$$

$$\tilde{\mathbf{x}}_{\mathbf{e},\nu(\mu)}(t) = (\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}(\mathbf{x}_{\mathbf{e},\mu,\nu}(t+1) - \mathbf{E}_\mu \mathbf{x}_{\mathbf{e},\mu,\nu}(t)), \tag{III.15}$$

where $\mathbf{E}_\nu - \mathbf{E}_\mu = \mathbf{A} - \mathbf{L}_\nu \mathbf{C}_\nu - \mathbf{A} + \mathbf{L}_\mu \mathbf{C}_\mu = \mathbf{L}_\mu \mathbf{C}_\mu - \mathbf{L}_\nu \mathbf{C}_\nu$.

Remark 2. *We design* $\mathbf{L}_\mu$ *and* $\mathbf{L}_\nu$ *to make* $\mathbf{E}_\nu - \mathbf{E}_\mu$ *invertible.*

Proof. 1) Under normal operation or under disturbance, the evolution of $\mathbf{x}_{\mathbf{e},\mu,\nu}$ (III.9) and $\mathbf{x}_{\mathbf{e},\mu,\nu}(t) =$

$\mathbf{x}_{\mathbf{e},\nu}(t) - \mathbf{x}_{\mathbf{e},\mu}(t)$ are substituted to (III.14),

$$\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t) = (\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}(\mathbf{E}_\nu \mathbf{x}_{\mathbf{e},\nu}(t) - \mathbf{E}_\mu \mathbf{x}_{\mathbf{e},\mu}(t) - \mathbf{E}_\nu(\mathbf{x}_{\mathbf{e},\nu}(t) - \mathbf{x}_{\mathbf{e},\mu}(t))) = \mathbf{x}_{\mathbf{e},\mu}(t). \tag{III.16}$$

Similarly, $\tilde{\mathbf{x}}_{\mathbf{e},\nu(\mu)}(t) = \mathbf{x}_{\mathbf{e},\nu}(t)$.

2) Under sensor anomaly, the evolution of $\mathbf{x}_{\mathbf{e},\mu,\nu}$ (III.13) and $\mathbf{x}_{\mathbf{e},\mu,\nu}(t) = \mathbf{x}_{\mathbf{e},\nu}(t) - \mathbf{x}_{\mathbf{e},\mu}(t)$ are substituted to (III.14),

$$\begin{aligned}\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t) =& (\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}(\mathbf{E}_\nu \mathbf{x}_{\mathbf{e},\nu}(t) - \mathbf{E}_\mu \mathbf{x}_{\mathbf{e},\mu}(t) - (\mathbf{L}_\nu \mathbf{\Gamma}_\nu - \mathbf{L}_\mu \mathbf{\Gamma}_\mu)\boldsymbol{\gamma}(t) - \mathbf{E}_\nu(\mathbf{x}_{\mathbf{e},\nu}(t) - \mathbf{x}_{\mathbf{e},\mu}(t))) \\ =& \mathbf{x}_{\mathbf{e},\mu}(t) - (\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}(\mathbf{L}_\nu \mathbf{\Gamma}_\nu - \mathbf{L}_\mu \mathbf{\Gamma}_\mu)\boldsymbol{\gamma}(t),\end{aligned} \tag{III.17}$$

if $\mathbf{L}_\nu \mathbf{\Gamma}_\nu \neq \mathbf{L}_\mu \mathbf{\Gamma}_\mu$, then $\tilde{\mathbf{x}}_{\mathbf{e},\mu}(t) \neq \mathbf{x}_{\mathbf{e},\mu}(t)$.

Similarly, $\tilde{\mathbf{x}}_{\mathbf{e},\nu(\mu)}(t) \neq \mathbf{x}_{\mathbf{e},\nu}(t)$ if $\mathbf{L}_\nu \mathbf{\Gamma}_\nu \neq \mathbf{L}_\mu \mathbf{\Gamma}_\mu$. □

Based on Lemma 1, m_{nc} estimation errors can be calculated for each observer. In ideal system case, these m_{nc} estimation errors are averaged to be the estimation error $\tilde{\mathbf{x}}_{\mathbf{e},i}$ of each observer. The combination of m_{nc} estimation errors for a noisy system is introduced in Section III.4.3.2.

After getting the estimation errors of all of the observers, the next step is to analyze the convergence behavior of the estimation error of each observer. For each observer, $\tilde{\mathbf{x}}_{\mathbf{e},i} \in \mathbb{R}^{n_x}$ contains n_x states. The evolution matrix $\mathbf{E}_i$ of the estimation error of observer i may not be a diagonal matrix. This causes the coupling of estimation errors between different states, which makes the ratio of estimation error of each state non-constant. Therefore, instead of using the estimation errors directly, we diagonalize the evolution matrix $\mathbf{E}_i$ using a basis of eigenvectors $\mathbf{V}_i$. The diagonal elements in the diagonalized matrix $\mathbf{E}_{\Lambda,i}$ (eigenvalues of $\mathbf{E}_i$), where $\mathbf{E}_{\Lambda,i} = (\mathbf{V}_i)^{-1}\mathbf{E}_i\mathbf{V}_i$, are the same as the time-invariant observer poles. Then, we can define the convergence ratio to specify the convergence of the estimation error for each state.

Definition 7. *Convergence ratio is the ratio of the absolute value of estimation error along with* *time step t* (III.18) *is called the convergence ratio.*

$$cr_{i,j}(t) = \frac{1}{\kappa_{CR}} \left[\left| \frac{\tilde{\mathbf{x}}^{(j)}_{\mathbf{e},\Lambda,i}(t)}{\tilde{\mathbf{x}}^{(j)}_{\mathbf{e},\Lambda,i}(t-1)} \right| + \sum_{k_i=2}^{\kappa_{CR}} \sqrt[k_i]{\left| \frac{\tilde{\mathbf{x}}^{(j)}_{\mathbf{e},\Lambda,i}(t)}{\tilde{\mathbf{x}}^{(j)}_{\mathbf{e},\Lambda,i}(t-k_i)} \right|} \right], \tag{III.18}$$

where $\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,i}(t) = (\mathbf{V}_i)^{-1}\tilde{\mathbf{x}}_{\mathbf{e},i}(t)$*,* $\tilde{\mathbf{x}}^{(j)}_{\mathbf{e},\Lambda,i}(t)$ *is the* j^{th} *element in* $\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,i}(t)$*, and* κ_{CR} *is a selected integer to average the convergence ratios over* κ_{CR} *time steps.*

Based on the above definition, the convergence ratio of each estimation error $cr_{i,j}$ is actually the same as the corresponding j^{th} observer pole under normal operation. This is also indicated by

$$\left| \frac{\tilde{\mathbf{x}}^{(j)}_{\mathbf{e},\Lambda,i}(t)}{\tilde{\mathbf{x}}^{(j)}_{\mathbf{e},\Lambda,i}(t-k_i)} \right| = \left| \frac{(\mathbf{E}^{j,j}_{\Lambda,i})^{k_i}\tilde{\mathbf{x}}^{(j)}_{\mathbf{e},\Lambda,i}(t-k_i)}{\tilde{\mathbf{x}}^{(j)}_{\mathbf{e},\Lambda,i}(t-k_i)} \right| = \left| (\mathbf{E}^{(j,j)}_{\Lambda,i})^{k_i} \right|, \tag{III.19}$$

where $\mathbf{E}^{(j,j)}_{\Lambda,i}$ is the j^{th} diagonal element of matrix $\mathbf{E}_{\Lambda,i}$Therefore,

$$cr_{i,j}(t) = \left| \mathbf{E}^{(j,j)}_{\Lambda,i} \right|, \forall t \geq 0. \tag{III.20}$$

An anomaly (a sensor anomaly or a disturbance) can change the convergence ratio of the estimation error in two possible cases. One case is that an anomaly makes the estimation error converge faster to zero. The other case is that an anomaly makes the estimation error converge slower or diverge to some other non-zero value. In ideal system case, the anomalies in both cases can be detected by comparing the convergence ratios with observer poles. If a convergence ratio is larger or smaller than its corresponding observer pole, then this convergence ratio indicates the occurrence of an anomaly. Definition 7 shows that $(m_{nc} + 1) \times n_x$ convergence ratios are calculated at each time step. Because of the system noise, it is possible that some of the convergence ratios indicate an anomaly even though there is no anomaly. So we define the system as an anomalous system if as least half of the convergence ratios indicate anomaly. A threshold is selected for noisy system as discussed in Section III.4.3.2.

To achieve robust anomaly detection, a disturbance should be distinguished from a sensor anomaly [41]. For this purpose, bias is defined

Definition 8. *The term* $\mathbf{b}(t)$ *in an affine function* $\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}(t)$ *is called* *bias.*

Under disturbance, the bias is $\mathbf{D}\mathbf{d}(t)$, which is the same for all observers. Under sensor anomaly, the bias is $-\mathbf{L}_i\mathbf{\Gamma}_i\boldsymbol{\gamma}(t)$, which is different for different observers. The disturbance signal $\mathbf{d}(t)$ can be correctly determined when the system is under disturbance because of the correct decoupled estimation error. In contrast, the sensor anomaly signal cannot be correctly determined because of the incorrect decoupled estimation error and unknown $\mathbf{\Gamma}_i$. Based on this analysis, the bias is calculated based on each observer according to (III.22) in Theorem 1.

Theorem 1. *Given an ideal control system* (III.1) *with* $\mathbf{w}(t) = \mathbf{0}$ *and* $\mathbf{v}(t) = \mathbf{0}$*, the biases* $\tilde{\mathbf{d}}_{\mu(\nu)}(t)$ *and* $\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t)$ *are calculated according to* (III.21) *and* (III.22) *respectively, with the following results:*

1. *When the system is under disturbance,*

$$\begin{aligned}\forall \mu, \nu =& 0, 1, ..., m_{nc} \wedge \mu \neq \nu,\\ \tilde{\mathbf{d}}_{\mu(\nu)}(t) =& \tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) = \mathbf{d}(t).\end{aligned}$$

2. *When the system is under sensor anomaly,*

$$\begin{aligned}\forall \mu, \nu =& 0, 1, ..., m_{nc} \wedge \mu \neq \nu,\\ \tilde{\mathbf{d}}_{\mu(\nu)}(t) =& \tilde{\mathbf{d}}_{\nu(\mu)}(t),\\ \tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) \neq& \tilde{\mathbf{d}}_{\Lambda,\nu(\mu)}(t) \quad if \quad \mathbf{V}_\mu \neq V_\nu.\end{aligned}$$

$$\tilde{\mathbf{d}}_{\mu(\nu)}(t) = (\mathbf{D}^\intercal\mathbf{D})^{-1}\mathbf{D}^\intercal[\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t+1) - \mathbf{E}_\mu\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t)], \tag{III.21}$$

$$\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) = ((\mathbf{D}_{\Lambda,\mu})^\intercal\mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^\intercal[\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu(\nu)}(t+1) - \mathbf{E}_{\Lambda,\mu}\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu(\nu)}(t)], \tag{III.22}$$

where $\mathbf{D}_{\Lambda,\mu} = (\mathbf{V}_\mu)^{-1}\mathbf{D}$, *and* $\mathbf{E}_{\Lambda,\mu} = (\mathbf{V}_\mu)^{-1}\mathbf{E}_\mu\mathbf{V}_\mu$.

For the proof of Theorem 1, see Appendix VI.2.

Theorem 1 shows that $(m_{nc}+1)\times m_{nc}$ biases are calculated at each time step. Each bias is compared with other biases. If any two biases disagree with each other, then the system is under sensor anomaly. If all of the biases agree with each other, indicating that the system is under disturbance, then we can determine the disturbance signal by averaging all of the biases. The combination of all of the biases for a noisy system is introduced in Section III.4.3.2.

III.4.3.2 Noisy System Case

Lemma 1 and Theorem 1 in Section III.4.3.1 show the effectiveness of the CR method in sensor anomaly detection when the system is ideal. In practice, we also need to consider system noise: process noise and sensor noise. When only process noise exists in the system, the output of the system can still be correctly measured, which means the state of the system can be exactly known. Therefore, process noise does not affect the accuracy of the estimation error calculation. However, when sensor noise contaminates the sensor measurements, the estimation error cannot be correctly calculated. The boundedness of sensor noise ensures the boundedness of the error of estimation error $||\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)} - \mathbf{x}_{\mathbf{e},\mu}||$. Lemma 2 and Lemma 3 give the impact of process noise and the impact of sensor noise on the estimation error calculation, respectively.

Lemma 2. *Given a control system* (III.1) *with bounded process noise and* $\mathbf{v}(t) = \mathbf{0}$, $\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t) = \mathbf{x}_{\mathbf{e},\mu}(t)$ *still holds when the system is under normal operation or under disturbance.*

Proof. When the system is subject to the process noise $\mathbf{w}(t)$, the estimation error evolution becomes:

$$\mathbf{x}_{\mathbf{e},\mu}(t+1) = \mathbf{E}_\mu \mathbf{x}_{\mathbf{e},\mu}(t) + \mathbf{w}(t). \tag{III.23}$$

Then the difference of the estimated states between two observers μ and ν is the same as (III.9).

By substituting (III.9) to (III.14), the calculated estimation error becomes

$$\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t) =(\mathbf{E}_{\nu} - \mathbf{E}_{\mu})^{-1}[\mathbf{E}_{\nu}\mathbf{x}_{\mathbf{e},\nu}(t) - \mathbf{E}_{\mu}\mathbf{x}_{\mathbf{e},\mu}(t) - \mathbf{E}_{\nu}(\mathbf{x}_{\mathbf{e},\nu}(t) - \mathbf{x}_{\mathbf{e},\mu}(t))] = \mathbf{x}_{\mathbf{e},\mu}(t). \tag{III.24}$$

□

Lemma 3. *Given a control system* (III.1) *with bounded sensor noise and* $\mathbf{w}(t) = \mathbf{0}$*,* $||\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t) - \mathbf{x}_{\mathbf{e},\mu}(t)||$ *is bounded by* $||(\mathbf{E}_{\nu} - \mathbf{E}_{\mu})^{-1}||(||\mathbf{L}_{\nu}|| + ||\mathbf{L}_{\mu}||)v$*.*

For the proof of Lemma 2, see Appendix VI.2.

Lemma 3 shows that the impact of sensor noise is different for estimation errors calculated based on different pairs of observers. Thus, when calculating the estimation error of each observer, we combine its m_{nc} decoupled estimation errors with different weighting ratios. The weighting ratio is determined based on the bound of $||\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)} - \mathbf{x}_{\mathbf{e},\mu}||$. If the bound of $||\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)} - \mathbf{x}_{\mathbf{e},\mu}||$ is larger, then the corresponding weighting ratio is smaller. The combined estimation error and the weighting ratio are shown as follows

$$\begin{aligned}
\tilde{\mathbf{x}}_{\mathbf{e},\mu}(t) &= \Sigma_{\nu=0,\nu\neq\mu}^{m_{nc}} \phi_{\nu} \tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t), \\
\Sigma_{\nu=0,\nu\neq\mu}^{m_{nc}} \phi_{\nu} &= 1, \\
\phi_{\nu} &= \frac{1}{m_{nc} - 1} \frac{\Sigma_{j=0,j\neq\mu,j\neq\nu}^{m_{nc}} \overline{e}_{\mu(j)}}{\Sigma_{j=0,j\neq\mu}^{m_{nc}} \overline{e}_{\mu(j)}}, \\
\overline{x}_{e,\mu(\nu)} &= ||(\mathbf{E}_{\nu} - \mathbf{E}_{\mu})^{-1}||(||\mathbf{L}_{\nu}|| + ||\mathbf{L}_{\mu}||)v.
\end{aligned} \tag{III.25}$$

The sensor noise affects the accuracy of estimation error decoupling, thus affecting the convergence ratios and sensor anomaly detection. Lemma 3 indicates that the impact of sensor noise can be mitigated by choosing the observer gains $\mathbf{L}_{\mu}$ and $\mathbf{L}_{\nu}$ with smaller norms. An observer gain with a smaller norm, however, may reduce the convergence speed of the estimation error. Thus, there is a trade-off in choosing observer gains. The impact of sensor noise on the convergence ratios can also be mitigated via averaging over κ_{CR} time steps as shown in Definition 7. In addition to techniques for mitigating the impact of sensor noise, a threshold θ_{CR} for convergence ratios should be selected to balance the tolerance of system noise and the ability to detect an anomaly. As discussed in Section III.4.3.1, the convergence ratios are the same as the observer poles under normal operation but they are different from observer poles under anomaly in ideal case. However, the observer poles are usually selected to be close to 0 to ensure fast observer's estimation error convergence and noise exists on the system. So we select a upper threshold θ_{CR}, which is larger than the largest observer pole but less than one. Then the sensor anomaly, which makes the estimation error converge faster, cannot be detected by the CR method. With the threshold θ_{CR}, the lower bound of the sensor anomaly signal that can be detected is ($\kappa_{CR} = 1$)

$$\begin{aligned} \|\boldsymbol{\gamma}(t)\| \geq & \|\{(\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}\}_j(\mathbf{L}_\nu\boldsymbol{\Gamma}_\nu - \mathbf{L}_\mu\boldsymbol{\Gamma}_\mu)\|^{-1} \\ & (\theta_{CR}\|\mathbf{x}^{(j)}_{\mathbf{e},\mu}(t-1)\| + (1+\theta_{CR})\|\{(\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}\}_j\|(\|\mathbf{L}_\nu\| + \|\mathbf{L}_\mu\|)v + \|\mathbf{x}^{(j)}_{\mathbf{e},\mu}(t)\|), \end{aligned} \tag{III.26}$$

where $\{(\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}\}_j$ is the j^{th} row of matrix $(\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}$. This lower bound is proportional to the threshold θ_{CR} and the bound of the sensor noise v.

Both the process noise and the sensor noise affect the accuracy of the bias calculation, thus affecting the ability to distinguish a disturbance from a sensor noise. Based on the boundedness of the process noise and the sensor noise, the error of the bias calculation $\|\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) - \mathbf{d}(t)\|$ is also bounded when the system is under disturbance. Lemma 4 and Lemma 5 give the bound of $\|\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) - \mathbf{d}(t)\|$ under disturbance when the system is subject to either the process noise or the sensor noise, respectively.

Lemma 4. *Given a control system* (III.1) *with bounded process noise and* $\mathbf{v}(t) = \mathbf{0}$*,* $\|\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) - \mathbf{d}(t)\|$ *is bounded by* $\|((\mathbf{D}_{\Lambda,\mu})^\intercal\mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^\intercal(\mathbf{V}_\mu)^{-1}\|w$.

For the proof of Lemma 4, see Appendix VI.2.

Lemma 5. *Given a control system* (III.1) *with bounded sensor noise and* $\mathbf{w}(t) = \mathbf{0}$*,* $\|\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) - \mathbf{d}(t)\|$ *is bounded* $\|((\mathbf{D}_{\Lambda,\mu})^\intercal\mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^\intercal(\mathbf{V}_\mu)^{-1}\|(1 + \|\mathbf{E}_\mu\|)\|(\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}\|(\|\mathbf{L}_\nu\| + \|\mathbf{L}_\mu\|)v$.

For the proof of Lemma 5, see Appendix VI.2.

Combining Lemma 4 and Lemma 5, the bound of the error of the bias calculation is

$$\|\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t)-\mathbf{d}(t)\| \leq \|((\mathbf{D}_{\Lambda,\mu})^{\mathsf{T}}\mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^{\mathsf{T}}(\mathbf{V}_{\mu})^{-1}\|(w+(1+\|\mathbf{E}_{\mu}\|)\|(\mathbf{E}_{\nu}-\mathbf{E}_{\mu})^{-1}\|(\|\mathbf{L}_{\mu}\|+\|\mathbf{L}_{\nu}\|)v). \tag{III.27}$$

Notice that the bounds are different for biases calculated based on different pairs of observers, and that they are all zero-mean. Based on the bounds, one specific threshold $\theta_{d,\mu(\nu),\zeta(\eta)}$ ($\mu,\nu,\zeta,\eta = 0,..,m_{nc} \wedge \mu \neq \nu \wedge \zeta \neq \eta$) can be selected to compare with the difference between any two biases averaged over κ_{CR} time steps, thus determining whether the system is under disturbance or sensor anomaly. If any one pair of the biases exceeds the corresponding threshold, then the system is under sensor noise. Otherwise, the system is under disturbance.

If the system is under disturbance, the combination of the weighted biases is considered as the disturbance signal. The weighting ratio of each bias is determined based on the bound of $\|\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t)-\mathbf{d}(t)\|$. If the bound is larger, then the corresponding weighting ratio is smaller. The combined bias and the weighting ratio are shown as follows

$$\begin{aligned}
\tilde{\mathbf{d}}(t) &= \Sigma_{\nu=0,\nu\neq\mu}^{m_{nc}}\Sigma_{\mu=0}^{m_{nc}}\psi_{\mu(\nu)}\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t),\\
\Sigma_{\nu=0,\nu\neq\mu}^{m_{nc}}\Sigma_{\mu=0}^{m_{nc}}\psi_{\mu(\nu)} &= 1,\\
\psi_{\mu(\nu)} &= \frac{1}{(m_{nc}+1)m_{nc}-1}\frac{\Sigma_{j=0,j\neq i}^{m_{nc}}\Sigma_{i=0,i\neq\mu}^{m_{nc}}\bar{d}_{i(j)}+\Sigma_{j=0,j\neq\nu}^{m_{nc}}\bar{d}_{\mu(j)}}{\Sigma_{j=0,j\neq i}^{m_{nc}}\Sigma_{i=0}^{m_{nc}}\bar{d}_{i(j)}},\\
\bar{d}_{\mu(\nu)} &= \|((\mathbf{D}_{\Lambda,\mu})^{\mathsf{T}}\mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^{\mathsf{T}}(\mathbf{V}_{\mu})^{-1}\|(w+(1+\|\mathbf{E}_{\mu}\|)\|(\mathbf{E}_{\nu}-\mathbf{E}_{\mu})^{-1}\|(\|\mathbf{L}_{\mu}\|+\|\mathbf{L}_{\nu}\|)v).
\end{aligned} \tag{III.28}$$

Algorithm 2 shows the procedure of the CR method. The CR method contains three steps. The first step is to calculate the estimation error for each observer. Then the convergence ratios of the estimation errors are used to detect the occurrence of an anomaly. If an anomaly is detected, biases are calculated and analyzed to determine whether the anomaly is a disturbance or a sensor anomaly.

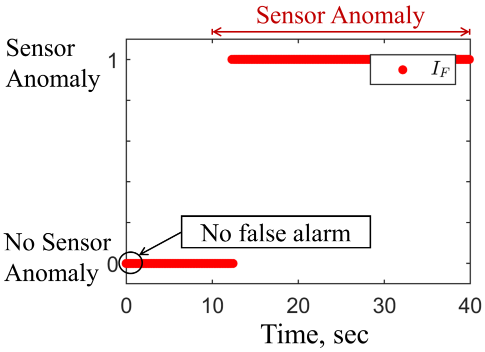


Figure III.7: Alarms I_F of the CR method under sensor anomaly α

Fig. III.7 shows the alarms generated by the CR method under sensor anomaly α. During the observers' transient state, false alarms are eliminated compared to Fig. III.4b, Fig. III.5b and Fig. III.6b. When the system is under sensor anomaly α, there is an $2s$ detection delay, which is caused by κ_{CR} for averaging the convergence ratio and the threshold θ_{CR}. The detection delay is decreased compared to the $8s$ detection delay in Fig. III.4.

III.4.4 MOLO Method for Critical Sensor Anomaly Diagnosis

The MOLO method has the potential to diagnose anomalies in critical sensors. It consists of multiple groups of open-loop observers. The states of the open-loop observers are updated periodically by the estimated state of the closed-loop observer using all of the sensor measurements. The open-loop observers in different groups have different update frequencies. Residuals are formed based on the difference between the measured outputs of the system and the estimated outputs calculated based on the estimated states by the open-loop observers. Then the averaged residual is analyzed to determine the occurrence of a critical sensor anomaly, and to isolate the anomalous sensor.

In noise-free case ($\mathbf{w}(t) = \mathbf{0}, \mathbf{v}(t) = \mathbf{0}$), the MOLO method only works if the open-loop system is stable or marginally stable. This is due to the fact that the estimation error of open-loop observer $\mathbf{x_{e,o}}(t)$ will diverge if the system is unstable, i.e., the eigenvalues of $\mathbf{A}$ lie outside of the unit circle,

Algorithm 2: CR method for sensor anomaly detection

```
function CR;
Input : $\tilde{\mathbf{x}}_i(t-\kappa_{CR}:t+1)(i=0,1,...,m_{nc})$ from time step $t-\kappa_{CR}$ to $t+1$
Output: $I_A, I_F, I_D, \tilde{\mathbf{d}}(t-1)$
//Estimation error calculation;
for $\mu = 0$ to $m_{nc}$ do
    for $\nu = 0$ to $m_{nc}$ do
        if $\mu \neq \nu$ then
            $\mathbf{x}_{\mathbf{e},\mu,\nu}(t) = \tilde{\mathbf{x}}_\mu(t) - \tilde{\mathbf{x}}_\nu(t)$;
            $\mathbf{x}_{\mathbf{e},\mu,\nu}(t+1) = \tilde{\mathbf{x}}_\mu(t+1) - \tilde{\mathbf{x}}_\nu(t+1)$;
            $\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t) = (\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}(\mathbf{x}_{\mathbf{e},\mu,\nu}(t+1) - \mathbf{E}_\nu \mathbf{x}_{\mathbf{e},\mu,\nu}(t))$;
            $\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu(\nu)}(t) = (\mathbf{V}_\mu)^{-1}\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t)$;
        end
    end
    $\tilde{\mathbf{x}}_{\mathbf{e},\mu}(t) = \Sigma_{\nu=0,\nu\neq\mu}^{m_{nc}} \phi_\nu \tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}$;
    $\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu}(t) = (\mathbf{V}_\mu)^{-1}\tilde{\mathbf{x}}_{\mathbf{e},\mu}(t)$;
    //Convergence ratio calculation;
    for $j = 1$ to $n_x$ do
        $cr_{\mu,j}(t) = \frac{1}{\kappa_{CR}}\left[\left|\frac{\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu}^{(j)}(t)}{\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu}^{(j)}(t-1)}\right| + \sum_{k_i=2}^{\kappa_{CR}} \sqrt[k_i]{\left|\frac{\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu}^{(j)}(t)}{\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu}^{(j)}(t-k_i)}\right|}\right]$;
        //Anomaly detection;
        if $cr_{\mu,j}(t) > \theta_{CR}$ then
            $I_A = I_A + 1$;
        end
    end
end
//Determine whether it is a sensor anomaly or a
disturbance;
if $I_A \geq \frac{(m_{nc}+1)\times n_x}{2}$ then
    for $i = 1$ to $m_{nc}$ do
        $\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t-1) = ((\mathbf{D}_{\Lambda,\mu})^\intercal \mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^\intercal[\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu(\nu)}(t) - \mathbf{E}_{\Lambda,\mu}\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu(\nu)}(t-1)]$;
    end
    if
    $Any\ avg(\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t-1-\kappa_{CR}:t-1) - \tilde{\mathbf{d}}_{\Lambda,\zeta(\eta)}(t-1-\kappa_{CR}:t-1)) > \theta_{d,\mu(\nu),\zeta(\eta)}$
    then
        $I_F = 1$;
    else
        $I_D = 1$;
        $\tilde{\mathbf{d}}(t-1) = \Sigma_{\nu=0,\nu\neq\mu}^{m_{nc}} \Sigma_{\mu=0}^{m_{nc}} \psi_{\mu(\nu)} \tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t-1)$;
    end
end
```

according to (III.29).

$$\mathbf{x_{e,o}}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}(t) = \mathbf{A}^t\mathbf{x_{e,o}}(t_0), \tag{III.29}$$

where $\mathbf{x_{e,o}}(t_0)$ is the initial estimation error. After introducing system noise, the condition for the estimation error of an open-loop observer to be bounded is given in Proposition 1.

Proposition 1. *Given a control system* (III.1)*, and an open-loop observer* (III.3)*, the following results can be drawn:*

1. *If all of the eigenvalues of* $\mathbf{A}$ *lie inside the unit circle, then the estimation error of an open-loop observer is bounded; and*

2. *If one or more of the eigenvalues of* $\mathbf{A}$ *lie on the unit circle and* $||\mathbf{A}|| = 1$*, then the estimation error of an open-loop observer is bounded.*

For the proof of proposition 1, see Appendix VI.2.

For systems that do not satisfy the conditions in Proposition 1, we need to periodically update the state of the open-loop observer with the state estimated by the closed-loop observer 0 which uses all of the sensor measurements when no sensor anomaly is detected. The initial estimation error of the open-loop observer is then the same as the estimation error of the closed-loop observer.

There is a trade-off between the estimation performance and the ability to detect a critical sensor anomaly. If the update frequency is fast, then the state estimated by the open-loop observer can track the state estimated by the closed-loop observer well, which is indicated by

$$\mathbf{x_{e,o}}(t) = \mathbf{A}^t\mathbf{x_e}(t_0) + \Sigma_{i=0}^{t-t_0-1}\mathbf{A}^i\mathbf{w}(t-1-i), \tag{III.30}$$

where $\mathbf{x_e}(t_0)$ is the estimation error of the closed-loop observer 0. If t is smaller, then the divergence of $\Sigma_{i=0}^{t-t_0-1}\mathbf{A}^i\mathbf{w}(t-1-i)$ is smaller, which means a better estimation under normal operation. However, fast update frequency can degrade the ability to detect a sensor anomaly, which is indicated by

$$\mathbf{r}(t) = \mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t) = \mathbf{C}(\mathbf{A}^t\mathbf{x}_\mathbf{e}(t_0) + \Sigma_{i=0}^{t-t_0-1}\mathbf{A}^i\mathbf{w}(t-1-i)) + \mathbf{v}(t) + \boldsymbol{\Gamma}\boldsymbol{\gamma}(t). \qquad \text{(III.31)}$$

The ramp sensor anomaly signal $\boldsymbol{\gamma}(t)$ is increasing with the time step t. At the time step that $\boldsymbol{\gamma}(t)$ is significant, the sensor anomaly can be detected.

The above discussion on the trade-off shows the necessity to have multiple open-loop observers for a marginally stable system with $||\mathbf{A}|| > 1$. In this chapter, we divide the multiple open-loop observers into M groups. Group 1 has the slowest update frequency and group M has the fastest update frequency. Each group has N observers with the same update frequency. Based on the trade-off, if one group triggers an alarm, then the groups with slower update frequencies generate alarms as well, but the groups with faster update frequencies may not generate alarms. So if all of the groups detect a sensor anomaly, then we can say that the sensor anomaly signal has a large slope. If only some of the groups detect a sensor anomaly, then we can say that the sensor anomaly signal has a small slope.

Although the estimated state under the case that $||\mathbf{A}|| > 1$ may diverge for a marginally stable system, we can mitigate the impact of the process noise via averaging because the process noise has zero mean. To average the residuals, we need to find the time steps that the open-loop observers have similar divergence caused by system noise. Taking one open-loop observer for example, the state of the open-loop observer is updated every $\kappa_{f,g}$ time steps and has been updated for j_N times. At time step $t + (j_N - 1)\kappa_{f,g}$, we need to average the residual at time steps $t + (j_N - j)\kappa_{f,g}$ $(j = 1, ..., j_N)$ to mitigate the impact of system noise. Proposition 2 validates the effectiveness of averaging.

Proposition 2. *Given a control system* (III.1)*, an open-loop observer is updated every $\kappa_{f,g}$ time steps. The impact of the system noise on the averaged residual* (VI.24) *is mitigated.*

$$\mathbf{r}_{avg,g}(t + (j_N - 1)\kappa_{f,g}) = \frac{1}{j_N}\Sigma_{j=1}^{j_N}\mathbf{r}_g(t + (j_N - j)\kappa_{f,g}), \qquad \text{(III.32)}$$

where j_N is a positive integer.

For the proof of proposition 2, see Appendix VI.2.

Proposition 2 shows the averaging method if we only have one open-loop observer in each group. Then the time steps that are needed for averaging is about $j_N \cdot \kappa_{f,g}$, which is large. To reduce the time steps for averaging, we have N ($N \leq \kappa_{f,g}$) open-loop observers in each group. We evenly distribute the time steps to update the states of the open-loop observers within the same group during one update period and we have

$$N = \left\lfloor \frac{\kappa_{f,g}}{\kappa_{\Delta,g}} \right\rfloor, \tag{III.33}$$

where $\kappa_{\Delta,g}$ is the update time step interval between two adjacent open-loop observers i and $i+1$ in group g. Then we calculate the average of the residuals generated by the open-loop observers in the same group.

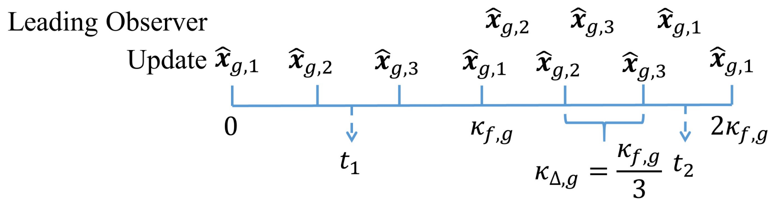
In order to average the residuals of N observers, we need the following definition

Definition 9. *The* <u>*leading observer*</u> *is the open-loop observer which has not been updated for the longest time steps among all of the observers in the same group during the time steps $(j-1) \cdot \kappa_{\Delta,g}$ and $j \cdot \kappa_{\Delta,g}$, where j is a positive integer. The leading observer could be found according to the following formula*

$$H_g = \lceil \frac{t - \kappa_{f,g} \lfloor \frac{t}{\kappa_{f,g}} \rfloor}{\kappa_{\Delta,g}} \rceil + 1. \tag{III.34}$$

Note that if $\lceil \frac{t - \kappa_{f,g} \lfloor \frac{t}{\kappa_{f,g}} \rfloor}{\kappa_{\Delta,g}} \rceil$ equals N, then set $H_g = 1$.

To average the residuals, the first step is to find the leading observer during the time steps $(j-1) \cdot \kappa_{\Delta,g}$ and $j \cdot \kappa_{\Delta,g}$. Fig. III.8 helps explain how we average the residuals generated by a group of three observers. Suppose we are at time step t_1, which is during the first update period $\kappa_{f,g}$. We simply average all the estimated states at time step t_1. Suppose we are at time step t_2. Observer $(g, 1)$ has not been updated for $t_2 - \kappa_{f,g}$ time steps, which is larger than that of observer


Leading Observer
Update
$\hat{\boldsymbol{x}}_{g,1}$
$\hat{\boldsymbol{x}}_{g,2}$
$\hat{\boldsymbol{x}}_{g,3}$
$\hat{\boldsymbol{x}}_{g,1}$
$\hat{\boldsymbol{x}}_{g,2}$
$\hat{\boldsymbol{x}}_{g,3}$
$\hat{\boldsymbol{x}}_{g,1}$
$\hat{\boldsymbol{x}}_{g,2}$
$\hat{\boldsymbol{x}}_{g,3}$
$\hat{\boldsymbol{x}}_{g,1}$
0
t_1
$\kappa_{f,g}$
$\kappa_{\Delta,g} = \frac{\kappa_{f,g}}{3}$
t_2
$2\kappa_{f,g}$

$$\boldsymbol{r}_{avg,g}(t_1) = \frac{1}{3}\sum_{i=1}^{3} \boldsymbol{r}_{g,i}(t_1)$$

$$\boldsymbol{r}_{avg,g}(t_2) = \frac{1}{3}[\sum_{i=2}^{3} \boldsymbol{r}_{g,i}\big(t_2 - \kappa_{f,g} + (i-1)\kappa_{\Delta,g}\big) + \boldsymbol{r}_{g,1}(t_2)]$$

Figure III.8: Residuals averaging

$(g,2)$ $(t_2 - \kappa_{f,g} - \kappa_{\Delta,g})$ and that of observer $(g,3)$ $(t_2 - \kappa_{f,g} - 2\kappa_{\Delta,g})$. Therefore, observer $(g,1)$ is the leading observer at time step t_2. Based on this leading observer, we find the corresponding time steps when the divergence is similar for the other two observers. After getting the three estimated states, we can calculate the averaged residual at time step t_2. It can been seen that the averaged residual is generated over $2\kappa_{f,g}$ time steps. The following formula shows the averaged residual at time step t

$$\mathbf{r}_{avg,g}(t) = \frac{1}{N}(\Sigma_{i=1}^{H_g}\mathbf{r}_{g,i}(t-(H_g-i)\kappa_{\Delta,g}) + \Sigma_{i=H_g+1}^{N}\mathbf{r}_{g,i}(t-\kappa_{f,g}+(i-H_g)\kappa_{\Delta,g})). \tag{III.35}$$

The average of the finite zero-mean random vector ($N < \infty$) does not exactly equal the zero vector. Based on the bounds of the system noise and update period $\kappa_{f,g}$, a threshold $\boldsymbol{\theta}_{MOLO,g}$ can be set for each group to compare with the averaged residual $\mathbf{r}_{avg,g}$. Notice that $\boldsymbol{\theta}_{MOLO,g} \in \mathbb{R}^{n_y}$ is a vector. We compare each element $\mathbf{r}^{(j)}_{avg,g}(t)$ in $\mathbf{r}_{avg,g}(t)$ with the corresponding element $\boldsymbol{\theta}^{(j)}_{MOLO,g}$ in $\boldsymbol{\theta}_{MOLO,g}$. If $\mathbf{r}^{(j)}_{avg,g}(t) \geq \theta^{(j)}_{MOLO,g}$, then group g triggers an alarm. Once the alarm is triggered, the states of the group of the open-loop observers are not updated by the closed-loop observer until the alarm is cleared.

Logic is applied to determine whether the system is under sensor anomaly or under normal operation based on which groups trigger alarms. Based on the discussion about the trade-off, if a group triggers an alarm, the groups with slower update frequencies should also trigger alarms theoretically. Therefore, we find the group g' which has the fastest update frequency among the groups that trigger alarms. If the majority of groups from 1 to g' trigger alarms, i.e., the inequality (III.36) holds, then the system is under sensor anomaly. Otherwise, it could be false alarms and the system is under normal operation.

$$\frac{1}{g'}\Sigma_{g=1}^{g'} I_{F,g}(t) \geq \theta_f, \tag{III.36}$$

where θ_f is a selected value with range 0.5 to 1. The sensor j, which makes the most of the groups that trigger alarms have $\mathbf{r}_{avg,g}^{(j)}(t) \geq \boldsymbol{\theta}_{MOLO,g}^{(j)}(g = 1, 2, ..., g')$, is identified as the anomalous sensor.

When the system is subject to a sensor anomaly on a critical sensor, the averaged residual is

$$\begin{aligned} \mathbf{r}_{avg,g}(t) =& \frac{1}{N}(\Sigma_{i=1}^{N}\mathbf{r}_{g,i}(t-(N-i)\kappa_{\Delta,g}) \\ =& \frac{1}{N}\Sigma_{i=1}^{N}\mathbf{C}\mathbf{A}^{t_3}e(t-t_3-(N-i)\kappa_{\Delta,g}) + \frac{1}{N}\Sigma_{i=1}^{N}\boldsymbol{\Gamma}\boldsymbol{\gamma}(t-(N-i)\kappa_{\Delta,g}). \end{aligned} \tag{III.37}$$

The above equation is drawn based on the assumption that observer N is the leading observer at time step t and it is updated at time step $t - t_3$. Suppose the sensor anomaly starts between time step $t-t_3$ and t. Theorem 1 in [61] indicates that $||\mathbf{x_e}(t-t_3-(N-i)\kappa_{\Delta,g})|| < \epsilon, \forall N$, where ϵ is a small positive number and it is related to system noise and initial estimation error. Therefore, the sensor anomaly signal could increase the averaged residual generated by multiple open-loop observers, thus detected by the MOLO method. If the slope of the ramp sensor anomaly signal is arbitrarily small, then the sensor anomaly signal can still bypass the MOLO method.

