Advances in Safety and Security of Cyber-Physical Systems

Sliding Mode Observers, Coalitional Control and Homomorphic Encryption

Keijzer, T.

DOI
10.4233/uuid:0e362a8b-b470-4660-a396-35726a2dca89

Publication date
2023

Document Version
Final published version

Citation (APA)
Keijzer, T. (2023). *Advances in Safety and Security of Cyber-Physical Systems: Sliding Mode Observers, Coalitional Control and Homomorphic Encryption*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:0e362a8b-b470-4660-a396-35726a2dca89

Advances in Safety and Security of Cyber-Physical Systems

Sliding Mode Observers, Coalitional Control and Homomorphic Encryption

Twan Keijzer

Advances in Safety and Security of Cyber-Physical Systems

Sliding Mode Observers, Coalitional Control and Homomorphic Encryption

Advances in Safety and Security of Cyber-Physical Systems

Sliding Mode Observers, Coalitional Control and Homomorphic Encryption

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board of Doctorates
to be defended publicly on
Monday 13 February 2023 at 10.00 o'clock

by

Twan KEIJZER

Master of Science in Aerospace Engineering,
Delft University of Technology, The Netherlands
Born in Amsterdam, The Netherlands.

This disseration has been approved by the promotors

Promotor:	Prof. dr. ir. J.W. van Wingerden
Copromotor:	Dr. R.M.G. Ferrari

Composition of the doctoral comittee:

Rector Magnificus,	chairperson
Prof. dr. ir. J.W. van Wingerden,	Delft University of Technology, promotor
Dr. R.M.G. Ferrari,	Delft University of Technology, copromotor
Independent members:	
Prof. dr. H. Sandberg,	KTH Royal Institute of Technology, Sweden
Prof. dr. T. Keviczky,	Delft University of Technology
Prof. dr. ir. P.H.A.J.M. van Gelder,	Delft University of Technology
Dr. ir. C.C. de Visser,	Delft University of Technology
Dr. M.S.T. Chong,	Eindhoven University of Technology

This dissertation has been completed in fulfilment of the requirements of the Dutch Institute of Systems and Control (DISC) for graduate study.

Keywords: Safety & Security, Cyber-Physical System, Sliding Mode Observer, Coalitional Control, Homomorphic Encryption, Collaborative Vehicle Platoon

Printed by: Glideprint

Cover: Twan Keijzer, DALL·E

ISBN: 978-94-6384-411-6

An electronic version of this dissertation is available at
`http://repository.tudelft.nl/`.

To my parents,
for always being strong in facing sadness.

In theory, the theory should work like practice.

Paraphrased from Benjamin Brewster,
Albert Einstein,
Richard P. Feynman,
Yogi Berra

Contents

Summary

Without us realising it, solutions for safety and security are present all around us. However, everyone has undoubtedly also experienced how inconvenient some safety and security measures can be. For example, think about security checks at the airport, the need to wear a bicycle helmet, or being asked to perform 2-factor authentication to log into an online account. Such inconveniences caused by safety and security measures can delay or even prevent their implementation, which is undesired. This reluctance to tolerate inconveniences for the sake of safety and security provides a challenge for engineers to find solutions with minimal impact on normal behaviour.

This challenge is especially pronounced in so-called cyber-physical systems (CPSs), in which digital automation is used to coordinate the actions of one or more physical systems. Examples of CPSs are airplanes, robotic arms or the power grid. Such CPSs have the combined advantages of the physical and cyber world, but are also subject to both threats to safety and security. In fact, the integration of physical and cyber parts in a CPS means that security issues can cause safety issues, and although less common safety issues can cause security issues.

Measures for safety and security of CPSs are categorised as prevention, resilience, and detection & accommodation. These different types of precautions can be used independently, but typically they need to be combined to provide adequate safety and security of a CPS. In this dissertation, three advances within safety and security of CPSs are presented which cover contributions on each of the different types of safety and security measures. Firstly, anomaly *detection* is addressed by extending existing sliding mode observer (SMO) based anomaly estimation methods with detection capability. To this end, two SMOs based anomaly detectors are presented, which are applicable to a large class of SMOs. These detectors, by design, have no false alarms and allow for strong theoretical guarantees on detectability.

Secondly, a topology-switching coalitional control technique which integrates *resilience, detection and accommodation* is designed for safe control of a collaborative vehicle platoon (CVP) subjected to man-in-the-middle (MITM) cyber-attacks. Here resilience to undetected attacks is achieved by means of scenario based model predictive control (MPC) and detected anomalies are accommodated by disabling the affected communication links. Lastly, a real-time implementation of encrypted control based on fully homomorphic encryption (FHE) is presented. FHE allows for manipulation of encrypted data, such that it can *prevent* confidentiality breaches during communication and computation.

Each contribution of this dissertation address a specific topic within safety and security of CPSs. By doing so, they demonstrate the potential of these methods to increase safety and security of CPSs while minimising their impact on normal behaviour. This will promote the adaptation of safety and security measures and allows for safety and security throughout the continued progress in automation.

Samenvatting

Zonder dat we het in de gaten hebben zijn er overal om ons heen veiligheidsmaatregelen aanwezig. Echter heeft iedereen ook wel eens ervaren hoe onhandig sommige veiligheidsmaatregelen kunnen zijn. Denk aan veiligheidscontroles op het vliegveld, het moeten dragen van een fietshelm, of het moeten uitvoeren van *2-factor authentication* voor online aanmelden. Zulke ongemakken die bij veiligheidsmaatregelen komen kijken kunnen hun implementatie vertragen of zelfs verhinderen, wat onwenselijk is. Deze onwil om ongemakken te tolereren voor extra veiligheid biedt een uitdaging voor ingenieurs om oplossingen te vinden met minimale impact tijdens normaal gebruik.

Deze uitdaging is extra interessant voor zogenoemde *cyber-physical systems (CPSs)*, waar digitale automatisering wordt gebruikt om de acties van een fysiek systeem aan te sturen, bijvoorbeeld in vliegtuigen, robotarmen of het stroomnetwerk. Zulke CPSs combineren de voordelen van de digitale en fysieke wereld, maar zijn ook onderworpen aan veiligheidsrisico's vanuit beide werelden. Daarnaast kunnen, door integratie van de digitale en fysieke systeemdelen, veiligheidsrisico's vanuit de digitale wereld effect hebben op de fysieke systeemdelen en vice versa.

Veiligheidsmaatregelen in CPSs kunnen gecategoriseerd worden als preventie, veerkracht, en detectie & aanpassing. Deze verschillende soorten maatregelen kunnen onafhankelijk gebruikt worden, maar moeten meestal gezamenlijk worden ingezet om de veiligheid voldoende te waarborgen. In dit proefschrift worden drie onderwerpen beschreven met als doel om de veiligheid van CPSs te verbeteren. Hierin wordt een bijdrage geleverd binnen elke soort veiligheidsmaatregelen. Eerst wordt detectie behandeld door het uitbreiden van bestaande observatiemethoden welke zijn gebaseerd op *sliding mode observers (SMOs)*. Hiervoor worden twee detectoren gepresenteerd, welke beide op een grote klasse SMOs toepasbaar zijn. Deze detectoren zijn ontworpen zodat ze geen vals alarm geven. Daarnaast geven we krachtige bewijzen voor wanneer afwijkingen wel detecteerbaar zijn.

Ten tweede wordt een zogenoemde *topology-switching coalitional* controle methode gepresenteerd, welke veerkracht, detectie en aanpassing combineert om veiligheid te waarborgen in een samenwerkend voertuigpeloton (CVP) dat onderworpen is aan zogenaamde *man-in-the-middle (MITM)* cyber-aanvallen. Hier wordt veerkracht tegen niet gedetecteerde aanvallen gerealiseerd door middel van een scenario-*model predictive control (MPC)* methode. Gedetecteerde aanvallen worden verholpen door communicatie op aangetaste kanalen te stoppen. Als laatste wordt een real-time implementatie van een versleutelde controle methode gebaseerd op *fully homomorphic encryption (FHE)* gepresenteerd. Door het gebruik van FHE kunnen versleutelde berichten worden bewerkt, zodat vertrouwelijkheid tijdens zowel communicatie als berekeningen kan worden gewaarborgd.

Elke bijdrage van dit proefschrift behelst een ander methode voor veiligheid in CPSs. Hiermee worden de mogelijkheden van deze methodes gedemonstreerd om de veiligheid te vergroten terwijl het effect op het normale gebruik minimaal is. Dit zal helpen de implementatie van veiligheidsmaatregelen te bevorderen en maakt het mogelijk om veiligheid te waarborgen in de doorgaande vooruitgang van de automatisering.

Acknowledgments

It was never the plan to pursue a PhD degree, yet here we are. After four years of collecting more questions than answers, I learned a lot about the pleasures and challenges of doing research. Due to the COVID-19 pandemic, the second year of the PhD was especially challenging. The lack of inspiration during this period of the PhD, where it might be needed most, took a heavy toll on my joy in doing research. Nevertheless, I look back on the last four years with mostly happy memories.

On a personal level, I first want to thank the love of my life, Josine. We have so many great plans ahead, which I'm looking forward to with all my heart. Thank you for showing me how to enjoy the little things, and for keeping me sane during the pandemic.

I also want to thank my family: Vonne, what I would give just to know how your life would have turned out in these last 12 years. Thank you for facing life with a smile. Len, thanks for all the sleepovers and haircuts, those memories will always be alive. Betty, thank you for always being unlimitedly proud of me, I'm struggling to live without it. Ronald, we have endured a lot of loss, thank you for always being there for me. We will stay strong together.

Furthermore, I would like to elevate to the rank of family Bert, Carola, Wil and Nanny. Thank you for your friendship which long predated my birth. Thank you for hearing out the long and winding explanations of my research. Thank you for the great discussions on travel. Thank you for the walks in the heathland.

Throughout my studies, I had the pleasure to make many friends in Amsterdam, Delft and beyond that made the process so much more enjoyable. Thank you Bryan and Dave for all the games of squash, all the dinner parties, and all the beers. Thank you Esmee for all the wine-infused conversations. And all the others I am so thankful to have in my life, including, but not limited to, Tijmen, Rutger, Kieran, Benjamin, Niels, Aeilt-Jan, Floris, Bart, Victor, Visakh, Marine, Bhushan, Jonathan, Pablo, Christian and Gustav.

I would also like to acknowledge the colleagues at DCSC who have inspired my research more than they might realise, just by getting a cup of coffee. I'd especially like to thank Cees, Gabriel, Giannis and Daniel who gave me a second home while Riccardo's group was still a two-man business; and Alex for being a great co-author, a patient roommate on business trips and for all his positive energy. Of course these are just a few amongst many DCSC colleagues, such as, Manuel, Manon, Kim, Sebastiaan, Clara, Atin, Tushar, Yichao, Jean, Zhixin, Tian, Steven, Wicak, Andrea, Amin, Mattia, Maarten, Leila, Pascal, Suad, Frederik, Eva, Claudia, Emanuel, Frida, Pedro, Wim, Wil, Heleen, Marieke and Francy. It was a pleasure working with all of you.

During my PhD I have also had the chance to interact with many DCSC master students of whom I supervised three throughout their MSc. thesis project. I truly enjoyed this process and I am proud of the resulting works. Thank you Vedang, Geert and Pieter! Especially, Pieter should be credited with getting his work published at the 2022 Conference on

Decision and Control. This work also forms the basis of Chapter 4 of this dissertation, and for this I owe him a debt of gratitude.

On this topic, I would also like to sincerely thank all my collaborators for their hard work. Without your contributions my thesis would not yet have been finished today. Paula, Pepe, Japie, Phillipe, and Fabian, it was a pleasure working with you and I hope we will have the chance to work together in the future.

Lastly, I owe a debt of gratitude to my promotors Jan-Willem and Riccardo. Thank you for this chance and the consecutive supervision. Riccardo, I still remember asking you during the last interview whether it was okay to take some holiday before starting the PhD. You replied: "That's okay, I'll ruin your holidays once you started your PhD". I'm grateful that you did not keep that promise. Thank you for your elaborate feedback, thank you for always keeping up with our regular meetings even during the pandemic, and thank you for sitting through our hour-long discussions on semantics.

Twan Keijzer
Delft, 13 February 2023

1

Introduction

The need for safety and security does not need much motivation. Without us realising it, solutions for safety and security are present in systems all around us. These are safety measures such as smoke detectors or the fuses that protect the devices in our homes from over-current, and security measures such as end-to-end encryption in messaging services or the password on your phone and/or computer. However, the way in which safety and security should be achieved is less straight forward.

Everyone has experiences showing how inconvenient some safety and security measures can be. For example, think about security checks at the airport, the need to fasten you seat-belt in a car, or being asked to perform 2-factor authentication to log into an online account. But, the adverse effects of the need for safety and security are not limited to inconveniences. Think for example about the need to stop gas-extraction due to the risk posed by related earthquakes.[1] Such adverse effects caused by safety and security measures can delay or even prevent their implementation, which is undesired.

Being born and raised in The Netherlands, for me personally, a striking example of this is the bicycle helmet. It has been extensively proven that wearing a helmet greatly reduces the risk of head injuries when you are in an accident. However, many Dutch cyclists still don't wear them because it requires a change that makes cycling less comfortable. This also holds for me, even after spending four years on research that is promoting safety through methods far more complex than a bicycle helmet, I am still very reluctant to put one on. This reluctance to tolerate inconveniences for the sake of safety and security provides a very interesting challenge for engineers, namely

Solutions for safety and security should be designed to minimize their impact on normal behaviour.

In the example of the bicycle helmet, an alternative with less impact on normal behaviour is the so-called airbag helmet. You can wear this helmet around your neck when cycling

[1] The gas-extraction related earthquakes are a big issue at the Groningen gas field in The Netherlands for years, especially since a 3.6-magnitude earthquake in 2012 it has become a permanent part of the public debate. https://energypolicy.columbia.edu/sites/default/files/pictures/CGEP_Groningen-Commentary_072518_0.pdf

1

and it automatically inflates around your head when you are involved in an accident. This allows the airbag helmet to reduce its impact during a normal cycling trip, while still offering the required protection when an accident occurs.

This improved design is made possible by the use of algorithms that automate detection of an accident and then reliably trigger the helmet to inflate. For many other systems automation has similarly played a critical role in improving safety while minimising the impact on nominal behaviour. Think for example about the automatic braking systems on modern cars or the SawStop[2] system that stops automatic saws as soon as it senses contact with your fingers.

The initial use of automation was, however, not to improve safety, but to improve performance. During the industrial revolution, such automation was first present in large self-powered machines which took over many tasks in the assembly line and in transportation. This marked a leap in complexity and size of the systems with respect to traditional tools. These large machines, if malfunctioning, could pose a threat to the people working amongst them, which gave rise to a greater concern for the safety of these systems.

The second (and third) industrial revolution brought the introduction of the computer and internet, i.e. the cyber space. This allowed for fast computation, communication and information distribution. With this new technology came many advances on which we rely so heavily today, such as email, video calling, and online shopping. But, just as physical systems are subject to safety issues, cyber systems are subject to threats to security. And, therefore, also appropriate measures need to be taken to address these threats to security.

The cyber space also allowed for many advances in automation of physical systems. Mainly, its ability to perform fast computation allowed for much more complex tasks to be automated. Additionally, using its communication abilities it also allowed for the automation of spatially distributed systems, such as smart grids, or remote control, as used for drones. Such systems all belong to a new class of systems called cyber-physical systems (CPSs).

Figure 1.1 shows some examples of CPSs based on the complexity of their automation and the degree to which they are spatially distributed. One can see that the term CPS covers a very broad range of systems from kitchen appliances to industrial robots and power grids. All of these systems have benefited from being automated, either by increased performance, increased ease of use, and/or reduced operation and production costs.

CPSs, however, not only have the combined advantages of the physical and cyber world, but are also subject to both threats to safety and security. In fact, the integration of physical and cyber systems in a CPS means that security issues can cause safety issues, and although less common safety issues can become security issues.

A famous example of a cyber threat that caused a risk to safety in a CPS is the Stuxnet worm that infected the Iranian Natanz nuclear-enrichment facility in 2010. [1, 2] Stuxnet was the first ever cyber-warfare weapon and was specifically targeted to disable the Natanz nuclear-enrichment facility. It did so by infecting three specific controllers within the plant and feeding them with pre-recorded unsuspicious data, while sending malicious input to the centrifuges they controlled. This attack led to a failure of the affected centrifuges and a temporary closure of the facility. [1]

[2] https://www.sawstop.com/

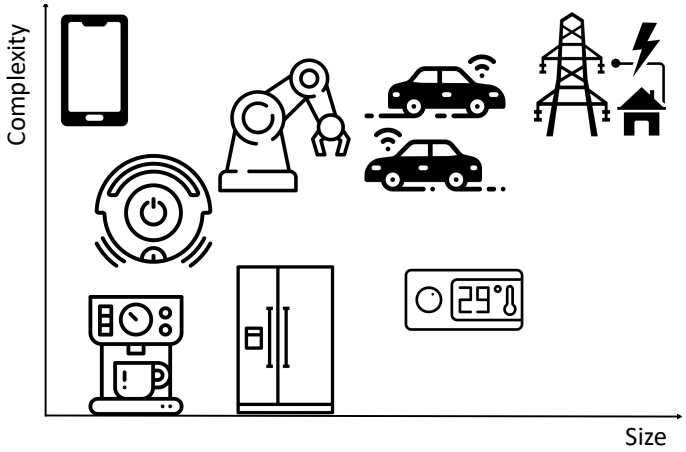


Figure 1.1: A few examples of CPSs ordered by size and complexity

Since Stuxnet many more cyber threats to CPSs have been exposed. Such as in [3], where it is shown how a whole range of modern vehicles can be hacked remotely through their infotainment system. Or the Blackenergy 3 malware which targeted the Ukrainian power grid and caused a power loss that affected almost a quarter million people [4]. This shows that safety and security for CPS is a very important topic that remains relevant for research to this day.