Remark 3. *The sensor anomaly signal could be designed to make* $\Sigma_{i=1}^{N}\boldsymbol{\Gamma}\boldsymbol{\gamma}(t-(N-i)\kappa_{\Delta,g}) = 0$ *in order to bypass the multiple open-loop observers. That means, however, the sensor anomaly signal is changing around zero every* $\kappa_{\Delta,g}$ *time steps. If the change is small, then the impact of the sensor anomaly is insignificant. If the change is large, then the sensor anomaly signal can cause a significant change in the residual generated by a closed-loop observer.*

Although this approach cannot guarantee the detection of a sensor anomaly with arbitrarily small slope, a sensor anomaly with a small slope would take a long time to disrupt the performance of the system. In addition, if the sensor anomaly is caused by an attack, this long time increases the cost of the attack implementation. During this time, other techniques, such as sensor fusion, may have already detected the sensor anomaly.

Algorithm 3 shows the procedure of the MOLO method. At each time step, we first find the leading observer. Then we average the residuals for each group. Then the averaged residual is analyzed to determine the occurrence of a critical sensor anomaly, and isolate the anomalous sensor. After the anomalous sensor is detected, if the system is stable or marginally stable with $||\mathbf{A}|| = 1$, then we can directly use the state estimated by an open-loop observer for the state feedback controller as indicated in Proposition 1. Otherwise, we need to replace the anomalous sensor.

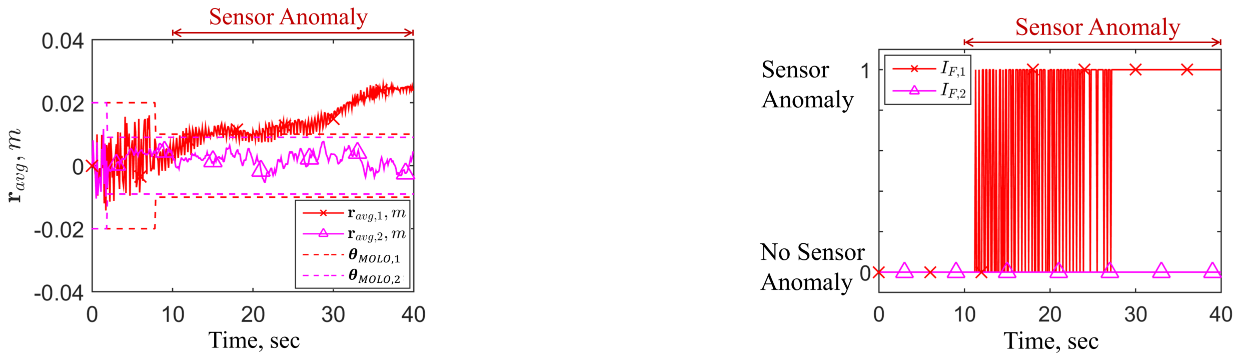
Fig. III.9 shows the performance of the MOLO method under sensor anomaly β. In this example, we have two groups of open-loop observers. Group 1 has update period $8s$ and group 2 has update period $2s$. There are 20 observers in each group and the update time steps are distributed evenly within one update period. Fig. III.9a shows the averaged residuals and Fig. III.9b shows the alarms of the two groups. After the first update period, the averaged residual is less noisy and the threshold of each group could be smaller. It can also be seen that the sensor anomaly is successfully detected by Group 1 at about $27s$ but bypasses Group 2. This is because the update period of Group 2 is too short compared to the slope of the sensor anomaly signal. Overall, sensor anomaly β is successfully detected by the MOLO method compared to Fig. III.5.

III.4.5 CCI Method for non-Critical Sensor Anomaly Mitigation

The CCI method can potentially mitigate the impact of an anomaly in a non-critical sensor during the diagnosis process. At each time step, this method selects the closed-loop observer, based on which the state feedback controller gives the smallest divergence of the control input. This divergence is defined as follows:

Algorithm 3: MOLO method for critical sensor anomaly diagnosis

```
function MOLO;
Input  : $\mathbf{y}(t), \mathbf{u}(t), \tilde{\mathbf{x}}_0(t), I_{F,g}(t-1), \hat{\mathbf{x}}_{g,i}$
Output: $I_F, I_{F,g}(t), i_f, \hat{\mathbf{x}}_{g,i}(t+1)$
for $g = 1$ to $M$ do
    $H_g = \lceil \frac{t-\kappa_{f,g}\lfloor \frac{t}{\kappa_{f,g}} \rfloor}{\kappa_{\Delta,g}} \rceil + 1$;
    if $H_g > N$ then
        $H_g = 1$;
    end
    for $i = 1$ to $N$ do
        if time to update $\hat{\mathbf{x}}_{g,i}$ then
            $\hat{\mathbf{x}}_{g,i}(t) = (1 - I_{F,g}(t-1))\tilde{\mathbf{x}}_0(t) + I_{F,g}(t-1)\hat{\mathbf{x}}_{g,i}(t)$;
            $\hat{\mathbf{x}}_{g,i}(t+1) = \mathbf{A}\hat{\mathbf{x}}_{g,i}(t) + \mathbf{B}\mathbf{u}(t)$;
        else
            $\hat{\mathbf{x}}_{g,i}(t+1) = \mathbf{A}\hat{\mathbf{x}}_{g,i}(t) + \mathbf{B}\mathbf{u}(t)$;
        end
        //Residuals generation;
        $\mathbf{r}_{g,i}(t) = \mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}_{g,i}(t)$
    end
    //Averaged residual;
    if $t \leq \kappa_{f,g}$ then
        $\mathbf{r}_{avg,g}(t) = \frac{1}{N}\Sigma_{i=1}^{N}\mathbf{r}_{g,i}(t)$;
    else
        $\mathbf{r}_{avg,g}(t) = \frac{1}{N}(\Sigma_{i=1}^{H_g}\mathbf{r}_{g,i}(t-(H_g-i)\kappa_{\Delta,g}) + \Sigma_{i=H_g+1}^{N}\mathbf{r}_{g,i}(t-\kappa_{f,g}+(i-H_g)\kappa_{\Delta,g}))$;
    end
end
//Sensor anomaly diagnosis;
$tmp = 0$; //The number of groups that trigger alarms;
$tmp_{sensor,j} = 0$; //The sensor that each group thinks it is
anomalous;
for $g = 1$ to $M$ do
    for $j = 1$ to $n_x$ do
        if $\mathbf{r}_{avg,g}^{(j)}(t) \geq \boldsymbol{\theta}_{MOLO,g}^{(j)}$ then
            $I_{F,g}(t) = 1$;
            $g' = g$;
            $tmp = tmp + 1$;
            $tmp_{sensor,j} = tmp_{sensor,j} + 1$;
        end
    end
end
if $\frac{1}{g'}\Sigma_{g=1}^{g'} I_{F,g}(t) \geq \theta_f$ then
    $I_F = 1$;
    $i_f = \max_j tmp_{sensor,j}$;
end
```



(a) The averaged residuals of the two groups of observers: Group 1 has update period $8s$ and Group 2 has update period $2s$

(b) Alarms of the two groups $I_{F,1}$ and $I_{F,2}$

Figure III.9: The performance of the MOLO method under sensor anomaly β

Definition 10. *Divergence of the control input* $||\boldsymbol{\Delta}\mathbf{u}_i||$ *is the absolute difference between the calculated control input based on the closed-loop observer and that based on an open-loop observer.*

$$||\boldsymbol{\Delta}\mathbf{u}_i(t)|| = ||\mathbf{F}\tilde{\mathbf{x}}_i(t) - \mathbf{F}\hat{\mathbf{x}}(t)||. \tag{III.38}$$

The open-loop observer in the CCI method is slightly different from those used in the MOLO method. Since the CCI method switches among several closed-loop observers from time to time, the state of the open-loop observer should be updated to be the estimated state by the closed-loop observer which is used for feedback at time step t. For example, if closed-loop observer i is used for feedback at time step t, then we need to calculate the estimated state $\hat{\mathbf{x}}(t+1)$ of the open-loop observer with the initial state $\tilde{\mathbf{x}}_i(t)$.

First, we analyze this method in ideal system, and give the lower bound of the sensor anomaly signal that the CCI method can switch to the observer without the anomalous sensor during the diagnosis process. Then, we analyze the impact of system noise on the lower bound of the sensor anomaly signal.

III.4.5.1 Ideal System Case

Under normal operation, the divergence of the control input calculated based on a closed-loop observer is a function of its estimation error. Under sensor anomaly, the closed-loop observer without anomalous sensor gives the best state estimation, thus the smallest divergence. Theorem 2 demonstrates that the divergence of the control input $||\mathbf{\Delta u}_{i_f}(t+1)||$ based on the closed-loop observer i_f without the anomalous sensor i_f is smaller than that based on other closed-loop observers with the anomalous sensor.

Theorem 2. *Given an ideal control system* (III.1) *with* $\mathbf{w}(t) = \mathbf{0}$ *and* $\mathbf{v}(t) = \mathbf{0}$*, and a sensor anomaly starting at time step t on sensor* i_f*, observer* i_f *gives the smallest divergence of the control input* $||\mathbf{\Delta u}_{i_f}(t+1)||$ *if the lower bound of the sensor anomaly signal satisfies* (III.39).

$$\begin{aligned} \forall i =& 0,1,...,m_{nc}, i \neq i_f, \\ ||\boldsymbol{\gamma}(t)|| \geq& ||\mathbf{F}\mathbf{L}_i\mathbf{\Gamma}_i||^{-1}[||\mathbf{F}\mathbf{L}_{i_f}\mathbf{C}_{i_f}\mathbf{x}_{\mathbf{e},i_f}(t)|| + ||\mathbf{F}\mathbf{L}_i\mathbf{C}_i\mathbf{x}_{\mathbf{e},i}(t)|| + ||\mathbf{F}\mathbf{A}\mathbf{x}_{\mathbf{e},i_f,i}(t)||]. \end{aligned} \tag{III.39}$$

Proof. With anomalous sensor i_f starting at time step t, observer i_f is not affected by the anomalous sensor. The estimated state $\tilde{\mathbf{x}}_i(t+1)$ of observer i ($i \neq i_f$) containing the anomalous sensor and the estimated state $\tilde{\mathbf{x}}_{i_f}(t+1)$ observer i_f are

$$\begin{aligned} \tilde{\mathbf{x}}_i(t+1) =& \mathbf{E}_i\tilde{\mathbf{x}}_i(t) + \mathbf{L}_i(\mathbf{C}_i\mathbf{x}(t) + \mathbf{\Gamma}_i\boldsymbol{\gamma}(t)) + \mathbf{B}\mathbf{u}(t), \\ \tilde{\mathbf{x}}_{i_f}(t+1) =& \mathbf{E}_{i_f}\tilde{\mathbf{x}}_{i_f}(t) + \mathbf{L}_{i_f}\mathbf{C}_{i_f}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t). \end{aligned} \tag{III.40}$$

Since the initial state of the open-loop observer is the same as the estimated state of the observer which is used for feedback at time step t, two cases should be considered:

1. At time step t, observer i ($i \neq i_f$) is used for feedback,

$$\hat{\mathbf{x}}(t+1) = \mathbf{A}\tilde{\mathbf{x}}_i(t) + \mathbf{B}\mathbf{u}(t). \tag{III.41}$$

2. At time step t, observer i_f is used for feedback,

$$\hat{\mathbf{x}}(t+1) = \mathbf{A}\tilde{\mathbf{x}}_{i_f}(t) + \mathbf{B}\mathbf{u}(t). \tag{III.42}$$

Under case 1), the divergence of the control input of observer i_f and observer i ($i \neq i_f$) are shown in (III.43) and (III.44), respectively.

$$||\boldsymbol{\Delta}\mathbf{u}_{i_f}(t+1)|| = ||\mathbf{F}\mathbf{A}\mathbf{x}_{\mathbf{e},i_f,i}(t) + \mathbf{F}\mathbf{L}_{i_f}\mathbf{C}_{i_f}\mathbf{x}_{\mathbf{e},i_f}(t)||, \tag{III.43}$$

$$||\boldsymbol{\Delta}\mathbf{u}_i(t+1)|| = ||\mathbf{F}\mathbf{L}_i\mathbf{C}_i\mathbf{x}_{\mathbf{e},i}(t) + \mathbf{F}\mathbf{L}_i\boldsymbol{\Gamma}_i\boldsymbol{\gamma}(t)||. \tag{III.44}$$

So when the lower bound of the sensor anomaly signal satisfies (III.39), observer i_f gives the smallest divergence of the control input, and is selected to provide feedback for the state feedback controller at time step $t+1$. The same result is also drawn for case 2). □

Based on Theorem 2, when the system is under non-critical sensor anomaly and the sensor anomaly signal satisfies (III.39), the CCI method can switch to the observer without the anomalous sensor before the anomalous sensor is identified. If the magnitude or the slope of the sensor anomaly signal is too small, then the CCI method may not be able to select the observer without the anomalous sensor to mitigate the impact of sensor anomaly; and the lower bound of the sensor anomaly signal during the observers' transient state is larger than that during steady state because of the relatively large estimation error. In order to reduce the lower bound of the sensor anomaly signal, horizon size κ_{CCI} is introduced to calculate the divergence of the control input to consider the impact of the integral of the sensor anomaly signal over κ_{CCI} steps. Therefore, at each time step t, we need to recalculate the state of the open-loop observer with initial state same as the estimated state $\tilde{\mathbf{x}}_i(t+1-\kappa_{CCI})$ of the selected closed-loop observer at time step $t+1-\kappa_{CCI}$. Then, the divergence of the control input of observer i_f and i are

$$||\mathbf{\Delta u}_{i_f}(t+1)|| = ||\mathbf{F}(-(\mathbf{E}_{i_f})^{\kappa_{CCI}}\mathbf{x}_{\mathbf{e},i_f}(t+1-\kappa_{CCI}) + \mathbf{A}^{\kappa_{CCI}}\mathbf{x}_{\mathbf{e},i_f}(t+1-\kappa_{CCI}))||, \tag{III.45}$$

$$||\mathbf{\Delta u}_i(t+1)|| = ||\mathbf{F}(-(\mathbf{E}_i)^{\kappa_{CCI}}\mathbf{x}_{\mathbf{e},i}(t+1-\kappa_{CCI}) - \Sigma_{j=0}^{\kappa_{CCI}-1}(\mathbf{E}_i)^j\mathbf{L}_i\mathbf{\Gamma}_i f(t-j) + \mathbf{A}^{\kappa_{CCI}}\mathbf{x}_{\mathbf{e},i_f}(t+1-\kappa_{CCI}))||. \tag{III.46}$$

Thus, the lower bound of the integral of the sensor anomaly signal is

$$\begin{aligned} ||\Sigma_{j=0}^{\kappa_{CCI}-1}\mathbf{F}(\mathbf{E}_i)^j\mathbf{L}_i\mathbf{\Gamma}_i\boldsymbol{\gamma}(t-j)|| \geq & 2||\mathbf{F}A^{\kappa_{CCI}}\mathbf{x}_{\mathbf{e},i_f}(t+1-\kappa_{CCI})|| \\ & + ||\mathbf{F}(\mathbf{E}_{i_f})^{\kappa_{CCI}}\mathbf{x}_{\mathbf{e},i_f}(t+1-\kappa_{CCI})|| + ||\mathbf{F}(\mathbf{E}_i)^{\kappa_{CCI}}\mathbf{x}_{\mathbf{e},i}(t+1-\kappa_{CCI})||. \end{aligned} \tag{III.47}$$

If the sensor anomaly starts between time steps $t+1-\kappa_{CCI}$ and t, $\mathbf{x}_{\mathbf{e},i_f}(t+1-\kappa_{CCI})$ and $\mathbf{x}_{\mathbf{e},i}(t+1-\kappa_{CCI})$ are very small. In addition, the absolute value of the eigenvalues of $\mathbf{E}_{i_f}$ and $\mathbf{E}_i$ are smaller than 1. Increasing the horizon step κ_{CCI} and placing the observer poles closer to the origin can reduce both $||\mathbf{F}(\mathbf{E}_{i_f})^{\kappa_{CCI}}\mathbf{x}_{\mathbf{e},i_f}(t+1-\kappa_{CCI})||$ and $||\mathbf{F}(\mathbf{E}_i)^{\kappa_{CCI}}\mathbf{x}_{\mathbf{e},i}(t+1-\kappa_{CCI})||$. For the term $||\mathbf{F}\mathbf{A}^{\kappa_{CCI}}\mathbf{x}_{\mathbf{e},i_f}(t+1-\kappa_{CCI})||$, however, we need to consider three conditions: $\mathbf{A}$ is stable, marginally stable and unstable. If the open-loop system is stable or marginally stable, i.e., the eigenvalues of $\mathbf{A}$ lie inside or on the unit circle, the term $||\mathbf{F}\mathbf{A}^{\kappa_{CCI}}\mathbf{x}_{\mathbf{e},i_f}(t+1-\kappa_{CCI})||$ is bounded. Thus, increasing κ_{CCI} can reduce the lower bound of the sensor anomaly signal and increase the ability of the CCI method to select the observer without the anomalous sensor. If the open-loop system is unstable, i.e., the one or more eigenvalues of $\mathbf{A}$ lie inside the unit circle, the term $||\mathbf{F}\mathbf{A}^{\kappa_{CCI}}\mathbf{x}_{\mathbf{e},i_f}(t+1-\kappa_{CCI})||$ is diverging, which reduces the ability of the CCI method. Therefore, the selection of the optimal horizon step κ_{CCI} depends on the property of the physical system.

III.4.5.2 Noisy System Case

With system noise, the lower bound of the sensor anomaly signal is increased as shown in Lemma 6 (the horizon step κ_{CCI} is not considered in Lemma 6).

Lemma 6. *Given a control system* (III.1)*, and a sensor anomaly starting at time step t on sensor i_f, observer i_f gives the smallest divergence of the control input if the lower bound of the sensor anomaly signal satisfies* (III.48).

$$\begin{aligned} &\forall i = 0, 1, ..., m_{nc}, i \neq i_f, \\ ||\boldsymbol{\gamma}(t)|| \geq &||\mathbf{F}\mathbf{L}_i\boldsymbol{\Gamma}_i||^{-1}(||\mathbf{F}\mathbf{L}_{i_f}\mathbf{C}_{i_f}\mathbf{x}_{\mathbf{e},i_f}(t)|| + ||\mathbf{F}\mathbf{L}_i\mathbf{C}_i\mathbf{x}_{\mathbf{e},i}(t)|| + ||\mathbf{F}\mathbf{A}\mathbf{x}_{\mathbf{e},i_f,i}(t)|| + ||\mathbf{F}\mathbf{L}_{i_f}||v_{i_f} + ||\mathbf{F}\mathbf{L}_i||v_i). \end{aligned} \tag{III.48}$$

The proof is similar to that for Theorem 2.

The transient dynamics caused by switching among observers may degrade the performance of the control system [53]. To avoid frequently switching, a threshold θ_{CCI} is used to decide when to enable or disable the switching. θ_{CCI} should be selected to balance the frequency of switching and the ability to mitigate the impact of the sensor anomaly.

Algorithm 4 gives the procedure of the CCI method. At each time step, the CCI method calculates the estimated state of an open-loop observer with the initial state the same as the selected observer at time step $t+1-\kappa_{CCI}$. Then it switches to the observer which gives the smallest divergence of the control input if the switching is enabled.

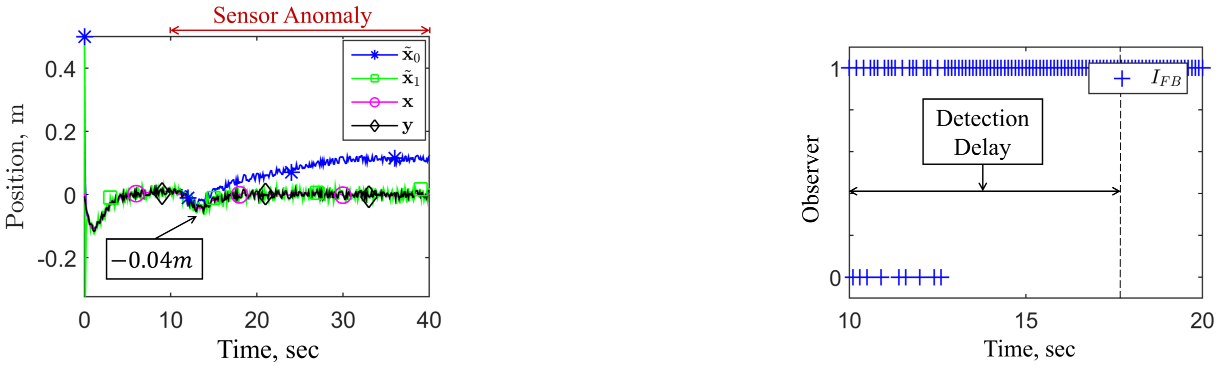
Fig. III.10 shows the system with the CCI method under sensor anomaly α. The maximum absolute value of position under sensor anomaly is $4cm$, which is smaller than that with the CO method as shown in Fig. III.4a. During the detection delay ($2s$), the CCI method has already switched to observer 1 for state estimation at $13s$, thus mitigating the impact of the sensor anomaly.

III.4.6 Integration of CO, CR, MOLO and CCI Methods

In this section, the three new methods, CR, MOLO, and CCI methods are introduced and compared with the CO methods through simulation to show the improvements.

Algorithm 4: CCI method for non-critical sensor anomaly mitigation

```
function CCI;
Input  : $t$, $\tilde{\mathbf{x}}_i(t+1), \tilde{\mathbf{x}}_i(t+1-\kappa_{CCI})$ $(i = 0, ..., m_{nc})$, $I_{FB}(t+1-\kappa_{CCI})$
Output: $I_{FB}(t+1)$
//Open-loop observer state estimation;
if $t > \kappa_{CCI}$ then
    $\hat{\mathbf{x}}(t+1) = \mathbf{A}^{\kappa_{CCI}} \tilde{\mathbf{x}}_{I_{FB}(t+1-\kappa_{CCI})}(t+1-\kappa_{CCI}) + \Sigma_{j=0}^{\kappa_{CCI}-1} \mathbf{A}^j \mathbf{B}\mathbf{u}(t-j)$
else
    $\hat{\mathbf{x}}(t+1) = \mathbf{A}^{t+1} \tilde{\mathbf{x}}_0(t_0) + \Sigma_{j=0}^{t} \mathbf{A}^j \mathbf{B}\mathbf{u}(t-j)$
end
//Control input calculation;
$\mathbf{u}_o(t+1) = \mathbf{F}\hat{\mathbf{x}}(t+1)$;
$\mathbf{u}_i(t+1) = \mathbf{F}\tilde{\mathbf{x}}_i(t+1)$;
$||\mathbf{\Delta u}_i(t+1)|| = ||\mathbf{u}_i(t+1) - \mathbf{u}_o(t+1)||$;
if $||\mathbf{\Delta u}_i(t+1)|| \geq \theta_{CCI}$ for all $i$ then
    $I_{FB}(t+1) = \min_i ||\mathbf{\Delta u}_i(t+1)||$;
else
    $I_{FB}(t+1) = I_{FB}(t)$;
end
```



(a) Estimated states of both observers $\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_1$, real state $\mathbf{x}$, and sensor measurement $\mathbf{y}$ of the system

(b) Selected observer index I_{FB} with the CCI method during time $10s$ and $20s$

Figure III.10: The performance of the CCI method under sensor anomaly α

- The CR method enables sensor anomaly detection during the observers' transient state, and no false alarms generated compared to CO method;
- The MOLO method successfully detects the critical sensor anomaly, while CO method fails; and
- The CCI method switches to the observer without the anomalous sensor during the diagnosis process, and the position of the object during sensor anomaly is reduced to $0.04m$ compared to $0.3m$ with CO method.

We systematically integrate all of the above methods to utilize their advantages, improving the overall performance of sensor anomaly diagnosis and mitigation. Algorithm 5 shows the integration of the CO, CR, MOLO, and CCI methods. At each time step, the CCI method is used to mitigate the impact of a potential sensor anomaly. Then the CR method determines whether there is a anomalous sensor on the system. If the CR method flags an alarm, and if the system observers have reached their steady state under normal operation ($t > t_{ss}$, where t_{ss} is the number of time steps that is needed for observers to reach their steady state), the CO method is used to isolate the anomalous sensor, and the system switches to the observer that can mitigate the impact of the sensor anomaly after the anomalous sensor is isolated. Meanwhile, the MOLO method detects whether there is an anomaly on a critical sensor. Robust control design in the presence of a disturbance is not within the scope of this dissertation.

III.5 Illustrative Example

A simplified suspension system (a two-mass-two-spring system) [83] is used to test the proposed algorithm with four methods. The system shown in Fig. III.11 has five states: position h_1 of mass 1, velocity $\dot{h}_1$ of mass 1, distance between two mass h, velocity $\dot{h}$, and integral of h, which is used to achieve zero steady-state error. The five states are measured by five sensors directly, as shown in Table III.2. A controller controls the system through $\mathbf{u}$. Potential disturbance comes from the ground. We want to maintain h to stay at 0m, which is also the reference signal of this system.

Algorithm 5: Integration of four methods

```
for $t = t_0$ to the end of simulation do
    //Estimated state of closed-loop observers;
    $\tilde{\mathbf{x}}_i(t+1) = \mathbf{E}_i\tilde{\mathbf{x}}_i(t) + \mathbf{L}_i\mathbf{y}_i(t) + \mathbf{B}\mathbf{u}(t)$;
    //Diagnosis and Mitigation begins;
    $I_{FB}(t+1) = CCI(t, \tilde{\mathbf{x}}_i(t+1), \tilde{\mathbf{x}}_i(t+1-\kappa_{CCI}), I_{FB}(t+1-\kappa_{CCI}))$;
    $\mathbf{u}(t+1) = \mathbf{F}\tilde{\mathbf{x}}_{I_{FB}(t+1)}(t+1)$;
    $[I_A, I_F, I_D, \tilde{\mathbf{d}}(t-1)] = CR(\tilde{\mathbf{x}}_i(t-k_{CR} : k+1))$;
    if $I_D = 0$ and $t \geq t_{ss}$ then
        $[I_F, I_{F,g}(t), i_f, \hat{\mathbf{x}}_{g,i}(t+1)] = MOLO(\mathbf{y}(t), \mathbf{u}(t), \tilde{\mathbf{x}}_0(t), I_{F,g}(t-1), \hat{\mathbf{x}}_{g,i})$;
    end
    if $I_F = 1$ and $t \geq t_{ss}$ then
        $[I_F, i_f] = CO(\mathbf{y}(t), \tilde{\mathbf{x}}_i)$;
        $I_{FB}(t+1) = i_f$;
        $\mathbf{u}(t+1) = \mathbf{F}\tilde{\mathbf{x}}_{I_{FB}(t+1)}$;
    else if $I_{F,g} = 1$ for any $g$ then
        if $\mathbf{A}$ is stable or $\mathbf{A}$ is marginally stable and $\|\mathbf{A}\| \leq 1)$ then
            $\mathbf{u}(t+1) = -\mathbf{F}\hat{\mathbf{x}}_{g,1}(t+1)$;
        else
            Replace the anomalous sensor $i_f$
        end
    else
        Robust control to tolerate disturbance
    end
end
```

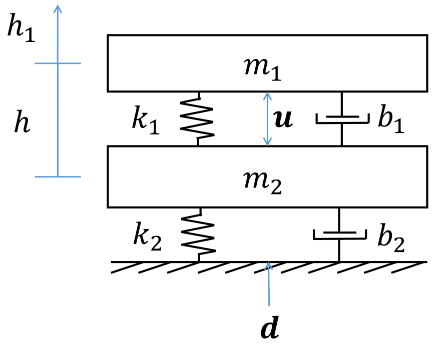


Figure III.11: Simplified suspension system

The system has sampling time $0.01s$, process noise bound 0.001 (m or m/s) and sensor noise bound 0.01 (m or m/s). The observers' transient state is about $0.1s$ (10 time steps). The initial state of the system is $(0,0,0,0,0)$. The initial state of the observers is $(0.02,0.01,1,0,0)$. Table III.3 shows part of parameters of the four methods.

Table III.2: Sensors of the Simplified Suspension System

	Variable	Set
Sensor 1	h_1	S_{nc}
Sensor 2	$\dot{h}_1$	S_{nc}
Sensor 3	h	S_{nc}
Sensor 4	$\dot{h}$	S_{nc}
Sensor 5	Σh	S_c

Table III.3: Part of parameters of the four methods

CO	θ_{CO}	0.012
CR	κ_{CR}	$0.1s$ (10 time steps)
	θ_{CR}	0.9
MOLO	M	2
	N	20
	$k_{f,1}$	$10s$ (1000 time steps)
	$k_{f,2}$	$0.4s$ (40 time steps)
	$\boldsymbol{\theta}^{(5)}_{MOLO,1}$	$0.025m$
	$\boldsymbol{\theta}^{(5)}_{MOLO,2}$	$0.015m$
CCI	κ_{CCI}	$10s$ (1000 time steps)
	θ_{CCI}	$0.001N$

Four scenarios are considered as examples:

- Scenario 1: A ramp anomaly signal with slope $1m/s$ ($0.01m$ per time step) added to sensor 3, saturating at $10m$;
- Scenario 2: A ramp anomaly signal with slope $1m/s$ ($0.01m$ per time step) added to sensor 3, saturating at $10m$;
- Scenario 3: A ramp anomaly signal with slope $0.01m/s$ ($0.0001m$ per time step) added to sensor 5, saturating at $10m$; and

- Scenario 4: A step disturbance from the ground with magnitude $0.2m$, starting at $t = 3000(30s)$.

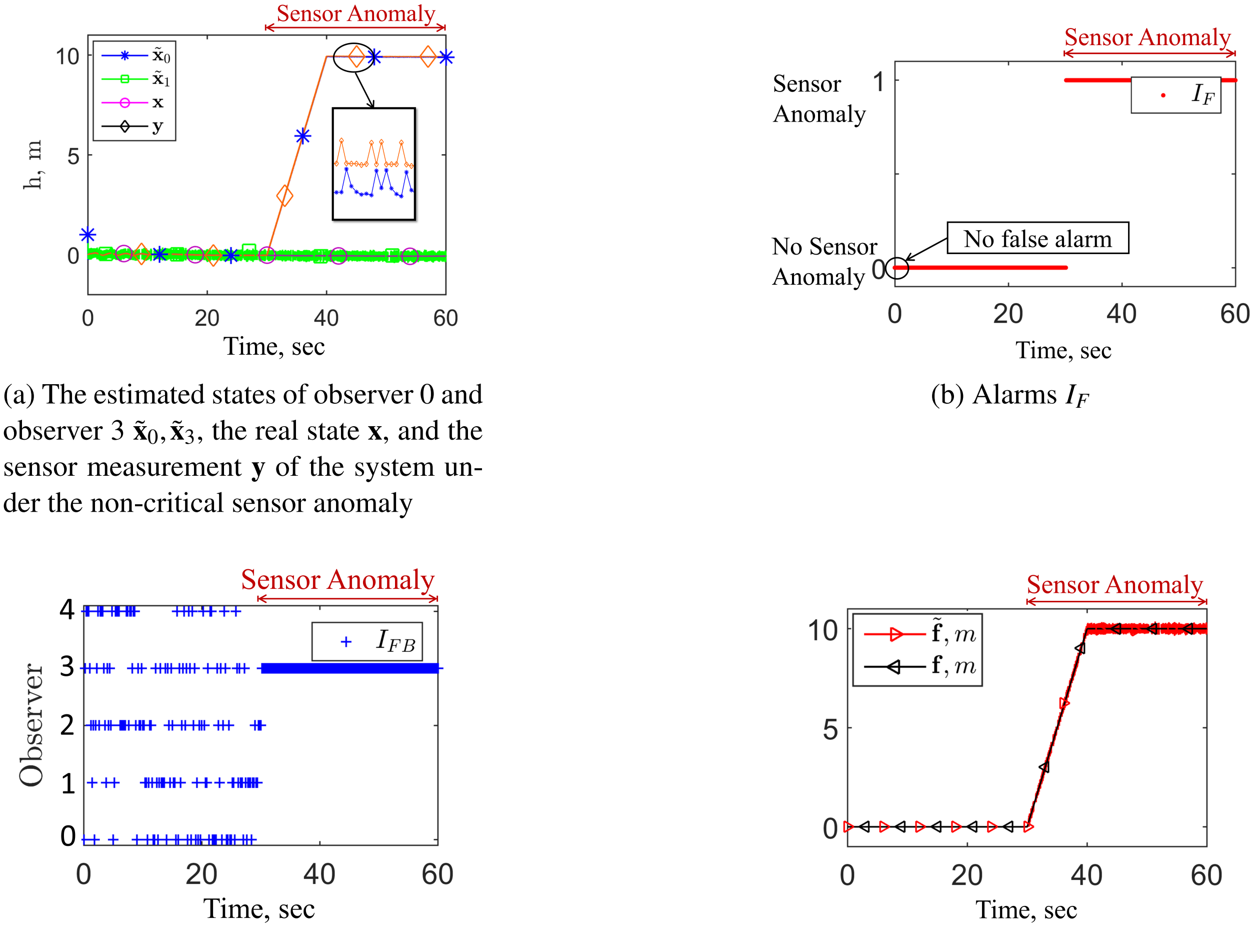
The sensor anomalies in Scenario 1 and 3 start at $t = 3000(30s)$. The sensor anomaly in Scenario 2 starts at $t = 5(0.05s)$.

Fig. III.12 shows the system under a non-critical sensor anomaly happening during the steady state of the system. During the observers' transient state, the CR method eliminates false alarms as shown in Fig. III.12b. At the time the sensor anomaly occurs, the CCI method switches to observer 3 for feedback as shown in Fig. III.12c, allowing more time for diagnosis. The CR method triggers an alarm after detecting the sensor anomaly. The CO method isolates the anomalous sensor, and calculates the sensor anomaly signal as shown in Fig. III.12d. The proposed algorithm integrating the four methods successfully protects the system from a non-critical sensor anomaly happening during the observers' steady state.

Fig. III.13 shows the system under a non-critical sensor anomaly happening during the observers' transient state. The CR method successfully detects the occurrence of the sensor anomaly with about $0.06s$ time delay as shown in Fig. III.13b, which is caused by relatively large θ_{CR} (0.9) compared to the observer poles (about 0.1). The CCI method switches to the observer without the anomalous sensor later than the time step that the CR method detects the sensor anomaly. This is because the observers cannot provide good state estimations during the observers' transient state, thus the observer without the anomalous sensor may not give the smallest divergence of the calculated control input. This scenario shows the effectiveness of the CR method for sensor anomaly detection during the observers' transient state.

Fig. III.14 shows the system is subject to a critical sensor anomaly. In Fig. III.14b, the averaged residuals are less noisy after the first update period. Group 1 successfully detects the occurrence of the sensor anomaly, while group 2 does not. This scenario shows the effectiveness of the MOLO method for a non-critical sensor anomaly diagnosis.

Fig. III.15 shows the system under disturbance from the ground. The CR method successfully distinguishes a disturbance from a sensor anomaly, and correctly estimates the disturbance



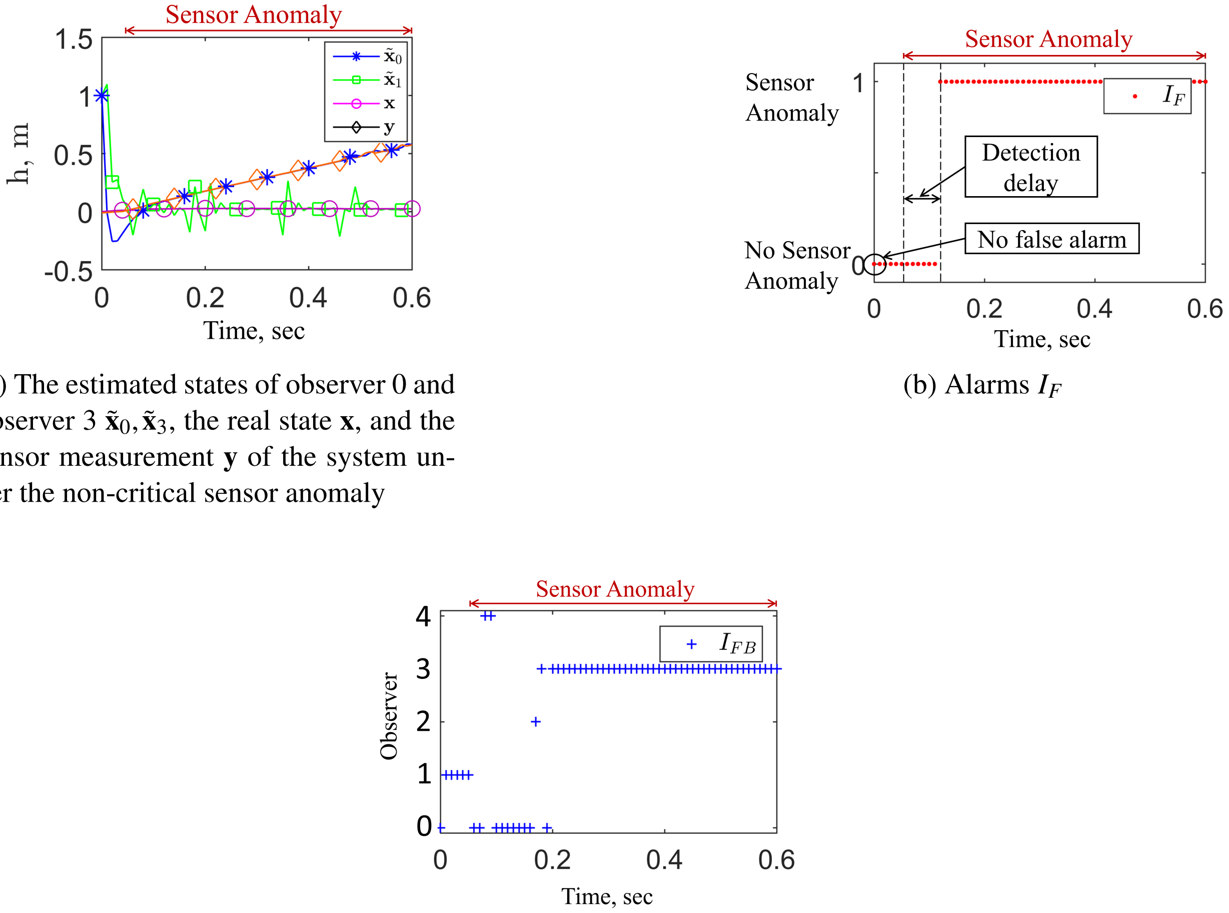
(a) The estimated states of observer 0 and observer 3 $\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_3$, the real state $\mathbf{x}$, and the sensor measurement $\mathbf{y}$ of the system under the non-critical sensor anomaly

(b) Alarms I_F

(c) Observer index I_{FB} selected for the state feedback controller

(d) Estimated sensor anomaly signal $\tilde{\boldsymbol{\gamma}}$ and the real sensor anomaly signal $\boldsymbol{\gamma}$

Figure III.12: The suspension system under a non-critical sensor anomaly happening during the steady state of the system

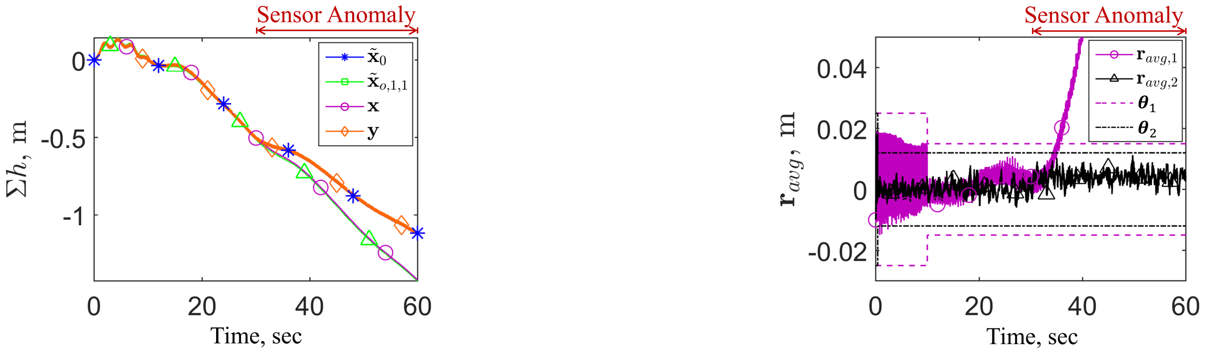


(a) The estimated states of observer 0 and observer 3 $\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_3$, the real state $\mathbf{x}$, and the sensor measurement $\mathbf{y}$ of the system under the non-critical sensor anomaly

(b) Alarms I_F

(c) Observer index I_{FB} selected for the state feedback controller

Figure III.13: The suspension system under a non-critical sensor anomaly happening during the observers' transient state



(a) The estimated states of closed-loop observer 0 and the open-loop observer $(1,1)$, $\tilde{\mathbf{x}}_0, \hat{\mathbf{x}}_{1,1}$ the real state $\mathbf{x}$, and the sensor measurement $\mathbf{y}$ of the system

(b) Averaged residuals of both groups of open-loop observers $\mathbf{r}_{avg,1}, \mathbf{r}_{avg,2}$

Figure III.14: The suspension system under a critical sensor anomaly

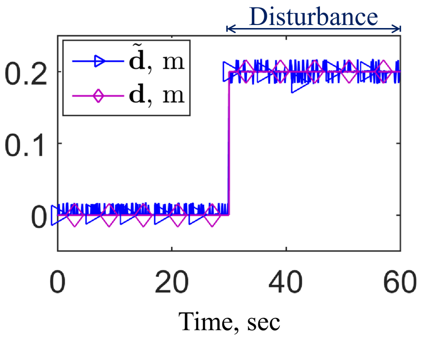


Figure III.15: Scenario 4: The calculated disturbance $\tilde{\mathbf{d}}$ and the real disturbance $\mathbf{d}$ of the system

signal.