By now, it may seem like we have strayed a long way from the initial example of the bicycle helmet. However, in the large scale CPS subjected to cyber threats the main conclusions derived from the bicycle helmet still hold. Firstly, to promote the willingness to implement measures to address these threats, it is important that their impact on normal operation of the system is minimised. And, secondly, just like with the airbag helmet, automated detection of a threat and an appropriate response represent an important approach to achieve this goal.

In general, precautions for safety and security are categorised as prevention, resilience, and detection & accommodation. Prevention methods aim to reduce the likelihood of a threat affecting the system and causing a risk to safety and security. In the bicycle example, prevention methods, such as bicycle infrastructure and traffic rules, aim to reduce the likelihood of bicycle accidents. For any threat that is not prevented, resilience is achieved when the impact on the system is small enough to avoid risk of lasting damage. For example, putting on a bicycle helmet reduces the risk of brain damage after a bicycle crash, and thus increases resilience. Lastly, detection & accommodation is the approach used in the airbag helmet. It is, just like resilience, used to reduce the risk of damage due to an unsafe or unsecure situation. However, contrary to resilience, it is an active method which only comes into action when an unsafe or unsecure situation occurs.

These different types of precautions can theoretically be used independently, but typically they need to be combined to provide adequate safety and security of a CPS. In this dissertation, three advances within safety and security of CPSs are presented which cover contributions on each of the different types of precautions.

1

In the remainder of this chapter, first two CPSs are presented in Section 1.1, which will be used to motivate the pursued researched topics and will serve as applications throughout the rest of the dissertation. Secondly, a formal discussion of safety and security of CPS is presented in Section 1.2. Then, in Section 1.3 an overview of existing literature on safety and security of CPS is given. Lastly, the contributions of this dissertation will be outlined in Section 1.4.

1.1 Applications Motivating the Research

Although the research presented in this dissertation is applicable to general classes of systems, I have taken inspiration from two real world applications to choose the research directions to pursue. These motivating applications are presented in this section. Firstly, Section 1.1.1 introduces the application of a man-in-the-middle (MITM) cyber-attack on a collaborative vehicle platoon (CVP). Secondly, Section 1.1.2 introduces the application of so-called oscillatory failure case (OFC) faults in the fly-by-wire (FBW) actuation system of civil aircraft.

1.1.1 Collaborative Vehicle Platoons

Road congestion and emissions due to road vehicles is increasingly becoming a concern [5]. In many places the congestion problem has been approached by increasing the lane count of congested road. This has also led to an increased complexity of the road network, which in combination with the high car density has led to an increase in road accidents. From a control perspective, all these problems can be solved by introducing an appropriate form of autonomous driving [6].

Different types of almost fully autonomous vehicles (AVs) are being researched by tech giants such as Google, Microsoft and Apple[3]. These AVs typically use computer vision based algorithms to be able to recognise visual road-side instructions (signs, crosswalks,traffic lights,etc.) and avoid obstacles. Such AVs only require the user to enter a destination and would then drive there fully autonomously. This technology, however, still has a long way to go before becoming reality [7].

Alternatively, as an evolution of cruise control (CC) and adaptive cruise control (ACC), so called collaborative adaptive cruise control (CACC) is introduced as an automation solution where vehicles on a highway collaboratively keep a small inter-vehicle distance forming a so-called collaborative vehicle platoon (CVP) [7, 8][4]. An illustration of such a CVP is shown in Figure 1.2. One can see that the vehicles are capable of measuring parts of the state of the preceding vehicle, while also communicating with their neighbours. In combination, this information allows the vehicles to increase tracking performance and reduce the minimum inter-vehicle distance, thus reducing pollution, congestion, and road accidents.

Because of the decreased inter-vehicle distance such CVPs can reduce road congestion and vehicle emissions on highways. This main advantage of autonomous driving can be achieved while significantly reducing complexity with respect to AVs. CVPs, however, are

[3] see for example https://waymo.com/

[4] In literature CACC is often used to refer to a specific control law to achieve collaborative platooning behaviour. CVP as used in this dissertation is more general and includes all possible methods of collaborative control.

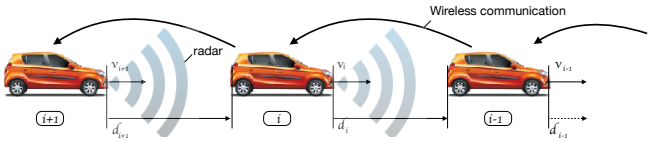


Figure 1.2: Three vehicles in a CVP with a typical predecessor-follower communication topology. [9]

limited to scenarios where the environment is *known*, as they are not equipped to deal with unexpected obstacles. Therefore, a CVP also requires all vehicles to have the required sensing and communication capability.

While the communication of a CVP is shown to greatly improve performance, they also bring with them an increased risk of cyber-attacks. As the communication between vehicles in a CVP is wireless, if unprotected, this is an easy target for cyber-attacks. In this dissertation we consider a so-called man-in-the-middle (MITM) cyber-attack on the inter-vehicle communication, where an attacker can intercept communicated data and replace it with any other data. This is initially a threat to security, but as the inter-vehicle communication is used to compute the vehicle's input it can also cause a threat to safety. In this work the considered threat to safety for a CVP is the occurrence of a crash between any two vehicles.

In order to protect the CVP from a loss of safety and security due to the MITM attack, a combination of prevention, resilient design, and detection & accommodation is required. Firstly, encryption of the communicated data can be used to prevent the MITM attack. However, as we have seen from many recent examples [3, 10, 11], attackers often find a way around such solutions, for example by getting access to one of the vehicles where typically the data is decrypted to perform calculations.

Therefore, a second line of defence is needed based on a combination of resilience, detection and accommodation. Resilience cannot offer a full solution without compromising the nominal performance as the cyber-attack is unpredictable and can possibly become excessively large. Therefore, in the considered CVP, resilience is always combined with detection & accommodation.

Consequently, resilience is required for any undetected attacks and the detection & accommodation should adhere to following design goals. Firstly, as the vehicles drive close together, typically with less than 1 s headway, detection & accommodation actions need to be fast. Secondly, the driver might well have a supervisory role within a CVP and be part of the accommodation after an attack is detected. Therefore, as any detection should be taken seriously by the supervisory driver, the scheme should have a very low false alarm rate.

1.1.2 Oscillatory Fault Case in Actuation of Civil Aircraft

The design of a commercial aircraft is a complex procedure involving many different requirements, but a common factor throughout it is the aim to minimize weight to obtain better fuel efficiency in operation. In this context, the aircraft is supported by a flexible structure that is designed to withstand a specified load envelope with minimal structural reinforcements. The load envelope specification is normally based on the expected struc-

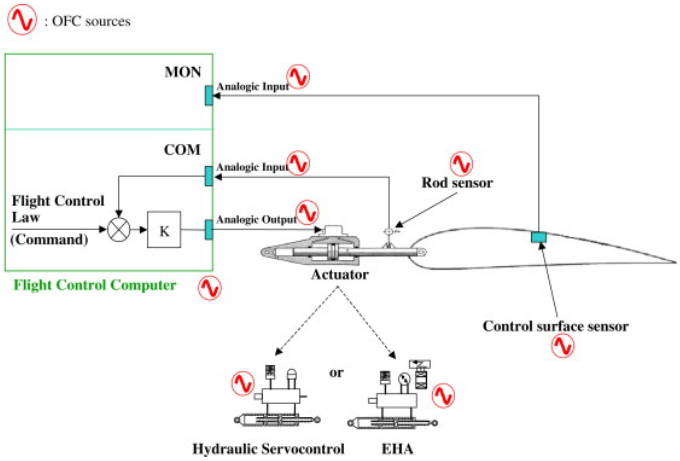


Figure 1.3: Schematic of an servo-control loop with potential sources of an OFC. [13]

tural loads on the aircraft resulting from atmospheric effects (such as turbulence) and maneuvers performed by the aircraft itself. However, certain system faults can lead to additional structural loads which must also be considered when performing the aircraft structural design.

Specifically, loss of safety through a so-called oscillatory failure case (OFC) is considered in this dissertation. An OFC starts as an oscillatory fault in the servo-loop control of an aircraft actuator at any of the locations indicated in Figure 1.3. Such faults rarely occur, but, if they go unmitigated, they can cause oscillations of the control surface, which in turn cause significant additional loads on the aircraft structure. [12, 13]

As show in Figure 1.3, a risk to safety through an OFC can have many different causes, such as malfunctioning sensors or analog input/output (I/O) modules, as well as unwanted oscillations from the command generation of the flight control computer (FCC). This wide variety of threats eventually all become a risk to safety, namely the oscillation of the control surface that causes the OFC if it is not mitigated.

As a result of the diversity of initial causes of oscillation, prevention is hard, requiring different methods for each possible cause. The same holds for resilience, detection and accommodation if they are applied at the individual sensors, I/O components or FCC. However, resilience, detection and accommodation can also be applied on the oscillation of the control surface that actually constitutes the OFC. Oscillations at this single point are easier to address than the variety of initial causes.

Partial resilience to this threat is currently achieved through structural reinforcement, which is unwanted as it is heavy and ultimately increases the operational costs of the aircraft. However, if the OFC is detected existing redundancy in the actuators can be utilized to accommodate the threat. Therefore, by increasing speed and reliability of detection, the required resilience is reduced, lowering the aircraft operating costs as well as its emissions.

Detection requirements for OFCs are typically set in terms of oscillation magnitude and number of oscillations until detection. The structural design is then made to be resilient to any undetected OFC. As OFCs occur only rarely, but can quickly cause large structural loads,

it is important that detection occurs fast and that no false detection occurs. Furthermore, if guaranteed detectability proofs are provided, they form a useful input to determine the resilience needed in the structural design.

1.2 Defining Safety and Security in CPSs

In this section we will formally introduce and define terms from the field of safety and security in CPSs. It must be noted that no single definition exists for most of these terms. The definitions used here are inspired by works dedicated to defining these terms, such as [14, 15]. However, the definitions presented here are kept short and nonrestrictive on purpose.

The class of CPSs contains many different types of systems, such as illustrated in Figure 1.1, where CPSs of varying size and complexity are shown. A CPS derives its name from the fact that it consists of a combination of physical and cyber systems. Here the physical system is a part of the CPS that interacts with the environment and other physical systems within the CPS. The cyber system is a part of the CPS that manipulates data and transfers it to other cyber systems either within or outside the CPS. Generally, in a control systems framework a physical system would consist of actuators, a plant, and sensors. A cyber system would then consist of digital communication and computation. A cyber-physical system can then be defined based on these parts as

Definition 1.1 (Cyber-Physical System)**.** A system that contains at least one cyber system connected to at least one physical system. ◁

This definition of a CPS is very broad and also includes systems that many would classify as mechatronic or embedded systems. Here mechatronic or embedded systems are typically of a smaller scale, whereas CPSs are characterized by larger size and include coordination of multiple physical systems by one or more cyber systems. These characteristics of CPSs are important to be considered as they make a CPS more complex and more vulnerable to faults and cyber-attacks. However, for the purpose of this dissertation it is not required limit the definition of a CPS to these larger, more complex systems as all results that will be presented are equally applicable to smaller, less complex CPS.

A CPS as defined above can be affected by threats to safety and security such as physical faults and cyber-attacks. For the purpose of this dissertation all threats will be classified as

Definition 1.2 (Malicious threat)**.** A threat is malicious if it is executed with intent to threaten safety or security of the CPS. ◁

Definition 1.3 (Accidental threat)**.** A threat is accidental if it occurs without intent to threaten safety or security of the CPS. ◁

Many other classifications of threats are possible based on whether they affect a cyber or physical system, or based on whether its effect is internal to the system or also extends into the environment. An overview of these definitions and how they are used can be found in [14]. The definitions given above are, however, deemed most suitable for CPSs by the author. This because most threats that are, for example, initially a threat in the cyber system, will also affect the physical system due to the connectivity between cyber and

physical parts of the CPS. Furthermore, due to this same connectivity threats that were initially internal to the system will often, in time, also extend into the environment.

The defined threats can induce anomalies in the CPS, which in turn cause a risk to security or safety. Anomalies present themselves differently in physical and cyber systems, so this distinction will be made in defining them. Firstly, define a physical anomaly as

Definition 1.4 (Physical Anomaly)**.** An undesirable change to the dynamics of the physical system. ◁

For the cyber system, anomalies can affect confidentiality, integrity, or availability of the data within the cyber system. Here confidentiality, integrity and availability (CIA) are defined as [16]

Definition 1.5 (Confidentiality)**.** Data can only be accessed by authorised entities. ◁

Definition 1.6 (Integrity)**.** Data is unaltered and trustworthy. ◁

Definition 1.7 (Availability)**.** Data is available when requested. ◁

Cyber anomalies can then be defined based on the so-called CIA-triad as

Definition 1.8 (Cyber Anomaly)**.** A loss of confidentiality, integrity or availability of data in the cyber system. ◁

Based on the presented classification of threats and anomalies, the concepts of safety and security will be defined. Also for safety and security multiple definitions exist based on different definitions of the threats and anomalies that cause them. Here the definition is chosen in line with the classification of threats.

Definition 1.9 (Safety)**.** The safety of a CPS is at risk from (and only from) all anomalies caused by accidental threats and physical anomalies caused by malicious threats. A CPS is safe if the risk these anomalies pose to the system and its environment is acceptable. ◁

Definition 1.10 (Security)**.** The security of a CPS is at risk from (and only from) all anomalies caused by malicious threats. A CPS is secure if the risk these anomalies pose to the system and its environment is acceptable. ◁

One can see that here, like is common in literature, the anomalies are defined as binary properties and the risk to safety or security associated with the anomalies is quantitative. Due to this representation, a threat that causes an anomaly, only causes loss of safety or security if the impact is sufficiently large. This realization can be used to define and understand the different methods of mitigation. A graphical representation of the relations between threats, anomalies and risks to safety and security is shown in Figure 1.4 where also the different mitigation methods, which will be defined next, are already shown. The figure also shows the interconnections between the cyber and physical anomalies that occur in CPSs as discussed before.

Now that the threats to safety and security have been properly defined, we can look at methods of mitigation. A first method of mitigation is through prevention of an anomaly, which can be formally defined as

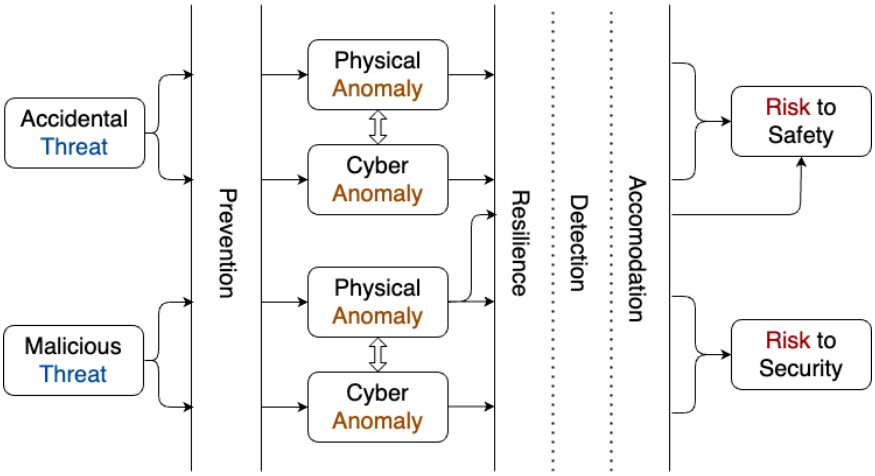


Figure 1.4: Flowchart indicating how threats, anomalies and risk to safety and security are related to each other and to methods of mitigation.

Definition 1.11 (Prevention)**.** An anomaly is prevented if it cannot be caused by a threat.◁

Examples of prevention methods are encryption of communication to prevent loss of confidentiality, airless tires to prevent flats or redundancy of power supply to prevent loss of power to critical systems such as nuclear reactors.

Secondly, mitigation can also be achieved by means of resilience of the CPS to an anomaly, which can be formally defined as

Definition 1.12 (Resilience)**.** A CPS is resilient to an anomaly if it can sufficiently contain the anomaly and the risk it causes without relying on a detection decision. ◁

Examples of methods for resilience are the use of a crumple zone in modern cars, the use of redundant components or the implementation of robust control laws. Lastly, detection and accommodation of an anomaly can be used as a means of mitigation.

Definition 1.13 (Detection & Accommodation)**.** An anomaly is successfully accommodated by a CPS if it is detected and the system behaviour is actively changed based on this detection decision such that the reconfigured system can sufficiently contain the anomaly and the risk it causes. ◁

Examples of methods for detection and accommodation are airbags which are deployed once a crash is detected, fire alarms, where detection is automatic and accommodation is manual, or the use of adaptive control.

These methods of mitigation are visualised in Figure 1.4 and each have their own advantages and disadvantages. Therefore, these methods are often used in combination to sufficiently reduce the risks to safety and security. Prevention typically is a highly robust method of mitigation as it can ensure that the threat has no effect at all on the CPS. Therefore, if possible one would prefer to always use prevention to ensure safety and security. However, prevention is often expensive, requiring extensive design changes or

compromises to operational efficiency. Prevention is therefore often used in combination with methods for resilience and detection & accommodation.

Resilience methods are, however, less robust than methods for prevention, but also often less expensive, and similarly detection & accommodation is less robust and less expensive than resilience. Therefore, to improve safety of CPSs while keeping them cost efficient it is very important to reduce the cost of methods of prevention and increase the robustness of methods for resilience and detection & accommodation. The research presented in this dissertation is in line with these goals.