III.6 Summary

In this chapter, the CO method and three new methods (CR, MOLO, and CCI) are integrated to solve the sensor anomaly diagnosis and mitigation problem using multiple closed-loop and open-loop observers. The closed-loop observers include one that uses all of the sensor measurements for state estimation, and others that exclude a non-critical sensor. The open-loop observers do not use any sensor measurements for state estimation. Based on these closed-loop and open-loop observers, new methods are proposed and integrated to improve sensor anomaly diagnosis and mitigation:

- The CR method can detect non-critical sensor anomalies during the observers' transient state;
- The MOLO method can detect and isolate critical sensor anomalies; and
- The CCI method can mitigate the impact of non-critical sensor anomalies during the anomaly diagnosis process.

The three new methods are integrated with a previously developed residual-based method (CO method) to collaboratively address the sensor anomaly diagnosis and mitigation problem in this chapter. The collaboration of the methods is illustrated in Figure III.1a and Table III.1. The proposed algorithm allows any residual-based method to be integrated besides the CO method. Simulation results show the effectiveness of our proposed framework.

This multi-observer sensor anomaly detection and mitigation approach can be easily extended to the multiple sensor anomalies case as long as the system observability still holds without the anomalous sensors. Note that it may be impossible in general to detect every kind of sensor anomaly. Some sensor anomaly may not be detected if the norm of the sensor anomaly signal is smaller than a certain lower bound. The aim of our sensor anomaly diagnosis and mitigation method is to decrease this lower bound to reduce the anomalies that we cannot detect, and to allow more time for other techniques to protect the system before it moves into a severe condition.

However, at a high level, the proposed framework has some limitations. For the critical sensor anomalies, the proposed MOLO method cannot detect a ramp sensor anomaly signal with arbitrary slope. The MOLO method does not provide detection guarantees for the critical sensor anomaly. In addition, this framework is designed specifically for sensor anomalies in continuous systems. However, many CPS cannot be simply modeled as continuous systems because many CPS are hybrid by nature, containing both discrete and continuous variables. In hybrid CPS, different types of anomalies may occur in either continuous process or discrete dynamics or both. Thus, the improved sensor anomaly diagnosis and mitigation framework is not sufficient to detect anomalies in hybrid CPS. All of these limitations are addressed in the next chapter, where we model CPS as hybrid systems. We classify the anomalies in hybrid systems into different types. By utilizing the relationship between the continuous and the discrete variables, we propose a new anomaly detection method that can provide detection guarantees for various types of anomalies in hybrid systems.

CHAPTER IV

Conflict-driven Anomaly Detection for Hybrid Systems with Current-State Observable Discrete Dynamics

IV.1 Introduction

In Chapter III, we improved the sensor anomaly diagnosis and mitigation framework to enhance the security of Cyber-Physical Systems (CPS). However, the improved framework has limitations when it is used to detect general anomalies. In addition, as mentioned in Chapter I, many CPS are hybrid systems consisting of both continuous and discrete variables. The improved sensor anomaly diagnosis and mitigation framework in Chapter III is developed for continuous systems, thus it is not sufficient to detect anomalies on discrete variables or both continuous and discrete variables in hybrid CPS.

In this chapter, we propose a new anomaly detection approach that addresses the limitations of the sensor anomaly diagnosis and mitigation framework in the detection of general anomalies in hybrid CPS. The proposed anomaly detection approach can provide formal detection guarantees for different types of anomalies in hybrid CPS. This approach, which is called the conflict-driven method, uses a hybrid observer to estimate both continuous and discrete variables. Based on all possible anomalies for hybrid systems and the fact that the continuous and discrete variables are related in hybrid systems, the conflict-driven method leverages the estimated discrete and continuous states along with their interrelationship to detect these challenging anomalies, as opposed to the traditional methods which analyze either the continuous system or the discrete system separately. The contributions of this chapter are as follows:

1. We propose a conflict-driven method to provide guarantees on the detection of some types of anomalies that are not detectable using traditional observer-based and residual-based methods in addition to the anomalies that can be detected by the traditional methods. In the conflict-driven method, we define three conflict types based on the relation between the discrete and continuous variables of the hybrid systems. The conflict-driven method detects anomalies by checking the occurrence of the conflicts.

2. We define a classification taxonomy for anomalies in hybrid systems. An anomaly in a hybrid system may affect the continuous variables or the discrete variables or both. Some anomalies are undetectable by only considering the continuous component of the system because the anomalous system may have consistent input-output data with the system model under normal operation. Some anomalies are undiagnosable by only considering the discrete component of the system because the observed discrete event sequence of the anomalous system is the same as the system under normal operation. In this dissertation, we classify the anomalies into eight different types based on the variables that are affected, input-output data consistency, and diagnosability of the anomaly.

3. We develop a new hybrid observer for anomaly detection. We use a Set-Valued Observer (SVO) as the continuous state observer of the hybrid observer. With the SVO, we can apply the conflict-driven method to hybrid systems with unobservable continuous components.

4. We provide a mapping between conflict types and anomaly types. Based on the occurrence of the conflict types, we can identify if the anomaly is related to the continuous component of the system, the discrete component or both.

The rest of this chapter is organized as follows. In Section IV.2, we give an introduction of the class of hybrid systems that are of interest. In Section IV.3, we briefly describe the hybrid observer used in the conflict-driven method. In Section IV.4, we propose a classification taxonomy of anomalies in hybrid systems. In Section IV.5, we introduce the conflict-driven anomaly detection method and demonstrate its effectiveness mathematically. In Section IV.6, we illustrate the effectiveness of the conflict-driven method using the simulated Positive Train Control (PTC) system as introduced in Chapter I.2.2.3. A summary of this chapter is provided in Section IV.7.

IV.2 Modeling Framework

In this section, we describe the class of hybrid systems to which the conflict-driven method can be applied. We formulate the class of hybrid systems mathematically. We partition the hybrid system into one nominal hybrid subsystem and at least one anomalous hybrid subsystem(s). The nominal hybrid subsystem corresponds to the system under normal operation and different anomalous hybrid subsystems correspond to the system under different anomalies.

IV.2.1 Notation

Let $||\cdot||$ denote ∞-norm, $\tilde{\cdot}$ denote estimated variables, $\dot{\cup}$ denote disjoint union, and $\Box\sigma$ denote the ∞-norm ball of center 0 and of radius σ. In addition, $\mathbf{x} \in \mathbb{R}^{n_x}$ represents a vector, where its i^{th} element is indicated by $\mathbf{x}^{(i)}$. $\mathbf{A} \in \mathbb{R}^{m \times n}$ represents a matrix. The linear span of a set of vectors is denoted by $span(\cdot)$. For a set $X \subset \mathbb{R}^{n_x}$, we denote its closure, interior, and boundary by $\overline{X}$, X^o and ∂X respectively. Clearly, $\partial X = \overline{X} \backslash X^o$. The volume of the closed set $\overline{X}$ is denoted by $Vol(\overline{X})$. A polyhedron X is represented by $X = Set(\mathbf{M}, \mathbf{m}) := \{\mathbf{x} : \mathbf{M}\mathbf{x} \leq \mathbf{m}\}$, where $\mathbf{M} \in \mathbb{R}^{m \times n_x}$ and $\mathbf{m} \in \mathbb{R}^m$. The detailed notations used in this chapter are described in Appendix B.

IV.2.2 System Modeling

Hybrid systems consist of a set of discrete states and a set of continuous states. In this chapter, we consider that each discrete state has an invariant which describes the set of allowable continuous states. We focus on hybrid systems which satisfy the following conditions.

System Conditions:

1. The hybrid system is deterministic. That is, the discrete transitions of the hybrid system are deterministic and the invariant in each discrete state is well-defined.

2. The number of discrete states is finite.

3. The hybrid system is memoryless. That is, the continuous dynamics in any discrete state are independent of the previous discrete state.

4. The discrete transitions of the hybrid systems are controlled switchings. That is, the continuous vector field may change discontinuously when the discrete state changes [10].

5. There is no discontinuity in continuous variables.

6. The discrete component of the system is current-state observable.

7. The invariant of each discrete state is static.

   Remark 4. *Conditions 1 to 4 limit our work to deterministic memoryless hybrid systems with finite discrete states and controlled switchings. Condition 5 is imposed without any reset maps. For example, a bouncing ball with autonomous jumps and discrete velocity changes or a system with hysteresis does not fit into our work. Condition 6 allows us to design a discrete state observer, which gives us a unique estimated discrete state after a finite number of observable discrete events occur. The definition of current-state observable is introduced later in Definition 12. Condition 7 does not limit the class of systems. We only need to model the system in a way such that the invariant of each discrete state is static. If a discrete state of the hybrid system has an invariant changing with time, we need to partition the discrete state into several discrete states with static invariants such that the system can fit into our work. Most hybrid CPS satisfy the above list of system conditions, such as the train system and a gantry system described in [59].*

Mathematically, a hybrid system can be modeled as a hybrid automaton

$$\mathcal{H} = (\mathcal{X},\mathcal{U},\mathcal{Y},Init,field,E,\phi,\eta), \tag{IV.1}$$

where each element is defined as

- $\mathcal{X} = Q \times X$: a set of discrete and continuous states
- $\mathcal{U} = \Psi \times U$: a set of discrete and continuous inputs
- $\mathcal{Y} = \Omega \times Y$: a set of discrete and continuous outputs
- $Init = (q(t_0), \mathbf{x}(t_0)) \in \mathcal{X}$: an initial state
- $field : \mathcal{X} \times U \to X$: a time-invariant vector field
- $E = \Psi \dot{\cup} \Omega$: a set of discrete events
- $\phi : Q \times \Psi \to Q$: a set of discrete transitions
- $\eta : \mathcal{X} \times \mathcal{U} \to \mathcal{Y}$: an output map consisting of a discrete output map ζ and a continuous output equation h
  - $\zeta : Q \times \Psi \to \Omega$: a discrete output map
  - $h : Q \times X \to Y$: a continuous output equation

The hybrid automaton captures both nominal system models with a set of nominal discrete states Q_n and anomaly models with a set of anomalous discrete states Q_f, where "*f*" indicates "fault". The set of all discrete states is defined as $Q = Q_n \dot{\cup} Q_f$. We consider that a hybrid system contains one nominal hybrid subsystem $\mathcal{H}_n$ and at least one anomalous hybrid subsystem(s) $\mathcal{H}_f$. An example hybrid system is shown in Fig. IV.1. The nominal hybrid subsystem corresponds to the system under normal operation and it contains the set of nominal discrete states. Different anomalous hybrid subsystems correspond to different anomalies. An anomalous hybrid subsystem contains a set of anomalous discrete states. An anomaly f transits the system from the nominal hybrid subsystem to the corresponding anomalous hybrid subsystem. The nominal hybrid subsystem $\mathcal{H}_n$ can be derived by removing Q_f and the events and transitions to and from Q_f. The initial state *Init*, which is a combination of initial discrete state $q(t_0) \in Q_n$ and initial continuous state $\mathbf{x}(t_0)$, is not required to be known.

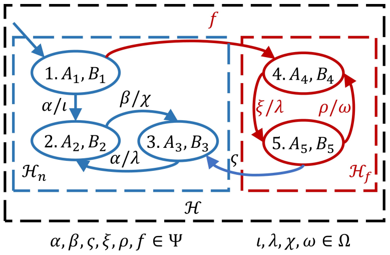


Figure IV.1: An example hybrid system $\mathcal{H}$ consisting of the nominal hybrid subsystem $\mathcal{H}_n$ and one anomalous hybrid subsystem $\mathcal{H}_f$

For each discrete state $q \in Q$, we associate continuous dynamics that can be represented by a Linear Time Invariant (LTI) model, subject to process and measurement noise.

$$\begin{aligned} field: \quad & \mathbf{x}(t+1) = \mathbf{A}_q \mathbf{x}(t) + \mathbf{B}_q \mathbf{u}(t) + \mathbf{w}(t), \\ h: \quad & \mathbf{y}(t) = \mathbf{C}_q \mathbf{x}(t) + \mathbf{v}(t), \end{aligned} \tag{IV.2}$$

where $\mathbf{A}_q \in \mathbb{R}^{n_x \times n_x}, \mathbf{B}_q \in \mathbb{R}^{n_x \times n_u}, \mathbf{C}_q \in \mathbb{R}^{n_y \times n_x}$ are system matrices, $\mathbf{x} \in X \subset \mathbb{R}^{n_x}$, $\mathbf{u} \in U \subset \mathbb{R}^{n_u}$ and $\mathbf{y} \in Y \subseteq \mathbb{R}^{n_y}$ are continuous states, inputs and outputs, respectively. The process and measurement noise are represented by $\mathbf{w} \in \mathbb{R}^{n_x}$ and $\mathbf{v} \in \mathbb{R}^{n_y}$, respectively. We define the system noise as $\mathbf{d}(t) = [\mathbf{w}(t) \quad \mathbf{v}(t)]^\intercal$ and assume:

Assumption 4. *The system noise at each time step is bounded, satisfying* $\forall i = t_0, ..., t, \mathbf{d}(i) \in \mathcal{B}_d$*, where* $\mathcal{B}_d := \{[\mathbf{w} \quad \mathbf{v}]^\intercal : ||\mathbf{w}|| \leq w, ||\mathbf{v}|| \leq v\}$ *and the initial condition is bounded, satisfying* $\mathbf{x}(t_0) \in \mathcal{B}_{x_o}$*, where* $\mathcal{B}_{x_o} := \{\mathbf{x}(t_0) : ||\mathbf{x}(t_0)|| \leq x\}$*.*

The continuous dynamical models of the system in anomalous discrete states are not required to be known for anomaly detection using the conflict-driven method.

Discrete events E consists of discrete input events Ψ and discrete output events Ω. Additionally, discrete events E can be partitioned into observable events E_o and unobservable events E_{uo}, i.e., $E = E_o \dot{\cup} E_{uo}$. Only observable events can be detected by an observer. We denote a set of observable input events as Ψ_o and a set of unobservable input events as Ψ_{uo}. All of the output events are observable.

The i^{th} discrete event occurs at time t_i. The continuous evolutions occur in time $t \in [t_{i-1}+1, t_i], \forall i = 1,2,....$ In reality, discrete events may occur between two adjacent sample times. We assume:

Assumption 5. *The occurrence of the discrete events can be captured at sample times. At most one input event occurs within one sampling period. An output event occurs simultaneously with an input event.*

Note that the discrete state is changed one time step after a discrete input event occurs, that is $\phi(q(t_i), \psi) = q'(t_i + 1)$, where $q(t_i), q'(t_i + 1) \in Q$.

To each discrete state $q \in Q$, we associate an invariant:

$$Inv_q = \{\mathbf{x} : \forall i = 1, ..., n_x, \underline{\beta}_i \leq \mathbf{x}^{(i)} \leq \overline{\beta}_i, \} \subseteq X, \tag{IV.3}$$

where $\underline{\beta}_i$ and $\overline{\beta}_i$ are constant values. An invariant is a hyperrectangle with a bounded interval on each continuous state variable.

To each discrete transition $\phi(q, \psi) = q'$, we associate a guard:

$$G(q, q', \psi) = \{\mathbf{x} \in Inv_q : s_G \mathbf{x}^{(i_G)} \geq c_G\}, \tag{IV.4}$$

where $i_G \in \{1, 2, ..., n_x\}$, c_G is a scalar and s_G is either -1 or 1. A guard is a hyperrectangle enclosed by the boundary of the invariant ∂Inv_q and the hyperplane:

$$\mathcal{P}(q, q', \psi) = \{\mathbf{x} : \mathbf{x}^{(i_G)} = s_G c_G\}. \tag{IV.5}$$

Our definitions of guard $G(q, q', \psi)$ and invariant Inv_q indicate that c_G is between the lower and upper bounds of the state variable $\mathbf{x}^{(i_G)}$ of the invariant Inv_q, i.e., $\underline{\beta}_{i_G} \leq c_G \leq \overline{\beta}_{i_G}$. A guard $G(q, q', \psi)$ indicates that the transition ψ will take place if and only if the i_G^{th} state variable of $s_G\mathbf{x}$ is greater than or equal to c_G in discrete state q. An invariant Inv_q indicates that the system can remain in discrete state q if and only if the continuous state $\mathbf{x} \in Inv_q \backslash \bigcup_j G(q, q_j, \psi_j)$.

We define a post-guard hyperplane of guard $G(q, q', \psi)$ as:

Definition 11. *Post-guard hyperplane of guard $G(q, q', \psi)$ is one of the hyperplanes forming the boundary of the invariant ∂Inv_q, which satisfies* (IV.6)*:*

$$\mathcal{L}(q, q', \psi) = \{\mathbf{x} \in X : s_G \mathbf{x}^{(i_G)} \geq c_G \wedge \mathbf{x}^{(i_G)} \in \{\underline{\beta}_{i_G}, \overline{\beta}_{i_G}\}\}. \tag{IV.6}$$

An example of post-guard hyperplane $\mathcal{L}(q, q', \psi)$ is shown in Fig. IV.2. To simplify notation, we denote $c_{\mathcal{L}}$ as the value of $\mathbf{x}^{(i_G)}$, where $\mathbf{x} \in \mathcal{L}(q, q', \psi)$. If $\mathcal{P}(q, q', \psi)$ forms one of the hyperplanes of ∂Inv_q, then $\mathcal{L}(q, q', \psi) = \mathcal{P}(q, q', \psi)$. Otherwise, $\mathcal{L}(q, q', \psi) \cap \mathcal{P}(q, q', \psi) = \emptyset$.

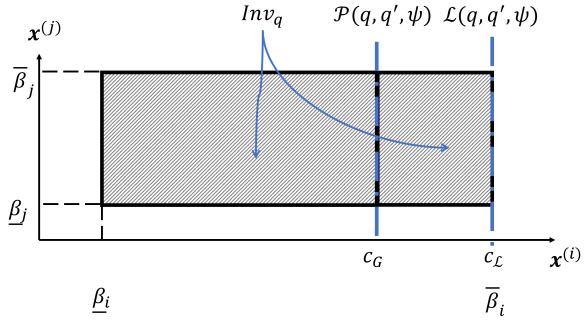


Figure IV.2: Visualization of the invariant Inv_q (the shaded rectangle), the hyperplane $\mathcal{P}(q, q', \psi)$ of guard $G(q, q', \psi)$, and the post-guard hyperplane $\mathcal{L}(q, q', \psi)$ of discrete state q in $2-D$ continuous state space

The proposed conflict-driven method utilizes a hybrid observer for state estimation. The framework of the hybrid observer is introduced in [5], which is designed based on the Finite State Machine (FSM) associated with the nominal hybrid subsystem. The FSM $\mathcal{M}_n$ is derived by extracting the discrete behavior from $\mathcal{H}_n$, which is represented by tuple $(Q, \Psi, \Omega, q(t_0), E, \phi, \zeta)$. In order to get a unique estimate of the discrete state with the hybrid observer after finite number of observable events, we assume:

Assumption 6. *The FSM $\mathcal{M}_n$ is current-state observable.*

Current-state observable is defined as:

Definition 12. *A FSM is* *current-state observable* *if there exists an integer k such that for any unknown initial discrete state, the discrete state, after the i^{th} observable discrete input event occurs, can be uniquely determined from the observed input/output event pairs sequence up to i, i.e., $i \geq k$ [5].*

Note that one observable input/output event pair is considered as one input event to the hybrid observer. Thus, after the k^{th} observable input/output event pair occurs, the hybrid observer can give a unique estimated discrete state. The necessary and sufficient condition of current state observability is given in [5]. The time that the k^{th} observable input/output event pair occurs is denoted as t_k.

IV.2.3 Nominal Hybrid Subsystem

In the nominal hybrid subsystem, we partition the invariant of each nominal discrete state into an intermediate region and a normal operating region. Before introducing intermediate region and normal operating region, we first define neighbor discrete state as

Definition 13. *Neighbor discrete state* $q_{\mathcal{N},q}$ *of nominal discrete state q is a nominal discrete state that can by reached via one observable discrete event $\psi \in \Psi$ from the discrete state q:*

$$q_{\mathcal{N},q} = \{q_j \in Q_n : \exists \mathcal{P}(q, q_j, \psi), q \neq q_j\}. \tag{IV.7}$$

The set of neighbor discrete states of nominal discrete state q is called neighbor set $\mathcal{N}_q$. Then the intermediate region $\mathcal{R}_{in,q,q_j}$ is the set of continuous states satisfy guard $G(q, q_j, \psi)$, where $q_j \in \mathcal{N}_q$:

$$\mathcal{R}_{in,q,q_j} = \{\mathbf{x} \in Inv_q : q_j \in \mathcal{N}_q \wedge s_G \mathbf{x}^{(i_G)} \geq c_G\}, \tag{IV.8}$$

where $\mathbf{x}^{(i_G)}$ is the state variable corresponding to guard $G(q, q_j, \psi)$. Since the invariant of each nominal discrete state is compact [1], the intermediate region $\mathcal{R}_{in,q,q_j}$ is compact as well. The in-

[1] A hyperrectangle is closed, compact and convex.

termediate region $\mathcal{R}_{in,q}$ is the union of the intermediate regions $\mathcal{R}_{in,q,q_j}$, where $q_j \in \mathcal{N}_q$. That is, $\mathcal{R}_{in,q} = \bigcup_{\forall q_j \in \mathcal{N}_q} \mathcal{R}_{in,q,q_j}$. The intermediate region $\mathcal{R}_{in,q}$ is not necessarily compact.

For each discrete state $q \in Q_n$, we define a normal operating region as the set of continuous states that are in the invariant but not the intermediate region $\mathcal{R}_{in,q}$,

$$\mathcal{R}_{no,q} = \overline{Inv_q} \backslash \overline{\mathcal{R}_{in,q}}. \tag{IV.9}$$

To have an appropriate hybrid model for which the conflict-driven method can provide detection guarantees, we pose the following assumption:

Assumption 7. *The intermediate region is bounded by the hyperplane $\mathcal{P}(q,q_j,\psi)$ corresponding to the guard $G(q,q_j,\psi)$ and the post-guard hyperplane $\mathcal{L}(q,q_j,\psi)$ and ∂Inv_q in each discrete state.*

$$\mathcal{R}_{in,q} \subset \{\mathbf{x} \in Inv_q : \forall q_j \in Q_n : \exists \mathcal{P}(q,q_j,\psi_j), \min(c_G, c_{\mathcal{L}}) \leq \mathbf{x}^{(i_G)} \leq \max(c_G, c_{\mathcal{L}})\}. \tag{IV.10}$$

The visualization of this assumption on $2-D$ space is shown in Fig. IV.3. The intermediate region of discrete state q is the union of $\mathcal{R}_{in,q,q_k}$ and $\mathcal{R}_{in,q,q_j}$. The intermediate region $\mathcal{R}_{in,q}$ is a subset of the region bounded by the hyperplane $\mathcal{P}(q,q_j,\psi_j)$ and the hyperplane $\mathcal{L}(q,q_j,\psi_j)$ and the region bounded by the hyperplane $\mathcal{P}(q,q_k,\psi_k)$ and the hyperplane $\mathcal{L}(q,q_k,\psi_k)$. System condition 5 and Assumption 7 indicate that $\mathcal{P}(q,q_k,\psi_k)$ is one of the hyperplanes forming ∂Inv_{q_k} and $\mathcal{P}(q,q_j,\psi_j)$ is one of the hyperplanes forming ∂Inv_{q_j}.

The basic principle of conflict-driven method is to check at each time step, whether or not the sets of current and future continuous states, which are calculated using reachability analysis based on the estimated set of continuous states, intersect with the invariant of the estimated discrete state. The sets of continuous states include an initial set given by the continuous state observer, and a forward reachable set which is the set of all continuous states that can be reached along trajectories starting in the initial set. Reachable set calculation requires the following assumption.

Assumption 8. *The continuous input signal is bounded, and the bound is known, i.e., $||\mathbf{u}|| \leq \mu$.*

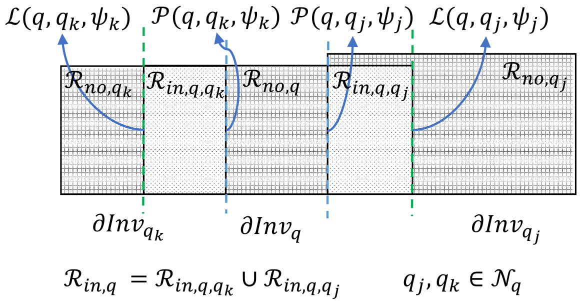

$\mathcal{L}(q,q_k,\psi_k)$
$\mathcal{P}(q,q_k,\psi_k)$
$\mathcal{P}(q,q_j,\psi_j)$
$\mathcal{L}(q,q_j,\psi_j)$
$\mathcal{R}_{no,q_k}$
$\mathcal{R}_{in,q,q_k}$
$\mathcal{R}_{no,q}$
$\mathcal{R}_{in,q,q_j}$
$\mathcal{R}_{no,q_j}$
∂Inv_{q_k}
∂Inv_q
∂Inv_{q_j}
$\mathcal{R}_{in,q} = \mathcal{R}_{in,q,q_k} \cup \mathcal{R}_{in,q,q_j}$
$q_j, q_k \in \mathcal{N}_q$

Figure IV.3: Normal operating and intermediate region. $\mathcal{P}(q,q_k,\psi_k)$ is the hyperplane of the guard $G(q,q_k,\psi_k)$, and $\mathcal{L}(q,q_k,\psi_k)$ is the post-guard hyperplane. $\mathcal{P}(q,q_j,\psi_j)$ is the hyperplane of the guard $G(q,q_j,\psi_j)$, and $\mathcal{L}(q,q_j,\psi_j)$ is the post-guard hyperplane

Remark 5. *Assumption 8 and system condition 5 indicate that the normal operating regions are connected. The reachability analysis needs the continuous system to be open-loop stable or marginally stable such that the reachable set is non-diverging. For anomaly detection purpose, we do not need the continuous system to be open-loop stable or marginally stable. If the time step for reachable set calculation is selected small enough, the reachable set will not diverge too much. The detail of calculating the time step for reachability analysis is described in Section IV.5.*

IV.2.4 Anomalous Hybrid Subsystem

An anomaly $f \in \Psi_{uo}$ is defined as an unobservable input event that transits the nominal hybrid subsystem $\mathcal{H}_n$ to one of the anomalous hybrid subsystems $\mathcal{H}_f$. Comparing to the nominal hybrid subsystem, we define that the anomalous hybrid subsystem should satisfy at least one of the following two conditions.

Anomaly Conditions:

1. Either the continuous dynamics or the invariant of at least one of the anomalous discrete states is different from any of the nominal discrete states.

2. The observable discrete event sequence in the anomalous hybrid subsystem is different from that in the nominal hybrid subsystem.

Otherwise, the anomalous hybrid subsystem behaves exactly the same as the nominal hybrid subsystem and the impact of the anomaly is insignificant.

According to the difference between the nominal hybrid subsystem and the anomalous hybrid subsystem, an anomaly f either affects the continuous variables or the discrete variables or both. The continuous dynamics that are changed under anomaly f have two different mathematical forms: multiplicative and additive [80]. A multiplicative anomaly is represented by the product of a variable with the anomaly itself, such as a parameter change within a process. Under an additive anomaly, a variable is influenced by an addition of the anomaly signal itself, such as offsets of sensor values. Arguably, a multiplicative anomaly can be represented by an additive anomaly model (e.g., Section 3.5 in [24]). Thus, we restrict our attention to additive anomaly models as follows.

$$\mathbf{x}(t+1) = \mathbf{A}_q\mathbf{x}(t) + \mathbf{B}_q\mathbf{u}(t) + \mathbf{w}(t) + \mathbf{\Gamma_1}\boldsymbol{\gamma_1}(t), \tag{IV.11}$$

$$\mathbf{y}(t) = \mathbf{C}_q\mathbf{x}(t) + \mathbf{v}(t) + \mathbf{\Gamma_2}\boldsymbol{\gamma_2}(t), \tag{IV.12}$$

where $\mathbf{\Gamma_1} \in \mathbb{R}^{n_x \times n_x}, \mathbf{\Gamma_2} \in \mathbb{R}^{n_y \times n_y}$ are diagonal anomaly matrices with binary variables. The i^{th} diagonal variable of $\mathbf{\Gamma_1}$ is 1 if and only if the i^{th} state variable is added with an anomalous signal $\boldsymbol{\gamma_1}(t) \in \mathbb{R}^{n_x}$. The i^{th} diagonal variable of $\mathbf{\Gamma_2}$ is 1 if and only if the i^{th} output is added with an anomalous signal $\boldsymbol{\gamma_2}(t) \in \mathbb{R}^{n_y}$. $\mathbf{\Gamma_1}, \boldsymbol{\gamma_1}$ and $\mathbf{\Gamma_2}, \boldsymbol{\gamma_2}$ are not required to be known for anomaly detection using conflict-driven method. Note that an additive anomaly $\boldsymbol{\gamma_1}$ on the state equation (IV.11) could be transformed into an equivalent additive anomaly $\boldsymbol{\gamma_2}$ on the output equation (IV.12)[2].

IV.3 Hybrid Observer

In this section, we give a brief overview of the hybrid observer, which consists of a discrete state observer and a continuous state observer [5]. Under the hybrid observer framework in [5], we propose to use a Set-Valued Observer (SVO) as the continuous state observer [79].

[2] Transformation between $\boldsymbol{\gamma_1}$ and $\boldsymbol{\gamma_2}$: $\mathbf{\Gamma_2} = \mathbf{C}\mathbf{\Gamma_1}, \boldsymbol{\gamma_2}(t) = \Sigma_{i=0}^{t-1}\mathbf{A}_q^i\boldsymbol{\gamma_1}(t-1-i)$.

Given the nominal hybrid model $\mathcal{H}_n$, we design a hybrid observer to estimate both the discrete state and the continuous state of the system. The hybrid observer $\mathcal{O}$ consists of a discrete state observer $\mathcal{D}$ and a continuous state observer $\mathcal{C}$, as shown in Fig. IV.4. The discrete state observer receives observable discrete input/output event pairs $(\psi(t), \omega(t))$ and gives $\tilde{q}(t)$. If no event pair is received by time t, then the estimated discrete state is the same as that at the previous time step, i.e., $\tilde{q}(t) = \tilde{q}(t-1)$. The estimated discrete state $\tilde{q}(t)$ may contain a set of estimated discrete states until the occurrence of the k^{th} observable input/output event pair. After the occurrence of the k^{th} observable input/output event pair, $\tilde{q}(t)$, which contains a unique estimate, is passed to the corresponding continuous state observer. The continuous state observer is designed as a SVO [79]. The SVO gives a set of estimated continuous states $\tilde{X}(\mathbf{y}, t)$ using the continuous input $\mathbf{u}(t)$ and output $\mathbf{y}(t)$.

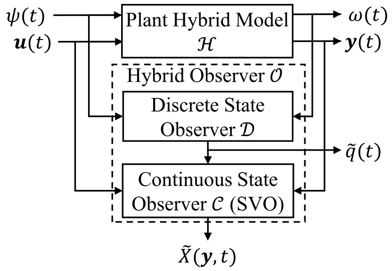


Figure IV.4: The structure of the hybrid observer $\mathcal{O}$

IV.3.1 Discrete State Observer

The discrete state observer is represented by a Finite State Machine (FSM) which is a tuple $\mathcal{D} = (\tilde{Q}, E_{\mathcal{D}}, -, \tilde{q}(t_0), E_{\mathcal{D}}, \tilde{\phi}, -)$, where $E_{\mathcal{D}} = (\Psi, \Omega)$ is the set of discrete input/output event pairs of $\mathcal{M}_n$ and "-" means that $\mathcal{D}$ does not contain the corresponding component as general FSM mentioned in Section IV.2.2. The discrete state observer is tracking the set of possible discrete states that the system can be in. Therefore, no discrete output events or discrete output map are defined for discrete state observer.

The construction of the discrete state observer $\mathcal{D}$ follows the steps presented in [14]. It starts from the initial estimated discrete state $\tilde{q}(t_0)$: with an unknown initial discrete state of $\mathcal{M}_n$, $\tilde{q}(t_0) = Q_n$. Then we need to determine the set of possible transitions out of the discrete state in the discrete state observer and the discrete states in the nominal discrete system $\mathcal{M}_n$ to which it can transition. For each estimated discrete state $\tilde{q} \in \tilde{Q}$, we identify the input/output event pairs (ψ, ω) that label all the transitions out of any state q' in $\tilde{q}$. These events are called the active event set of $\tilde{q}$. For each pair (ψ, ω) in the active event set, we identify $q \in Q_n$ that can be reached from $q' \in \tilde{q}$, and these states return as a new $\tilde{q}$ in $\tilde{Q}$. This transition is added to $\tilde{\phi}$ satisfying:

$$\tilde{\phi} := \{q \in Q_n : \exists q' \in \tilde{q}, q \in \phi(q', \psi) \wedge \omega = \zeta(q', \psi)\}. \tag{IV.13}$$

Repeat this step until no new $\tilde{q}$ and $\tilde{\phi}$ can be added to $\mathcal{D}$.

IV.3.2 Continuous State Observer

We use the SVO as the continuous state observer to construct a *set* of estimated continuous states at each time step. The set given by the SVO is a polyhedron and can be represented by a pair of matrices. In order to use the SVO we first describe an operator *Rack* which allows us to calculate matrix pairs of a polyhedron. In the remainder of this section, we describe the set-valued estimation of the SVO and discuss the estimation accuracy of the SVO by introducing a central estimator.

IV.3.2.1 *Rack* Operator

Rack gives a *set* of possible matrix pairs which directly characterize a set S [79]. For $\mathbf{M_1} \in \mathbb{R}^{m \times n_x}, \mathbf{M_2} \in \mathbb{R}^m$, and $\mathbf{m_3} \in \mathbb{R}^m$, define $S \subseteq \mathbb{R}^{n_x}$ as $S = \{\mathbf{x} : \mathbf{M_1}\mathbf{x} + \mathbf{M_2} z \leq \mathbf{m_3}$ for some $z \in \mathbb{R}\}$. Define:

$$Rack[(\mathbf{M_1} \quad \mathbf{M_2}), \mathbf{m_3}] = \{(\mathbf{M}, \mathbf{m}) \in \mathbb{R}^{\tilde{m} \times n_x} \times \mathbb{R}^{\tilde{m}} : S = Set(\mathbf{M}, \mathbf{m})\}, \tag{IV.14}$$

where $Set(\mathbf{M}, \mathbf{m}) = \{\mathbf{x} : \mathbf{M}\mathbf{x} \leq \mathbf{m}\}$.

Note that the set S is unique but its matrix representation is not because different matrix representations $(\mathbf{M}, \mathbf{m})$ may contain different redundant constraints. For the sake of computational efficiency, redundant constraints are removed after selecting one of the matrix representations. Refer to [66] for different redundant constraints removing methods. The construction of the elements $\mathbf{M}, \mathbf{m}$ can be achieved by eliminating the variable z through the Fourier-Motzkin algorithm [43]. If k variables need to be eliminated, we can use *Rack* iteratively. That is, $Rack^k[(\mathbf{M_1} \quad \mathbf{M_2}), \mathbf{m_3}] = Rack[Rack^{k-1}[(\mathbf{M_1}' \quad \mathbf{M_2}'), \mathbf{m_3}']]$, which is a multivariable form of $Rack[(\mathbf{M_1} \quad \mathbf{M_2}), \mathbf{m_3}]$. Note that when we use *Rack* recursively for k times, then $\mathbf{M_2}$ is defined as a $\mathbb{R}^{m \times k}$ matrix corresponding to the k variables that need to be eliminated.

IV.3.2.2 Set-Valued Estimation

The SVO starts with the bound of the initial state $\mathcal{B}_{x_o}$ and then at each time step computes a set of estimated state-vectors, which is denoted as $\tilde{X}(\mathbf{y}, t-1)$, based on the initial condition, the measured output, and the (known) input. Fig. IV.5 shows how the SVO works for a $2-D$ system. At each time step, the SVO computes a one-time step forward reachable set (enclosed by the green dash-dotted line) based on the set of estimated continuous states ($\tilde{X}(\mathbf{y}, t-1)$, enclosed by the blue solid line) at the previous time step, known system continuous dynamics, process noise bound, and continuous input. Then using an output measurement and bounded measurement noise, the SVO also computes a set of continuous states (denoted as $\hat{X}(\mathbf{y}, t)$ and enclosed by the black dotted line). The intersection of the one-time step forward reachable set and $\hat{X}(\mathbf{y}, t)$ becomes the set of estimated continuous states at the current time step $\tilde{X}(\mathbf{y}, t)$.

The set of state-vectors $\hat{X}(\mathbf{y}, t)$ at time t based on a single measurement can be represented by $Set(\hat{\mathbf{M}}(t), \hat{\mathbf{m}}(t))$:

$$\hat{X}(\mathbf{y}, t) = \{\mathbf{x} : \hat{\mathbf{M}}(t)\mathbf{x} \leq \hat{\mathbf{m}}(t)\}, \tag{IV.15}$$

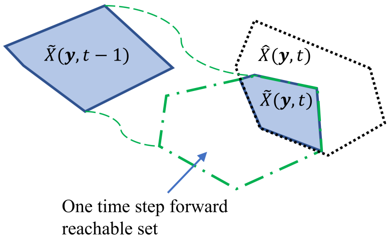

$\tilde{X}(\boldsymbol{y}, t-1)$
$\hat{X}(\boldsymbol{y}, t)$
$\tilde{X}(\boldsymbol{y}, t)$
One time step forward reachable set

Figure IV.5: The visualization of the SVO in a 2-D system

where

$$\hat{\mathbf{M}}(t) = \begin{bmatrix} \mathbf{C}_q \\ -\mathbf{C}_q \end{bmatrix}, \hat{\mathbf{m}}(t) = \begin{bmatrix} v\mathbf{1} + \mathbf{y}(t) \\ v\mathbf{1} - \mathbf{y}(t) \end{bmatrix}, \tag{IV.16}$$

where **1** is a vector with all entires as “1”.

The set of estimated state-vectors $\tilde{X}(\mathbf{y}, t)$ at time t is represented by $Set(\tilde{\mathbf{M}}(t), \tilde{\mathbf{m}}(t))$:

$$\tilde{X}(\mathbf{y}, t) = \{\mathbf{x} : \tilde{\mathbf{M}}(t)\mathbf{x} \leq \tilde{\mathbf{m}}(t)\}. \tag{IV.17}$$

The following three steps give the computational implementation of the SVO [79].