1.3 State of the Art

In this chapter an overview of the recent literature on safety and security of CPSs will be presented. First, nominal control of CPSs will be discussed in Section 1.3.1, followed by an overview of threats to their operation in Section 1.3.2. These threats give us a motivation to look into the methods to mitigate the anomalies they cause such that they no longer cause a loss of safety and/or security. Therefore, Section 1.3.3 will discuss prevention, Section 1.3.4 resilience, Section 1.3.5 detection, and Section 1.3.6 accommodation of anomalies. Other recent surveys on safety and security in CPSs can be found in [17–22].

1.3.1 Nominal Control of CPSs

In the introduction to this chapter, CPSs were introduced as the driving force behind many technological advances. Ones that we all take for granted, such as mobile phones, the power grid, and the automatic assembly line; but also many technologies that are still under heavy development, such as autonomous or collaborative vehicles and smart (micro) girds. Due to the size of these systems, they require a control that is at least partially distributed and relies on, often wireless, communication for collaboration between the subsystems within the CPS. Therefore, although any automatic control system can be classified as a CPS, this section will primarily focus on forms of control for large scale CPS.

Many different control methods have been adapted and developed to apply to large scale CPS. Such adaptations typically revolve around the way information is shared within the CPS, i.e. the communication topology. As one can imagine, in a large scale CPS this opens up a practically unlimited amount of possible topologies to explore, ranging from fully decentralised to fully centralised control.

A broadly used term within this field is distributed control, which was founded as the middle ground between centralised and decentralised control, allowing for a trade-off between communication & computation requirements and performance. As such, distributed control encompasses almost the full range of topologies. However, in distributed control the communication between subsystems is typically only used to share information and not for one subsystem to dictate the actions of other subsystems. In other words, there is no hierarchy between the subsystems.

By efficiently using communication between subsystems, distributed control allows for performance close to the (optimal) centralised control, while retaining the flexibility and scalability offered by decentralised control [23]. A well documented type of distributed control is distributed model predictive control (DMPC) [24–26], which has also been extensively researched for control of CVPs [27, 28]. Examples of other control methods

proposed for distributed control are sliding mode controller (SMC) [29], linear-quadratic regulator (LQR) [30], and event-triggered control [31]. In [32] Dual decomposition was used to synthesize an optimal distributed controller.

As an extension to distributed control, also methods have been proposed that include the topology design in the controller synthesis [33] or even allow for changing communication topologies online to trade-off between communication costs and performance [34]. Such topology-switching communication architectures are aided by so-called plug-and-play control [35, 36], which are designed to be applicable regardless of the communication topology. Examples of topology-switching control of CVPs can be found in [37–39].

On the other end of the spectrum, hierarchical control uses the communication topologies to introduce a hierarchical structure between subsystems. Such architectures typically consist of simple *local* control with little information and a high update rate, while information is accumulated higher up the hierarchy where more complex *global* or *supervisory* control is performed with a lower update rate. Such hierarchical control is commonly used in industrial applications. Such as in [40], where multi-rate model predictive control (MPC) in a fully hierarchical topology is proposed for control of a chemical plant, or in [41], where a similar method is used for high level platoon speed control based on traffic models. To gain the advantages of both the distributed and hierarchical communication topologies, the approaches can also be combined. As, for example, in [42], where a topology consisting of a number of distributed hierarchical controllers is proposed for the control of a chemical plant. A more extensive description of CPS topology types is given in [43].

Related to the example applications used within this research, see Section 1.1, also some research will be presented specific to distributed control for CVPs. Here many works focus on the so-called string stability property that ensures disturbances will not grow along the CVP. [44] uses frequency domain analysis to prove a linear controller can be used to achieve string stability. This result was then validated on a real vehicle platoon [8]. This control law, the so-called collaborative adaptive cruise control (CACC), remains a popular topic of research to this day [45]. However, other control methods are also proposed to achieve string stability, such as [46] using control matching MPC, [47] using DMPC, and [48, 49] using an artificial potential field. Another topic of research in the control of CVPs is the impact of the communication protocols. In [50, 51] realistic communication protocols for CVPs are discussed. A string stable linear matrix inequality (LMI) based controller design considering such realistic communication can be found in [52].

1.3.2 Threats to CPSs

In Section 1.2, threats to CPS have been classified as accidental and malicious threats, which can both cause physical or cyber anomalies. This distinction is theoretically more general than the distinction between faults and cyber-attacks often made in literature. However, in practice the distinctions can be used similarly. Faults are typically defined in literature as accidental threats that cause physical anomalies, and cyber-attacks as malicious threats causing cyber anomalies. This neglects the existence of two types of threats that are considered in the classification presented in Section 1.2, namely accidental threats causing cyber anomalies and malicious threats causing physical anomalies. [5]

[5] For a recent example of the latter see `https://www.bloomberg.com/news/articles/2022-10-08/trains-in-northern-germany-halted-for-hours-after-cables-cut`

Theoretically, this makes the classification in faults and cyber-attacks incomplete, however, in practice these types of threats occur far less often. Therefore, these threats are also of less concern than faults or cyber-attacks. Furthermore, resilience, detection and accommodation methods developed for faults and cyber-attacks can typically similarly be applied to anomalies caused by these threats. Therefore, we will use the distinction fault vs. cyber-attack to classify the literature in the remainder of this section.

In the *Global Risk Report 2018* [11] the risk of cyber-attacks is estimated to be second only to environmental disaster. Furthermore, the possibilities of cyber-attacks are very broad. Surveys of the risks of cyber-attacks on public transport, vehicle to vehicle communication and unmanned aerial vehicles (UAVs) can be found in respectively [53], [54], and [55, 56]. In [10], using two 2009 cars, it was demonstrated that attackers can remotely brake or disable the gas pedal. Similarly [57] shows results from actual attacks performed via a mobile app connected to the car. Well-known examples of a real-world cyber-attacks are the Stuxnet attack [58], the Maroochy water breach [59], and also, recently in 2020, Honda was the target of a cyber-attack, likely through the remote desktop client [60].

In literature many studies can be found on the potential resources of cyber-attackers [61]. The survey [61] assigns some commonly studied cyber-attacks on scales of (1) disclosure resources, (2) disruption resources, and (3) model knowledge. In this framework, which is shown in Figure 1.5, the following attacks types are defined:

- For eavesdropping attacks the attacker only needs disclosure resources.
- For denial of service (DoS) attacks [62, 63] the attacker only needs disruption resources.
- For Replay attacks [64–66] the attacker needs both disclosure and disruption resources, but no model knowledge.
- For bias-injection attacks [67] limited model knowledge and disruption resources are required.
- For zero-dynamics (undetectable) attacks [68, 69] full model knowledge and disruption resources are needed, but no disclosure resources. Recent extensions of this notion aim to quantify the area where attacks are nearly undetectable using the concepts of security index [70, 71] and weak detectability [72, 73].
- For covert attacks [74] the attacker needs full model knowledge as well as disclosure and disruption resources.

Using these classifications, the effect of these attacks is studied for specific plants and controller structures, such as power networks [75–78], wireless sensor networks [79], and DMPC [80, 81]. Additionally, various attack scenarios have been studied for CVPs, a recent survey can be found in [82]. [83–85] identify various vulnerabilities in CVPs, [86] studies the impact of various attacks on a CVP, [87] also studies impact of attacks, but compares controller robustness, and [88] quantifies the trade-off between security and quality of service for CVP. Lastly, I would like to mention the so-called *topology attacks* or *rewiring attacks* as considered in [89–91], which are cyber-attacks that aim at changing the connection topology of CPS and by doing so change the physical behaviour of the system.

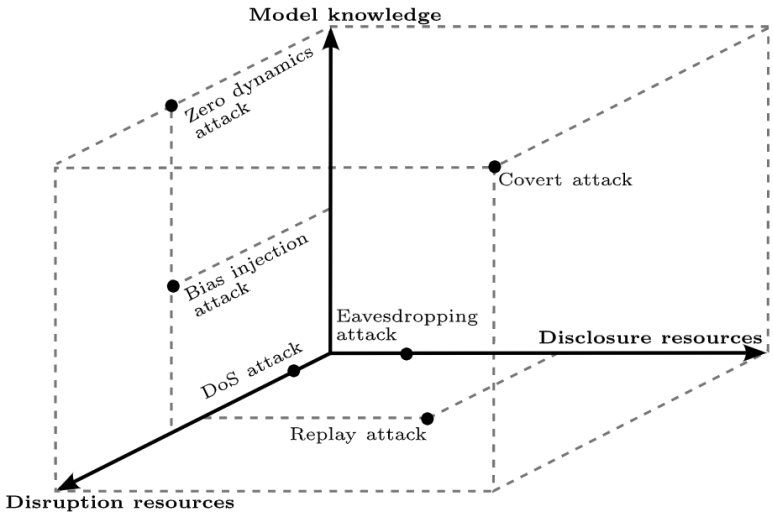


Figure 1.5: Classification of attack types from [61]

Contrary to cyber-attacks, which are often considered for general systems, and only classified based on the attacker capabilities, faults or failure modes are often system-specific. For example, [92] considers actuator faults in an HVAC system, [93] considers faults in Lithium-Ion batteries, [94] considers failure modes of transistors, [95] considers a maritime application, [96–98] considers communication faults in a CVP, [99–101] consider reduced actuator effectiveness on an aircraft, and [12, 13] consider the OFC which was introduced earlier in Section 1.1. Such system specific analysis of faults often results in a description of the fault behaviour which can then be used in its mitigation.

A more general distinction between fault types is however commonly used in work on resilience, detection and accommodation of faults [102]. Here, however, only the location of the fault is considered and not the fault behaviour. Firstly, a commonly considered fault type is the actuator fault. This fault is typically modeled as an additional input in the state transition equation. Secondly, system faults change the response of the system to inputs. Such faults are typically modeled as a change to the state transition matrix. Lastly, sensor faults are typically modeled as an additive term to the output equation.

1.3.3 Prevention of anomalies

Threats can occur in many forms and can affect the CPS in many ways. Physical anomalies are the result of hardware malfunction and can therefore not be prevented directly by means of a software solution. Notable research, however outside the scope of this dissertation, is done in the field of predictive maintenance [103] for prevention of such physical anomalies. The research on resilience, detection and accommodation of physical anomalies is more abundant and will be discussed in the following sections.

Most work on anomaly prevention focuses on cyber anomalies within which the largest research field is that of encryption. Encryption was originally developed to provide confidentiality in information technology (IT) systems. Modern day encryption was pioneered

by [104] with the Diffie-Hellman encryption scheme. Since then the field has matured [105], giving rise to amongst others the famous Rivest–Shamir–Adleman (RSA) encryption scheme [106]. These encryption schemes have served well in IT systems, but have a major drawback for use on CPSs. Namely, they don't allow for operations to be performed on encrypted numbers such that the data needs to be decrypted and re-encrypted in the cyber system to perform control.

An important development for the applicability of encryption to CPS is the introduction of homomorphic encryption (HME), which allows to perform operations on encrypted data [107–109]. These early HME schemes are only partially homomorphic, meaning they only support either addition or multiplication, but not both. To this day, these early HME schemes are used in research on secure control [110–114] using various methods to overcome the inherent limitations of the underlying HME schemes.

Later, fully homomorphic encryption (FHE) schemes were introduced using the lattice-based learning with errors (LWE) problem [115, 116] as a basis for encryption. These FHE schemes do support both addition and multiplication. One of the main drawbacks of FHE, however, is the limited multiplicative depth, i.e. the limited amount of allowed multiplications before getting erroneous results. The GSW FHE scheme [117, 118] allows for extended multiplicative depth, but it still remains an issue for application in CPSs if the controller has an internal state. Solutions have been proposed like using a periodic reset of internal controller states [119] or scaling the control system to work exclusively with integer values [120–122]. The challenge of multiplicative depth in FHE together with other challenges in encrypted control are surveyed in [123]. However, regardless of the open challenges recently some implementations of FHE on real systems have been published [124–126]. For a more extensive overview of FHE based on LWE please refer to [127].

Another interesting avenue of research that allows for private computation without using encryption is called secure multi-party computation (sMPC) which divides the computation task over two or more (non-colluding) parties such that neither party can infer any relevant information [128–130]. Loosely related are the fields of differential privacy and privacy-preserving consensus which are being researched for distributed computation [131–135].

1.3.4 Resilience to anomalies

Resilience against anomalies is needed for any anomaly which cannot be prevented, or detected and accommodated. Resilience can be achieved through hardware solutions using, for example, physical redundancy. But, resilience is also an integral part of many control laws, in fact all functional control laws have some resilience (also often called robustness) against uncertainty. However, to keep the scope limited, in this section we will focus on controllers that are explicitly designed to increase resilience against anomalies.

An anomaly that is considered often for research on resilience is loss of availability of communicated signals, either through accidental packet loss or through malicious DoS attacks. A recent survey of resilience against such anomalies can be found in [136]. Such methods include delay estimators [63], adaptive control [137], event-triggered control [39, 138, 139], semidefinite programming based control [140] and game theory [62].

Another threat that is commonly considered in distributed control and estimation literature is that of non-compliant agents. By the definitions in Section 1.2 this malicious

threat can either cause a physical anomaly if the agents maliciously change their control input [141–145] or a cyber anomaly if the non-compliance is in the form of distributing wrong information [146–152]. Notably DMPC has been extensively studied for control resilient against non-compliant agents [142–145]. For distributed estimation a pioneering work was on the Byzantine generals problem [153], which has been built upon for application in consensus and state estimation. For example [149] considers distributed state estimation for nonlinear agents, [146, 148] consider distributed consensus, [150] considers asynchronous communication and time-delays, and [154] extends resilience by considering a small number of designated trusted nodes. Common in all these works is that the network topology plays an important role in the resilience of the CPS [155].

A rich body of literature also exists on resilience against physical damage, which is often referred to as fault tolerant control (FTC), which can be further subdivided in passive and active FTC. Active FTC often includes online parameter estimation as in [156, 157]. Passive FTC is achieved through a broad range of control methods such as MPC[158, 159], SMC [160–162], Backstepping [163] and Fuzzy control [164]. Specifically I would like to mention the field of incremental control which has been used to reduce model dependence, and therefore increase fault tolerance of nonlinear dynamic inversion [99, 165], Backstepping [100], Dual Heuristic Programming [166] and SMC [167, 168].

1.3.5 Detection of anomalies

Anomaly detection is an integral part of almost any scheme for safe and secure control. However, the aim of a detection algorithm is distinctly different from those that attain resilience or perform accommodation. Resilience and accommodation are forms of control, i.e. they aim at finding a control input sequence that will reduce the risks to safety and security. Detection, however, is a form of observation, i.e., it focuses on monitoring the system behaviour and distinguishing (uncertain) nominal behaviour from anomalous behaviour. To this end, detection employs many different principles and approaches than those used for resilience and accommodation.

For detection, system behaviour is often monitored based on a so-called residual, which is designed to be around zero under nominal behaviour and deviates from zero only under anomalous behaviour. This residual is then evaluated, for example by comparing it to a threshold, for detection of the anomaly. Detection methods can thus be described by how they generate the residual and how the residual is evaluated.

In model-based anomaly detection, usually an observer is used to generate a residual. Traditionally this residual is the output estimation error, or some simple function thereof, [169–171]. A framework to generate thresholds for these residuals, in the linear case, was introduced as early as 1988 [172]. These residuals are still used extensively to this day [173–177]. However, as signified by recent surveys on fault diagnosis in industrial control systems (ICSs) [178], swarms [179] and cyber security in power grids [180] the field has produced many notable advancements. Anomaly detection using an output estimation error based residual have been extended to non-linear [176, 181, 182] and large scale systems using distributed techniques [89, 175, 176, 183–188] or reduced order models [182, 189] for detection. Further advancements have been made using set-based methods which allow for integrated estimation and detection [190–192]. And using active fault detection, where the plant input is changed specifically to improve detection performance [92, 190, 192, 193].

However, there is a fundamental limitation to all methods that base detection on the output estimation error of the observer. An anomaly can only be detected once the state estimation becomes incorrect. This means that the state estimate cannot be correct in anomalous conditions. This is an undesirable property that can be avoided by using an anomaly estimate based residual that can be obtained using e.g. unknown input observer (UIO) or sliding mode observer (SMO) based methods.

SMOs can be used to generate an anomaly estimate based on the so-called equivalent output injection (EOI) which is invariant to *matched uncertainties* [194]. This condition of invariance became known as the *observer matching condition*, which has been relaxed in many different ways [195–199] to improve applicability of the method. At the same time, robustness against unmatched uncertainties was increased [200, 201]. Initially the anomalies that could be detected were only actuator anomalies [202, 203], but later also extensions of the method were proposed to allow for detection of sensor anomalies [204–206]. Furthermore, SMO-based detection schemes have now also been proposed for large scale and connected systems [207–209], and have been applied in aerospace [210], maritime [95], and CVPs [209, 211].

Apart from the works that have been mentioned before, which have a general applicability to anomalies caused by accidental or malicious threats, much work has been specifically devoted to detecting malicious cyber-attacks. As shown in Section 1.3.2, such cyber-attacks come in many forms and can be designed using model knowledge and disclosure resources to become hard or even impossible to be detected using generally applicable schemes. [68, 212, 213] identify certain fundamental limitations of linear detectors against malicious cyber-attacks. Amongst others, work on the security index [70, 71] and weak detectability [72, 73] have attempted to quantify detectability close to these fundamental limits. [35, 36] proposed a detection method using disturbances between physically coupled systems to overcome some of these limitations. [214] identified conditions on systems that are robust to stealthy attacks. For replay attacks detection is proposed using so-called *watermarking* approaches. These approaches make small changes to the input [65, 193, 215] or output [64, 66] of a system that do not disrupt the control or supervision, but can help to distinguish true real-time data from replayed data.