1. Initialization:

$$\tilde{X}(\mathbf{y}, t_0) = \hat{X}(\mathbf{y}, t_0) \cap \mathcal{B}_{x_o}, \tag{IV.18}$$

where $\hat{X}(\mathbf{y}, t_0) = Set(\hat{\mathbf{M}}(t_0), \hat{\mathbf{m}}(t_0))$ and $\hat{\mathbf{M}}(t_0), \hat{\mathbf{m}}(t_0)$ are calculated using (IV.16) at time t_0, and $\mathcal{B}_{x_o} = \{\mathbf{x}(t_0) : ||\mathbf{x}(t_0)|| \leq x\}$ as mentioned in Assumption 4.

2. Calculate the set $\hat{X}(\mathbf{y}, t)$ with the measured output:

$$\hat{X}(\mathbf{y}, t) = Set(\hat{\mathbf{M}}(t), \hat{\mathbf{m}}(t)), \tag{IV.19}$$

where $\hat{\mathbf{M}}(t), \hat{\mathbf{m}}(t)$ are calculated using (IV.16).

3. Calculate the set of the estimated continuous states at time step t, i.e., $\tilde{X}(\mathbf{y}, t) = Set(\tilde{\mathbf{M}}(t), \tilde{\mathbf{m}}(t))$. $\tilde{X}(\mathbf{y}, t)$ is the intersection of $\hat{X}(\mathbf{y}, t)$ and all of the states evolved from $\tilde{X}(\mathbf{y}, t-1)$ according to the

known continuous dynamics, process noise bound and continuous input:

$$\tilde{X}(\mathbf{y},t) = \left\{ \mathbf{x}(t) : \begin{bmatrix} \mathbf{I} & -\mathbf{A}_q & -\mathbf{I} \\ -\mathbf{I} & \mathbf{A}_q & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I} \\ \hat{\mathbf{M}}(t) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{M}}(t-1) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{x}(t-1) \\ \mathbf{w}(t-1) \end{bmatrix} \leq \begin{bmatrix} \mathbf{B}_q\mathbf{u}(t-1) \\ -\mathbf{B}_q\mathbf{u}(t-1) \\ w\mathbf{1} \\ w\mathbf{1} \\ \hat{\mathbf{m}}(t) \\ \tilde{\mathbf{m}}(t-1) \end{bmatrix} \right\}, \tag{IV.20}$$

where $\mathbf{I}$ is an identity matrix with an appropriate dimension. The first two constraints in (IV.20) correspond to the continuous dynamics of the system defined in (IV.2), third and fourth constraints correspond to the bound of the process noise $\mathbf{w}$, fifth constraint corresponds to the set $\hat{X}(\mathbf{y},t)$ and sixth constraint corresponds to the set $\tilde{X}(\mathbf{y},t-1)$. Note that the constraints in (IV.20) have the following form:

$$\mathbf{M_1}(t)\mathbf{x}(t) + \mathbf{M_2}(t) \begin{bmatrix} \mathbf{x}(t-1) \\ \mathbf{w}(t-1) \end{bmatrix} \leq \mathbf{m_3}(t). \tag{IV.21}$$

In order to get the set of continuous states which satisfy (IV.20), we need to eliminate the variables $[\mathbf{x}(t-1) \quad \mathbf{w}(t-1)]^\intercal \in \mathbb{R}^{2n_x}$ and represent $\tilde{X}(\mathbf{y},t)$in the form of $\tilde{\mathbf{M}}(t)\mathbf{x}(t) \leq \tilde{\mathbf{m}}(t)$. Therefore, we use *Rack* operator for $2n_x$ times iteratively:

$$(\tilde{\mathbf{M}}(t), \tilde{\mathbf{m}}(t)) \in Rack^{2n_x} \left[\mathbf{M_1}(t)\mathbf{x}(t) + \mathbf{M_2}(t) \begin{bmatrix} \mathbf{x}(t-1) \\ \mathbf{w}(t-1) \end{bmatrix} \leq \mathbf{m_3}(t) \right]. \tag{IV.22}$$

Remark 6. *Note that if the order of the continuous system is large (say the order is over ten) which largely increases the number of constraints and reduces the computation speed when removing variables using Rack, then we can use a hyperrectangle with bounded intervals on continuous state variables to overapproximate $\tilde{X}(\mathbf{y},t)$ and reduce the number of constraints, improving the computation efficiency while introducing some conservatism.*

Theorem 3.1 in [79] demonstrates the effectiveness of the computational implementation of the SVO. This theorem indicates that the real system state at time t is guaranteed to be within $\tilde{X}(\mathbf{y},t)$ if the system is in one of the nominal discrete states.

IV.3.2.3 Central Estimator and Estimation Error

In order to understand the estimation accuracy of the SVO, a central estimator Φ_c is defined in [79], which gives a state-vector with each entry as the mean of the upper bound and the lower bound of the set $\tilde{X}(\mathbf{y},t)$ for each state variable.

$$(\Phi_c \mathbf{y})(t) = \tilde{\mathbf{x}}_c(t), \tag{IV.23}$$

where $\tilde{\mathbf{x}}_c^{(i)}(t) = \frac{1}{2}(\tilde{\mathbf{x}}_{max}^{(i)}(t) + \tilde{\mathbf{x}}_{min}^{(i)}(t))$ and $i = 1, 2, ..., n_x$, $\tilde{\mathbf{x}}_{max}^{(i)}(t) = \max\{\tilde{\mathbf{x}}^{(i)} : \tilde{\mathbf{x}} \in \tilde{X}(\mathbf{y},t)\}$ and $\tilde{\mathbf{x}}_{min}^{(i)}(t) = \min\{\tilde{\mathbf{x}}^{(i)} : \tilde{\mathbf{x}} \in \tilde{X}(\mathbf{y},t)\}$. Theorem 3.2 in [79] demonstrates that the performance of the central estimator is pointwise optimal, which means that the current estimation error is the smallest possible for the current measurement trajectory. Suppose the upper bound of the estimation error of the central estimator is $\boldsymbol{\theta} \in \mathbb{R}^{n_x}$, i.e., $|\mathbf{x}^{(i)}(t) - \tilde{\mathbf{x}}_c^{(i)}(t)| \leq \boldsymbol{\theta}^{(i)}, \forall t$, then we have:

$$\begin{aligned} |\mathbf{x}^{(i)}(t) - \tilde{\mathbf{x}}_{max}^{(i)}(t)| &\leq \boldsymbol{\theta}^{(i)}, \\ |\mathbf{x}^{(i)}(t) - \tilde{\mathbf{x}}_{min}^{(i)}(t)| &\leq \boldsymbol{\theta}^{(i)}, \end{aligned} \tag{IV.24}$$

for all $i = 1, 2, ..., n_x$ [79]. Note that with overapproximation using hyperrectangles, the maximum and minimum estimated state along each state variable are unchanged and $\boldsymbol{\theta}$ is not affected by overapproximation.

For convenience, we define the estimation error $\mathbf{x_e}(t)$ as the difference between the real continuous state $\mathbf{x}(t)$ and the central estimate $\tilde{\mathbf{x}}_c(t)$, i.e., $\mathbf{x_e}(t) = \mathbf{x}(t) - \tilde{\mathbf{x}}_c(t)$.

Remark 7. *Note that the continuous state observer could also be designed using another observer/estimator due to the flexibility of the hybrid observer framework. With a continuous state* *observer which has a better estimation accuracy, we can get a tighter lower bound of the anomalous signal with which the conflict-driven method can provide detection guarantees.*

As mentioned in Section II.2, the SVO delivers a non-empty estimated set if the system is in a nominal discrete state. If the SVO gives an empty estimated set, the system is identified to be in an anomalous discrete state [76]. If the input-output data sequence generated by the anomalous hybrid system can also be generated by the nominal hybrid subsystem, the SVO gives a non-empty estimated set and fails to detect the anomaly. In the following section, we give a classification taxonomy of anomalies in hybrid systems.

IV.4 Types of Anomalies

In this section, we propose a classification taxonomy of anomalies f in hybrid systems considered in this chapter. We only focus on anomalies happening after the nominal hybrid subsystem is in its steady state and assume that:

Assumption 9. *An anomaly f occurs after the discrete state observer enters its steady state, i.e., $t_f \geq t_k$.*

Based on the types of variables that are affected, we classify the anomalies into three different types: Type-C, Type-D, and Type-B anomalies, where C represents "continuous", D represents "discrete", B represents "both".

As mentioned in Section IV.2.4, the anomalous hybrid subsystem should satisfy at least one of two anomaly conditions. Under Type-C anomalies, the continuous dynamics of the anomalous hybrid subsystem are different from those of the nominal hybrid subsystem, satisfying anomaly condition 1. Some Type-C anomalies are easy to detect because the SVO gives an empty estimated continuous state set. But some Type-C anomalies cannot be detected by existing methods because the input-output data sequence of the anomalous discrete state in the anomalous hybrid subsystem satisfies the continuous dynamics of the current estimated nominal discrete state. That is, with the same input and the initial state, the difference of the output of the anomalous hybrid subsystem and the output of the nominal hybrid subsystem is always smaller than a certain threshold, but the difference of the real system states in two subsystems is larger than the upper bound of the estimation error of the nominal hybrid subsystem. An example anomaly can be the anomaly caused by False Data Injection Attack introduced in [61]. We define this type of anomaly as an invisible anomaly:

Definition 14. *A Type-C anomaly f is an invisible anomaly if the input-output data sequence of the anomalous discrete state in the anomalous hybrid subsystem satisfies the continuous dynamics of the current estimated nominal discrete state. Otherwise, we say f is an visible anomaly.*

Based on the definition of visible/invisible anomaly, we additionally classify Type-C anomalies into two different types: Type-C_v and Type-C_{iv} anomalies, where the subscript v represents "visible" and iv represents "invisible". Type-C_v anomalies can be easily detected by existing methods. The utilized hybrid observer contains a SVO. We can use the SVO to help detect Type-C_v anomalies. The SVO gives an empty estimated continuous state set under Type-C_v anomalies. Type-C_{iv} is challenging to detect because the SVO still delivers a non-empty estimated continuous state set at each time step.

Under Type-C anomaly, the norm of the estimation error is increased because the continuous dynamics of the anomalous hybrid subsystem are different from those used for state estimation. If the change of the estimation error is small, the impact of the anomaly is insignificant. We only consider the anomalies under which the norm of the estimation error is larger than the upper bound of the estimation error in nominal hybrid subsystem:

$$\forall t \geq t_f, q \in Q_f \implies ||\mathbf{x_e}(t)|| > ||\boldsymbol{\theta}||. \tag{IV.25}$$

Under Type-D anomalies, the discrete behavior of the anomalous hybrid subsystem is different from that of the nominal hybrid subsystem. In discrete systems, Type-D anomalies are classified into two different types: diagnosable anomalies Type-D_d and undiagnosable anomalies

Type-D_{ud}, where d represents "diagnosable" and ud represents "undiagnosable" [14].

Definition 15. *A Type-D anomaly f is undiagnosable if there exist two discrete event sequences e_A (the subscript represents "anomalous") and e_N (the subscript represents "nominal") in system $\mathcal{M}$ that satisfy the following conditions:*

1. *e_A contains f and e_N does not;*
2. *e_A is of arbitrarily long length after f; and*
3. *$Proj(e_A) = Proj(e_N)$, where operator $Proj(\cdot)$ projects a list of discrete event sequence to observable discrete event sequence.*

When no such pair of strings exists, f is said to be diagnosable in system $\mathcal{M}$.

Diagnosable anomalies can be detected using the discrete state observer because the observed discrete event sequence of the anomalous hybrid subsystem is different from the discrete event sequence of the nominal hybrid subsystem after some time. Because the hybrid system we consider is deterministic (system condition 1) and the nominal hybrid subsystem is known, all of the possible discrete state sequences of the nominal hybrid subsystem are known based on the observed discrete event sequences of the nominal hybrid subsystem. The hybrid observer we use in this paper contains a discrete state observer. We can compare the estimated discrete state sequence with the possible discrete state sequences to detect whether or not a Type-D_d anomaly occurs.

Under some anomalies, both the continuous dynamics and the discrete behavior of the nominal hybrid subsystem are changed. We call them Type-B anomalies. A Type-B anomaly can be considered as a combination of Type-C and Type-D anomalies. As shown in Fig. IV.6, Type-B anomalies can be classified into four different types: Type-$B_{v,d}$, Type-$B_{iv,d}$, Type-$B_{v,ud}$, and Type-$B_{iv,ud}$ anomalies.

Based on our classification taxonomy, there are eight types of anomalies: Type-C_v, Type-C_{iv}, Type-D_d, Type-D_{ud}, Type-$B_{v,d}$, Type-$B_{iv,d}$, Type-$B_{v,ud}$, and Type-$B_{iv,ud}$. For Type-C_v, Type-C_{iv}, Type-D_d, Type-D_{ud}, Type-$B_{v,d}$, Type-$B_{iv,d}$ and Type-$B_{v,ud}$ anomalies, we provide detection guar-

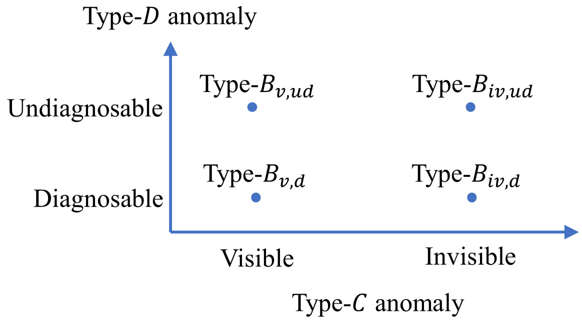


Figure IV.6: Classification taxonomy of anomalies in hybrid systems

antees with the proposed conflict-driven method. For Type-$B_{iv,ud}$ anomaly, the proposed conflict-driven method can detect in some cases but it cannot provide detection guarantees.

IV.5 Conflict-driven Anomaly Detection

In this section, we first give a brief introduction of conflict-driven anomaly detection. Then we give theoretic analysis for anomaly detection. Finally, we talk about the relationship between the conflict types and the anomaly types.

IV.5.1 Method Description

In the conflict-driven method, we define three conflict types and check for the occurrence of the conflicts to detect anomalies. The work flow diagram is shown in Fig. IV.7. Note that this method is used after the hybrid observer is in the steady state, i.e., $t \geq t_k$.

The conflict-driven method has five steps at each time step:

1. Check the estimated discrete state sequence:

After the discrete state observer gives the estimated discrete state $\tilde{q}(t)$ at time t, an updated estimated discrete state sequence can be formed up to time t. Then the method compares the updated estimated discrete state sequence with all of the possible discrete state sequences. If the updated

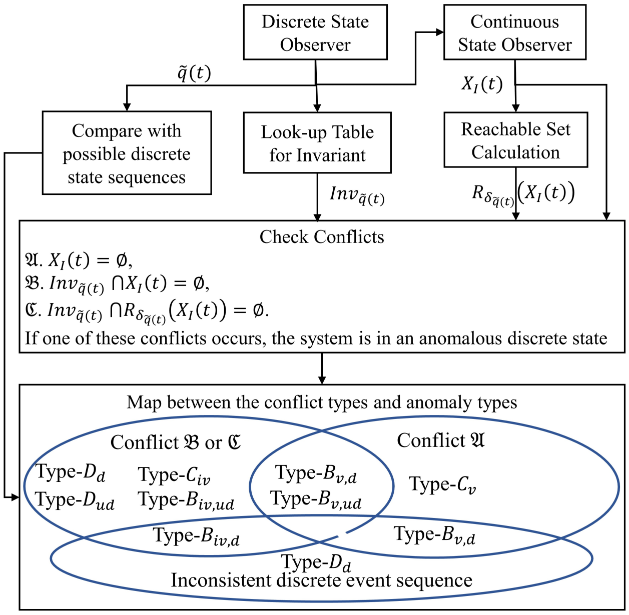

Discrete State Observer
Continuous State Observer
$\tilde{q}(t)$
$X_I(t)$
Compare with possible discrete state sequences
Look-up Table for Invariant
Reachable Set Calculation
$Inv_{\tilde{q}(t)}$
$R_{\delta_{\tilde{q}(t)}}(X_I(t))$
Check Conflicts
$\mathfrak{A}$. $X_I(t) = \emptyset$,
$\mathfrak{B}$. $Inv_{\tilde{q}(t)} \cap X_I(t) = \emptyset$,
$\mathfrak{C}$. $Inv_{\tilde{q}(t)} \cap R_{\delta_{\tilde{q}(t)}}(X_I(t)) = \emptyset$.
If one of these conflicts occurs, the system is in an anomalous discrete state
Map between the conflict types and anomaly types
Conflict $\mathfrak{B}$ or $\mathfrak{C}$
Conflict $\mathfrak{A}$
Type-D_d
Type-D_{ud}
Type-C_{iv}
Type-$B_{iv,ud}$
Type-$B_{v,d}$
Type-$B_{v,ud}$
Type-C_v
Type-$B_{iv,d}$
Type-$B_{v,d}$
Type-D_d
Inconsistent discrete event sequence

Figure IV.7: Conflict-driven anomaly detection

estimated discrete state sequence does not satisfy any of the possible discrete state sequences, an anomaly is detected. Note that for the estimated discrete state sequence, we only use the estimate with a unique estimated discrete state.

2. Calculate an initial set $X_I(t)$:

Given $\tilde{X}(\mathbf{y},t)$ from the SVO, define $X_I(t)$ to be a zonotope[3] that overapproximates $\tilde{X}(\mathbf{y},t)$. A zonotope, which is computationally efficient for reachability analysis of hybrid system [30], is a Minkowski sum of a finite set of line segments and defined as

Definition 16. *Zonotope Z is a set such that:*

$$Z = (\tilde{\mathbf{x}}_c, <\mathbf{g_1}, ..., \mathbf{g_p}>) = \{\mathbf{x} \in \mathbb{R}^{n_x} : \mathbf{x} = \tilde{\mathbf{x}}_c + \Sigma_{i=1}^{i=p} b_i \mathbf{g_i}, -1 \leq b_i \leq 1\}, p \geq n_x, \tag{IV.26}$$

where $\tilde{\mathbf{x}}_c, \mathbf{g_i} \in \mathbb{R}^{n_x}$ *are the center and generators, respectively.*

Both p and n_x determine the maximum number of vertices and facets.

[3] Note that the zonotope $X_I(t)$ may not be unique. In this paper, we use the smallest hyperrectangle to overapproximate $\tilde{X}(\mathbf{y},t)$ and express the hyperrectangle in the form of zonotope. In this way, the zonotope $X_I(t)$ is unique.

3. Calculate the reachable set $R_{\delta_{\tilde{q}(t)}}(X_I(t))$:

   The $\delta_{\tilde{q}(t)}$ time-step forward reachable set $R_{\delta_{\tilde{q}(t)}}(X_I(t))$ starting from $X_I(t)$ is defined as

$$\begin{aligned} R_{\delta_{\tilde{q}(t)}}(X_I(t)) := \{\mathbf{x_R} \in \mathbb{R}^{n_x} : \exists \mathbf{x}_{(t_0, t_0+\delta_{\tilde{q}(t)})}, \mathbf{u}_{(t_0, t_0+\delta_{\tilde{q}(t)}-1)}, \\ \mathbf{x}(t+1) = \mathbf{A}_{\tilde{q}(t)}\mathbf{x}(t) + \mathbf{B}_{\tilde{q}(t)}\mathbf{u}(t), ||\mathbf{u}|| \leq \mu, \\ (\mathbf{x}(t_0) \in X_I(t)) \wedge (\mathbf{x}(t_0 + \delta_{\tilde{q}(t)}) = \mathbf{x_R})\}, \end{aligned} \tag{IV.27}$$

   where $\mathbf{x_R}$ is continuous state, $\mathbf{x}_{(t_0, t_0+\delta_{\tilde{q}(t)})}$ is the trajectory of the continuous state from time t_0 to $t_0 + \delta_{\tilde{q}(t)}$, $\mathbf{u}_{(t_0, t_0+\delta_{\tilde{q}(t)}-1)}$ is the trajectory of the input signal from time t_0 to $t_0 + \delta_{\tilde{q}(t)} - 1$, and the time-step $\delta_{\tilde{q}(t)} \in \mathbb{Z}_{\geq 0}$ is determined off-line and will be introduced later. The reachable set satisfies

$$R_{\delta_{\tilde{q}(t)}}(X_I(t)) \subseteq \mathbf{A}_{\tilde{q}(t)}^{\delta_{\tilde{q}(t)}} X_I(t) + \square \sigma_{\tilde{q}(t)}, \tag{IV.28}$$

   where $\sigma_{\tilde{q}(t)} = \frac{1-||\mathbf{A}_{\tilde{q}(t)}||^{\delta_{\tilde{q}(t)}}}{1-||\mathbf{A}_{\tilde{q}(t)}||}(||\mathbf{B}_{\tilde{q}(t)}||\mu + w)$. Refer to [30] for more details about reachable set calculation using zonotopes.

4. Check for conflicts:

   We define three conflict types in this paper, as shown in Fig. IV.8.

   Conflict $\mathfrak{A}$. The initial set is an empty set, i.e., $X_I(t) = \emptyset$. This is equivalent to $\tilde{X}(\mathbf{y}, t) = \emptyset$.

   Conflict $\mathfrak{B}$. The initial set has no intersection with the invariant of the estimated discrete state ($X_I(t) \cap Inv_{\tilde{q}(t)} = \emptyset$).

   Conflict $\mathfrak{C}$. The $\delta_{\tilde{q}(t)}$ time steps forward reachable set has no intersection with the invariant of the estimated discrete state, i.e., $R_{\delta_{\tilde{q}(t)}}(X_I(t)) \cap Inv_{\tilde{q}(t)} = \emptyset$.

   If one of these conflicts occurs, the system is in an anomalous discrete state.

5. Map conflict types to anomaly types:

   Based on the observed conflict types and estimated discrete state sequence, we can determine the possible types of anomalies (The detailed mapping is shown in Fig. IV.12 in Section IV.5.3).

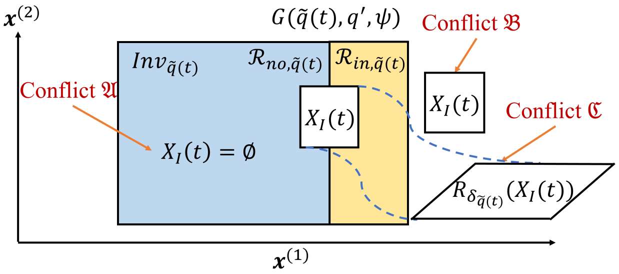


Figure IV.8: Three conflict types shown in a $2-D$ system. The large rectangle (with blue and yellow regions) is the invariant $Inv_{\tilde{q}(t)}$ with guard $G(\tilde{q}(t), q', \psi)$. The blue region is the normal operating region $\mathcal{R}_{no,\tilde{q}(t)}$ and the yellow region is the intermediate region $\mathcal{R}_{in,\tilde{q}(t)}$. $X_I(t)$ is the initial set at time step t (three $X_I(t)$s in the figure show the three different conflict types). $R_{\delta_{\tilde{q}}}(X_I(t))$ is the $\delta_{\tilde{q}}$-reachable set starting from the initial set $X_I(t)$

Note that for Step 3, we need to appropriately determine $\delta_{\tilde{q}(t)}$ for each discrete state *offline* to avoid false alarms[4] and provide detection guarantees according to the following two steps:

1. In Inv_q, starting from the intersection of the invariant Inv_q and the hyperplane corresponding to the i^{th} guard $G(q, q_i, \psi_i)$ as defined by (IV.5) and Inv_q, we find the minimum time steps $\delta_{q,i}$ which satisfies

$$R_{\delta_{q,i}+1}(\mathcal{P}(q, q_i, \psi_i) \cap Inv_q) \cap \mathcal{L}(q, q_i, \psi_i) \neq \emptyset. \tag{IV.29}$$

If (IV.29) does not have a solution, it means that Conflict $\mathfrak{C}$ never occurs when the real continuous state is approaching to $\mathcal{P}(q, q_i, \psi_i) \cap Inv_q$. Then we set $\delta_{q,i} = 0$. Note that $\delta_{q,i}$ may be different for different guards in the same discrete state. The reason we use $\delta_{q,i} + 1$ is that the continuous system is a discrete-time model and we want to ensure the $\delta_{q,i}$ time-step forward reachable set, starting from any possible real continuous state when a transition occurs, has intersection with Inv_q in nominal discrete state q.

2. Let $\delta_q = \min_i(\delta_{q,i})$. Note that δ_q may be 0. Then we only need to check Conflicts $\mathfrak{A}$ and $\mathfrak{B}$ in discrete state q.

[4] If $\delta_{\tilde{q}(t)}$ is too large, the reachable set could be completely outside the invariant, causing false alarms.

IV.5.2 Anomaly Detection

In this subsection, we demonstrate the effectiveness of the conflict-driven method in the detection of different types of anomalies: Type-C, Type-D and Type-B anomalies. To simplify the discussion, we focus on representing the anomaly as an additive anomaly on the output equation $\boldsymbol{\gamma_2}$.

IV.5.2.1 Type-C Anomaly

Under Type-C_v anomaly, Conflict $\mathfrak{A}$ occurs if the norm of the anomalous signal exceeds a certain threshold, which is given in Proposition 3. The occurrence of Conflict $\mathfrak{A}$ indicates that the constraints of the set of estimated continuous states $\tilde{X}(\mathbf{y},t)$ described in (IV.20) are infeasible. That is, if the norm of the anomalous signal is larger than the lower bound, then at least one of the conditions in $(\tilde{\mathbf{M}}(t), \tilde{\mathbf{m}}(t))$ in (IV.20) is violated. Among the constraints in $(\tilde{\mathbf{M}}(t), \tilde{\mathbf{m}}(t))$, the constraints in $(\hat{\mathbf{M}}(t), \hat{\mathbf{m}}(t))$ (IV.16) describing $\hat{X}(\mathbf{y},t)$ are directly related to the anomalous signal $\boldsymbol{\gamma_2}$ based on anomaly model (IV.12). Therefore, we can get the lower bound of the anomalous signal $\boldsymbol{\gamma_2}$ (IV.31) by violating the constraints corresponding to $\hat{X}(\mathbf{y},t)$, that is,

$$\hat{\mathbf{M}}(t)\mathbf{x}(t) > \hat{\mathbf{m}}(t). \tag{IV.30}$$

Now we can state Proposition 3 as follows.

Proposition 3. *Given a hybrid automaton $\mathcal{H}$ and assume the nominal hybrid automaton $\mathcal{H}_n$ is available to build a hybrid observer $\mathcal{O}$, if a Type-C_v anomaly occurs at time $t_f > t_k$ satisfying* (IV.31)*, then Conflict $\mathfrak{A}$ occurs.*

$$||\boldsymbol{\Gamma_2}\boldsymbol{\gamma_2}(t)|| > 2v. \tag{IV.31}$$

Proof. If the anomalous signal $\boldsymbol{\gamma}_{\mathbf{2}}(t)$ satisfies (IV.31), then

$$\|\mathbf{y}(t) - \mathbf{C}_q\mathbf{x}(t) - \mathbf{v}(t)\| > 2v. \tag{IV.32}$$

Let the i^{th} entry of $\mathbf{y}(t) - \mathbf{C}_q\mathbf{x}(t) - \mathbf{v}(t)$ equal to $\|\mathbf{y}(t) - \mathbf{C}_q\mathbf{x}(t) - \mathbf{v}(t)\|$. Then we have the following two possibilities:

1) If $\mathbf{y}^{(i)}(t) - \mathbf{C}_q^{(i,:)}\mathbf{x}(t) - \mathbf{v}^{(i)}(t) > 0$, then

$$\mathbf{y}^{(i)}(t) - \mathbf{C}_q^{(i,:)}\mathbf{x}(t) - \mathbf{v}^{(i)}(t) > 2v. \tag{IV.33}$$

By rearranging, we can get

$$\begin{aligned} -\mathbf{C}_q^{(i,:)}\mathbf{x}(t) &> -\mathbf{y}^{(i)}(t) + \mathbf{v}^{(i)}(t) + 2v \\ -\mathbf{C}_q^{(i,:)}\mathbf{x}(t) &> -\mathbf{y}^{(i)}(t) + v. \end{aligned} \tag{IV.34}$$

2) Similarly, if $\mathbf{y}^{(i)}(t) - \mathbf{C}_q^{(i,:)}\mathbf{x}(t) - \mathbf{v}^{(i)}(t) < 0$, then we have

$$\mathbf{C}_q^{(i,:)}\mathbf{x}^{(i)}(t) > \mathbf{y}^{(i)}(t) + v. \tag{IV.35}$$

Then, (IV.20) is infeasible and Conflict $\mathfrak{A}$ occurs, which means the input-output sequence is inconsistent with the nominal hybrid subsystem. □

As discussed before, the continuous variables are affected under Type-C_{iv} anomaly, but the SVO can still provide non-empty estimated continuous state set. In order to detect this type of anomaly, we utilize the estimated states from both continuous and discrete state observers, and take advantage of observation of a discrete event. This enables us to employ the contradictions among estimated continuous and discrete states and the model parameters such as guards and invariants to detect these challenging anomalies. These contradictions are formalized in Conflicts

$\mathfrak{B}$ and $\mathfrak{C}$. In what follows, we set the stage to present the main theorem-Theorem 3. This theorem provides sufficient conditions on the lower bound of the anomalous signal under which the conflict-driven method is guaranteed to detect Type-C_{iv} anomalies. This lower bound is smaller than the one presented in our previous work [86]. Similar to [86], we first find the lower bound of the estimation error that creates either Conflict $\mathfrak{B}$ or $\mathfrak{C}$, and then relate this bound to the lower bound on the anomalous signal.

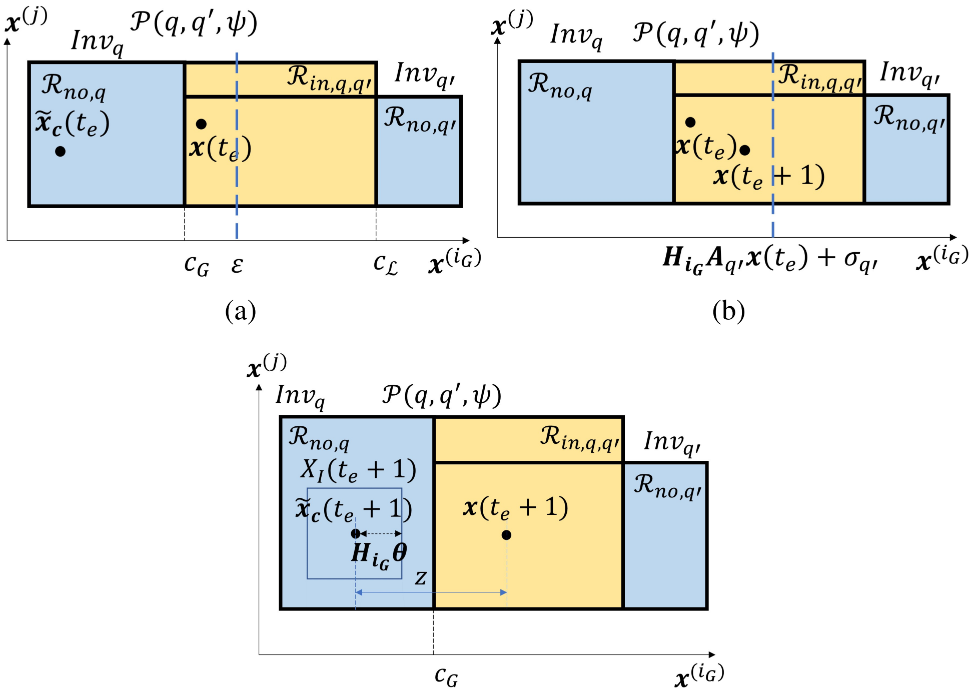
Suppose a Type-C_{iv} anomaly occurs at time t_f under which there is a large estimation error on the i_G^{th} state variable, i.e., $|\mathbf{x_e}^{(i_G)}| > \boldsymbol{\theta}^{(i_G)}$, and a discrete event ψ occurs at time $t_e > t_f$ which associates a guard with condition on the i_G^{th} state variable, i.e., $\{\mathbf{x} \in Inv_q : s_G \mathbf{x}^{(i_G)} \geq c_G\}$. Without loss of generality, we assume that the projection of $\mathcal{R}_{no,q}$ onto the i_G^{th} state variable is bounded above by c_G, i.e., $\mathbf{H}_{i_G}\mathcal{R}_{no,q} \leq c_G$ (because $s_G = 1$), where $\mathbf{H}_{i_G} \in \mathbb{R}^{n_x}$ is the projection row vector with the i_G^{th} entry “1” and “0” elsewhere. The procedure for the case where $\mathbf{H}_{i_G}\mathcal{R}_{no,q} \geq -c_G$ (because $s_G = -1$) is identical. When this event occurs, transitioning the system from q to q', we can only have two possibilities for the central estimate at time t_e, either $\tilde{\mathbf{x}}_\mathbf{c}(t_e) \in \mathcal{R}^o_{no,q}$, or $\tilde{\mathbf{x}}_\mathbf{c}(t_e) \in \mathcal{R}^o_{in,q,q'}$. Based on the definitions of guard, invariant, post-guard hyperplane, and Assumption 7, along the i_G^{th} state variable the upper bound of Inv_q is $c_\mathcal{L}$ and the lower bound of $Inv_{q'}$ is c_G. For brevity in notation and as in this section we mainly consider $G(q,q',\psi)$, we refer to it as G.

As mentioned above, the central estimate $\tilde{\mathbf{x}}_\mathbf{c}(t_e)$ is either in the normal operating region or the intermediate region when an observable event occurs. Consider the first possibility where $\tilde{\mathbf{x}}_\mathbf{c}(t_e) \in \mathcal{R}^o_{no,q}$, that is, when the real continuous state satisfies the guard, the central estimate is in the normal operating region of discrete state q. The visualization of this case is shown in Fig. IV.9.

The goal is to find the lower bound of the estimation error along the i_G^{th} state variable, such that:

- The initial set $X_I(t_e+1)$ has no intersection with $Inv_{q'}$.

We denote such minimum estimation error corresponding to G by z_G^*. To find z_G^*, it suffices to find the minimum z such that for all $\tilde{\mathbf{x}}_\mathbf{c}(t_e+1)$ the upper bound of $X_I(t_e+1)$ is smaller than the lower



(a)

(b)

(c)

Figure IV.9: Visualization in $2-D$ when $\tilde{\mathbf{x}}(t_e) \in \mathcal{R}^o_{no,q}$ and Conflict $\mathfrak{B}$ occurs under Type-C_{iv} anomaly: (a) At time t_e, $c_G \leq \mathbf{H}_{i_G} < \varepsilon$, discrete event ψ occurs. (b) At time t_e+1, the discrete state is changed to q' and the real continuous state is evolved to $\mathbf{x}(t_e+1)$, satisfying the continuous dynamics described by state matrices $(\mathbf{A}_{q'}, \mathbf{B}_{q'})$. Thus, $\mathbf{x}(t_e+1) \in R_1(\mathbf{x}(t_e))$ and $\mathbf{H}_{i_G}R_1(\mathbf{x}(t_e)) \leq \mathbf{H}_{i_G}\mathbf{A}_{q'}\mathbf{x}(t_e)+\sigma_{q'}$

bound of $Inv_{q'}$ along the i_G^{th} state variable,

$$\mathbf{H}_{i_G}\tilde{\mathbf{x}}_{\mathbf{c}}(t_e+1)+\mathbf{H}_{i_G}\boldsymbol{\theta} < c_G. \tag{IV.36}$$

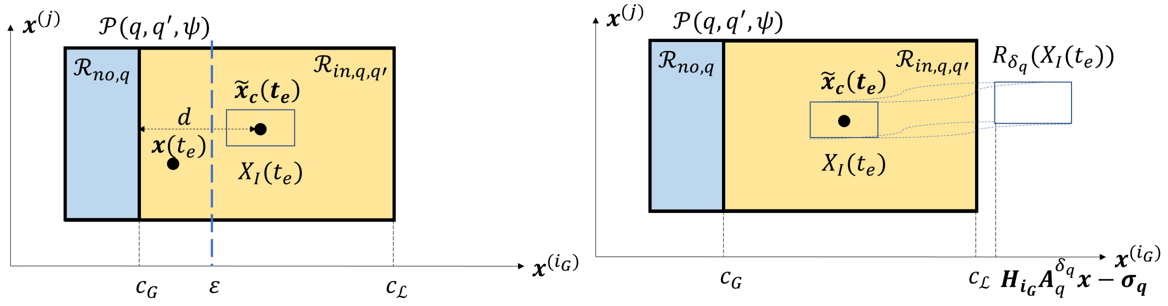
Note that at time t_e, the continuous state of the system along the i_G^{th} state variable $\mathbf{H}_{i_G}\mathbf{x}(t_e)$ is greater than or equal to c_G, i.e., $\mathbf{H}_{i_G}\mathbf{x}(t_e) \geq c_G$. Meanwhile, $\mathbf{H}_{i_G}\mathbf{x}(t_e)$ is smaller than the maximum value of the one time step forward reachable set from $\mathcal{P}(q,q',\psi)\cap Inv_q$ along the i_G^{th} state variable, i.e., $\mathbf{H}_{i_G}\mathbf{x}(t_e) < \epsilon$, where $\epsilon = \max(\mathbf{H}_{i_G}R_1(\mathcal{P}(q,q',\psi)\cap Inv_q))$, as shown in Fig. IV.9a. After the occurrence of event ψ, the estimated discrete state is changed to q' at time t_e+1. If the state matrices of the anomalous discrete state are still $(\mathbf{A}_q,\mathbf{B}_q)$, which are different from the estimated state matrices $(\mathbf{A}_{q'},\mathbf{B}_{q'})$, then the Type-$C_{iv}$ anomaly becomes Type-C_v anomaly which can be detected by observing the occurrence of Conflict $\mathfrak{A}$. If the state matrices of the anomalous discrete state are changed to $(\mathbf{A}_{q'},\mathbf{B}_{q'})$ after the occurrence of event ψ, then the SVO still provides non-empty set of the estimated continuous states under the Type-C_{iv} anomaly. In the following we consider the case that the state matrices of the anomalous discrete state are changed to $(\mathbf{A}_{q'},\mathbf{B}_{q'})$ after the occurrence of event ψ. As shown in Fig. IV.9b, the set of all possible continuous states at time t_e+1 can be represented by:

$$\begin{aligned}&\forall\mathbf{x}(t_e)\in Inv_q, c_G \leq \mathbf{H}_{i_G}\mathbf{x}(t_e) < \epsilon,\\ &\qquad\qquad \mathbf{x}(t_e+1)\in R_1(\mathbf{x}(t_e)) \subseteq \mathbf{A}_{q'}\mathbf{x}(t_e)+\Box\sigma_{q'},\end{aligned} \tag{IV.37}$$

where $\sigma_{q'} = ||\mathbf{B}_{q'}||\mu + w$.

Combining (IV.36), (IV.37) and the fact that $z = \mathbf{H}_{i_G}\mathbf{x}-\mathbf{H}_{i_G}\tilde{\mathbf{x}}_{\mathbf{e}}$, we can pose the problem of finding z_G^* as a robust optimization problem.

$$\begin{aligned} z_G^* = &\min_z \quad z\\ &\text{s. t.} \quad z\geq 0,\ z\geq \mathbf{H}_{i_G}\mathbf{A}_{q'}\mathbf{x}+\sigma_{q'}+\mathbf{H}_{i_G}\boldsymbol{\theta}-c_G\\ &\qquad\quad \forall\mathbf{x}\in Inv_{q'}, c_G\leq \mathbf{H}_{i_G}\mathbf{x}\leq\epsilon,\end{aligned} \tag{IV.38}$$



(a)

(b)

Figure IV.10: Visualization in 2D when $\tilde{\mathbf{x}}_\mathbf{c}(t_e) \in \mathcal{R}^o_{in,q,q'}$ and Conflict $\mathfrak{C}$ occurs under Type-C_{iv} anomaly: (a) At time t_e, the maximum estimation error along $\mathbf{x}^{(i_G)}$ state variable is d, i.e., $d = |\mathbf{H}_{i_G}\tilde{\mathbf{x}}_\mathbf{c}(t_e) - c_G|$. (b) The reachable set at time t_e is $R_{\delta_q}(X_I(t_e))$ and its lower bound is $\mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q}\mathbf{x} - \sigma_q$

By utilizing robust optimization method [7], we can convert (IV.38) to a linear programming problem as follows:

$$\begin{aligned} z_G^* = & \min_{\mathbf{J},z} \quad z \\ & \text{s. t.} \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix} z - \begin{bmatrix} \mathbf{J}^\intercal \boldsymbol{\rho}_\mathbf{1} \\ 0 \end{bmatrix} \geq \begin{bmatrix} \sigma_{q'} + \mathbf{H}_{i_G}\boldsymbol{\theta} - c_G \\ 0 \end{bmatrix} \\ & \boldsymbol{\Lambda}^\intercal \mathbf{J} \geq (\mathbf{H}_{i_G}\mathbf{A}_{q'})^\intercal, \mathbf{J} \geq \mathbf{0}, \end{aligned} \tag{IV.39}$$

where $\mathbf{0} \in \mathbb{R}^{2n\times 1}$ is a zero vector. $\mathbf{x}$ is in a polytopic uncertain set, i.e., $\boldsymbol{\Lambda}\mathbf{x} \leq \boldsymbol{\rho}_\mathbf{1}$ for problem (IV.38), where $\boldsymbol{\Lambda} \in \mathbb{R}^{2n\times n}$, $\boldsymbol{\rho}_\mathbf{1} \in \mathbb{R}^{2n\times 1}$ and $\mathbf{J} \in \mathbb{R}^{2n\times 1}$ is a variable of the optimization problem.

For the second possibility, i.e., $\tilde{\mathbf{x}}_\mathbf{c}(t_e) \in \mathcal{R}^o_{in,q,q'}$, we are seeking the lower bound of the estimation error along the i_G^{th} state variable such that it satisfies the following:

- The reachable set for δ_q time steps from any point within the initial set $X_I(t_e)$ has no intersection with Inv_q.

The visualization of this case is shown in Fig. IV.10. Considering the worst case that the continuous state is the furthest to the upper bound of ∂Inv_q along the i_G^{th} state variable, i.e., $\mathbf{H}_{i_G}\mathbf{x}(t_e) = c_G$, our objective can be equivalently changed to find the minimum distance between c_G and $\mathbf{H}_{i_G}\tilde{\mathbf{x}}_\mathbf{c}(t_e)$. We denote this minimum distance by d_G^*. Define $d = |\mathbf{H}_{i_G}\tilde{\mathbf{x}}_\mathbf{c}(t_e) - c_G|$ as the distance between $\mathcal{P}(q,q',\psi)$

and the estimated state along the i_G^{th} state variable, which is also the maximum estimation error at time t_e given the central estimated continuous state $\tilde{\mathbf{x}}_{\mathbf{c}}(t_e)$. With this definition, the initial set at time t_e can be represented as $X_I(t_e) = \{\mathbf{x} : \mathbf{H}_{i_G}\mathbf{x} \in [c_G + d - \mathbf{H}_{i_G}\boldsymbol{\theta}, c_G + d + \mathbf{H}_{i_G}\boldsymbol{\theta}]\}$ as shown in Fig. IV.10a. Starting from this initial set $X_I(t_e)$, the projection of the reachable set for δ_q time steps forward onto the i_G^{th} state variable becomes $\mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q}\mathbf{x} \pm \sigma_q$, $\forall \mathbf{x} \in X_I(t_e)$, where $\sigma_q = \frac{1-||\mathbf{A}_q||^{\delta_q}}{1-||\mathbf{A}_q||}(||\mathbf{B}_q||\mu + w)$. As shown in Fig. IV.10b, if the lower bound of the reachable set $R_{\delta_q}(X_I(t_e))$ is greater than $c_{\mathcal{L}}$, i.e., $\mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q}\mathbf{x} - \sigma_q > c_{\mathcal{L}}$, $\forall \mathbf{x} \in X_I(t_e)$, then it is guaranteed that the δ_q time-step forward reachable set starting from this initial set $X_I(t_e)$ has no intersection with the invariant Inv_q. We can pose the problem of finding d_G^* as the following robust optimization problem.

$$\begin{aligned} d_G^* = &\min_{d} \quad d \\ &\text{s. t.} \quad d \geq 0,\ \mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q}\mathbf{x} - \sigma_q \geq c_{\mathcal{L}} \\ &\qquad\quad \forall \mathbf{x} \in Inv_q, \mathbf{x} \in X_I(t_e). \end{aligned} \tag{IV.40}$$

With a change of variables and by employing the robust optimization techniques [7], we can write an equivalent problem to (IV.40) as a linear program.

$$\begin{aligned} d_G^* = &\min_{\mathbf{D},\mathbf{J}} \quad \mathbf{H}_{i_G}\mathbf{D} \\ &\text{s. t.} \quad \begin{bmatrix} \mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q} \\ \mathbf{H}_{i_G} \end{bmatrix}\mathbf{D} - \begin{bmatrix} \mathbf{J}^{\mathsf{T}}\boldsymbol{\rho}_{\mathbf{2}} \\ 0 \end{bmatrix} \geq \begin{bmatrix} \sigma_q + c_{\mathcal{L}} \\ 0 \end{bmatrix} \\ &\qquad\quad \boldsymbol{\Lambda}^{\mathsf{T}}\mathbf{J} \geq -(\mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q})^{\mathsf{T}}, \mathbf{J} \geq \mathbf{0}, \mathbf{D} \geq \mathbf{0}, \end{aligned} \tag{IV.41}$$

where $\mathbf{0}$ is a zero vector with proper dimension, and $\mathbf{D} \in \mathbb{R}^n$ is a vector with the i_G^{th} entry d and other entries “0”. $\mathbf{x}$ is in a polytopic uncertain set, i.e., $\boldsymbol{\Lambda}\mathbf{x} \leq \boldsymbol{\rho}_{\mathbf{2}}$, where $\boldsymbol{\rho}_{\mathbf{2}} \in \mathbb{R}^{2n\times 1}$ and $\mathbf{J} \in \mathbb{R}^{2n\times 1}$ is the dual variable.

Now that we have introduced z_G^* and d_G^*, we can present the main result of the paper.

Theorem 3. *Given a hybrid automaton $\mathcal{H}$ with the nominal hybrid automaton $\mathcal{H}_n$ available to* *build a hybrid observer $\mathcal{O}$. Suppose a Type-C_{iv} anomaly f occurs at time t_f. If an event $\psi \in \Psi_o$ occurs at $t_e > t_f$, which is supposed to transit the system from discrete state q to q', and the guard $G(q,q',\psi)$ is a condition on the real continuous state which is affected by the anomaly f, i.e., $G(q,q',\psi) : s_G\mathbf{x}^{(i_G)} \geq c_G$ and $|\mathbf{x_e}^{(i_G)}| \geq \boldsymbol{\theta}^{(i_G)}$, then the conflict-driven method is guaranteed to detect the anomaly, if the anomaly f satisfies:*

$$||\boldsymbol{\Gamma_2\gamma_2}(t)|| > \max(||\mathbf{C}_q||z_q^* + ||\mathbf{C}_q\boldsymbol{\theta}|| + 2v, ||\mathbf{C}_q||d_q^* + ||\mathbf{C}_q\boldsymbol{\theta}|| + 2v), \tag{IV.42}$$

where $z_q^* = \max_{q'} z_G^*$ and $d_q^* = \max_{q'} d_G^*$ can be derived by solving the robust optimization problems (IV.38) and (IV.40), respectively for all possible q'.

Proof. The solution z_G^* is the lower bound of the estimation error which ensures $X_I(t_e+1) \cap Inv_{q'} = \emptyset$, i.e. Conflict $\mathfrak{B}$. The value of z_G^* varies from one guard to another. Therefore, by considering z_q^*, we guarantee that at the discrete state q, regardless of guard, Conflict $\mathfrak{B}$ occurs, if the estimation error is larger than z_q^*. On the other hand, the solution d_G^* is the lower bound of the estimation error, which ensures $R_{\delta_q}(X_I(t_e)) \cap Inv_q = \emptyset$, i.e., Conflict $\mathfrak{C}$. The value of d_G^* varies for different guards, hence, we similarly take the maximum of these values for all possible q', which is d_q^*. Then it is guaranteed that if the estimation error is larger than d_q^*, regardless of guard, Conflict $\mathfrak{C}$ occurs. By combining the two conditions, we can conclude that if the estimation error $\mathbf{x_e}(t)$ is larger than the maximum value of z_q^* and d_q^*, then Conflict $\mathfrak{B}$ or C is guaranteed to occur.

With the lower bound of the estimation error, we can get the lower bound of the anomalous signal of Type-C_{iv} anomaly based on (IV.12) such that Conflict $\mathfrak{B}$ or C is guaranteed to occur.

Since the SVO is still giving non-empty estimated set under Type-C_{iv} anomaly,

$$||\mathbf{y}(t) - \tilde{\mathbf{y}}_c(t)|| \leq ||\mathbf{C}_q\boldsymbol{\theta}|| + v, \tag{IV.43}$$

where $\tilde{\mathbf{y}}_c(t) = \mathbf{C}_q\tilde{\mathbf{x}}_c(t)$ is the estimated output based on the central estimate. Then we have

$$
\begin{aligned}
||\mathbf{C}_q\mathbf{x}(t) + \mathbf{v}(t) + \mathbf{\Gamma_2}\boldsymbol{\gamma_2}(t) - \mathbf{C}_q\tilde{\mathbf{x}}_c(t)|| &\leq ||\mathbf{C}_q\boldsymbol{\theta}|| + v \\
||\mathbf{C}_q\mathbf{x_e}(t) + \mathbf{v}(t) + \mathbf{\Gamma_2}\boldsymbol{\gamma_2}(t)|| &\leq ||\mathbf{C}_q\boldsymbol{\theta}|| + v \qquad \text{(IV.44)} \\
||\mathbf{\Gamma_2}\boldsymbol{\gamma_2}(t)|| - ||\mathbf{C}_q\mathbf{x_e}(t)|| - ||\mathbf{v}(t)|| &\leq ||\mathbf{C}_q\boldsymbol{\theta}|| + v.
\end{aligned}
$$

Then we have

$$
||\mathbf{C}_q\mathbf{x_e}(t)|| \geq ||\mathbf{\Gamma_2}\boldsymbol{\gamma_2}(t)|| - ||\mathbf{v}(t)|| - ||\mathbf{C}_q\boldsymbol{\theta}|| - v. \qquad \text{(IV.45)}
$$

Based on (IV.42), it is clear that

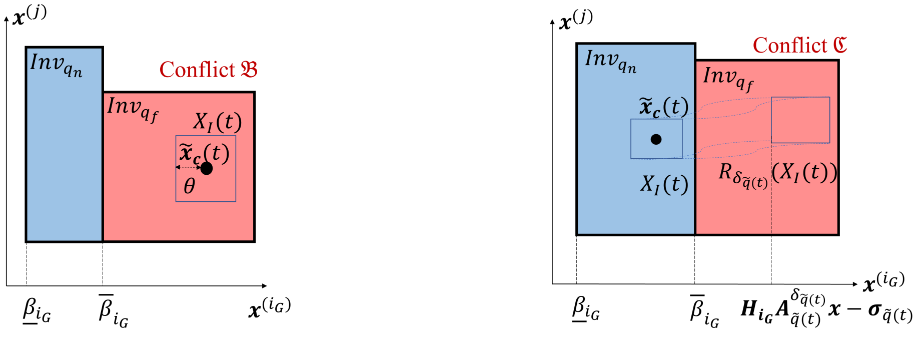
$$
\begin{aligned}
||\mathbf{C}_q||||\mathbf{x_e}(t)|| &\geq \max(||\mathbf{C}_q||z_q^*, ||\mathbf{C}_q||d_q^*) \\
||\mathbf{x_e}(t)|| &\geq \max(z_q^*, d_q^*).
\end{aligned} \qquad \text{(IV.46)}
$$

Therefore the conflict-driven method provides guarantees on the detection of anomalous signals that satisfy condition (IV.42), regardless of where the estimated state is located in the Inv_q at the time of event. This concludes the proof. □

Remark 8. *Note that with the SVO as the continuous state observer, the value of z_q^* and d_q^* are smaller than those in [86]. Therefore, the lower bound of the anomalous signal with which the conflict-driven method can provide detection guarantee is smaller than that in [86].*

IV.5.2.2 Type-*D* Anomaly

Under Type-*D* anomaly, the continuous dynamics of the anomalous discrete state are the same as those of one of the nominal discrete states. There are two cases under a Type-*D* anomaly: 1) The continuous dynamics are the same as those of the current estimated discrete state; 2) The continuous dynamics are different from those of the current estimated discrete state. Based on the difference between the anomalous hybrid subsystem and the nominal hybrid subsystem mentioned



(a)

(b)

Figure IV.11: Visualization in $2-D$ under Type-D anomaly: (a) Conflict 𝔅 occurs under the case that $Inv^o_{q_f} \cap Inv^o_{q_n} = \emptyset$; (b) Conflict ℭ occurs under the case that $Inv^o_{q_f} \cap Inv^o_{q_n} = \emptyset$

in Section IV.2.4, the invariant of at least one of the anomalous discrete states is different from the estimated discrete state under the first case. Therefore, it is possible that the initial set or the reachable set does not intersect with the invariants of the estimated discrete states and Conflict 𝔅 or ℭ occurs as shown in Fig. IV.11. The conflict-driven method can guarantee to detect the first case of Type-D_{ud} anomalies if the open set of the invariants of the anomalous discrete states do not intersect with those of the nominal discrete states, i.e., $Inv^o_{q_n} \cap Inv^o_{q_f} = \emptyset$, as demonstrated in Proposition 4. Under the second case, Conflict 𝔄 will occur because the continuous dynamics of the anomalous discrete states are different from the continuous dynamics of the estimated discrete states, which is demonstrated in Proposition 5.

Proposition 4. *Given a hybrid automaton $\mathcal{H}$ and assume the nominal hybrid automaton $\mathcal{H}_n$ is available to build a hybrid observer $\mathcal{O}$. Suppose a Type-D anomaly f occurs such that the discrete state is changed from $q_n \in Q_n$ to $q_f \in Q_f$, then the conflict-driven method is guaranteed to detect the anomaly if $Inv^o_{q_n} \cap Inv^o_{q_f} = \emptyset$ and one of the following is true:*

$$\begin{aligned} &\exists i \in [1, n_x] \\ &\min_{\mathbf{x}\in Inv_{\tilde{q}(t)}} \|\mathbf{x}^{(i)} - \tilde{\mathbf{x}}^{(i)}_{\mathbf{c}}(t)\| \geq \boldsymbol{\theta}^{(i)}, \end{aligned} \tag{IV.47}$$

or

$$\exists i_G \in [1, n_x], \forall \mathbf{x} \in X_I(t)$$
$$\mathbf{H}_{i_G}\mathbf{A}_{\tilde{q}(t)}^{\delta_{\tilde{q}(t)}}\mathbf{x} - \sigma_{\tilde{q}(t)} > \overline{\beta}_{i_G}$$
$$\text{or} \tag{IV.48}$$
$$\mathbf{H}_{i_G}\mathbf{A}_{\tilde{q}(t)}^{\delta_{\tilde{q}(t)}}\mathbf{x} + \sigma_{\tilde{q}(t)} < \underline{\beta}_{i_G}.$$

Proof. Before Type-D anomaly occurs, the estimated discrete state $\tilde{q}(t)$ is the nominal discrete state q_n, i.e., $\tilde{q}(t) = q_n$. Under Type-D anomaly and the case that the continuous dynamics of the anomalous discrete state are the same as the continuous dynamics of the current estimated nominal discrete state, we have $|\mathbf{x_e}(t)| \leq \boldsymbol{\theta}$. If (IV.47) is true, then the initial set $X_I(t)$ with the central estimate $\tilde{\mathbf{x}}_c(t)$ does not have any intersection with $Inv_{\tilde{q}(t)}$, i.e., $Inv_{\tilde{q}(t)} \cap X_I(t) = \emptyset$. Conflict $\mathfrak{B}$ occurs and the anomaly is detected. In (IV.48), $\mathbf{H}_{i_G}\mathbf{A}_{\tilde{q}(t)}^{\delta_{\tilde{q}(t)}}\mathbf{x} - \sigma_{\tilde{q}(t)}$ is the lower bound of the reachable set along the i_G^{th} state variable and $\mathbf{H}_{i_G}\mathbf{A}_{\tilde{q}(t)}^{\delta_{\tilde{q}(t)}}\mathbf{x} + \sigma_{\tilde{q}(t)}$ is the upper bound of the reachable set along the i_G^{th} state variable. If (IV.48) is true, then the reachable set $R_{\delta_{\tilde{q}(t)}}(X_I(t))$ has no intersection with the invariant $Inv_{\tilde{q}(t)}$. Conflict $\mathfrak{C}$ occurs and the anomaly is detected. □

Proposition 5. *Given a hybrid automaton $\mathcal{H}$ and assume the nominal hybrid automaton $\mathcal{H}_n$ is available to build a hybrid observer $\mathcal{O}$. Suppose a Type-D anomaly f occurs such that the discrete state is changed from $q_n \in Q_n$ to $q_f \in Q_f$ and $(\mathbf{A}_{q_n}, \mathbf{B}_{q_n})$ of q_n is different from $(\mathbf{A}_{q_f}, \mathbf{B}_{q_f})$ of q_n, then the conflict-driven method is guaranteed to detect the anomaly if the following is infeasible:*

$$\begin{bmatrix} \mathbf{I} & -\mathbf{A}_{q_n} & -\mathbf{I} \\ -\mathbf{I} & \mathbf{A}_{q_n} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I} \\ \mathbf{0} & -\mathbf{C}_{q_f}\mathbf{A}_{q_f} & -\mathbf{C}_{q_f} \\ \mathbf{0} & \mathbf{C}_{q_f}\mathbf{A}_{q_f} & \mathbf{C}_{q_f} \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{x}(t-1) \\ \mathbf{w}(t-1) \end{bmatrix} \leq \begin{bmatrix} \mathbf{B}_{q_n}\mathbf{u}(t-1) \\ -\mathbf{B}_{q_n}\mathbf{u}(t-1) \\ w\mathbf{1} \\ w\mathbf{1} \\ v\mathbf{1} - (\mathbf{y}(t) - \mathbf{C}_{q_f}\mathbf{B}_{q_f}\mathbf{u}(t-1)) \\ v\mathbf{1} + (\mathbf{y}(t) - \mathbf{C}_{q_f}\mathbf{B}_{q_f}\mathbf{u}(t-1)) \end{bmatrix}, \tag{IV.49}$$

where **I**, **0** *and* **1** *have appropriate dimensions.*

Proof. The first two constraints in (IV.49) correspond to the continuous evolution in estimated discrete state q_n. Third and fourth constraints in (IV.49) correspond to the bound of the process noise. Fifth and sixth constraints in (IV.49) correspond to the measurement in anomalous discrete state q_f. The infeasibility of (IV.49) means the input-output data sequence of the anomalous discrete state is different from that of the current estimated discrete state and the SVO gives an empty estimated continuous state set. Therefore, Conflict $\mathfrak{A}$ occurs and the Type-D anomaly is detected. □

Note that by utilizing both continuous and discrete dynamics of the system to detect anomalies, the conflict-driven method may detect a Type-D anomaly whether or not it is diagnosable. According to Propositions 4 and 5, the conflict-driven method can provide detection guarantees for Type-D anomalies, including undiagnosable anomalies, if 1) the invariants of the anomalous discrete states do not intersect with the invariants of the nominal discrete states; or 2) the continuous dynamics of the anomalous discrete states are different from those of the current estimated discrete state.

IV.5.2.3 Type-B Anomaly

We have demonstrated the detection of Type-C_v, Type-C_{iv}, Type-D_d, and Type-D_{ud} anomalies, which are summarized in Table. IV.1. Note that for Type-D_d anomaly, we use the discrete state observer instead of checking the occurrence of the conflicts. Type-B anomalies combine Type-C and Type-D anomalies. The conflict-driven method can also guarantee to detect Type-$B_{v,d}$, Type-$B_{v,ud}$ and Type-$B_{iv,d}$ anomalies. This method can detect but cannot provide detection guarantees for Type-$B_{iv,ud}$ anomaly. Under Type-$B_{iv,ud}$ anomaly, if the incorrect estimated continuous state set is consistent with the current estimated discrete state, then the conflict-driven method fails to detect it.

Table IV.1: Summary of Anomaly Detection

Anomaly Type	Detection Guarantee?
Type-C_v	If the anomalous signal satisfies (IV.31)
Type-C_{iv}	If an observable discrete input event occurs and the anomalous signal satisfies (IV.42)
Type-D_d	Yes
Type-D_{ud}	If the invariants of the anomalous discrete states do not intersect with the invariants of the nominal discrete states except the boundaries or the continuous dynamics of the anomalous discrete state are different from those of the current estimated state.
Type-$B_{v,d}$	Yes
Type-$B_{v,ud}$	Yes
Type-$B_{iv,d}$	Yes
Type-$B_{iv,ud}$	Not guaranteed. But it may be detected using the conflict-driven method

IV.5.3 Relationship Between Conflict Types and Anomaly Types

With the occurrence of different conflict types and the estimated discrete state sequence, we can determine the types of associated anomalies as well.

A mapping between the conflict types and the types of anomalies is shown in Fig. IV.12. The inconsistency of the discrete event sequence means the anomaly is Type-D_d, Type-$B_{v,d}$ or Type-$B_{iv,d}$. The occurrence of Conflict $\mathfrak{A}$ means that the input-output sequence from system in the anomalous discrete state is different from that in a nominal discrete state. Then we can conclude that the anomaly type could be either Type-C_v, Type-$B_{v,d}$ or Type-$B_{v,ud}$. The occurrence of Conflict $\mathfrak{B}$ or Conflict $\mathfrak{C}$ means that the anomaly could be any type except Type-C_v. Note that the possible types of anomalies are the same when Conflict $\mathfrak{B}$ or Conflict $\mathfrak{C}$ occurs.

Under some anomalies, multiple conflict types occur. If the discrete event sequence is inconsistent and Conflict $\mathfrak{A}$ occurs, the anomaly is Type-$B_{v,d}$. If the discrete event sequence is inconsistent and Conflict $\mathfrak{B}$ or $\mathfrak{C}$ occurs, the anomaly is Type-$B_{iv,d}$. If Conflict $\mathfrak{B}$ or $\mathfrak{C}$ occurs and Conflict $\mathfrak{A}$ follows, the anomaly is Type-$B_{v,d}$ or Type-$B_{v,ud}$. If Conflict $\mathfrak{B}$ or $\mathfrak{C}$ occurs, and Conflict $\mathfrak{A}$ and the inconsistent discrete event sequence occur, then the anomaly is Type-$B_{v,d}$, which is shown as the dash line in Fig.IV.12. The timing of the occurrence of different conflict types and inconsistent discrete event sequence depends on the system and anomaly specifications. Note that Conflicts $\mathfrak{B}$ and $\mathfrak{C}$ cannot occur after Conflict $\mathfrak{A}$ because the initial set $X_I(t)$ is empty after the occurrence of Conflict $\mathfrak{A}$.

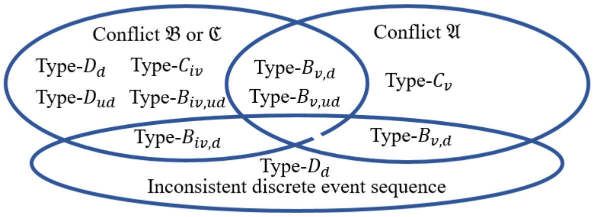


Figure IV.12: Mapping between the anomaly types and the conflict types

Remark 9. *With the types of conflicts and the observed events, we can have some information about the anomalies in the system. If we know all of the possible anomalies in the system and the corresponding anomalous hybrid subsystem, then we can end up doing anomaly isolation.*

IV.6 Simulation Results

In this section, we revisit the Positive Train Control (PTC) system introduced in Chapter I.2.2.3 with a specific scenario. We present the hybrid model, and select five different types of anomalies in the system. Then we show the effectiveness of the conflict-driven method.

IV.6.1 Positive Train Control system

We give a simple representation of the train dynamics and Radio Block Controller (RBC) reflecting the informal PTC cooperation protocol [23, 68]. The system diagram is shown in Fig. IV.13. Note that the "conflict-driven monitor" is what we contribute in this chapter. We model the train and its local controller as a hybrid system. The error signal is the continuous input to the system and the measured train position and velocity are the continuous outputs of the system. The RBC takes discrete output events from the system and commands discrete input events to the system. Both the discrete and the continuous input and output signals are sent to the conflict-driven monitor to detect anomalies. If an anomaly is detected, the monitor notifies the Automatic Train Protection (ATP) to brake the train. The PTC protocol consists of four discrete states (operating modes) as shown in Fig. IV.14, which corresponds to the nominal Finite State Machine (FSM) $\mathcal{M}_n$:

1. *far* (faraway): the local train controller regulates the train speed freely, which is called Movement Authority (MA).

2. *neg* (negotiation): the train communicates with the RBC asking for MA-extension.

3. *cor* (correcting): the train is braking.

4. *fsa* (failstate): the train is at full stop and awaits for manual clearance by the train operator.

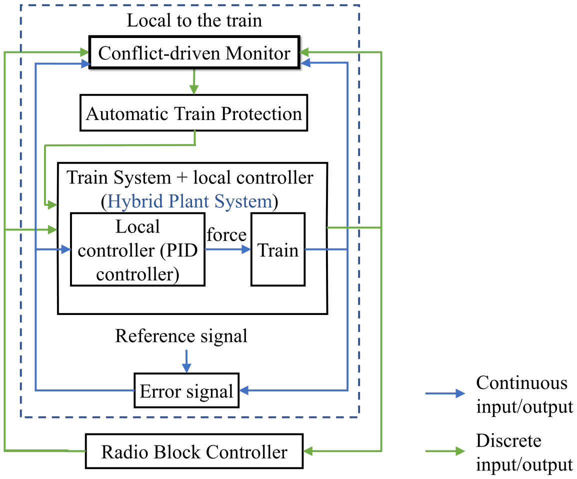


Figure IV.13: Positive Train Control system diagram

Based on the hybrid system modeling framework in Section IV.2.2, we additionally partition each mode of the PTC protocol into several discrete states such that the invariant of each discrete state is a hyperrectangle. The continuous dynamics and the invariant in each discrete state depend on the specific scenario. In this chapter, we consider that a train is scheduled to stop at the next

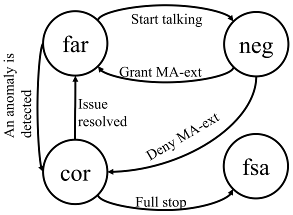


Figure IV.14: Finite State Machine $\mathcal{M}_n$ of the Positive Train Control protocol

station. Suppose the train is equipped with a GPS receiver and a speedometer and has two different local controllers with different parameters: one speed controller and one position controller. If the train is in the *far* or *neg* mode, the train is controlled by the speed controller and the continuous dynamics are described by matrices $(\mathbf{A_v}, \mathbf{B_v})$. The difference between the *far* and the *neg* modes is that the speed in the *neg* mode is unchanged. If the train is commanded to stop at the station, the PTC protocol transits to the *cor* mode. In the *cor* mode, the train is controlled by the position controller and the continuous dynamics are described by $(\mathbf{A_p}, \mathbf{B_p})$ under normal operation. The continuous state of the train is $\mathbf{x} = [x_p \quad x_v \quad x_f]^\intercal$, where x_p, x_v, x_f are the train position, speed and force, respectively. The continuous output of the train is $\mathbf{y} = [y_p \quad y_v]^\intercal$, where y_p, y_v are the measured train position and speed, respectively. If the train is in the *far* mode, the reference speed is $45m/s$. If the train is in the *cor* mode, the reference position of the train is a ramp signal ending at the station location and the train stops at the station under normal operation.

An example Type-C_{iv} anomaly is shown in Fig. IV.15, which is in the *anomaly* mode corresponding to the anomalous hybrid subsystem. When the anomaly occurs, the system transits to the *anomaly* mode with continuous dynamics described by matrices $(\mathbf{A_v}, \mathbf{B_v}, \mathbf{\Gamma_2})$. If this anomaly is detected, the system transits into the *cor* mode with the continuous dynamics $(\mathbf{A_b}, \mathbf{B_b})$. During braking, if the anomaly is resolved, the system follows the command from the RBC to stop at the station. Otherwise, the train comes to a full stop and then transits to the *fsa* mode, waiting for the manual clearance by the train operator.

The system parameters are shown in Table IV.2. The normal operating regions and the intermediate regions of the nominal discrete states are shown in Fig. IV.16. The green regions are the

Table IV.2: PTC system parameters

Parameters	Values
One track segment length	$2000m$
Location of sensors	$1500m, 3500m$
Station location	$5500m$
Desired train speed in *far* mode	$45m/s$
Sampling time	$0.1s$

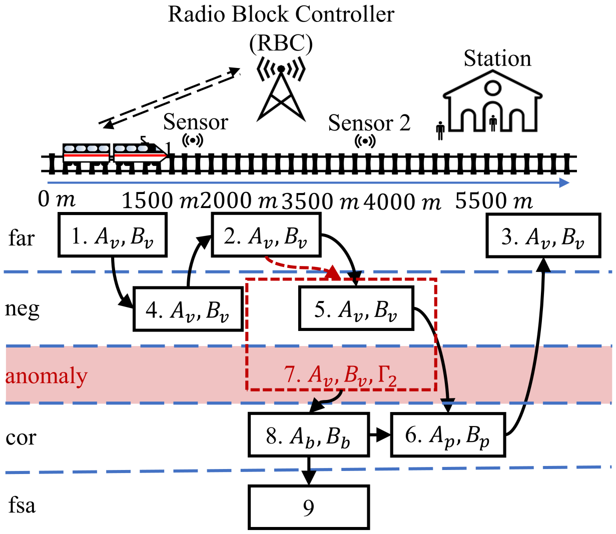


Figure IV.15: Nominal hybrid automaton $\mathcal{H}_n$ (in black solid lines) and an example Type-C_{iv} anomaly (in red dotted lines) of the PTC system. The numbers show the discrete states and state matrices represent the continuous dynamics in the discrete states. No continuous dynamics in discrete state 9 because the train is at full stop. In the *anomaly* mode, there are other anomalous discrete states which are not shown in the figure

normal operating regions when the train is in *far* mode. The violet regions are the normal operating regions when the train is in the *neg* mode. The orange region is the normal operating region when the train is in *cor* mode. The yellow regions are the intermediate regions in the nominal discrete states.

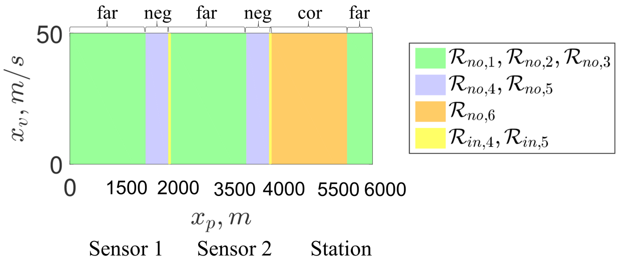


Figure IV.16: The normal operating regions (the green, violet and orange regions) and the intermediate regions (yellow regions) of the PTC system

The time step for reachability analysis of each discrete state is $\delta_4 = \delta_5 = 9$ and the rest are $\delta_i = 0$ when we follow the steps described in Section IV.5.

IV.6.2 Anomaly Detection Result

We choose the following five anomaly scenarios to illustrate the effectiveness of the conflict-driven method. The first four anomalies are Type-C_v, Type-C_{iv}, Type-D_d and Type-D_{ud}. As discussed in Section IV.5, Type-$B_{v,d}$, Type-$B_{v,ud}$ and Type-$B_{iv,d}$ anomalies are easy to detect because they combine either Type-C_v or Type-D_d anomaly, which are easy to detect. For Type-B anomaly, we choose the Type-$B_{iv,ud}$ anomaly.

- Type-C_v: A ramp anomalous signal with slope $100N/s$ is added to the state variable x_f and this anomaly could be due to an unexpected injection on the driving force from the train controller.
- Type-C_{iv}: A ramp anomalous signal with slope $0.05m/s$ is added to the measured train position y_p.
- Type-D_d: The RBC always grants the MA-extension.

- Type-D_{ud}: The train received commands from the RBC 40 seconds later than that under normal operation.
- Type-$B_{iv,ud}$: Combination of Type-C_{iv} and Type-D_{ud} anomalies.

The above anomalies start at $50s$, transiting the system from the *far* mode (discrete state 2) to the *anomaly* mode, and run until the end of the simulation. The effect of these anomalies if undetected are as follows: Under the Type-C_v anomaly, the train arrives at the station earlier. Under the Type-C_{iv} anomaly, the train stops at a wrong location. Under the Type-D_d anomaly, the train remains in the *far* mode and passes the station without stopping. Under the Type-D_{ud} anomaly, the train first passes the station and then the train position controller makes the train come back to the station. Under the Type-$B_{iv,ud}$ anomaly, the train stops at a wrong location.

1) Type-C_v Anomaly: The SVO gives an empty estimated continuous state set under this anomaly. Fig. IV.17 shows the detection performance of the conflict-driven method. After time $t = 58.4s$, the estimated sets $\tilde{X}(\mathbf{y}, t)$ given by the SVO are empty and Conflict $\mathfrak{A}$ occurs. This is because after time $t = 58.4s$, the set of continuous states calculated based on the input signal (one time-step forward reachable set starting from the estimated continuous state set at the previous time step) and the set of continuous states calculated based on the output measurement do not intersect with each other, i.e., the input-output data of the anomalous hybrid subsystem is inconsistent with that of the nominal hybrid subsystem.

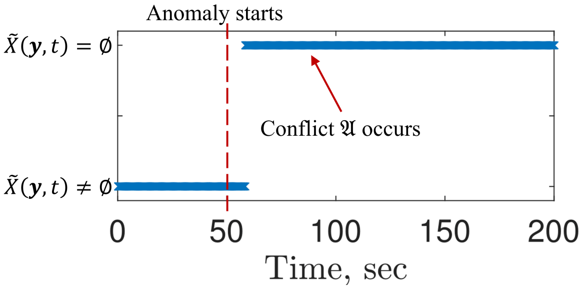


Figure IV.17: Simulation result under the Type-C_v anomaly with the occurrence of Conflict $\mathfrak{A}$

2) Type-C_{iv} Anomaly: This anomaly cannot be detected by traditional methods. As shown in Fig. IV.18, the conflict-driven method detects this anomaly at time $90.1s$ when Conflict $\mathfrak{C}$ occurs.

The estimated discrete state is 5, indicating that the train is communicating with the RBC to ask for MA-extension. But the reachable set $R_{\delta_5}(X_I(901))$ indicates that the train has already received commands from the RBC. Therefore, Conflict $\mathfrak{C}$ occurs, i.e., $R_{\delta_5}(X_I(901)) \cap Inv_5 = \emptyset$.

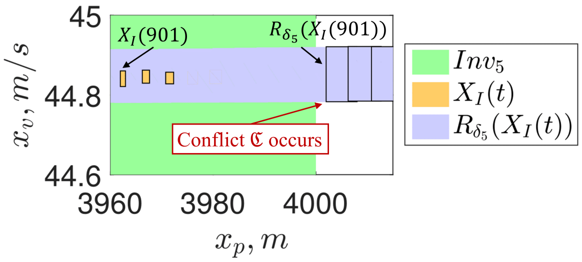
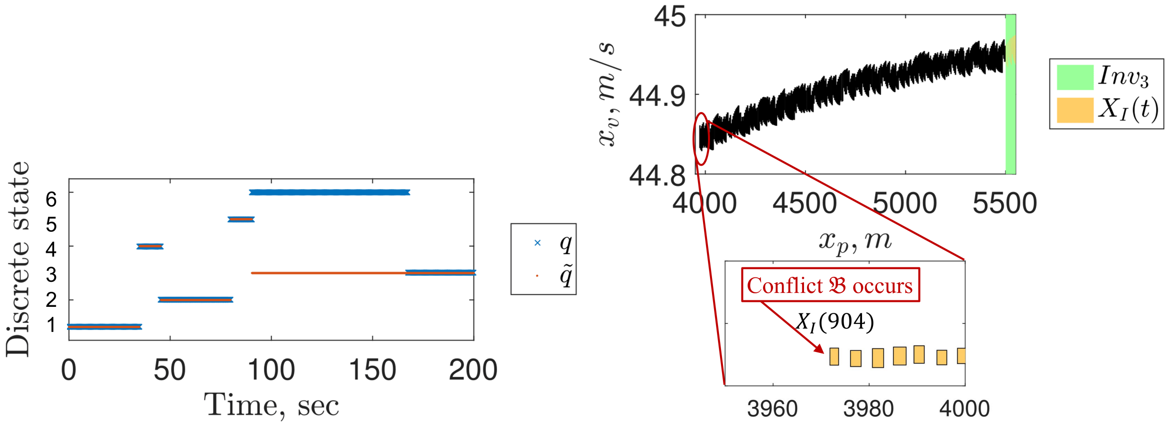


Figure IV.18: Simulation result under the Type-C_{iv} anomaly with the occurrence of Conflict $\mathfrak{C}$

3) Type-D_d Anomaly: This anomaly can be detected by existing discrete model-based methods. Fig. IV.19a shows that the anomaly is detected by the discrete state observer at time 90.4s. At time 90.4s when Conflict $\mathfrak{B}$ occurs, the estimated discrete state changed from 5 to 3 because the RBC grants the MA-extension. But the discrete state under normal operation should change to 6, i.e., the RBC should deny the MA-extension. At time 90.4s Conflict $\mathfrak{B}$ also occurs as shown in Fig. IV.19b. The invariant of discrete state 3 along the position x_p state variable is from 5500m to 7500m but the upper bound of the initial set $X_I(904)$ at time 90.4s is 3978m, i.e., $X_I(904) \cap Inv_3 = \emptyset$.



(a) Inconsistent discrete state sequence

(b) Conflict $\mathfrak{B}$ occurs

Figure IV.19: Simulation result under the Type-D_d anomaly

4) Type-D_{ud} Anomaly: This anomaly cannot be detected by existing discrete model-based methods. As shown in Fig. IV.20, the conflict-driven method detects the anomaly at time $t = 90.1s$ and Conflict $\mathfrak{C}$ occurs. The estimated discrete state is 5. Because it takes 40 seconds more for the train to receive command from the RBC than under normal operation, the reachable set $R_{\delta_5}(X_I(901))$ does not intersect with the invariant of the discrete state 5.

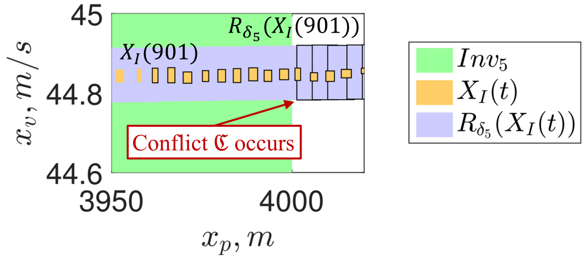


Figure IV.20: Simulation result under the Type-D_{ud} anomaly with the occurrence of Conflict $\mathfrak{C}$

5) Type-$B_{iv,ud}$ Anomaly: Existing methods cannot detect general Type-$B_{iv,ud}$ anomalies. The conflict-driven method cannot provide detection guarantees but has the possibility to detect a Type-$B_{iv,ud}$ anomaly. Fig. IV.21 shows the detection performance of the conflict-driven method under the Type-$B_{iv,ud}$ anomaly. This anomaly is detected by the conflict-driven method at time $90.1s$ when Conflict $\mathfrak{C}$ occurs. The reason that this method can detect this anomaly is that the estimated continuous state set is inconsistent with the estimated discrete state. Under Type-$B_{iv,ud}$ anomaly, the estimated continuous state is larger than the actual continuous state. If the discrete event "deny the MA-extension" occurs earlier than the system under normal operation and the estimated discrete state changes from 5 to 6 earlier, then Conflicts $\mathfrak{B}$ and $\mathfrak{C}$ may not occur. Finding the conditions for Type-$B_{iv,ud}$ anomaly that can be detected by the conflict-driven method is part of future work.