1.3.6 Accommodation of anomalies

The accommodation approaches presented in this section are, by definition, always accompanied by a detection approach. The control methods presented here will have a large overlap with those discussed for resilience in Section 1.3.4. However, it is important to note that the approach taken in accommodation is fundamentally different as it is only active after an anomaly is detected. This means the nominal system performance is unaffected by the accommodation, and at the time of detection there will be a sudden change in behaviour of the controlled system. Therefore, this section will focus on how the behaviour of the controlled system is changed as a reaction to an anomaly detection.

There are several approaches to anomaly accommodation presented in literature. A common approach to accommodate anomalies in the actuation signal or plant behaviour is to start using a changed or augmented controller after detection. In [216] a robust SMC module is implemented that augments the nominal control action after an anomaly is detected. Similarly, in [217] a bank of observers is used to detemine which fault scenario is

active, and this information is used to augment the setpoint of an MPC controller. In [101] FTC of a Boeing 747 aircraft is achieved using an adaptive SMC, where the adaptation is triggered by an anomaly detection and uses an anomaly estimate in the adaptation algorithm.

A similar approach also exists for anomalies in communicated and measured signals for (distributed) estimation, where the estimate is augmented [67] or replaced [218, 219] using model-based estimates. This reconfiguration of the estimator, can also be used in combination with the nominal controller to do accommodation for control, as is done in [220, 221]. Similar, but notably different, [97] uses the input buffer from the nominal MPC controller to generate model-based inputs to overcome short term communication loss in a CVP.

Fully integrated safe control solutions which guarantee safety through a combination of resilience, detection and accommodation are limited. A general approach to do so is proposed in [222]. Some more work exist for specific applications, such as for CVPs. For a CVP subject to faulty or malicious agents, or communication loss such safe control structures are proposed in respectively [223] and [98]. For further reading and a more complete survey on anomaly accommodation, one is referred to [224, 225].

1.4 Research Goals

In Section 1.3 an overview has been given of many topics within safety and security of CPSs according to the different types of cyber-defences. Although this overview is far from extensive, it covers a plethora of different topics in prevention, resilience, detection and accommodation of anomalies. It has however also been shown that there are still many contributions to be made within each type of mitigation. Therefore, the research presented in this dissertation provides contributions to prevention, resilience, detection and accommodation by addressing three open issues within the field of safety and security of CPSs. In the remainder of this section the chosen topics will be presented and motivated based on the CVP and aircraft servo loop applications from Section 1.1.

1.4.1 Contributions on Detection

Chapter 2
Design of two robust anomaly *detection* techniques applicable to a large class of sliding mode observer based anomaly estimators.

In Chapter 2, two novel sliding mode observer based detectors are presented. The choice for a sliding mode observer based approach can be motivated based on the applications from Section 1.1.[6] Firstly, only model-based detection methods have been considered as in both the automotive and aerospace industry typically good models of the designed systems are available. Secondly, a detection method that is applicable to non-linear models is needed to address both applications. This because the servo-loop control in Equation (2.58) is highly non-linear and even though a linear model for the vehicles in a CVP is presented

[6] while the method is designed with specific applications in mind, it will be presented in Chapter 2 in a form that is applicable to many other systems too.

in Section 1.1.1, any more accurate models for these vehicles are also non-linear. Lastly, both applications are time-sensitive and require fast detection (and accommodation) of the anomaly to avoid a loss of safety. Therefore, the detection method must have fast dynamics. From the anomaly detection methods reviewed in Section 1.3.5 it has been found that SMO based anomaly detection best fits all these requirements.

SMO based detectors have been shown to be widely applicable to linear and nonlinear systems and provide finite-time convergence making them suitable for fast detection of anomalies in a wide variety of systems. SMOs have however mostly been developed for anomaly estimation and not much research has been dedicated to detection. This can partially be attributed to the fact that in an ideal system, with only matched uncertainty, detection based on a SMO is trivial, as anomaly estimation is perfect. However, due to the fast dynamics of the SMO also small unmatched uncertainties, such as measurement noise, can result in relatively large anomaly estimation errors.

Therefore, in Chapter 2, two SMO based detectors are presented which are applicable to a range of existing SMOs, preserving their developed anomaly estimation capabilities. The first detector uses the anomaly estimate and provides a threshold to perform detection. This threshold is designed such that (1) no false detections can occur and (2) any anomaly that is sufficiently large for a sufficient time can be detected. The second detector generates two bounds on the SMO state estimation error based on healthy behaviour. If these bounds cross each other it means the current value of the SMO state estimation error could not have been caused by healthy behaviour, which is then used to perform detection. For this detector the same detection guarantees are presented as for the first method.

1.4.2 Contributions on Resilience and Accommodation

Chapter 3
Design of a topology-switching coalitional control technique which integrates *resilience, detection and accommodation* for integrated safe control of a collaborative vehicle platoon.

In Chapter 3 an integrated method for resilience, detection and accommodation is presented to guarantee safety in a CVP under MITM attack. The chapter focuses on how to achieve resilience against undetected attacks and accommodation of detected attacks. Robust control, providing resilience, is achieved through MPC while the accommodation approach has been inspired by the field of topology-switching control. Topology-switching control was introduced in Section 1.3.1 as a nominal control method for large scale CPSs. However, its flexibility in changing communication topology also allows for accommodation of MITM attacks by disabling affected communication links.

It has been chosen to use an MPC based control law in this work as it allows for handling the safety and string stability constraints of the CVP directly in the form of constraints to the MPC problem. Furthermore, much literature is available on topology-switching MPC as well as on applications of MPC to CVPs. These concepts had however never been utilised as a cyber-defence for CVPs. To this end in Chapter 3 a topology-switching MPC problem is derived which can guarantee safety from crashes even if the system is under MITM attack. Furthermore, it is proven that the CVP is string stable in all healthy conditions. Main

contribution within this integrated design is the development of an MPC problem that is robust to any undetected attacks and to topology switches caused by anomaly detection and accommodation.

1.4.3 Contributions on Prevention

Chapter 4
Design of a real-time implementation of encrypted control based on fully homomorphic encryption for *prevention* of confidentiality breaches.

In Chapter 4, a real-time implementation of an encrypted control law is presented that can secure connected CPSs, such as CVPs, from both cyber-attacks targeting communication as well as those targeting the individual subsystems. In this work an academic example of an inverted double-pendulum is used to demonstrate the effectiveness of the method, however in future work application to any connected CPS is possible. It has been chosen to address the encryption method known as FHE. Using such FHE schemes one can perform calculations on encrypted data, allowing one to keep data secure, also while it is being processed. This property is very promising for use in control systems as it allows for constructing encrypted controllers, closing the loop in terms of encrypted control.

These FHE schemes are, however highly computationally expensive which means that so far all implementations have been non-real-time or on slow systems. In Chapter 4 a real-time implementation of an FHE scheme is presented that can stabilize an inverted double-pendulum. To achieve this, two main contributions are made by adapting and implementing Gentry's FHE scheme. Firstly, a novel so-called *reduced cipher* is introduced that reduces the computational complexity of the FHE scheme. Secondly, the adapted scheme is implemented on a set of two field-programmable gate arrays (FPGAs) for real-time performance. It is shown that this implementation of FHE can stabilize the inverted double-pendulum in real-time.

2

Anomaly Detection using Sliding Mode Observers

Detection of anomalies in cyber-physical systems (CPSs) allows for automated accommodation of the original anomaly, i.e. by means of repairs or reconfigurations in the cyber or physical system. Sliding mode observers (SMOs) have been proposed for exact anomaly estimation for a class of ideal systems without unmatched uncertainties and measurement noise. For such ideal systems anomaly detection is trivial, however for systems with unmatched uncertainties or measurement noise a dedicated detector is needed. In this chapter two of such robust anomaly detectors are presented, which extend the anomaly detection capability of a large class of SMOs to include systems with unmatched uncertainties and measurement noise. Theoretical guarantees on robustness and detectability are presented for both detectors. The first detector is based on the so-called equivalent output injection (EOI), which is closely related to the anomaly estimate. The second detector is directly based on the SMO state estimation error. Doing so, the second detector bypasses the low-pass filter generating the EOI allowing for faster detection of anomalies and making it possible to detect smaller magnitude anomalies. The obtained theoretical results are illustrated by application of the detectors to (1) detect a man-in-the-middle (MITM) attack on a collaborative vehicle platoon (CVP) and to (2) detect an oscillatory failure case (OFC) in the servo loop control of a commercial aircraft.

This chapter is based on

Twan Keijzer and Riccardo M.G. Ferrari. Threshold design for fault detection with first order sliding mode observers. Automatica, *146:110600, 2022.*

Twan Keijzer, Japie A.A. Engelbrecht, Phillipe Goupil, and Riccardo M.G. Ferrari. A sliding mode observer approach to oscillatory fault detection in commercial aircraft. Control Engineering Practice, *under review.*

Twan Keijzer, Fabian Jarmolowitz, and Riccardo M.G. Ferrari. Detection of cyber-attacks in collaborative intersection control. In European Control Conference, *2021.*

Detection of anomalies is an integral part of the mitigation of anomalies in cyber-physical systems (CPSs). Most importantly, fast detection allows for a quick response of automated anomaly accommodation. However, even if an equivalent automated response could be achieved by means of resilience, detection can also provide important information on the frequency of anomalies that can be used to prevent more occurrences in the future. Such response might be automated too, but more often it requires manual intervention in the form of i.e. repairs (physical), reconfigurations (cyber), or redesigns (both).