IV.7 Summary

Hybrid CPS can be modeled as hybrid systems since hybrid CPS contain continuous dynamics and discrete behavior. In this chapter, we propose a classification taxonomy of anomalies

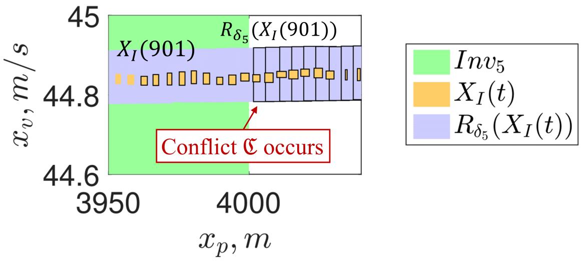


Figure IV.21: Simulation result under the Type-$B_{iv,ud}$ anomaly with the occurrence of Conflict $\mathfrak{C}$

in hybrid systems based on the variables that are affected by the anomalies and the input-output data consistency. We classify the anomalies in hybrid systems into eight different types: Type-C_v, Type-C_{iv}, Type-D_d, Type-D_{ud}, Type-$B_{v,d}$, Type-$B_{v,ud}$, Type-$B_{iv,d}$ and Type-$B_{iv,ud}$. To detect these challenging anomalies in hybrid systems, we utilize the relation between the continuous and discrete variables and propose a novel anomaly detection method: conflict-driven method.

This method utilizes a hybrid observer which consists of a discrete state observer and a continuous state observer to detect various types of anomalies in hybrid systems. The discrete state observer is designed as a finite-state automaton to estimate the discrete state of the system. The continuous state observer is designed as a Set-Valued Observer (SVO) to estimate a set of continuous states of the hybrid system. Based on the relation between the discrete and continuous variables in hybrid systems, we define three different conflict types. We demonstrate that the conflict-driven method is guaranteed to detect Type-C_v, Type-C_{iv}, Type-D_d, Type-D_{ud}, Type-$B_{v,d}$, Type-$B_{v,ud}$, and Type-$B_{iv,d}$ anomalies under certain conditions. Additionally, we give a mapping between the types of conflicts and the types of anomalies.

We used a simplified Positive Train Control (PTC) system to illustrate the effectiveness of the conflict-driven method. Based on the simulation results, we showed that the conflict-driven method can also detect some Type-$B_{iv,ud}$ anomaly even though the conflict-driven method cannot provide detection guarantees for Type-$B_{iv,ud}$ anomalies.

The conflict-method expands the capabilities of anomaly detection in hybrid systems, however, this method has some limitations when applied to more general hybrid systems. The hybrid observer used in the conflict-driven method requires the discrete component of the hybrid system to be current-state observable in order to give a unique estimated discrete state. However, some hybrid systems contain unobservable discrete events such that the discrete components are not current-state observable. Moreover, this method requires the knowledge of guard conditions is known *a priori*. Sometimes it may be impossible to know the guard conditions *a priori*. In addition, the hybrid observer uses the SVO as the continuous state observer to estimate the continuous state of the system. The SVO requires that no discontinuity exists in continuous variables. Thus, the hybrid observer cannot be used for hybrid systems with discontinuity in continuous variables during discrete transitions. To address these limitations, we propose a new observer framework which only uses the continuous measurements for state estimation and anomaly detection. This framework is presented in the next chapter.

CHAPTER V

A Novel Hybrid Observer Design for State Estimation and Anomaly Diagnosis Applied to Hybrid Systems with Unobservable Discrete Events

V.1 Introduction

In Chapter IV, we proposed the conflict-driven anomaly detection method for hybrid systems. However, the conflict-driven method has some limitations when applied to a wider class of hybrid systems as discussed in Section IV.7. In the hybrid system formalism in Section IV.2.2, each discrete transition is associated with a discrete event and a guard condition. For some hybrid systems, each discrete transition is also associated with a reset function, which resets the value of the continuous state of the system when a discrete transition occurs. In addition, the guard conditions and reset functions corresponding to the discrete transitions may be unknown *a priori* and the discrete events are unobservable. Estimating the state and diagnosing anomalies for these hybrid systems can be challenging.

In this chapter, we propose a new observer framework which consists of two continuous state observers to estimate state and diagnose anomalies for hybrid systems with unobservable discrete events, such as the microgrid system with unplanned islanding mentioned in Section I.2.3.3. The two continuous state observers use different sets of sensors and the same continuous system model associated with the current estimated discrete state (assuming that the initial discrete state of the system is given) to estimate the continuous state of the system. Based on the estimated continuous state trajectories, the Recursive Least Squares (RLS) method is used to estimate the current continuous model of the system. To determine the current discrete state of the system, we run multiple continuous models in parallel for a finite time period, including the known continuous models of the system and the estimated continuous model provided by the RLS. The discrete state can be uniquely determined if the continuous dynamics of different discrete states are distinguishable[1]. We call this framework the Convergence Ratio Multi-model Hybrid Observer (CRMMHO) framework. The contributions of this chapter are as follows:

1. We propose the CRMMHO framework to estimate both the discrete and the continuous variables for hybrid systems with unobservable discrete events;
2. We use the CRMMHO framework to diagnose anomalies in more general hybrid systems; and
3. We apply the CRMMHO framework in the simulated microgrid system to validate its effectiveness.

V.2 Problem Formulation

In this section, we describe the class of hybrid systems of interest and formally state the problem.

V.2.1 Notation

Let $||\cdot||$ denote ∞-norm, $\tilde{\cdot}$ denote estimated variables. In addition, $\mathbf{x} \in \mathbb{R}^n$ represents a vector, where its i^{th} element is indicated by $\mathbf{x}^{(i)}$. $\mathbf{A} \in \mathbb{R}^{n \times m}$ represents a matrix, where its i^{th} row and j^{th} column are indicated by $\mathbf{A}^{(i,:)}$ and $\mathbf{A}^{(:,j)}$, respectively. The detailed notations are shown in Appendix C.

[1] Two continuous systems are distinguishable if for any non-zero initial continuous states and the same input, the outputs of the two continuous systems are not identical for a finite time period [58].

V.2.2 Hybrid Systems

In this chapter, we consider the class of hybrid systems that can be represented by a tuple $\mathcal{H} = (\mathcal{X}, \mathcal{U}, Y, Init, field, \phi, h, f_r)$, where each element is defined as

- $\mathcal{X} = Q \times X$: a set of discrete and continuous states
- $\mathcal{U} = \Psi \times U$: a set of discrete and continuous inputs
- Y: a set of continuous outputs
- $Init = (q(t_0), \mathbf{x}(t_0)) \in \mathcal{X}$: an initial state
- $field : \mathcal{X} \times U \to X$: a time-invariant vector field
- $\phi : Q \times \Psi \to Q$: a set of discrete transitions
- $h : Q \times X \to Y$: a continuous output equation
- $f_r : \mathbf{x}(t_i) = f_r(\mathbf{x}(t_i - 1), \psi(i))$: a reset function, where t_i is the time when the i^{th} discrete event occurs

For each discrete state $q \in Q$, we associate continuous dynamics that can be represented by a Linear Time Invariant (LTI) model, subject to process and measurement noise.

$$\begin{aligned} field : \quad & \mathbf{x}(t+1) = \mathbf{A}_q \mathbf{x}(t) + \mathbf{B}_q \mathbf{u}(t) + \mathbf{w}(t) \\ h : \quad & \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{v}(t) \end{aligned} \tag{V.1}$$

where $\mathbf{x} \in \mathbb{R}^{n_x}, \mathbf{u} \in \mathbb{R}^{n_u}, \mathbf{y} \in \mathbb{R}^{n_y}$ are the system continuous state, continuous input and continuous output, $\mathbf{A}_q \in \mathbb{R}^{n_x \times n_x}, \mathbf{B}_q \in \mathbb{R}^{n_x \times n_u}$ are system matrices which depend on the discrete state q, $\mathbf{C} \in \mathbb{R}^{n_y \times n_x}$ is the system output matrix, which is the same for all $q \in Q$. The process and measurement noise are represented by $\mathbf{w} \sim \mathcal{N}(0, \mathbf{W})$ and $\mathbf{v} \sim \mathcal{N}(0, \mathbf{V})$, respectively, where $||\mathbf{w}|| \leq w$ and $||\mathbf{v}|| \leq v$. The initial continuous state $\mathbf{x}(t_0)$ is unique but unknown.

To each discrete transition $\phi(q, \psi) = q'$, we associate a guard condition:

$$G(q, q', \psi) = \{\mathbf{x} \in \mathbb{R}^{n_x} : \mathbf{s_G x} \geq \mathbf{c_G}\} \tag{V.2}$$

where $\mathbf{c_G}$ is a vector and $\mathbf{s_G}$ is a matrix. The guard condition indicates that the discrete transition ψ occurs if the system is in discrete state q and the continuous state satisfies $\mathbf{s_G x} \geq \mathbf{c_G}$.

In this chapter, we assume

Assumption 10. *The initial discrete state $q(t_0)$ is known.*

The i^{th} discrete event $\psi(i)$ occurs at time t_i. The continuous evolutions of the system in one discrete state occur in time $t \in [t_{i-1} + 1, t_i]$, for all $i = 1, 2, ...$ We assume

Assumption 11. *No discrete event occurs between two adjacent sample time boundaries and at most one discrete event occurs at one sample time boundary.*

Assumption 12. *The matrix pair $(\mathbf{A}_q, \mathbf{C})$ is observable for the initial discrete state $q(t_0) \in Q$.*

As mentioned in Section V.1, we use Kalman filters in the proposed CRMMHO framework. To ensure the convergence of Kalman filter gains, we assume:

Assumption 13. *The continuous system in each discrete state is stable or marginally stable, i.e., $\lambda(\mathbf{A}_q) \leq 1, \forall q \in Q$, where $\lambda(\mathbf{A}_q)$ is the eigenvalues of matrix $\mathbf{A}_q$.*

The objective of this chapter is to design an observer for state estimation and anomaly diagnosis for hybrid systems that

1. not all discrete events ψ are observable; and
2. not all guard conditions $G(q, q', \psi)$ and reset functions f_r are known *a priori*.

Remark 10. *Note that the proposed CRMMHO framework still works under the extreme case that there are no observable discrete events ψ and we do not have any knowledge of guard conditions $G(q, q', \psi)$ and reset functions f_r.*

V.3 Observer Design

In this section, we first introduce the proposed Convergence Ratio Multi-model Hybrid Observer (CRMMHO) framework for the class of hybrid systems described in Section V.2. Then we demonstrate the effectiveness of the CRMMHO framework. Finally, we show how to use the CRMMHO framework to diagnose anomalies.

V.3.1 Observer Framework Description

As shown in Fig. V.1, the CRMMHO framework for state estimation and anomaly diagnosis consists of two continuous state observers and integrates three methods: the Convergence Ratio (CR) [85], the Recursive Least Squares (RLS) [37] and the model selection. The two continuous state observers are designed to be Kalman filters[2], which estimate the continuous state of the system using different sets of measured outputs. The continuous state observer 0 uses all of the measured outputs. The continuous state observer 1 uses any subset of measured outputs, which still ensure the observability of the continuous system. Based on the estimated continuous states $\hat{\mathbf{x}}_0$ and $\hat{\mathbf{x}}_1$ by the two continuous state observers, the CR method calculates the estimation errors $\tilde{\mathbf{x}}_{\mathbf{e},0}$ and $\tilde{\mathbf{x}}_{\mathbf{e},1}$ for both of the continuous state observers. If the norm of $\hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1$ is larger than a threshold θ_{CR}, it indicates that a discrete transition has occurred and the continuous state observers enter their transient state. Then the estimated state $\tilde{\mathbf{x}}$ is updated based on $\tilde{\mathbf{x}}_{\mathbf{e},0}$. Otherwise, $\tilde{\mathbf{x}} = \hat{\mathbf{x}}_0$. After some time steps, the norm of $\hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1$ will reach a steady state and not change much at each time step. Then, we record the time step t_{ss} and start to use an on-line system identification method, which is the RLS in this chapter[3], to recursively estimate the current system matrices at each time step. When the norm of *a posteriori* error of the RLS is smaller than a certain threshold θ_{RLS}, it means that the matrix estimation has converged and we record the number of time steps Δt_{RLS} needed for the estimated matrices to converge. Then the estimated system matrices are used by the model

[2] The continuous state observers can also be designed as Luenberger observers [85].

[3] Note that we can use other on-line system identification method under the CRMMHO framework.

selection method to determine the current discrete state. In the model selection method, multiple continuous system models run in parallel for Δt time steps (the derivation of Δt is given in Step 7 in Section V.3.1.3) with $\tilde{\mathbf{x}}(t_{ss})$ as the initial continuous state. The multiple running models include all known continuous models and the estimated continuous model by the RLS. Based on each model, we can calculate a residual, which is the difference between the measured output from the system and the output calculated from the model. The norms of the residuals from the running models are compared. The model with the smallest norm of the residual corresponds to the estimated discrete state if the continuous dynamics of the current discrete state are distinguishable from all of other known continuous dynamics. However, it is possible that the system enters some unknown discrete state with unknown continuous dynamics. If the norm of the smallest residual is much larger than the norm of the residual calculated based on the estimated continuous dynamics from the RLS, then the system is considered to be subject to some unknown anomaly.

Remark 11. *Note that the model selection method in the CRMMHO framework only runs multiple models for a short period of time Δt. If the system stays in one discrete state much longer than Δ_{RLS}, then the maximum value of Δt is given by Equation* (V.15). *In addition, the model selection method only calculates a single output vector for each model at one time step. But set-membership methods give a set of output vectors (or state vectors) for one model at one time step. The amount of computation time running the multiple continuous models for Δt time steps is less than the amount of computation time used by set-membership methods running multiple models all the time. Thus, the proposed CRMMHO framework is more computationally efficient.*

Remark 12. *In the model selection method, we compare the norm of the residuals from multiple continuous models to determine the current discrete state. The residuals are functions of system dynamics. The concepts of large or small residual norm are used for purpose of development, selection and solution of the model selection method. We assume that the dynamics of different continuous models are similar to the extent that the concept of large and small residual norm can be universally applied. The validation of this discussion is system-specific and part of future work.*

First, we briefly describe how Kalman filter and the RLS work in Sections V.3.1.1 and

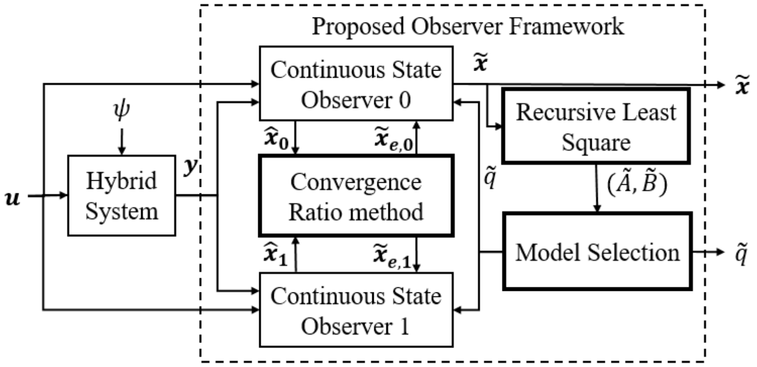


Figure V.1: The CRMMHO framework

V.3.1.2, respectively. Then we give a detailed description of the CRMMHO framework in Section V.3.1.3.

V.3.1.1 Kalman Filter

The Kalman filter is the most widely used state estimator for linear systems with known uncertainties. The standard Kalman filter derivation is given here

$$
\begin{aligned}
\hat{\mathbf{x}}(t|t-1) &= \mathbf{A}_q\hat{\mathbf{x}}(t-1)+\mathbf{B}_q\mathbf{u}(t-1)\\
\mathbf{P}(t|t-1) &= \mathbf{A}_q\mathbf{P}(t-1)\mathbf{A}_q^{\mathsf{T}}+\mathbf{W}\\
\mathbf{K}(t) &= \mathbf{P}(t|t-1)\mathbf{C}^{\mathsf{T}}(\mathbf{V}+\mathbf{C}\mathbf{P}(t|t-1)\mathbf{C}^{\mathsf{T}})^{-1}\\
\hat{\mathbf{x}}(t) &= \hat{\mathbf{x}}(t|t-1)+\mathbf{K}(t)(\mathbf{y}(t)-\mathbf{C}\hat{\mathbf{x}}(t|t-1))\\
\mathbf{P}(t) &= (\mathbf{I}-\mathbf{K}(t)\mathbf{C})\mathbf{P}(t|t-1)(\mathbf{I}-\mathbf{K}(t)\mathbf{C})^{\mathsf{T}}+\mathbf{K}(t)\mathbf{V}\mathbf{K}^{\mathsf{T}}(t)
\end{aligned}
\tag{V.3}
$$

where **K** is the Kalman gain and **P** is the covariant matrix reflecting the accuracy of estimates. It is well known that the Kalman gain **K** will converge if the system is open-loop stable [61]. In practice the Kalman gain usually converges in a few steps. Without loss of generality, we assume the Kalman filter to be already in the steady state and define

$$
\mathbf{P} \triangleq \lim_{t\to\infty}\mathbf{P}(t|t-1), \quad \mathbf{K} \triangleq \mathbf{P}\mathbf{C}^{\mathsf{T}}(\mathbf{V}+\mathbf{C}\mathbf{P}\mathbf{C}^{\mathsf{T}})^{-1}. \tag{V.4}
$$

Based on Assumption 13, we use the constant Kalman gains in the CRMMHO framework and its demonstration.

V.3.1.2 Recursive Least Squares

The RLS is an on-line system identification method which can use given measurements to estimate unknown system parameters. We use the RLS to estimate the continuous system matrices $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$.

In our case, we have

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{A}_q & \mathbf{B}_q \end{bmatrix} \begin{bmatrix} \mathbf{x}(t-1) \\ \mathbf{u}(t-1) \end{bmatrix} \tag{V.5}$$

where $\mathbf{u}(t-1)$ is known, and we use $\tilde{\mathbf{x}}$ as an estimate of $\mathbf{x}$. Let $\Xi = \begin{bmatrix} \tilde{\mathbf{A}}^\intercal \\ \tilde{\mathbf{B}}^\intercal \end{bmatrix}$ and $\boldsymbol{\xi} = \begin{bmatrix} \tilde{\mathbf{x}} \\ \mathbf{u} \end{bmatrix}$. RLS finds the best linear function Ξ that computes the value of $\tilde{\mathbf{x}}$ from the values of $\boldsymbol{\xi}$. Then Ξ can be used to estimate $\begin{bmatrix} \mathbf{A}_q & \mathbf{B}_q \end{bmatrix}$. The RLS procedure is as follows:

$$\begin{aligned}
\boldsymbol{\varepsilon}_{\mathbf{o}}(t) &= \tilde{\mathbf{x}}_0(t) - \Xi^\intercal(t-1)\boldsymbol{\xi}(t-1) \\
\Xi(t) &= \Xi(t-1) + \frac{\mathbf{G}_{\mathbf{RLS}}(t-1)\boldsymbol{\xi}(t-1)}{1+\boldsymbol{\xi}^\intercal(t-1)\mathbf{G}_{\mathbf{RLS}}(t-1)\boldsymbol{\xi}(t-1)}\boldsymbol{\varepsilon}_{\mathbf{o}}^\intercal(t) \\
\mathbf{G}_{\mathbf{RLS}}(t) &= \mathbf{G}_{\mathbf{RLS}}(t-1) - \frac{\mathbf{G}_{\mathbf{RLS}}(t-1)\boldsymbol{\xi}(t-1)\boldsymbol{\xi}^\intercal(t-1)\mathbf{G}_{\mathbf{RLS}}(t-1)}{1+\boldsymbol{\xi}^\intercal(t-1)\mathbf{G}_{\mathbf{RLS}}(t-1)\boldsymbol{\xi}(t-1)}
\end{aligned} \tag{V.6}$$

where $\boldsymbol{\varepsilon}_{\mathbf{o}}$ is *a priori* error and $\mathbf{G}_{\mathbf{RLS}}$ is the adaptation gain. Initially, define $\mathbf{G}_{\mathbf{RLS}}(t_0) = c\mathbf{I}$, where c is a large positive constant (e.g. 1000) and $\mathbf{I}$ is an identify matrix.

V.3.1.3 Work Flow of the CRMMHO Framework

As shown in Fig. V.2, the detailed work flow of the CRMMHO framework is described as follows:

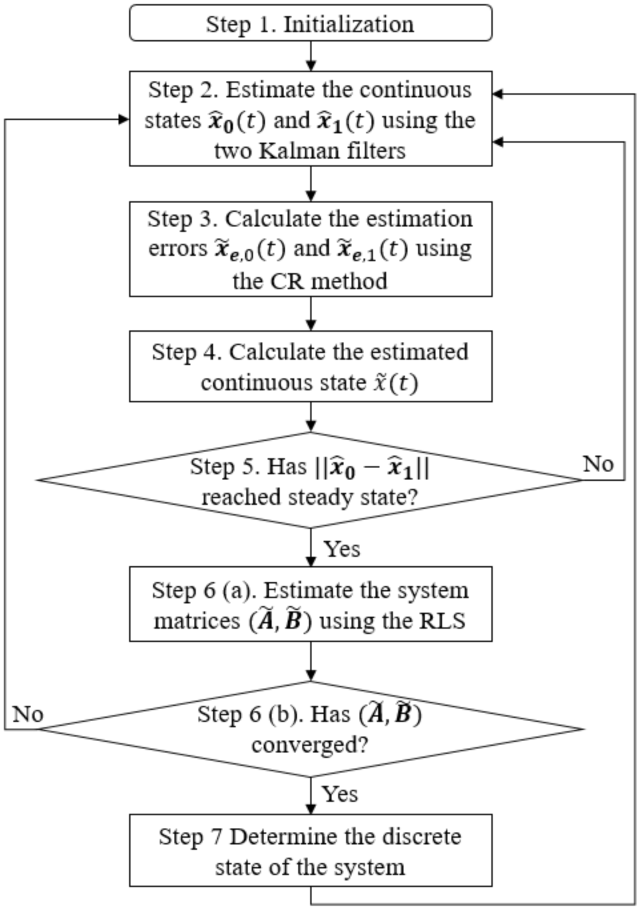

Step 1. Initialization
Step 2. Estimate the continuous states $\hat{\boldsymbol{x}}_{\mathbf{0}}(t)$ and $\hat{\boldsymbol{x}}_{\mathbf{1}}(t)$ using the two Kalman filters
Step 3. Calculate the estimation errors $\tilde{\boldsymbol{x}}_{\boldsymbol{e},0}(t)$ and $\tilde{\boldsymbol{x}}_{\boldsymbol{e},1}(t)$ using the CR method
Step 4. Calculate the estimated continuous state $\tilde{x}(t)$
Step 5. Has $||\hat{\boldsymbol{x}}_{\mathbf{0}} - \hat{\boldsymbol{x}}_{\mathbf{1}}||$ reached steady state?
No
Yes
Step 6 (a). Estimate the system matrices $(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}})$ using the RLS
No
Step 6 (b). Has $(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}})$ converged?
Yes
Step 7 Determine the discrete state of the system

Figure V.2: Flow chart of the CRMMHO framework

1. Initialization:

   The estimated initial discrete state is the same as the real initial discrete state $\tilde{q}(t_0) = q(t_0) = q_0$ based on Assumption 10. The initial guess for the RLS is $\mathbf{\Xi}(0) = \begin{bmatrix} \mathbf{A}_{q_0}^{\top} \\ \mathbf{B}_{q_0}^{\top} \end{bmatrix}$ and $\mathbf{G_{RLS}}(t_0) = c\mathbf{I}$.

2. Estimate $\hat{\mathbf{x}}_0(t)$ and $\hat{\mathbf{x}}_1(t)$ using two Kalman filters:

   Note that the two Kalman filters are using the same estimated continuous state at the previous time step $\tilde{\mathbf{x}}(t-1)$ for state estimation. With constant Kalman gains, the estimated continuous states by the two Kalman filters are given as follows:

$$\begin{aligned}
\hat{\mathbf{x}}_0(t) &= \mathbf{A}_{\tilde{q}}\tilde{\mathbf{x}}(t-1) + \mathbf{B}_{\tilde{q}}\mathbf{u}(t-1) + \mathbf{K}_0(\mathbf{y}(t) - \mathbf{C}(\mathbf{A}_{\tilde{q}}\tilde{\mathbf{x}}(t-1) + \mathbf{B}_{\tilde{q}}\mathbf{u}(t-1)))\\
&= \hat{\mathbf{x}}(t|t-1) + \mathbf{K}_0(\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t|t-1))\\
&= (\mathbf{I} - \mathbf{K}_0\mathbf{C})\hat{\mathbf{x}}(t|t-1) + \mathbf{K}_0\mathbf{y}(t)\\
\hat{\mathbf{x}}_1(t) &= \mathbf{A}_{\tilde{q}}\tilde{\mathbf{x}}(t-1) + \mathbf{B}_{\tilde{q}}\mathbf{u}(t-1) + \mathbf{K}_1(\mathbf{y}_1(t) - \mathbf{C}_1(\mathbf{A}_{\tilde{q}}\tilde{\mathbf{x}}(t-1) + \mathbf{B}_{\tilde{q}}\mathbf{u}(t-1)))\\
&= \hat{\mathbf{x}}(t|t-1) + \mathbf{K}_1(\mathbf{y}_1(t) - \mathbf{C}_1\hat{\mathbf{x}}(t|t-1))\\
&= (\mathbf{I} - \mathbf{K}_1\mathbf{C}_1)\hat{\mathbf{x}}(t|t-1) + \mathbf{K}_1\mathbf{y}_1(t)
\end{aligned} \tag{V.7}$$

where $\hat{\mathbf{x}}_0 \in \mathbb{R}^{n_x}, \hat{\mathbf{x}}_1 \in \mathbb{R}^{n_x}$ are the estimated continuous states by continuous state observers 0 and 1, respectively, $\mathbf{C}_1 \in \mathbb{R}^{n'_y \times n_x}$ is the output matrix used by continuous state observer 1, $\mathbf{y}_1 \in \mathbb{R}^{n'_y}$ is the output variable used by continuous state observer 1, and $\hat{\mathbf{x}}(t|t-1) = \mathbf{A}_{\tilde{q}}\tilde{\mathbf{x}}(t-1) + \mathbf{B}_{\tilde{q}}\mathbf{u}(t-1)$. Note that $n'_y < n_y$.

Then compare the norm of the difference of the estimated continuous states $||\hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1||$ with the threshold θ_{CR}. If $||\hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1|| > \theta_{CR}$, then continue on Step 3. Otherwise, stay in Step 2 for the next time step. Note that the threshold θ_{CR} may be different for different discrete states because the impact of the noise on state estimation is different in different discrete states. Adaptive θ_{CR} should be used to make sure the change of discrete state can be successfully detected. Refer to [77] on how to determine an adaptive threshold using a data-driven method.

3. Calculate the estimation errors $\tilde{\mathbf{x}}_{\mathbf{e},0}(t), \tilde{\mathbf{x}}_{\mathbf{e},1}(t)$ using the CR method:

$$\begin{aligned}
\tilde{\mathbf{x}}_{\mathbf{e},0}(t) &= (\mathbf{I} - \mathbf{K}_0\mathbf{C})(\mathbf{K}_0\mathbf{C} - \mathbf{K}_1\mathbf{C}_1)^{\dagger}(\hat{\mathbf{x}}_0(t) - \hat{\mathbf{x}}_1(t))\\
\tilde{\mathbf{x}}_{\mathbf{e},1}(t) &= (\mathbf{I} - \mathbf{K}_1\mathbf{C}_1)(\mathbf{K}_0\mathbf{C} - \mathbf{K}_1\mathbf{C}_1)^{\dagger}(\hat{\mathbf{x}}_0(t) - \hat{\mathbf{x}}_1(t)),
\end{aligned} \tag{V.8}$$

where $(\mathbf{K}_0\mathbf{C} - \mathbf{K}_1\mathbf{C}_1)^{\dagger}$ is the pseudo-inverse of matrix $(\mathbf{K}_0\mathbf{C} - \mathbf{K}_1\mathbf{C}_1)$, that is $(\mathbf{K}_0\mathbf{C} - \mathbf{K}_1\mathbf{C}_1)^{\dagger} = [(\mathbf{K}_0\mathbf{C} - \mathbf{K}_1\mathbf{C}_1)^{\intercal}(\mathbf{K}_0\mathbf{C} - \mathbf{K}_1\mathbf{C}_1)]^{-1}(\mathbf{K}_0\mathbf{C} - \mathbf{K}_1\mathbf{C}_1)^{\intercal}$.

4. Calculate the estimated continuous state $\tilde{\mathbf{x}}(t)$:

$$\tilde{\mathbf{x}}(t) = \hat{\mathbf{x}}_0(t) + \tilde{\mathbf{x}}_{\mathbf{e},0}(t) \tag{V.9}$$

5. Check whether or not the continuous state estimation has entered a new steady state:

   When the system enters a new discrete state, $||\hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1||$ reaches a new steady state which is larger than θ_{CR}. We compare the following

   $$\frac{|||\hat{\mathbf{x}}_0(t+1) - \hat{\mathbf{x}}_1(t+1)|| - ||\hat{\mathbf{x}}_0(t) - \hat{\mathbf{x}}_1(t)|||}{||\hat{\mathbf{x}}_0(t) - \hat{\mathbf{x}}_1(t)||} \tag{V.10}$$

   which indicates the change of $||\hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1||$, with a threshold θ_{diff} to determine whether or not the continuous state estimation has entered a new steady state. If (V.10) is smaller than θ_{diff}, record the current time step t_{ss} and continue on Step 6. Otherwise, go back to Step 2 for the next time step.

6. Estimate the system matrices using the RLS following the procedure introduced in Section V.3.1.2. Then the estimated system matrices $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ are as follows

   $$\begin{aligned} \tilde{\mathbf{A}} &= [\Xi^{(1:n_x,1:n_x)}]^\intercal \\ \tilde{\mathbf{B}} &= [\Xi^{(n_x+1:n_x+n_u,1:n_x)}]^\intercal \end{aligned} \tag{V.11}$$

   In addition, we use *a posteriori* error $\boldsymbol{\varepsilon}$ to track the convergence of the system matrices estimation

   $$\boldsymbol{\varepsilon}(t) = \tilde{\mathbf{x}}(t) - \Xi^\intercal(t)\boldsymbol{\xi}(t) \tag{V.12}$$

   If $||\boldsymbol{\varepsilon}(t+1)|| \leq \theta_{RLS}$, then we record the time steps Δt_{RLS} for convergence and continue on Step 7. Otherwise, we go back to Step 2 for the next time step.

7. Determine the discrete state of the system:

   For each known continuous model and the estimated continuous model, we can calculate the output starting from time t_{ss}:

$$\begin{aligned}\mathbf{y}_{o,\hat{q}}(t_{ss}+j) &= \mathbf{C}\mathbf{x}_{o,\hat{q}}(t_{ss}+j) \\ &= \mathbf{C}\mathbf{A}_{\hat{q}}^{j}\mathbf{x}(t_{ss}) + \Sigma_{i=0}^{j-1}\mathbf{C}\mathbf{A}_{\hat{q}}^{i}\mathbf{B}_{\hat{q}}\mathbf{u}(t_{ss}+j-1-i)\end{aligned} \tag{V.13}$$

where $\hat{q} = 0, 1, 2, 3...$, 0 indicates the continuous model estimated by the RLS, $1, 2, 3...$ indicate the continuous models of discrete states $1, 2, 3...$, respectively, $\mathbf{x}_{o,\hat{q}}$ and $\mathbf{y}_{o,\hat{q}}$ are the state and output calculated using the continuous model.

The residual is the difference between the measured output from the system and the calculated output from each continuous model:

$$\mathbf{r}_{\hat{q}}(t_{ss}+j) = \mathbf{y}(t_{ss}+j) - \mathbf{y}_{o,\hat{q}}(t_{ss}+j) \tag{V.14}$$

We introduce a similarity index for each continuous model to help find the continuous model which has the closest behavior to the current continuous dynamics of the system. The similarity index for each continuous model is defined as the mean of the norm of the residual calculated based on the continuous model:

$$I_{\hat{q}} = \frac{1}{\Delta t}\Sigma_{j=1}^{\Delta t}||\mathbf{r}_{\hat{q}}(t_{ss}+j)||, \quad \text{where } \Delta t = \max(\Delta t_{RLS}, \frac{2n_x - n_y}{n_y - n_u}) \tag{V.15}$$

Then we compare the smallest $I_{\hat{q}}$, where $\hat{q} \neq 0$ with I_0. If $I_{\hat{q}} < \theta_I I_0$, where θ_I is a pre-defined threshold, then the system is currently in discrete state $\hat{q}$ and the estimated discrete state $\tilde{q}$ equals to $\hat{q}$. Otherwise, the system is in an unknown discrete state.

Note that two discrete events may occur close to each other or even in adjacent time steps. For example, discrete event ψ_1 makes the system transition from discrete state 1 to 2 and discrete event ψ_2 makes the system to transition from discrete state 2 to 3. Suppose ψ_2 occurs right after the occurrence of ψ_1 and the matrix estimation by the RLS has not converged. Then the CRMMHO cannot identify the discrete state 2 after the occurrence of discrete event ψ_1. The CRMMHO only estimates the current discrete state, which is 3. Suppose ψ_2 occurs after the matrix estimation by the RLS has converged. Then the CRMMHO can identify both the discrete state 2 and discrete state 3 with some time delay. An extreme case is that the system never stays in one discrete state for a long time such that the matrix estimation by the RLS never converges. Then the CRMMHO framework can only estimate the continuous state of the system.

V.3.2 Demonstration

Among the seven steps in the CRMMHO framework mentioned in Section V.3.1, there are three key steps. One is the CR method which estimates the continuous state of the system. The second is the RLS method which estimates the continuous model of the current discrete state. The third is the model selection which estimates the current discrete state of the system. In this section, we demonstrate the effectiveness of the CR, the RLS and the model selection methods mathematically, respectively.

The CR method uses two continuous state observers to calculate the estimation errors of the two observers. Then the calculated estimation errors are used to correct the continuous state estimated by the plain Kalman filter. In [85], it is demonstrated that the CR method can provide a good estimation error calculation using Luenberger observers when the system is subject to process noise. Theorem 4 demonstrates that the same result remains for Kalman filters in the case that the Kalman filters are using incorrect state matrices for state estimation. The reason that the CR method is not affected by incorrect state matrices is that the two continuous state observers (Kalman filters) are using the same state matrices for state estimation. The CR method takes the difference of the estimated states of the two continuous state observers to calculate the estimation errors, then the effect of the incorrect state matrices are canceled out.

Theorem 4. *Given a hybrid system $\mathcal{H}$, suppose the difference of the state matrices between the discrete states q_1 and q_2 is ($\mathbf{\Delta A}_{q_1,q_2}, \mathbf{\Delta B}_{q_1,q_2}$), where $\mathbf{\Delta A}_{q_1,q_2} = \mathbf{A}_{q_1} - \mathbf{A}_{q_2}$ and $\mathbf{\Delta B}_{q_1,q_2} = \mathbf{B}_{q_1} - \mathbf{B}_{q_2}$,* *then the estimation errors* $\tilde{\mathbf{x}}_{\mathbf{e},0}, \tilde{\mathbf{x}}_{\mathbf{e},1}$ *calculated using Equation* (V.8) *are the same as the actual estimation errors of the two observers in sensor noise free case, i.e.,* $\mathbf{v} = 0$.

Proof. The linear system of discrete states q_1 and q_2 is

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{A}_q\mathbf{x}(t) + \mathbf{B}_q\mathbf{u}(t) + \mathbf{w}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) \end{aligned} \tag{V.16}$$

where $q = q_1, q_2$.

At time t_i when a discrete transition occurs, the discrete state is changed from q_1 to q_2. Since the discrete event is unobservable, the observers are still using the state matrices $(\mathbf{A}_{q_1}, \mathbf{B}_{q_1})$ for state estimation after time t_i.

Let us focus on observer 0 in this proof. The proof is similar for observer 1.

The estimated continuous state given by observer 0 is shown in Equation (V.7). Then the estimation error of observer 0 is

$$\begin{aligned} \mathbf{x}_{\mathbf{e},0}(t) &= \mathbf{x}(t) - \hat{\mathbf{x}}_0(t) \\ &= \mathbf{x}(t) - (\mathbf{I} - \mathbf{K}_0\mathbf{C})\hat{\mathbf{x}}(t|t-1) - \mathbf{K}_0\mathbf{y}(t) \\ &= (\mathbf{I} - \mathbf{K}_0\mathbf{C})(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t-1)) \end{aligned} \tag{V.17}$$

The estimation difference of the two observers is

$$\begin{aligned} \hat{\mathbf{x}}_0(t) - \hat{\mathbf{x}}_1(t) &= \mathbf{x}_{\mathbf{e},1}(t) - \mathbf{x}_{\mathbf{e},0}(t) \\ &= (\mathbf{K}_0\mathbf{C} - \mathbf{K}_1\mathbf{C}_1)(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t-1)) \end{aligned} \tag{V.18}$$

Then the calculated estimation error for observer 0 based on Equation (V.8) is

$$\begin{aligned}\tilde{\mathbf{x}}_{\mathbf{e},0}(t) &= (\mathbf{I}-\mathbf{K}_0\mathbf{C})(\mathbf{K}_0\mathbf{C}-\mathbf{K}_1\mathbf{C}_1)^{\dagger}(\hat{\mathbf{x}}_0(t)-\hat{\mathbf{x}}_1(t)) \\ &= (\mathbf{I}-\mathbf{K}_0\mathbf{C})(\mathbf{x}(t)-\hat{\mathbf{x}}(t|t-1)) \\ &= \mathbf{x}_{\mathbf{e},0}(t)\end{aligned} \tag{V.19}$$

Therefore, with inaccurate state matrices, the estimation error can still be correctly calculated in the sensor noise free case. □

According to [85], the CR method is affected by sensor noise. Here we give the upper bound of the calculation error of the estimation error under bounded sensor noise.

$$\begin{aligned}\mathbf{x}_{\mathbf{e},0}(t)-\tilde{\mathbf{x}}_{\mathbf{e},0}(t) &= \mathbf{x}_{\mathbf{e},0}(t)-(\mathbf{I}-\mathbf{K}_0\mathbf{C})(\mathbf{K}_0\mathbf{C}-\mathbf{K}_1\mathbf{C}_1)^{\dagger}(\hat{\mathbf{x}}_0(t)-\hat{\mathbf{x}}_1(t)) \\ &= (\mathbf{I}-\mathbf{K}_0\mathbf{C})(\mathbf{K}_0\mathbf{C}-\mathbf{K}_1\mathbf{C}_1)^{\dagger}(\mathbf{K}_1\mathbf{v}_1(t)-\mathbf{K}_0\mathbf{v}(t))\end{aligned} \tag{V.20}$$

So the upper bound is $||(\mathbf{I}-\mathbf{K}_0\mathbf{C})(\mathbf{K}_0\mathbf{C}-\mathbf{K}_1\mathbf{C}_1)^{\dagger}(\mathbf{K}_1+\mathbf{K}_0)||v$.

Remark 13. *According to Theorem 4, the calculation of the estimation error does not require the continuous dynamics of discrete state q_2 to be observable. Therefore, we only assume that the continuous dynamics of the initial discrete state $q(t_0)$ are observable in Assumption 12 to ensure the estimation error convergence in the initial discrete state.*

With the calculated estimation error, we can update the estimated continuous state of the system $\tilde{\mathbf{x}}(t)$ according to Equation (V.9). With the updated estimated continuous state, we use the RLS to calculate the state matrices of discrete state q_2. Proposition 6 demonstrates the convergence of the estimated matrices by proving that the norm of *a posteriori* error $\boldsymbol{\varepsilon}(t)$ (V.12) is smaller than the norm of *a priori* error $\boldsymbol{\varepsilon}_{\mathbf{0}}$ (V.6) in the noise free case. We state the proposition here for convenience. For the detailed proof, refer to [16].

Proposition 6. *In the RLS algorithm* (V.6)*, the norm of a posteriori error $\boldsymbol{\varepsilon}(t)$ is smaller than the norm of a priori error $\boldsymbol{\varepsilon}_{\mathbf{0}}$ at each time step.*

With the estimated continuous system matrices by the RLS method, we can use the estimated matrices as a reference to help determine the current discrete state of the system. As mentioned in Section V.1, we can uniquely determine the discrete state of the system if the current continuous dynamics are distinguishable from the continuous dynamics of other discrete states. The formal definition of distinguishability is given in [58].

Definition 17. *The continuous dynamics* $(\mathbf{A}_{q_1}, \mathbf{B}_{q_1})$ *and* $(\mathbf{A}_{q_2}, \mathbf{B}_{q_2})$ *are said to be distinguishable on* $[t_0, t_0+T]$*, if for any non-zero*

$$(\mathbf{x}_{q_1}(t_0), \mathbf{x}_{q_2}(t_0), u(\cdot)) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u}, \tag{V.21}$$

the outputs $\mathbf{y}_{q_1}(\cdot)$ *and* $\mathbf{y}_{q_2}(\cdot)$ *are not identical to each other on* $[t_0, t_0+T]$.

The necessary and sufficient condition for distinguishability in the noise free case in continuous time is given in [58]. The theorem and its proof for the discrete-time continuous system is similar to that in [58]. For convenience, we restate the theorem and give its proof under our modeling formalism.

Theorem 5. *The continuous models described by Equation* (V.1) *(*$\mathbf{w}=0, \mathbf{v}=0$*) of two discrete states* q_1 *and* q_2 *with the same initial condition* $\mathbf{x}(t_i)$ *are distinguishable in time* $[t_i, t_i+\Delta t]$ *if and only if*

$$\Delta t \geq \frac{2n_x - n_y}{n_y - n_u} \quad \textit{and} \quad 2n_x > n_y \quad \textit{and} \quad n_y > n_u$$

and

$$\mathfrak{D}_{q_1,q_2} = \begin{bmatrix} \mathbf{C}_{q_1,q_2} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{C}_{q_1,q_2}\mathbf{A}_{q_1,q_2} & \mathbf{C}_{q_1,q_2}\mathbf{B}_{q_1,q_2} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{C}_{q_1,q_2}\mathbf{A}^2_{q_1,q_2} & \mathbf{C}_{q_1,q_2}\mathbf{A}_{q_1,q_2}\mathbf{B}_{q_1,q_2} & \mathbf{C}_{q_1,q_2}\mathbf{B}_{q_1,q_2} & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots & \mathbf{0} \\ \mathbf{C}_{q_1,q_2}\mathbf{A}^{\Delta t}_{q_1,q_2} & \mathbf{C}_{q_1,q_2}\mathbf{A}^{\Delta t-1}_{q_1,q_2}\mathbf{B}_{q_1,q_2} & \dots & \dots & \mathbf{C}_{q_1,q_2}\mathbf{B}_{q_1,q_2} \end{bmatrix} \tag{V.22}$$

has full column rank, where

$$\mathbf{A}_{q_1,q_2} = \begin{bmatrix} \mathbf{A}_{q_1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{A}_{q_2} \end{bmatrix}$$
$$\mathbf{B}_{q_1,q_2} = \begin{bmatrix} \mathbf{B}_{q_1} \\ \mathbf{B}_{q_2} \end{bmatrix} \tag{V.23}$$
$$\mathbf{C}_{q_1,q_2} = \begin{bmatrix} \mathbf{C} & -\mathbf{C} \end{bmatrix}$$

Proof. The outputs of $(\mathbf{A}_{q_1}, \mathbf{B}_{q_1})$ and $(\mathbf{A}_{q_2}, \mathbf{B}_{q_2})$ with the same initial continuous state $\mathbf{x}(t_i)$ are

$$\begin{aligned} \mathbf{y}_{q1}(t_i + j) &= \mathbf{C}\mathbf{A}_{q1}^{j}\mathbf{x}(t_i) + \mathbf{C}\Sigma_{i=0}^{j-1}\mathbf{A}_{q1}^{i}\mathbf{B}_{q1}\mathbf{u}(t_i + j - 1 - i) \\ \mathbf{y}_{q2}(t_i + j) &= \mathbf{C}\mathbf{A}_{q2}^{j}\mathbf{x}(t_i) + \mathbf{C}\Sigma_{i=0}^{j-1}\mathbf{A}_{q2}^{i}\mathbf{B}_{q2}\mathbf{u}(t_i + j - 1 - i) \end{aligned} \tag{V.24}$$

Suppose $q1$ is the actual current discrete state and $q2$ is one of the possible discrete states of the system. The residual $\mathbf{r}_{q_2}(t_i + j)$ is

$$\begin{aligned} \mathbf{r}_{q2}(t_i + j) &= \mathbf{y}_{q1}(t_i + j) - \mathbf{y}_{q2}(t_i + j) \\ &= \mathbf{C}(\mathbf{A}_{q1}^{j} - \mathbf{A}_{q2}^{j})\mathbf{x}(t_i) + \mathbf{C}\Sigma_{i=0}^{j-1}(\mathbf{A}_{q1}^{i}\mathbf{B}_{q1} - \mathbf{A}_{q2}^{i}\mathbf{B}_{q2})\mathbf{u}(t_i + j - 1 - i) \end{aligned} \tag{V.25}$$

If the continuous dynamics of discrete states $q1$ and $q2$ are distinguishable, then $\mathbf{r}_{q_2}(t + j) = \mathbf{0}, \forall j = 0, 1, ..., \Delta t$ holds only when the initial continuous state $\mathbf{x}(t_i)$ and all of the inputs $\mathbf{u}$ are $\mathbf{0}$, i.e.,

$$\mathfrak{D}_{q1,q2} \begin{bmatrix} \tilde{\mathbf{x}}(t_i) \\ \tilde{\mathbf{x}}(t_i) \\ \mathbf{u}(t_i) \\ \cdots \\ \mathbf{u}(t_i + \Delta t - 1) \end{bmatrix} = \mathbf{0} \tag{V.26}$$

admits only trivia solution. That is, the matrix $\mathfrak{D}_{q1,q2} \in \mathbb{R}^{((\Delta t+1)\times n_y)\times(2n_x+\Delta t\times n_u)}$ is a square or tall matrix $((\Delta t + 1) \times n_y \geq 2n_x + \Delta t \times n_u)$ and has full column rank.

□

Suppose the current discrete state of the system is q_1 during the model selection step. If the continuous dynamics of two discrete states q_1 and q_2 satisfy Theorem 5, then the residual calculated based on the continuous models of q_2 is non-zero in time $[t_{ss}, t_{ss} + \Delta t]$ in the noise free case. If the continuous dynamics of all of the discrete states are distinguishable, then we can uniquely determine the current discrete state. Otherwise, the discrete state which has the smallest norm of residual is not guaranteed to be the actual discrete state of the system. Theorem 5 provides the condition of distinguishability in the noise free case. However, the hybrid system we consider in this chapter contains process and sensor noise as described in Section V.2. If the continuous dynamics of discrete states q_1 and q_2 are distinguishable, the residual $\mathbf{r}_{q_2}$ calculated in the noisy system case may still equal to $\mathbf{0}$ because of the system noise. Therefore, we provide a condition to address the noisy system case. Since the discrete state q_1 is the current discrete state of the system, the continuous dynamics of q_1 are subject to system noise. The discrete state q_2 is one of the possible discrete states of the system and the continuous dynamics of q_2 are ideal ($\mathbf{w} = 0, \mathbf{v} = 0$). Then, the outputs of $(\mathbf{A}_{q_1}, \mathbf{B}_{q_1})$ and $(\mathbf{A}_{q_2}, \mathbf{B}_{q_2})$ with the same initial continuous state $\mathbf{x}(t_{ss})$ are

$$\begin{aligned}
\mathbf{y}_{q_1}(t_{ss}+j) &= \mathbf{C}\mathbf{A}_{q_1}^j\mathbf{x}(t_{ss}) + \mathbf{C}\Sigma_{i=0}^{j-1}\mathbf{A}_{q_1}^i\mathbf{B}_{q_1}\mathbf{u}(t_{ss}+j-1-i) + \mathbf{C}\Sigma_{i=0}^{j-1}\mathbf{A}_{q_1}^i\mathbf{w}(t_{ss}+j-1-i) + \mathbf{v}(t_{ss}+j)\\
\mathbf{y}_{q_2}(t_{ss}+j) &= \mathbf{C}\mathbf{A}_{q_2}^j\mathbf{x}(t_{ss}) + \mathbf{C}\Sigma_{i=0}^{j-1}\mathbf{A}_{q_2}^i\mathbf{B}_{q_2}\mathbf{u}(t_{ss}+j-1-i)
\end{aligned} \tag{V.27}$$

The residual $\mathbf{r}_{q_2}(t_{ss}+j)$ is

$$\begin{aligned}
\mathbf{r}_{q_2}(t_{ss}+j) &= \mathbf{y}_{q_1}(t_{ss}+j) - \mathbf{y}_{q_2}(t_{ss}+j)\\
&= \mathbf{C}(\mathbf{A}_{q_1}^j - \mathbf{A}_{q_2}^j)\mathbf{x}(t_{ss}) + \mathbf{C}\Sigma_{i=0}^{j-1}(\mathbf{A}_{q_1}^i\mathbf{B}_{q_1} - \mathbf{A}_{q_2}^i\mathbf{B}_{q_2})\mathbf{u}(t_{ss}+j-1-i)\\
&+ \mathbf{C}\Sigma_{i=0}^{j-1}\mathbf{A}_{q_1}^i\mathbf{w}(t_{ss}+j-1-i) + \mathbf{v}(t_{ss}+j)
\end{aligned} \tag{V.28}$$

If $\mathbf{r}_{q_2}(t_{ss}+j) \neq \mathbf{0}, \forall j = 0, 1, ..., \Delta t$, then it is guaranteed that the continuous dynamics of q_1 and q_2 are not the same. That means if the following holds from time t_{ss} to time $t_{ss} + \Delta t$, then the continuous dynamics of q_1 and q_2 are different:

$$\begin{aligned} &\forall j = 0, 1, ..., \Delta t \\ \|\mathbf{C}(\mathbf{A}_{q_1}^{j} - \mathbf{A}_{q_2}^{j})\mathbf{x}(t_{ss}) + \mathbf{C}\Sigma_{i=0}^{j-1}(\mathbf{A}_{q_1}^{i}\mathbf{B}_{q_1} - \mathbf{A}_{q_2}^{i}\mathbf{B}_{q_2})\mathbf{u}(t_{ss} + j - 1 - i)\| &\geq \|\mathbf{C}\Sigma_{i=0}^{j-1}\mathbf{A}_{q_1}^{i}\mathbf{w}(t_{ss} + j - 1 - i) + \mathbf{v}(t_{ss} + j)\| \\ &\geq \|\mathbf{C}\|\Sigma_{i=0}^{j-1}\|\mathbf{A}_{q_1}\|^{i} w + v \end{aligned} \tag{V.29}$$

Note that the condition (V.29) is a sufficient condition. It is possible that we can distinguish the two continuous models of discrete states q_1 and q_2 although (V.29) is not satisfied.

V.3.3 Anomaly Diagnosis

Anomalies may occur in hybrid systems, causing a change in continuous dynamics. Some anomalies have been studied before and we know how those anomalies affect the continuous dynamics of the system, especially the anomalies caused by traditional faults. For the anomalies that we have *a priori* knowledge about, we can model them as discrete states which are part of the hybrid systems associated with unobservable discrete events. The CRMMHO framework can detect anomalies by detecting the occurrence of the discrete event and isolate anomalies by identifying the current discrete state. For some anomalies, however, we do not have *a priori* knowledge, especially the anomalies caused by attacks. For the anomalies that we do not have *a priori* knowledge about, we do not model them as part of the hybrid system model. The CRMMHO framework can detect anomalies by detecting a change in the continuous dynamics of the system, and the continuous dynamics estimated by the RLS can provide some insight about the anomalies.

V.4 Simulation Result

In this section, we show the simulation result of our motivating example - the microgrid system. First, we present the hybrid model of the microgrid system. Then, we compare the state estimation performance of the proposed CRMMHO framework with a plain Kalman filter.

V.4.1 Microgrid System

As introduced in Section V.1, unplanned islanding may occur within the system, threatening worker safety and interrupting Distributed Energy Resource (DER) management. Traditional Islanding Detection Methodology (IDM) can detect the occurrence of the unplanned islanding, but they cannot provide state estimation during the diagnosis.

As shown in Fig. I.2, the microgrid system we consider in this chapter contains two DERs, two buses, one transmission line and two local loads. The hybrid model of the microgrid system contains three discrete states under normal operation: grid-tied, islanded and synchronization, as illustrated in Fig. V.3. Suppose initially the system is in the islanded discrete state and it is commanded to connect to the grid; then the system transitions to the synchronization discrete state to synchronize the amplitude and phase of the voltage in the microgrid with that in the grid. After the synchronization is finished, the microgrid is connected to the grid and the system transitions to the grid-tied discrete state. Under normal operation, the islanding is scheduled, which transitions the system from the grid-tied discrete state to the islanded discrete state. The continuous model in each discrete state contains 40 continuous states, all of which are measured. For the detailed modeling of the microgrid, refer to [72].

The proposed observer framework CRMMHO contains two Kalman filters. Suppose all of the continuous state variables can be directly measured. Kalman filter 0 uses all 40 measurements and Kalman filter 1 uses 24 measurements for state estimation. The parameters of the proposed observer framework are shown in Table V.1. Note that in order to compare the following thresholds with corresponding variables (the difference of the estimated continuous states between the two continuous state observers and residual), we normalize the variables based on the mean and standard deviation of the variables under normal operation, as shown in (V.30).

$$x_{norm} = \frac{x - \overline{x}}{\sigma_x} \tag{V.30}$$

where x_{norm} is the variable, $\overline{x}$ is the mean of the variable, and σ_x is standard deviation of the variable.

Table V.1: Parameters of the CRMMHO framework

Parameters	Value
θ_{CR}	4
θ_{diff}	4
θ_{RLS}	4
θ_I	100

To illustrate the effectiveness of the CRMMHO framework, we consider the unplanned islanding scenario. We run the simulation for $10s$. An unplanned islanding occurs at time $3s$ and the system stays in the islanded discrete state until the end of the simulation.

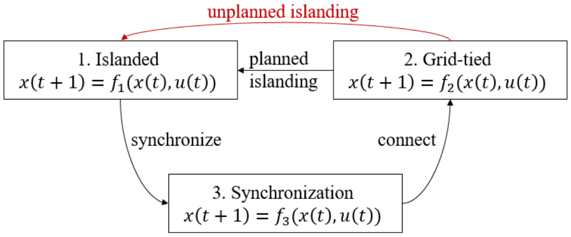


Figure V.3: Hybrid model of the microgrid system

V.4.2 Simulation Result

We compare the result of the CRMMHO framework with the plain Kalman filter 0 to show the effectiveness of the CRMMHO framework in state estimation and anomaly diagnosis. For continuous state estimation, we compare the 15^{th} estimated state variable, the current of bus 1 on the q-axis[4], which is used by the power controller to control the DER on bus 1 when the system is in the islanded discrete state. If the estimation of the current is bad, the controller cannot provide good control performance for the system, and may even damage the system under severe cases. Fig. V.4 shows the real and the estimated currents by the CRMMHO and the plain Kalman filter 0. From the simulation result we can see that after the unplanned islanding occurs, the estimated current by the plain Kalman filter has an offset compared to the real current in the steady state. The reason that the plain Kalman filter cannot provide a good state estimation after the occurrence of the unplanned islanding is because the continuous model used by the observer is the continuous model in the grid-tied discrete state as opposed to the continuous model in the islanded discrete state. In contrast, the estimated current by the CRMMHO can track the real current well. Fig. V.5 shows the real and the estimated discrete state by the CRMMHO. The estimated discrete state is uniquely determined at $4.1s$, which is $1.1s$ after the occurrence of the unplanned islanding. The $1.1s$ detection delay is caused by the convergence time of $||\hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1||$ and the RLS. In summary, the anomaly caused by the unplanned islanding is successfully diagnosed and we can provide a good state estimation during the diagnosis of the unplanned islanding.

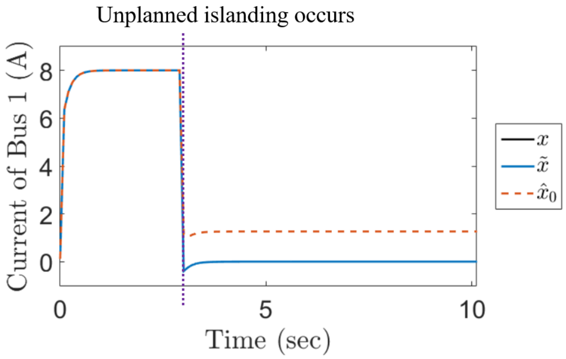


Figure V.4: The real current $\mathbf{x}^{(15)}$, the estimated continuous currents $\tilde{\mathbf{x}}^{(15)}, \hat{\mathbf{x}}_0^{(15)}$ by the CRMMHO and the plain Kalman filter 0 under unplanned islanding (note that the real current $\mathbf{x}^{(15)}$ overlaps with the estimated continuous state $\tilde{\mathbf{x}}^{(15)}$ by the CRMMHO)

[4] In a synchronous machines, the axis of the field winding in the direction of the DC field is the d-axis. 90 degrees later than the d-axis is the q-axis [38].

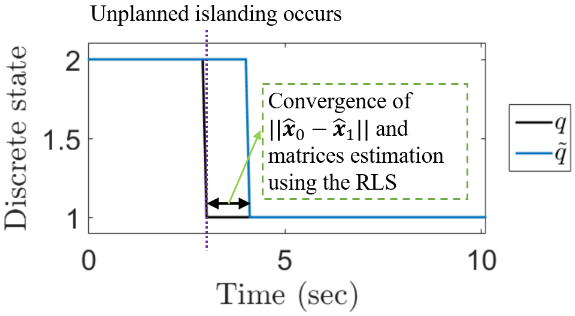


Figure V.5: The real discrete state q, estimated discrete state $\tilde{q}$ by the proposed observer under unplanned islanding

V.5 Conclusion & Future Work

In this chapter, we proposed a novel observer framework, the Convergence Ratio Multi-model Hybrid Observer (CRMMHO), for state estimation and anomaly diagnosis for hybrid systems with unobservable discrete events. The CRMMHO consists of two continuous state observers and three major methods: the Convergence Ratio (CR), the Recursive Least Squares (RLS) and the model selection. First, the CR method estimates the continuous state of the system. Then, the RLS method estimates the continuous system matrices. Finally, the model selection method uniquely determines the current discrete state. We demonstrated the CRMMHO framework in state estimation and anomaly detection mathematically and showed its effectiveness using the simulated microgrid system.

The CRMMHO has the following advantages compared to existing methods. It does not require *a priori* knowledge of the discrete transitions. The discontinuity in continuous variables is allowed. The CRMMHO is more computationally effective because: 1) it only uses two continuous state observers to estimate the continuous state of the system; and 2) the multiple continuous models only run for a short period of time and give a single output vector for each model per time step to determine the discrete state of the system. Note that this framework is flexible. The RLS method can be replaced with other on-line system identification methods and the model selection method can be replaced with any model falsification [74,75] or model invalidation methods [35].

More work needs to be done on improving the CRMMHO framework. The CRMMHO framework assumes that the initial discrete state of the system is given. However, the initial discrete state may not be known *a priori*. To relax the assumption of the known initial discrete state, we can run multiple continuous state observers in parallel, each corresponding to one known continuous models. By checking the convergence of the residual of each continuous state observer, we can determine the initial discrete state. In addition, in order to determine the discrete state of the system by analyzing residuals, we assume that the dynamics of different continuous models are similar to the extent that the concept of large and small residual norm can be universally applied. This assumption needs to be validated for specific system in the future. Under the CRMMHO framework, there are some thresholds, such as θ_{CR}, θ_I that we need to determine based on the specific system that we apply to. In order to determine θ_I, we need quantify the level of distinguishability based on the noise level. The impact of different levels of distinguishability on the residual should be studied, providing guidance on setting θ_I.

CHAPTER VI

Conclusions and Future Work

VI.1 Conclusions

This dissertation focuses on improving the security of Cyber-Physical Systems (CPS) by enhancing the utility of traditional observer-based anomaly diagnosis and mitigation methods. With the integration of the cyber world and the physical world, CPS have fast responses to the surrounding environment, such as changing markets and disturbances. As the world transitions to full integration of the cyber and physical realms, the unique challenges that CPS security face include not only traditional physical faults or attacks but also faults or attacks from the cyber domain. These cross-domain faults or attacks can cause anomalies in the physical systems, or even threaten public safety. In this dissertation, we enhance the security of CPS by improving on the traditional anomaly diagnosis and mitigation approaches that address anomalies caused by faults or attacks. We leverage the observer-based approach which is one of the most widely used anomaly diagnosis and mitigation techniques. An observer-based approach has the advantage that it can not only diagnose anomalies but also estimate the current state of the system. The estimated system state can help human operators understand how to resolve the anomaly as well as ensure the safety of the system.

The work done in the dissertation can be divided into three major parts, which are presented in Chapters III, IV and V, respectively. We have summarized the contributions of each chapter using a microgrid system as an example.

In Chapter III, we focus on anomalies in sensors and model CPS as continuous systems.

Under the multi-observer framework, three new methods are proposed and integrated to improve sensor anomaly diagnosis and mitigation. The three new methods respectively:

1. enable anomaly detection for some sensor anomalies during the observers' transient state;
2. detect some anomalies on critical sensors; and
3. potentially mitigate the impact of the anomalous sensor during the diagnosis process.

If we take one operation mode of the microgrid system as an example, such as the islanded operation mode, then the multi-observer sensor anomaly diagnosis and mitigation framework can be used to diagnose sensor anomalies and mitigate the impact of these sensor anomalies in the microgrid system.

In Chapter IV, we extend the work in Chapter III to hybrid systems, which consist of both continuous dynamics and discrete behavior. Additionally, the anomaly type is not limited to sensor anomalies. Assuming that the discrete behavior of hybrid systems is current-state observable, the contributions of this chapter are:

1. We propose a conflict-driven method to provide guarantees on the detection of some types of anomalies that are not detectable using traditional observer-based and residual-based methods in addition to the anomalies that can be detected by the traditional methods.
2. We define a classification taxonomy for anomalies in hybrid systems based on the variables that are affected, input-output data consistency, and diagnosability of the anomaly.
3. We develop a new hybrid observer, which uses a Set-Valued Observer (SVO) as the continuous state observer, for anomaly detection. With the SVO, we can apply the conflict-driven method to hybrid systems with unobservable continuous components.
4. We provide a mapping between conflict types and anomaly types. Based on the occurrence of the conflict types, we can identify if the anomaly is related to the continuous component of the system, the discrete component or both.

Using the microgrid system as an example, we can take all three operation modes (islanded, grid-tied and synchronization) into consideration. Assuming that we know when and which discrete event occurs if the system is under normal operation (the discrete events of the system are observable), we also know the invariants (allowable continuous state space) of the system in different nominal discrete states. Suppose the microgrid system is accidentally connected to the grid before the microgrid voltage is synchronized with the voltage in the grid. Then the conflict-driven method can detect the anomaly because the phase and the frequency of the islanded microgrid voltage are outside the invariant of the system in grid-tied operation mode.

In Chapter V, we consider a wider class of hybrid systems, including hybrid systems with unobservable discrete events. With unobservable discrete events, the discrete behavior of the hybrid systems may not be current-state observable and the conflict-driven anomaly detection method is not applicable. We address this in Chapter V with the contributions being:

1. We propose the Convergence Ratio Multi-model Hybrid Observer (CRMMHO) framework to estimate both the discrete and the continuous variables for hybrid systems with unobservable discrete events;
2. We use the CRMMHO framework to diagnose anomalies in more general hybrid systems; and
3. We apply the CRMMHO framework in the simulated microgrid system to validate its effectiveness.

For the microgrid system, if the discrete events of the system are not observable, then the conflict-driven method proposed in Chapter IV because the hybrid observer used in the conflict-driven method cannot estimate the discrete state of the system. The CRMMHO framework proposed in Chapter V is able to address this issue. If unexpected and unobservable islanding occurs to the grid-tied microgrid system, the CRMMHO framework can detect and isolate this anomaly by identifying that the discrete state of the microgrid is changed to islanded.

VI.2 Future Work

This dissertation enhances the utility of traditional observer-based anomaly diagnosis and mitigation methods in securing CPS. The methods proposed in this dissertation present opportunities for future work.

For hybrid system anomaly diagnosis, the invariant set is crucial since it describes the relationship between the continuous and discrete variables. In Chapter IV, the invariant set is formed based on both the continuous dynamics and safety constraints of the system for the corresponding discrete state. We use a hyperrectangle to do overapproximation which introduces some conservatism. The hyperrectangle may contain some continuous states which can never be reached when the system is under normal operation. In the future, we can use polyhedrons with fewer constraints to do overapproximation during set calculation to balance the estimation accuracy and the computation speed. In addition, an invariant of a discrete state may be a function of the continuous dynamics of the previous discrete state. Using a pre-determined invariant may also require an overapproximation. This can be addressed by determining the invariant of each discrete state dynamically. If we use a model predictive controller to control the system, we would have knowledge of the control input that we will apply to the system for next several time steps. With the known control input, we can calculate more conservative reachable sets as well as invariant sets dynamically to improve the estimation accuracy. In the future, we can integrate the conflict-driven method and the model predictive control to improve hybrid system anomaly diagnosis.

We have already done anomaly mitigation for continuous CPS in Chapter III. However, no work is done for anomaly mitigation for hybrid CPS in this dissertation. A majority of previous literature introduces off-line anomaly mitigation algorithms developed for hybrid systems [88]. However, anomalies may change system behavior abruptly, and anomaly mitigation strategies must be used on-line or even used before the anomaly is diagnosed so that the stability and safety of the system is maintained under anomalies. The results of anomaly mitigation for continuous systems or discrete systems may not be directly applied to hybrid systems since some hybrid systems may contain unique properties, such as Zeno behavior or instability caused by fast switchings. Additionally, anomalies in hybrid system could disrupt both the continuous dynamics and discrete behavior. So the anomaly mitigation strategies should maintain both the continuous performance including various stabilities, such as Lyapunov stability and input-to-state stability, and the discrete specifications that the hybrid system should follow. Developing an on-line anomaly mitigation algorithm for hybrid systems would represent an important extension of current capabilities.

In observer-based anomaly diagnosis, the fidelity of the system model is crucial to the diagnosis performance. If the system model is perfectly known, observer-based anomaly diagnosis can guarantee the diagnosis of anomalies and provide zero false positives. However, in reality, there are always system uncertainties, such as system noise and system modeling error and large parts of systems are poorly or not observable. Acquiring a high fidelity system model is never an easy task. Moreover, a high fidelity model may consume a lot of computational power. However, a low fidelity model can cause significant false positives and false negatives. To apply the proposed observer-based anomaly diagnosis proposed in this dissertation, it is important to understand and measure the diagnosis performance. Receiver Operating Characteristic (ROC) curve, which is a graphical plot illustrating the diagnostic ability of an anomaly diagnosis algorithm, could be used to help human operators choose the optimal threshold used in the diagnosis algorithm which can balance false positives or false negatives and the fidelity of the system model. Measuring the performance of different diagnosis algorithms for a specific system is one of our future works.

The main assumption of observer-based anomaly diagnosis is that the physics-based system model is known *a priori*. However, in real applications, the physics-based system model may not be available. Or maybe only part of the system has a known physics-based system model. There are other types of models besides physics-based models that can be used to represent a system and support anomaly diagnosis. For example, a statistical model could be developed using a data-driven method. Phenomenological models define relationships of variables to capture aspects of the physics or chemistry of the system; these model forms are often tuned statistically. Different types of models may capture different aspects of the system. The models developed using these methods can be used as a replacement for the models used in the observer-based anomaly diagnosis methods. However, as these models are non-deterministic, there may be false positives or negatives, which may be partially addressed by integrating these different types of models together to achieve better anomaly diagnosis. Intuitively, there are two ways of integration: a structural way and a non-structural way. In a structural way, we need to translate different models using a standardized model template in such a way so that the information provided by different models can be shared among the models. In a non-structural way, different types of models can be integrated using neural networks or other clustering approaches. Designing a flexible framework to effectively integrate various system models to diagnose anomalies in CPS is a future research direction.

As mentioned in Chapter I, CPS integrate the cyber world and the physical world to form large scale systems. CPS are composed of multi-domain subsystems. Modeling large scale CPS is challenging. One potential solution is to build CPS models in a distributed way. In distributed modeling, the system is decomposed into sub-systems spatially with weak interactions between them. Then those sub-systems are integrated together concurrently [18]. Another potential solution could be combinatorial modeling. Similar to distributed modeling, in combinatorial modeling, the system is also decomposed into sub-systems. But the decomposition may not be based on spatial distribution and strong coupling may still exist between sub-systems. When combining the sub-system models, the coupling between the sub-system models should be taken into consideration. Additionally, the level of abstraction of the sub-system models should be consistent. Lastly, the CPS model should be verified such that it can provide sufficient resolution for different usage purposes, such as control and anomaly diagnosis. Modeling and verification of large scale CPS model is another future research direction.

In this dissertation, we demonstrated the effectiveness of the proposed methods using different simulated systems, such as the suspension system, the Positive Train Control system and the microgrid system. By using different simulated systems, we have illustrated that our proposed methods have wide applications in various CPS. The application domain of the proposed methods is broader than the investigated applications. To make an actual contribution to the industry, we need to implement the proposed methods in real systems. For each proposed observer-based anomaly diagnosis and mitigation method, there are some thresholds that need to be pre-defined by human operators with some knowledge of the system. The values of the thresholds are system-specific. When implementing the proposed methods in a real system, how to tune the thresholds using the ROC curve to achieve satisfactory anomaly diagnosis performance is a future research direction.

As CPS are prevalent in critical infrastructures, the security of CPS faces different kinds of challenges in real applications. Sophisticated anomaly diagnosis is needed to enhance the security of CPS compared to either physical systems or cyber systems. It is impossible to make CPS impregnable since there may always be a smart attacker designing an intelligent attack on the system. The aim of proposing new anomaly diagnosis is to increase the cost of launching an attack and decrease the losses caused by an anomaly.

Appendix A

Appendices of Chapter III

Proof of CO method

Theorem 0 *Given an ideal control system (III.1) with* $\mathbf{w}(t) = \mathbf{0}$ *and* $\mathbf{v}(t) = \mathbf{0}$*, when sensor* i_f *is anomalous at time step t, the observer* i_f *gives the smallest norm estimation error if the anomaly signal satisfies*

$$||\boldsymbol{\gamma}(t)|| > ||\mathbf{L}_i\boldsymbol{\Gamma}_i||^{-1}(||\mathbf{E}_{i_f}\mathbf{x}_{\mathbf{e},i_f}(t)|| + ||\mathbf{E}_i\mathbf{x}_{\mathbf{e},i}(t)||), i \neq i_f. \tag{VI.1}$$

Proof. When sensor i_f is anomalous, $\boldsymbol{\Gamma}_{i_f} = 0^{(m-1)\times 1}$. Then the estimation error of observer i_f is

$$\mathbf{x}_{\mathbf{e},i_f}(t+1) = \mathbf{x}(t+1) - \tilde{\mathbf{x}}_{i_f}(t+1) = \mathbf{E}_{i_f}\mathbf{x}_{\mathbf{e},i_f}(t). \tag{VI.2}$$

In contrast, the estimation error of observer i ($i \neq i_f$) is

$$\mathbf{x}_{\mathbf{e},i}(t+1) = \mathbf{x}(t+1) - \tilde{\mathbf{x}}_i(t+1) = \mathbf{E}_i\mathbf{x}_{\mathbf{e},i}(t) - \mathbf{L}_i\boldsymbol{\Gamma}_i\boldsymbol{\gamma}(t). \tag{VI.3}$$

Therefore, if (VI.1) holds, the following is true

$$||\mathbf{x}_{\mathbf{e},i_f}(t+1)|| < ||\mathbf{x}_{\mathbf{e},i}(t+1)|| \quad \forall i = 0, 1..., m_{nc} \wedge i \neq i_f. \tag{VI.4}$$

□

Remark 14. *Remark: There is no physical meaning for* $\|\boldsymbol{\gamma}(t)\|$*. Theorem* 0 *gives a lower bound of* $\boldsymbol{\gamma}(t)$ *that the residual-based detection method could be used to select observer* i_f *, which is the one without the anomalous sensor* i_f*.*

Proof of Theorem 1

Theorem 1 *Given an ideal control system (III.1) with* $\mathbf{w}(t) = \mathbf{0}$ *and* $\mathbf{v}(t) = \mathbf{0}$*, the biases* $\tilde{\mathbf{d}}_{\mu(\nu)}(t)$ *and* $\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t)$ *are calculated according to* (III.21) *and* (III.22) *respectively, with the following results:*

1. *When the system is under disturbance,*

$$\begin{aligned}\forall \mu, \nu =& 0, 1, ..., m_{nc} \wedge \mu \neq \nu,\\ \tilde{\mathbf{d}}_{\mu(\nu)}(t) =& \tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) = \mathbf{d}(t).\end{aligned}$$

2. *When the system is under sensor anomaly,*

$$\begin{aligned}\forall \mu, \nu =& 0, 1, ..., m_{nc} \wedge \mu \neq \nu,\\ \tilde{\mathbf{d}}_{\mu(\nu)}(t) =& \tilde{\mathbf{d}}_{\nu(\mu)}(t),\\ \tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) \neq& \tilde{\mathbf{d}}_{\Lambda,\nu(\mu)}(t) \quad if \quad \mathbf{V}_\mu \neq \mathbf{V}_\nu.\end{aligned}$$

$$\tilde{\mathbf{d}}_{\mu(\nu)}(t) = (\mathbf{D}^\intercal\mathbf{D})^{-1}\mathbf{D}^\intercal[\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t+1) - \mathbf{E}_\mu\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t)], \tag{VI.5}$$

$$\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) = ((\mathbf{D}_{\Lambda,\mu})^\intercal\mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^\intercal[\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu(\nu)}(t+1) - \mathbf{E}_{\Lambda,\mu}\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu(\nu)}(t)], \tag{VI.6}$$

where $\mathbf{D}_{\Lambda,\mu} = (\mathbf{V}_\mu)^{-1}\mathbf{D}$*, and* $\mathbf{E}_{\Lambda,\mu} = (\mathbf{V}_\mu)^{-1}\mathbf{E}_\mu\mathbf{V}_\mu$*.*

Proof. 1) According to Lemma 1, $\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)} = \mathbf{x}_{\mathbf{e},\mu}$ if a disturbance exists. By substituting (III.10) to

(III.21), the calculated bias becomes

$$\tilde{\mathbf{d}}_{\mu(\nu)}(t) =(\mathbf{D}^{\top}\mathbf{D})^{-1}\mathbf{D}^{\top}[\mathbf{E}_{\mu}\mathbf{x}_{\mathbf{e},\mu}(t)+\mathbf{D}\mathbf{d}(t)-\mathbf{E}_{\mu}\mathbf{x}_{\mathbf{e},\mu}(t)] = \mathbf{d}(t). \tag{VI.7}$$

Similarly,

$$\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) =((\mathbf{D}_{\Lambda,\mu,})^{\top}\mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^{\top}(\mathbf{V}_{\mu})^{-1}[\mathbf{E}_{\mu}e_{\mu}(t)+\mathbf{D}\mathbf{d}(t)-\mathbf{E}_{\mu}\mathbf{x}_{\mathbf{e},\mu}(t)] = \mathbf{d}(t). \tag{VI.8}$$

2) Under sensor anomaly, the estimation error cannot be correctly calculated. Therefore, $\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}$ in (III.17) and $\mathbf{x}_{\mathbf{e},\mu}$ in (III.12) are substituted to (III.21) to calculate the difference between two biases based on two observers,

$$\begin{aligned}\tilde{\mathbf{d}}_{\mu(\nu)}(t)-\tilde{\mathbf{d}}_{\nu(\mu)}(t) = {} & (\mathbf{D}^{\top}\mathbf{D})^{-1}\mathbf{D}^{\top}[\mathbf{x}_{\mathbf{e},\mu}(t+1)-\mathbf{E}_{\mu}e^{\mu}(t)+\mathbf{E}_{\mu}(\mathbf{E}_{\nu}-\mathbf{E}_{\mu})^{-1}(\mathbf{L}_{\nu}\mathbf{\Gamma}_{\nu}-\mathbf{L}_{\mu}\mathbf{\Gamma}_{\mu})\boldsymbol{\gamma}(t) \\ & -\mathbf{x}_{\mathbf{e},\nu}(t+1)+\mathbf{E}_{\nu}\mathbf{x}_{\mathbf{e},\nu}(t)-\mathbf{E}_{\nu}(\mathbf{E}_{\nu}-\mathbf{E}_{\mu})^{-1}(\mathbf{L}_{\nu}\mathbf{\Gamma}_{\nu}-\mathbf{L}_{\mu}\mathbf{\Gamma}_{\mu})\boldsymbol{\gamma}(t)] \\ = {} & \mathbf{0}.\end{aligned} \tag{VI.9}$$

When the biases are calculated based on (III.22), then

$$\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) =((\mathbf{D}_{\Lambda,\mu})^{\top}\mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^{\top}(\mathbf{V}_{\mu})^{-1}[\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t+1)-\mathbf{E}_{\mu}\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t)], \tag{VI.10}$$

$$\tilde{\mathbf{d}}_{\Lambda,\nu(\mu)}(t) =((\mathbf{D}_{\Lambda,\nu})^{\top}\mathbf{D}_{\Lambda,\nu})^{-1}(\mathbf{D}_{\Lambda,\nu})^{\top}(\mathbf{V}_{\nu})^{-1}[\tilde{\mathbf{x}}_{\mathbf{e},\nu(\mu)}(t+1)-\mathbf{E}_{\nu}\tilde{\mathbf{x}}_{\mathbf{e},\nu(\mu)}(t)], \tag{VI.11}$$

are obtained for observer μ and ν, respectively. Based on (VI.9), the following is true

$$\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t+1)-\mathbf{E}_{\mu}\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t) = \tilde{\mathbf{x}}_{\mathbf{e},\nu(\mu)}(t+1)-\mathbf{E}_{\nu}\tilde{\mathbf{x}}_{\mathbf{e},\nu(\mu)}(t). \tag{VI.12}$$

So if $\mathbf{V}_{\mu} \neq \mathbf{V}_{\nu}$, then $((\mathbf{D}_{\Lambda,\mu})^{\top}\mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^{\top}(\mathbf{V}_{\mu})^{-1} \neq ((\mathbf{D}_{\Lambda,\nu})^{\top}\mathbf{D}_{\Lambda,\nu})^{-1}(\mathbf{D}_{\Lambda,\nu})^{\top}(\mathbf{V}_{\nu})^{-1}$. Thus $\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) \neq \tilde{\mathbf{d}}_{\Lambda,\nu(\mu)}(t)$.

□

Proof of Lemma 3

Lemma 3 *Given a control system (III.1) with bounded sensor noise and* $\mathbf{w}(t) = \mathbf{0}$*,* $\|\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t) - \mathbf{x}_{\mathbf{e},\mu}(t)\|$ *is bounded by* $\|(\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}\|(\|\mathbf{L}_\nu\| + \|\mathbf{L}_\mu\|)v$*.*

Proof. When sensor noise exists in the system, the estimation error evolution becomes

$$\mathbf{x}_{\mathbf{e},\mu}(t+1) = \mathbf{E}_\mu \mathbf{x}_{\mathbf{e},\mu}(t) - \mathbf{L}_\mu \mathbf{v}_\mu(t). \tag{VI.13}$$

Then, the difference of the estimated states between two observers μ and ν becomes

$$\mathbf{x}_{\mathbf{e},\mu,\nu}(t+1) = \mathbf{E}_\nu \mathbf{x}_{\mathbf{e},\nu}(t) - \mathbf{E}_\mu \mathbf{x}_{\mathbf{e},\mu}(t) - \mathbf{L}_\nu \mathbf{v}_\nu(t) + \mathbf{L}_\mu \mathbf{v}_\mu(t). \tag{VI.14}$$

Therefore, the calculated estimation error becomes

$$\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t) = \mathbf{x}_{\mathbf{e},\mu}(t) - (\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}(\mathbf{L}_\nu \mathbf{v}_\nu(t) - \mathbf{L}_\mu \mathbf{v}_\mu(t)). \tag{VI.15}$$

So, $\|\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}(t) - \mathbf{x}_{\mathbf{e},\mu}(t)\|$ is bounded by $\|(\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}\|(\|\mathbf{L}_\nu\| + \|\mathbf{L}_\mu\|)v$. □

Proof of Lemma 4

Lemma 4 *Given a control system (III.1) with bounded process noise and* $\mathbf{v}(t) = \mathbf{0}$*,* $\|\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) - \mathbf{d}(t)\|$ *is bounded.*

Proof. Estimation error can still be correctly calculated when the system is subject to process noise as proved in Lemma 2.

Then the bias calculated based on (III.22) becomes

$$\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) = \mathbf{d}(t) + ((\mathbf{D}_{\Lambda,\mu})^\intercal \mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^\intercal (\mathbf{V}_\mu)^{-1}\mathbf{w}(t). \tag{VI.16}$$

Therefore, $||\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) - \mathbf{d}(t)||$ is bounded by $||((\mathbf{D}_{\Lambda,\mu})^\intercal \mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^\intercal (\mathbf{V}_\mu)^{-1}||w$. □

Proof of Lemma 5

Lemma 5 *Given a control system (III.1) with bounded sensor noise and* $\mathbf{w}(t) = \mathbf{0}$*,* $||\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) - \mathbf{d}(t)||$ *is bounded.*

Proof. When the system has sensor noise, by substituting (VI.15) to (III.22),

$$\begin{aligned}\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) = &\mathbf{d}(t) - ((\mathbf{D}_{\Lambda,\mu})^\intercal \mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^\intercal (\mathbf{V}_\mu)^{-1}[\\ &(\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}(\mathbf{L}_\nu \mathbf{v}_\nu(t+1) - \mathbf{L}_\mu \mathbf{v}_\mu(t+1)) - \mathbf{E}_\mu(\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}(\mathbf{L}_\nu \mathbf{v}_\nu(t) - \mathbf{L}_\mu \mathbf{v}_\mu(t))].\end{aligned} \tag{VI.17}$$

Therefore, $||\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}(t) - \mathbf{d}(t)||$ is bounded by $||((\mathbf{D}_{\Lambda,\mu})^\intercal \mathbf{D}_{\Lambda,\mu})^{-1}(\mathbf{D}_{\Lambda,\mu})^\intercal (\mathbf{V}_\mu)^{-1}||(1 + ||\mathbf{E}_\mu||)||(\mathbf{E}_\nu - \mathbf{E}_\mu)^{-1}||(||\mathbf{L}_\nu|| + ||\mathbf{L}_\mu||)v$. □

Proof of Proposition 1

Proposition 1 *Given a control system (III.1), and an open-loop observer (III.3), the following results can be drawn:*

1. *If all of the eigenvalues of* $\mathbf{A}$ *lie inside the unit circle, then the estimation error of an open-loop observer is bounded;*
2. *If one or more of the eigenvalues of* $\mathbf{A}$ *lie on the unit circle and* $||\mathbf{A}|| = 1$*, then the estimation error of an open-loop observer is bounded.*

Proof. The real state of the system is

$$\mathbf{x}(t) = \mathbf{A}^t\mathbf{x}(t_0) + \Sigma_{i=0}^{t-1}\mathbf{A}^i\mathbf{B}\mathbf{u}(t-1-i) + \Sigma_{i=0}^{t-1}\mathbf{A}^i\mathbf{w}(t-1-i). \tag{VI.18}$$

The state estimated by the open-loop observer is

$$\hat{\mathbf{x}}(t) = \mathbf{A}^t\hat{\mathbf{x}}(t_0) + \Sigma_{i=0}^{t-1}\mathbf{A}^i\mathbf{B}\mathbf{u}(t-1-i). \tag{VI.19}$$

Then, the estimation error of the open-loop observer is

$$\mathbf{x_{e,o}}(t) = \mathbf{A}^t\mathbf{x_{e,o}}(t_0) + \Sigma_{i=0}^{t-1}\mathbf{A}^i\mathbf{w}(t-1-i). \tag{VI.20}$$

1. If all of the eigenvalues of $\mathbf{A}$ lie inside the unit circle, then $\mathbf{A}^t\mathbf{x_{e,o}}(t_0)$ is converging and according to [47]

$$\lim_{i\to\infty}\{\mathbf{A}^i\}_{j_1,j_2} = 0 \quad j_1, j_2 = 1, ..., n, \tag{VI.21}$$

where $\{\mathbf{A}^i\}_{j_1,j_2}$ is the element at the j_1^{th} row and the j_2^{th} column of $\mathbf{A}^i$. Let $\bar{\mathbf{A}}^{(j_1,j_2)} = max(\{\mathbf{A}^i\}_{j_1,j_2})$, where $i = 0, 1, ..., t-1$ and $\bar{\mathbf{A}}$ is formed by $\bar{\mathbf{A}}^{(j_1,j_2)}$. Then,

$$\Sigma_{i=0}^{t-1}\mathbf{A}^i\mathbf{w}(t-1-i) \leq \bar{\mathbf{A}}\Sigma_{i=0}^{t-1}\mathbf{w}(t-1-i). \tag{VI.22}$$

Since the random process noise w has zero-mean and bound w, $\Sigma_{i=0}^{t-1}\mathbf{A}^i\mathbf{w}(t-1-i)$ is bounded as well.

2. If one or more of the eigenvalues of $\mathbf{A}$ lie on the unit circle, then $\mathbf{A}^t\mathbf{x_{e,o}}(t_0)$ is bounded. The other term $\Sigma_{i=0}^{t-1}\mathbf{A}^i\mathbf{w}(t-1-i)$ is a linear combination of the random vector $\mathbf{A}^i\mathbf{w}(t-1-i)$. For a vector $\mathbf{A}\mathbf{w}(t)$, each element is a linear combination of zero-mean random variables in vector $\mathbf{w}(t)$ with the elements in the same row of $\mathbf{A}$ as coefficients

$$\mathbf{A}^{(j,:)}\mathbf{w}(t) = \Sigma_{i=1}^{n}\mathbf{A}^{(j,i)}\mathbf{w}^{(j)}(t). \tag{VI.23}$$

Since $||\mathbf{A}|| = 1$, i.e., $\Sigma_{i=1}^{n}|\mathbf{A}^{(j,i)}| \leq 1$ based on the definition of ∞-norm, $\mathbf{A}\mathbf{w}(t)$ is a zero-mean random vector with bound w. Thus, $\mathbf{A}^{i}\mathbf{w}(t-1-i)$ is also a zero-mean random vector with bound w. Therefore, $\Sigma_{i=0}^{t-1}\mathbf{A}^{i}\mathbf{w}(t-1-i)$ is bounded.

□

Proof of Proposition 2

Proposition 2 *Given a control system (III.1), an open-loop observer is updated every $\kappa_{f,g}$ time steps. The impact of the system noise on the averaged residual (VI.24) is mitigated.*

$$\mathbf{r}_{avg,g}(t+(j_N-1)\kappa_{f,g}) = \frac{1}{j_N}\Sigma_{j=1}^{j_N}\mathbf{r}_g(t+(j_N-j)\kappa_{f,g}), \tag{VI.24}$$

where j_N is a positive integer.

Proof. Since the process noise and sensor noise are zero-mean vectors,

$$\begin{aligned}\Sigma_{i=0}^{\infty}\mathbf{w}(i) =& \mathbf{0}^{n_x\times 1},\\ \Sigma_{i=0}^{\infty}\mathbf{v}(i) =& \mathbf{0}^{n_y\times 1}.\end{aligned} \tag{VI.25}$$

The residual generated by a single open-loop observer over one update period is

$$\begin{aligned}\mathbf{r}_g(t+(j_N-j)\kappa_{f,g}) = &\mathbf{y}(t+(j_N-j)\kappa_{f,g}) - \mathbf{C}\hat{\mathbf{x}}_g(t+(j_N-j)\kappa_{f,g})\\ =&\mathbf{C}\mathbf{x}(t+(j_N-j)\kappa_{f,g}) + \mathbf{v}(t+(j_N-j)\kappa_{f,g}) - \mathbf{C}\hat{\mathbf{x}}_g(t+(j_N-j)\kappa_{f,g})\\ =&\mathbf{C}\mathbf{A}^{t}e((j_N-j)\kappa_{f,g}) + \mathbf{v}(t+(j_N-j)\kappa_{f,g}) + \Sigma_{i=0}^{t-1}\mathbf{C}\mathbf{A}^{i}\mathbf{w}(t-1+(j_N-j)\kappa_{f,g}-i).\end{aligned} \tag{VI.26}$$

Then the averaged residual is

$$
\begin{aligned}
\mathbf{r}_{avg,g}(t+(j_N-1)\kappa_{f,g}) &= \frac{1}{j_N}\Sigma_{j=1}^{j_N}\mathbf{r}_g(t+(j_N-j)\kappa_{f,g}) \\
&= \frac{1}{j_N}\Sigma_{j=1}^{j_N}(\mathbf{CA}^t e((j_N-j)\kappa_{f,g})+\mathbf{v}(t+(j_N-j)\kappa_{f,g})+\Sigma_{i=0}^{t-1}\mathbf{CA}^i\mathbf{w}(t-1+(j_N-j)\kappa_{f,g}-i)).
\end{aligned}
\tag{VI.27}
$$

If $j_N \to \infty$, then

$$
\mathbf{r}_{avg,g}(t+(j_N-1)\kappa_{f,g}) = \frac{1}{j_N}\Sigma_{j=1}^{j_N}\mathbf{CA}^t\mathbf{x_e}((j_N-j)\kappa_{f,g}). \tag{VI.28}
$$

Therefore, the impact of system noise is mitigated. □

Tables of Notations in Chapter III

Table A.1: Table of Matrices

Matrices	Meaning
$\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\mathbf{F},\mathbf{\Gamma}$	System matrices, controller gain, and sensor anomaly vector
$\mathbf{C}_i,\mathbf{L}_i,\mathbf{\Gamma}_i,\mathbf{E}_i$	Output matrix, observer gain, anomaly vector, state matrix for observer i, $\mathbf{E}_i = \mathbf{A}-\mathbf{L}_i\mathbf{C}_i$
$\mathbf{V}_i$	A collection of eigenvectors of matrix $\mathbf{E}_i$
$\mathbf{E}_{\Lambda,i},\mathbf{B}_{\Lambda.i},\mathbf{D}_{\Lambda,i}$	Transformed matrices for observer i

Table A.2: Table of Variables

Variables	Meaning
$\mathbf{x},\mathbf{y},\mathbf{u},\mathbf{w},\mathbf{v},\mathbf{d},\boldsymbol{\gamma}$	System state, output, input, process noise, sensor noise, disturbance and sensor anomaly signal
$\mathbf{w},\mathbf{v}$	Bounds of the process noise and the sensor noise
m_{nc}	Number of the non-critical sensors
n_x,n_y,n_u,n_d	Dimensions of system state, output, input, and disturbance
S_{nc},S_c	Sets of non-critical sensors and critical sensors
$\mathbf{y}_i,\mathbf{v}_i$	Output and sensor noise for observer i
$\mathbf{x}_{\mathbf{e},i},\mathbf{x}_{\mathbf{e},\mathbf{o}}$	Estimation error of closed-loop observer i and an open-loop observer
$\tilde{\mathbf{x}}_i$	Estimated state by closed-loop observer i
$\hat{\mathbf{x}}_{g,i}$	Estimated state by open-loop observer i in group g
t_{ss}	Time steps for a closed-loop observer to reach its steady state

Table A.3: Table of Variables

Indicators, index	Meaning
I_A, I_F, I_D	Alarms for anomaly, sensor anomaly, and disturbance
i_f	Anomalous sensor index
I_{FB}	Index of the closed-loop observer for feedback

Table A.4: Table of Notations for the CO Method

	Meaning
$\mathbf{Q}_i$	Weighting matrix for observer i
θ_{CO}	Threshold for the CO method
$\mathbf{r}_i$	The residual generated by closed-loop observer i
$\tilde{\boldsymbol{\gamma}}$	Calculated anomaly signal

Table A.5: Table of Notations for the CR Method

	Meaning
$\mathbf{x}_{\mathbf{e},\mu,\nu}$	The difference of estimated states of two observers
$\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}, \bar{x}_{e,\mu(\nu)}$	Estimation error of observer μ calculated based on observers μ and ν and its upper bound
$\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu(\nu)}$	The calculated estimation error of observer μ after changing the coordinates
$\tilde{\mathbf{x}}_{\mathbf{e},\mu}$	Overall estimation error of observer μ, which is a function of $\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}, \nu = 0, 1, ..., m_{nc} \wedge \nu \neq \mu$
$\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu}$	Overall estimation error of observer μ after changing the coordinates
$cr_{i,j}$	Convergence ratio of the j^{th} state estimation error of observer i
$\tilde{\mathbf{d}}_{\mu(\nu)}$	The bias based on the calculated estimation error $\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}$
$\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}, \bar{d}_{\mu(\nu)}$	The bias based on the calculated estimation error $\tilde{\mathbf{x}}_{\mathbf{e},\Lambda,\mu(\nu)}$ and its upper bound
κ_{CR}	Time steps for the CR method
θ_{CR}	Threshold to determine the occurrence of an anomaly
$\theta_{d,\mu(\nu),\zeta(\eta)}$	Threshold to distinguish a sensor anomaly from a disturbance
ϕ_ν	Weighting ratio of calculated estimation error $\tilde{\mathbf{x}}_{\mathbf{e},\mu(\nu)}$
$\psi_{\mu(\nu)}$	Weighting ratio of calculated bias $\tilde{\mathbf{d}}_{\Lambda,\mu(\nu)}$

Table A.6: Table of Notations for the MOLO Method

	Meaning
M, N	The number of open-loop observers groups and the number of open-loop observers in one group
$\kappa_{f,g}, \kappa_{\Delta,g}$	Update period, update interval between two adjacent open-loop observers for group g
$\mathbf{r}_{g,i}$	Residual signal of observer i in group g
$H_g, \mathbf{r}_{avg,g}$	Leading observer, averaged residual in group g
$\boldsymbol{\theta}_{MOLO,g}$	Threshold for the MOLO

Table A.7: Table of Notations for the CCI Method

	Meaning
$\kappa_{CCI}, \theta_{CCI}$	Horizontal window, threshold for the CCI method
$\boldsymbol{\Delta}\mathbf{u}_i$	Control input difference of closed-loop observer i

Table A.8: Table of Other Notations

	Meaning
$\mathbf{x}^{(j)}$	The j^{th} element of a vector **x**
$\mathbf{A}^{(j,:)}$	The j^{th} row of a matrix **A**
$\mathbf{A}^{(j_1,j_2)}$	The element at j_1^{th} row j_2^{th} column of a matrix **A**
$\lVert \cdot \rVert$	The infinity norm $\lVert \cdot \rVert_\infty$

Appendix B

Appendices of Chapter IV

Tables of Notations in Chapter IV

Table B.1: Table of Systems

	Meaning
$\mathcal{H},\mathcal{H}_n,\mathcal{H}_f,\mathcal{M}_n$	Hybrid automaton, nominal hybrid automaton, anomalous hybrid automaton and nominal Finite State Machine
$\mathcal{X}$, $\mathcal{U}$, $\mathcal{Y}$, $Init$, $field$, E, ϕ, η	State, input, output, initial state, field vector, discrete events, discrete transitions, and output map of hybrid automaton $\mathcal{H}$
X,U,Y	A set of continuous states, inputs and outputs
Q,Ψ,Ω	A set of discrete states, inputs and outputs
Ψ_o,Ψ_{uo}	A set of observable input events and a set of unobservable input events
E,E_o,E_{uo}	A set of discrete events, $E=\Psi\dot{\cup}\Omega$, a set of observable events and a set of unobservable events
ζ	A discrete output map
h	A continuous output equation
Q_n,Q_f	A set of nominal discrete states and a set of anomalous discrete states
Inv_q,$\underline{\beta}_i$,$\overline{\beta}_i$	Invariant of discrete state q, the lower bound and the upper bound of the invariant along the i^{th} state variable
$G(q,q',\psi)$, $\mathcal{P}(q,q',\psi)$	Guard condition and its hyperplane corresponding to discrete transition $\phi(q,\psi)=q'$
$\mathcal{L}(q,q',\psi)$	Post-guard hyperplane of guard $G(q,q',\psi)$
$q_{\mathcal{N},q}$	Neighbor discrete state of nominal discrete state q
$\mathcal{N}_q$	Neighbor set of nominal discrete state q
$\mathcal{R}_{in,q},\mathcal{R}_{no,q}$	Intermediate region and normal operating region of discrete state q

Table B.2: Table of System Matrices

	Meaning
$\mathbf{A}_q, \mathbf{B}_q, \mathbf{C}_q, \mathbf{\Gamma}_1, \mathbf{\Gamma}_2$	System matrices of discrete state q and anomaly vector on state equation and output equation, respectively

Table B.3: Table of System Variables

	Meaning
$\mathbf{x}$,$\mathbf{y}$,$\mathbf{u}$,$\mathbf{w}$,$\mathbf{v}$, $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, $\mathbf{d}$	System state, output, input, process noise, sensor noise, anomaly signals added to the state equation and output equation, respectively, and system noise $\mathbf{d}$ = $[\mathbf{w} \quad \mathbf{v}]^\intercal$
w, v, μ	Upper bounds of the norm of process noise, sensor noise and input, respectively
$\mathcal{B}_d, \mathcal{B}_{x_o}$	Bounds of system noise and uncertain initial continuous state, respectively
n_x, n_y, n_u	Dimensions of system state, output, and input

Table B.4: Table of Observer Variables

	Meaning
$\mathcal{O}, \mathcal{C}, \mathcal{D}$	Hybrid observer, continuous state observer and discrete state observer
$\tilde{q}(t)$, $\hat{X}(\mathbf{y}, t)$, $\tilde{X}(\mathbf{y}, t)$, $\tilde{\mathbf{x}}_c(t)$	Estimated discrete state, a set of possible continuous states based on a single measurement, estimated continuous state set, and the central estimated continuous state at time step t
$\mathbf{x_e}$	Estimation error of the continuous state observer
$\boldsymbol{\theta}$	Upper bound of the estimation error of the central estimated continuous state
Rack	*Rack* operator used by the Set-Valued Observer (continuous state observer)

Table B.5: Table of the Conflict-driven Method

	Meaning
$X_I(t)$	Initial set at time step t
$Z, \tilde{\mathbf{x}}_c, \mathbf{g_i}$	Zonotope and its center and generators
$R_{\delta_{\tilde{q}(t)}}(X_I(t))$	$\delta_{\tilde{q}(t)}$-time step forward reachable set starting from initial set $X_I(t)$

Table B.6: Table of Variables Used in Robust Optimization

	Meaning
$\mathbf{H}_{i_G}$	Projection row vector with the i_G^{th} entry “i” and “0” elsewhere
$\mathbf{J}$,$\boldsymbol{\rho_1}$,$\boldsymbol{\rho_2}$,$\mathbf{\Lambda}$	Matrices and vectors used in robust optimization after changing variables

Table B.7: Table of Positive Train Control System

	Meaning
x_p, x_v, x_f	Train position, speed and force
y_p, y_v	Measured train position, speed and force
$(\mathbf{A_v}, \mathbf{B_v})$, $(\mathbf{A_p}, \mathbf{B_p})$, $(\mathbf{A_b}, \mathbf{B_b})$	Continuous dynamics under speed control and position control, and during braking, respectively

Appendix C

Appendices of Chapter V

Tables of Notations in Chapter V

Table C.1: Table of Systems

	Meaning
$\mathcal{H}$	Hybrid automaton
$\mathcal{X}$, $\mathcal{U}$, $\mathcal{Y}$, $Init$, $field$, ϕ, h, f_r	State, input, output, initial state, field vector, discrete transitions, continuous output equation, and reset function of hybrid automaton $\mathcal{H}$
X, U, Y	A set of continuous states, inputs and outputs
Q, Ψ	A set of discrete states and inputs
$G(q, q', \psi)$	Guard condition corresponding to discrete transition $\phi(q, \psi) = q'$
$\mathbf{A}_q, \mathbf{B}_q, \mathbf{C}$	System matrices of discrete state q

Table C.2: Table of System Variables

	Meaning
$\mathbf{x}$,$\mathbf{y}$,$\mathbf{u}$,$\mathbf{w}$,$\mathbf{v}$	System continuous state, output, input, process noise and sensor noise
w, v	Upper bounds of the norm of process noise and sensor noise, respectively
$\mathbf{W}$, $\mathbf{V}$	Covariance matrix of process noise and sensor noise, respectively
n_x, n_y, n_u	Dimensions of system continuous state, output, and input
$\mathbf{\Delta A}_{q_1,q_2}$, $\mathbf{\Delta B}_{q_1,q_2}$	Difference of state matrices between discrete states q_1 and q_2

Table C.3: Table of Observer and the Convergence Ratio (CR) Variables

	Meaning
$\tilde{q}$, $\hat{\mathbf{x}}_0$, $\hat{\mathbf{x}}_1$, $\tilde{\mathbf{x}}$	Estimated discrete state, estimated continuous state by continuous state observer 0, estimated continuous state by continuous state observer 1, estimated continuous state by the CRMMHO
$\mathbf{x}_{\mathbf{e},0}$,$\mathbf{x}_{\mathbf{e},1}$	Estimation errors of the continuous state observers 0 and 1, respectively
$\tilde{\mathbf{x}}_{\mathbf{e},0}$,$\tilde{\mathbf{x}}_{\mathbf{e},1}$	Calculated estimation errors of the continuous state observers 0 and 1, respectively
$\mathbf{P}$	Covariance matrix of Kalman filter
$\mathbf{K}$	Kalman filter gain
t_{ss}	The steady state time step of the CR method
θ_{CR}	Threshold for the difference of the estimated continuous states of the two continuous state observers
θ_{diff}	Threshold for the change of the difference of the estimated continuous states of the two continuous state observers

Table C.4: Table of the Recursive Least Squares (RLS) Variables

	Meaning
$\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$	Estimated system matrices
$\mathbf{\Xi}$, $\boldsymbol{\xi}$	Estimated matrix $\mathbf{\Xi}^{\top} = [\tilde{\mathbf{A}} \quad \tilde{\mathbf{B}}]$ and input vector
$\mathbf{G_{RLS}}$	The RLS adaptation gain
$\boldsymbol{\varepsilon_0}$, $\boldsymbol{\varepsilon}$	*A priori* error and *a posteriori* error
Δt_{RLS}	Time steps used for the RLS to converge

Table C.5: Table of the Model Selection Variables

	Meaning
$\mathbf{y}_{o,\tilde{q}}$	Output of continuous model $\tilde{q}$
$\mathbf{r}_{\tilde{q}}$	Residual signal of continuous model $\tilde{q}$
$I_{\tilde{q}}$	Similarity index of continuous model $\tilde{q}$
θ_I	Threshold for the model selection method
Δt	Time steps used to run the models

BIBLIOGRAPHY

[1] Abdelkader Akhenak, Mohammed Chadli, Didier Maquin, and José Ragot. Sliding mode multiple observer for fault detection and isolation. In *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, volume 1, pages 953–958. IEEE, 2003.

[2] C Aubrun, D Sauter, H Noura, and M Robert. Fault diagnosis and reconfiguration of systems using fuzzy logic: application to a thermal plant. *International Journal of Systems Science*, 24(10):1945–1954, 1993.

[3] Christel Baier and Joost-Pieter Katoen. *Principles of model checking*. MIT press, 2008.

[4] Irvin J Balaguer, Heung-Geun Kim, Fang Z Peng, and Eduardo I Ortiz. Survey of photovoltaic power systems islanding detection methods. In *Industrial Electronics, 2008. IECON 2008. 34th Annual Conference of IEEE*, pages 2247–2252. IEEE, 2008.

[5] Andrea Balluchi, Luca Benvenuti, Maria D Di Benedetto, and Alberto L Sangiovanni-Vincentelli. Design of observers for hybrid systems. In *HSCC*, pages 76–89. Springer, 2002.

[6] JP Barbot, H Saadaoui, M Djemai, and N Manamanni. Nonlinear observer for autonomous switching systems with jumps. *Nonlinear Analysis: Hybrid Systems*, 1(4):537–547, 2007.

[7] Dimitris Bertsimas and Frans JCT de Ruiter. Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds. *INFORMS Journal on Computing*, 28(3):500-511, 2016.

[8] Rakesh B Bobba, Katherine M Rogers, Qiyan Wang, Himanshu Khurana, Klara Nahrstedt, and Thomas J Overbye. Detecting false data injection attacks on dc state estimation. In *Preprints of the First Workshop on Secure Control Systems, CPSWEEK*, volume 2010, 2010.

[9] Kamel Bouibed, Lynda Seddiki, Kevin Guelton, and Herman Akdag. Actuator and sensor fault detection and isolation of an actuated seat via nonlinear multi-observers. *Systems Science & Control Engineering: An Open Access Journal*, 2(1):150–160, 2014.

[10] Michael Stephen Branicky, Vivek S Borkar, and Sanjoy K Mitter. A unified framework for hybrid control. In *Decision and Control, 1994., Proceedings of the 33rd IEEE Conference on*, volume 4, pages 4228–4234. IEEE, 1994.

[11] Stephen L Campbell and Ramine Nikoukhah. *Auxiliary signal design for failure detection*. Princeton University Press, 2015.

[12] Alvaro Cardenas, Saurabh Amin, Bruno Sinopoli, Annarita Giani, Adrian Perrig, and Shankar Sastry. Challenges for securing cyber physical systems. In *CPSSW*, page 5, 2009.

[13] Alvaro A Cardenas, Saurabh Amin, and Shankar Sastry. Secure control: Towards survivable cyber-physical systems. In *Distributed Computing Systems Workshops, 2008. ICDCS'08. 28th International Conference on*, pages 495–500. IEEE, 2008.

[14] Christos G Cassandras and Stephane Lafortune. *Introduction to discrete event systems*. Springer Science & Business Media, 2009.

[15] Jie Chen, Ron J Patton, and Hong-Yue Zhang. Design of unknown input observers and robust fault detection filters. *International Journal of Control*, 63(1):85–105, 1996.

[16] Xu Chen. Lecture 15: System identification and recursive least squares. *pdf file available at `http://www.engr.uconn.edu/~xchen/Teaching/ME233Sp2014/_static/UCB_ME233_15_sysid_rls.pdf`*, 2014.

[17] Sumith Choy and Erik Weyer. Reconfiguration schemes to mitigate faults in automated irrigation channels. *Control Engineering Practice*, 16(10):1184–1194, 2008.

[18] Dobre Ciprian, Cristea Valentin, and Iosif C Legrand. Simulation framework for modeling large-scale distributed systems. *arXiv preprint arXiv:1106.6122*, 2011.

[19] Robert N Clark. Instrument fault detection. *IEEE Transactions on Aerospace and electronic systems*, (3):456–465, 1978.

[20] Robert N Clark. A simplified instrument failure detection scheme. *IEEE Transactions on Aerospace and Electronic Systems*, 14(4):558–563, 1978.

[21] Critical Infrastructure Cybersecurity. Framework for improving critical infrastructure cybersecurity. *Framework*, 1:11, 2014.

[22] Jonny Carlos da Silva, Abhinav Saxena, Edward Balaban, and Kai Goebel. A knowledge-based system approach for sensor fault modeling, detection and mitigation. *Expert Systems with Applications*, 39(12):10977–10989, 2012.

[23] Werner Damm, Alfred Mikschl, Jens Oehlerking, Ernst-Rüdiger Olderog, Jun Pang, André Platzer, Marc Segelken, and Boris Wirtz. Automating verification of cooperation, control, and design in traffic applications. In *Formal Methods and Hybrid Real-Time Systems*, pages 115–169. Springer, 2007.

[24] Steven Ding. *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2008.

[25] Abhishek Dubey, Steve Nordstrom, Turker Keskinpala, Sandeep Neema, Ted Bapty, and Gabor Karsai. Towards a verifiable real-time, autonomic, fault mitigation framework for large scale real-time systems. *Innovations in Systems and Software Engineering*, 3(1):33–52, 2007.

[26] Christopher Edwards and Chee Pin Tan. Sensor fault tolerant control using sliding mode observers. *Control Engineering Practice*, 14(8):897–908, 2006.

[27] Al Faruque, Mohammad Abdullah, Sujit Rokka Chhetri, Arquimedes Canedo, and Jiang Wan. Acoustic side-channel attacks on additive manufacturing systems. In *Proceedings of the 7th International Conference on Cyber-Physical Systems*, page 19. IEEE Press, 2016.

[28] Thierry Floquet, Diego Mincarelli, Alessandro Pisano, and Elio Usai. Continuous and discrete state estimation in linear switched systems by sliding mode observers with residuals' projection. *IFAC Proceedings Volumes*, 45(9):265–270, 2012.

[29] Paul M Frank and Ralf Seliger. Fault detection and isolation in automatic processes. *Control and Dynamic Systems*, 49:241–287, 2012.

[30] Antoine Girard. Reachability of uncertain linear systems using zonotopes. In *HSCC*, volume 5, pages 291–305. Springer, 2005.

[31] Charles H Goodrich and James Kurien. Continuous measurements and quantitative constraints: Challenge problems for discrete modeling techniques. 2001.

[32] Navin Gupta and Edward J Williams. Simulation improves service and profitability of an automobile service garage. In *Proceedings of the 16th European Simulation Symposium*, 2004.

[33] Peter D Hanlon and Peter S Maybeck. Multiple-model adaptive estimation using a residual correlation kalman filter bank. *IEEE Transactions on Aerospace and Electronic Systems*, 36(2):393–406, 2000.

[34] Farshad Harirchi, Zheng Luo, and Necmiye Ozay. Model (in) validation and fault detection for systems with polynomial state-space models. In *ACC*, pages 1017–1023, 2016.

[35] Farshad Harirchi and Necmiye Ozay. Guaranteed model-based fault detection in cyber–physical systems: A model invalidation approach. *Automatica*, 93:476–488, 2018.

[36] Farshad Harirchi, Sze Zheng Yong, Emil Jacobsen, and Necmiye Ozay. Active model discrimination with applications to fraud detection in smart buildings. In *IFAC WC, Toulouse, France*, 2017.

[37] Monson H Hayes. 9.4: Recursive least squares. *Statistical Digital Signal Processing and Modeling*, page 541, 1996.

[38] G. Heydt, S. Kalsi, and E. Kyriakides. A short course on synchronous machines and synchronous condensers. *pdf file available at* `https://pserc.wisc.edu/documents/general_information/presentations/presentations_by_pserc_university_members/heydt_synchronous_mach_sep03.pdf`, 2003.

[39] Michael W Hofbaur and Brian C Williams. Hybrid estimation of complex systems. *IEEE Trans. on Systems, Man, and Cybernetics-B*, 34(5):2178–2191, 2004.

[40] Abdulmalik Humayed, Jingqiang Lin, Fengjun Li, and Bo Luo. Cyber-physical systems security—a survey. *IEEE Internet of Things Journal*, 4(6):1802–1831, 2017.

[41] Inseok Hwang, Sungwan Kim, Youdan Kim, and Chze Eng Seah. A survey of fault detection, isolation, and reconfiguration methods. *IEEE transactions on control systems technology*, 18(3):636–653, 2010.

[42] Rolf Isermann. Model-based fault-detection and diagnosis–status and applications. *Annual Reviews in control*, 29(1):71–85, 2005.

[43] S Keerthi and E Gilbert. Computation of minimum-time feedback control laws for discrete-time systems with state-control constraints. *IEEE Transactions on Automatic Control*, 32(5):432–435, 1987.

[44] Xenofon Koutsoukos, James Kurien, and Feng Zhao. Estimation of distributed hybrid systems using particle filtering methods. In *International Workshop on Hybrid Systems: Computation and Control*, pages 298–313. Springer, 2003.

[45] David Kushner. The real story of stuxnet. *ieee Spectrum*, 50(3):48–53, 2013.

[46] Frédéric Lafont, Jean-Paul Gauthier, Tarak Damak, and Ahmed Toumi Salowa Methnani. Actuator and sensor fault detection, isolation and identification in nonlinear dynamical systems, with an application to a waste water treatment plant. 2013.

[47] J. Lambers. Mat 610 summer session 2009-10 [lecture notes]. 2009.

[48] Jonathan M Lee. *Islanding detection methods for microgrids*. PhD thesis, Ms. Thesis, University of Wisconsin-Madison, 2011.

[49] Wen-Shing Lee, Doris L Grosh, Frank A Tillman, and Chang H Lie. Fault tree analysis, methods, and applications-a review. *IEEE transactions on reliability*, 34(3):194–203, 1985.

[50] Dimitri Lefebvre. Fault diagnosis and prognosis with partially observed petri nets. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(10):1413–1424, 2014.

[51] X Rong Li, Youmin Zhang, and Xiaorong Zhi. Multiple-model estimation with variable structure. iv. design and evaluation of model-group switching algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 35(1):242–254, 1999.

[52] Xiao-Rong Li and Yaakov Bar-Shalom. Multiple-model estimation with variable structure. *IEEE Transactions on Automatic control*, 41(4):478–493, 1996.

[53] Daniel Liberzon and A Stephen Morse. Basic problems in stability and design of switched systems. *IEEE Control systems*, 19(5):59–70, 1999.

[54] Feng Lin. Diagnosability of discrete event systems and its applications. *Discrete Event Dynamic Systems*, 4(2):197–212, 1994.

[55] Yao Liu, Peng Ning, and Michael K Reiter. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security (TISSEC)*, 14(1):13, 2011.

[56] Alvaro Llaria, Octavian Curea, Jaime Jiménez, and Haritza Camblong. Survey on microgrids: unplanned islanding and related inverter control techniques. *Renewable energy*, 36(8):2052–2061, 2011.

[57] Felipe Lopez, Miguel Saez, Yuru Shao, Efe Balta, James Moyne, Z. Morley Mao, Kira Barton, and Dawn Tilbury. Categorization of anomalies in smart manufacturing systems to sup-

port the selection of detection mechanisms. In *13th IEEE Conference on Automation Science and Engineering, submitted*, 2017.

[58] Hongwei Lou and Pengna Si. The distinguishability of linear control systems. *Nonlinear Analysis: Hybrid Systems*, 3(1):21–38, 2009.

[59] John Lygeros. Lecture notes on hybrid systems. In *Notes for an ENSIETA workshop*, 2004.

[60] Charlie Miller and Chris Valasek. Hackers remotely kill a jeep on the highway—with me in it.

[61] Yilin Mo and Bruno Sinopoli. False data injection attacks in control systems. In *1st Workshop on SCS*, 2010.

[62] Yilin Mo and Bruno Sinopoli. On the performance degradation of cyber-physical systems under stealthy integrity attacks. *IEEE Trans. on Automatic Control*, 61(9):2618–2624, 2016.

[63] Pieter J Mosterman and Gautam Biswas. Building hybrid observers for complex dynamic systems using model abstractions. In *International Workshop on Hybrid Systems: Computation and Control*, pages 178–192. Springer, 1999.

[64] Pieter J Mosterman and Gautam Biswas. A hybrid modeling and simulation methodology for dynamic physical systems. *Simulation*, 78(1):5–17, 2002.

[65] Ramine Nikoukhah and Stephen L Campbell. Auxiliary signal design for active failure detection in uncertain linear systems with a priori information. *Automatica*, 42(2):219–228, 2006.

[66] S Paulraj and P Sumathi. A comparative study of redundant constraints identification methods in linear programming problems. *Mathematical Problems in Engineering*, 2010, 2010.

[67] Charles L Phillips and H Troy Nagle. *Digital control system analysis and design*. Prentice Hall Press, 2007.

[68] André Platzer and Jan-David Quesel. European train control system: A case study in formal verification. In *ICFEM*, volume 5885, pages 246–265. Springer, 2009.

[69] Jana Price and James A Southworth. Positive train control systems. *Journal of Accident Investigation*, 2(1), 2006.

[70] S Joe Qin. Survey on data-driven industrial process monitoring and diagnosis. *Annual reviews in control*, 36(2):220–234, 2012.

[71] Nacim Ramdani, Louise Travé-Massuyès, and Carine Jauberthie. Mode discernibility and bounded-error state estimation for nonlinear hybrid systems. *Automatica*, 91:118–125, 2018.

[72] Md Rasheduzzaman. *Small signal modeling and analysis of microgrid systems*. Missouri University of Science and Technology, 2015.

[73] Paulo Rosa and Carlos Silvestre. Fault detection and isolation of lpv systems using set-valued observers: An application to a fixed-wing aircraft. *Control Engineering Practice*, 21(3):242–252, 2013.

[74] Paulo Rosa, Carlos Silvestre, and Michael Athans. Model falsification of lpv systems using set-valued observers. *IFAC Proceedings Volumes*, 44(1):1546–1551, 2011.

[75] Paulo Rosa, Carlos Silvestre, and Michael Athans. Model falsification using set-valued observers for a class of discrete-time dynamic systems: a coprime factorization approach. *International Journal of Robust and Nonlinear Control*, 24(17):2928–2942, 2014.

[76] Paulo Rosa, Carlos Silvestre, Jeff S Shamma, and Michael Athans. Fault detection and isolation of ltv systems using set-valued observers. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 768–773. IEEE, 2010.

[77] Miguel Saez, Francisco Maturana, Kira Barton, and Dawn Tilbury. Anomaly detection and productivity analysis for cyber-physical systems in manufacturing. In *Automation Science and Engineering (CASE), 2017 13th IEEE Conference on*, pages 23–29. IEEE, 2017.

[78] Meera Sampath, Raja Sengupta, Stephane Lafortune, Kasim Sinnamohideen, and Demosthenis C Teneketzis. Failure diagnosis using discrete-event models. *IEEE Trans. on control systems technology*, 4(2):105–124, 1996.

[79] Jeff S Shamma and Kuang-Yang Tu. Set-valued observers and optimal disturbance rejection. *IEEE Transactions on Automatic Control*, 44(2):253–264, 1999.

[80] Silvio Simani, Cesare Fantuzzi, and Ronald Jon Patton. Model-based fault diagnosis techniques. In *Model-based Fault Diagnosis in Dynamic Systems Using Identification Techniques*, pages 19–60. Springer, 2003.

[81] Jill Slay and Michael Miller. Lessons learned from the maroochy water breach. In *International Conference on Critical Infrastructure Protection*, pages 73–82. Springer, 2007.

[82] Oliver Stecklina and Peter Langendörfer. Ubiquitous computing asks for ubiquitous line of defense. 2011.

[83] Dawn Tilbury, Jonathan Luntz, and William Messner. Controls education on the www: Tutorials for matlab and simulink. In *American Control Conference, 1998. Proceedings of the 1998*, volume 2, pages 1304–1308. IEEE, 1998.

[84] Wenhui Wang, Linglai Li, Donghua Zhou, and Kaidi Liu. Robust state estimation and fault diagnosis for uncertain hybrid nonlinear systems. *Nonlinear analysis: Hybrid systems*, 1(1):2–15, 2007.

[85] Zheng Wang, DM Anand, J Moyne, and DM Tilbury. Improved sensor fault detection, isolation, and mitigation using multiple observers approach. *Systems Science & Control Engineering*, 5(1):70–96, 2017.

[86] Zheng Wang, Farshad Harirchi, Dhananjay Anand, CheeYee Tang, James Moyne, and Dawn Tilbury. Conflict-driven hybrid observer-based anomaly detection. In *ACC*, pages 5793–5800, 2018.

[87] Alan S Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12(6):601–611, 1976.

[88] Hao Yang, Bin Jiang, and Vincent Cocquempot. Fault tolerant control and hybrid systems. In *Fault Tolerant Control Design for Hybrid Systems*, pages 1–9. Springer, 2010.

[89] Qingsong Yang. *Model-based and data driven fault diagnosis methods with applications to process monitoring*. PhD thesis, Case Western Reserve University, 2004.

[90] S Hashtrudi Zad, Raymond H Kwong, and W Murray Wonham. Fault diagnosis in discrete-event systems: Framework and model reduction. *IEEE Transactions on Automatic Control*, 48(7):1199–1212, 2003.

[91] Feng Zhao, Xenofon Koutsoukos, Horst Haussecker, James Reich, and Patrick Cheung. Monitoring and fault diagnosis of hybrid systems. *IEEE Trans. on Systems, Man, and Cybernetics-B*, 35(6):1225–1240, 2005.

[92] Xiaorong Zhi and Youmin Zhang. Multiple-model estimation with variable structure part iii: Model-group switching algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 35(1):225, 1999.