Unknown input observers (UIOs) have been applied extensively for this purpose, allowing for anomaly estimation and detection [221, 229–233] for a class of systems as defined in [234, 235]. More recently, also sliding mode observers (SMOs) were adopted for this purpose [101, 202–204, 236–239]. These SMO-based anomaly estimation methods are applicable to a larger class of systems and have, in certain applications, better performance [240, 241].

SMO-based anomaly estimation methods have furthermore been developed to allow for even broader applicability. Methods have been proposed to achieve this using higher order exact differentiators [238, 242–245] or by using multiple cascaded SMOs [246]. However, also methods exist where a single first order sliding mode observer (FOSMO) is used, while still relaxing the matching condition [9, 196, 204, 228, 247], the non-minimum phase condition [248–250], or both [162, 199, 251].

Nevertheless, a challenge that still needs to be addressed is the design of SMO-based anomaly detectors when unmatched uncertainties and measurement noise are present. Such effects prevent ideal sliding motion to be reached, which causes the anomaly estimation results to no longer be exact. Therefore, existing detection methods that do not consider measurement noise and unmatched uncertainties cannot lead to robust detection. In this chapter, we will address the anomaly detection problem for systems with measurement noise and (un)matched uncertainties, by developing two robust SMO-based anomaly detectors.

Some works consider the effects of measurement noise on SMO-based state and fault estimation using higher order SMOs, such as [242, 252, 253] providing the relation between the noise magnitude and accuracy using the big O notation, or [245] giving bounds on the time-averaged accuracy. However, the works considering the effect of measurement noise on FOSMO-based anomaly estimation are very limited. In [93] a threshold is determined based on hypothesis testing and in [254] a threshold based on Monte Carlo analysis is proposed. Furthermore, some works do consider measurement noise, but without addressing the detection problem. However in [250] it is required that measurement errors directly affect the state equation, whereas [255] assumes the measurement noise derivatives to be bounded. Both these noise representations are restrictive and may limit the practical applicability. Lastly, several other works [256–258] mention a threshold for detection but present no method to determine its value.

In this chapter the SMO-based anomaly detection problem is addressed by designing two robust and deterministic anomaly detectors, applicable to a large class of FOSMOs, including existing designs such as [9, 101, 200, 203, 204, 228, 251, 259]. Specifically, it will be proven that the detectors are applicable to the SMOs from [204] and [9]. The designed detectors allow for robust anomaly detection on systems with measurement noise and (un)matched uncertainties. Furthermore, sufficient conditions will be presented for which

(1) there exists a realisation of the uncertainty and measurement noise such that detection occurs and (2) detection is guaranteed for all uncertainty and noise realisations.

Notation

For a vector x, $x_{(i)}$ denotes the i^{th} element of x. Inequalities for vectors are evaluated element-wise. Superscript ^{0} denotes nominal behaviour. $\text{diag}(X)$ denotes a column vector containing the diagonal elements of a square matrix X. $|x|$ denotes the element-wise absolute value of a matrix or vector x. Lastly, when $x = 0$, it is considered $\text{sign}(x) = -\text{sign}(x^{nz})$ where x^{nz} is the last non-zero x.

2.1 Problem Formulation

The aim of this chapter is to present the design of two robust anomaly detectors that are applicable to a large class of FOSMO based anomaly estimation schemes. The class of systems to which the detectors are applicable will be characterised in this section. Specifically, Sections 2.1.1 and 2.1.2 present general forms of the considered system and SMO respectively. Then, in Section 2.1.3 three propositions are presented which together form a sufficient condition for the detectors to be applicable. This class of systems is not restrictive, as it can be proven to encompass many existing SMOs, such as [9, 101, 200, 203, 204, 228, 251, 259].

Remark 2.1. The general forms of the system and SMO forms presented in this section are not directly applicable to the application of aircraft servo loop control. The modifications needed to make the scheme applicable to a class of nonlinear systems will be discussed in Section 2.5. Here this modified form is also applied to the aircraft servo loop problem. ◁

2.1.1 General System Description

Let us consider a general dynamical system with the form

$$\begin{cases} \dot{x}_1 = A_{11}x_1 + A_{12}^s x_2 + h_1(y,u) + E_{11}\zeta_1 + E_{12}^s\zeta_2 + N_1 f \\ \dot{x}_2 = A_{21}x_1 + A_{22}^s x_2 + h_2(y,u) + E_{21}\zeta_1 + E_{22}^s\zeta_2 + N_2 f \\ y = C_2 x_2 + F\zeta_2\,, \end{cases} \tag{2.1}$$

where $x_1 \in \mathbb{R}^{n-p}$ and $x_2 \in \mathbb{R}^p$ are partitions of the system state; $y \in \mathbb{R}^p$ is the system output; $u \in \mathbb{R}^w$ is the system input; $f \in \mathbb{R}^r$ is a time-varying term representing the anomaly to be detected; $\zeta_1 \in \mathbb{R}^{q_1}$ is the system uncertainty; $\zeta_2 \in \mathbb{R}^{q_2}$ is the measurement noise; and $h_1 : \mathbb{R}^{p\times w} \longrightarrow \mathbb{R}^{n-p}$ and $h_2 : \mathbb{R}^{p\times w} \longrightarrow \mathbb{R}^p$ are known, possibly nonlinear functions. The following common assumptions characterize the anomaly and the uncertainties.

Assumption 2.1. ζ_1, ζ_2, and f are bounded as $\bar{\zeta}_1 \geq \tilde{\zeta}_1 \triangleq \max_t(|\zeta_1|)$, $\bar{\zeta}_2 \geq \tilde{\zeta}_2 \triangleq \max_t(|\zeta_2|)$, and $\bar{f} \geq \tilde{f} \triangleq \max_t(|f|)$. Here $\bar{\zeta}_1$, $\bar{\zeta}_2$ and $\bar{f}$ are known, deterministic values. ◁

2.1.2 A General Form for First Order Sliding Mode Observers

We will consider an SMO of the general form

$$\begin{cases} \dot{\hat{x}}_1 = A_{11}\hat{x}_1 + A_{12}^s\hat{x}_2 + h_1(y,u) - (A_{12}^s - A_{12})C_2^{-1}e_y - K_1\nu, \\ \dot{\hat{x}}_2 = A_{21}\hat{x}_1 + A_{22}^s\hat{x}_2 + h_2(y,u) - (A_{22}^s - A_{22})C_2^{-1}e_y - K_2\nu, \\ \nu \triangleq -\text{sign}(Pe_y) \\ \hat{y} = C_2\hat{x}_2, \end{cases} \tag{2.2}$$

where $\hat{x}_1 \in \mathbb{R}^{n-p}$, $\hat{x}_2 \in \mathbb{R}^p$ and $\hat{y} \in \mathbb{R}^p$ are the state and output estimates; $e_y \triangleq y - \hat{y}$; and $\nu \in \mathbb{R}^p$ is the switching output feedback. The error dynamics then becomes

$$\begin{cases} \dot{e}_1 = A_{11}e_1 + A_{12}e_2 + E_{11}\zeta_1 + E_{12}\zeta_2 + N_1 f + K_1\nu, \\ \dot{e}_2 = A_{21}e_1 + A_{22}e_2 + E_{21}\zeta_1 + E_{22}\zeta_2 + N_2 f + K_2\nu, \\ e_y = C_2e_2 + F\zeta_2, \end{cases} \tag{2.3}$$

where $e_1 \triangleq x_1 - \hat{x}_1 \in \mathbb{R}^{n-p}$ and $e_2 \triangleq x_2 - \hat{x}_2 \in \mathbb{R}^p$ are the state estimation errors, $E_{12} = E_{12}^s - (A_{12}^s - A_{12})C_2^{-1}F$, and $E_{22} = E_{22}^s - (A_{22}^s - A_{22})C_2^{-1}F$. The anomaly is then estimated by $\hat{f}$ based on the switching term ν via

$$\begin{aligned} \dot{\nu}_{\text{eq}} &= -K_\nu(\nu_{\text{eq}} - \nu) \\ \hat{f} &= g(\nu_{\text{eq}}) \end{aligned} \tag{2.4}$$

where $K_\nu \succ 0 \in \mathbb{R}^{p\times p}$ is the gain matrix of a stable filter, ν_{eq} is the so-called equivalent output injection (EOI), and $g : \mathbb{R}^p \longrightarrow \mathbb{R}^r$ is the *anomaly estimation function*.

Remark 2.2. The function $g(\nu_{\text{eq}})$ can vary and depends on the specific SMO which is used. However, its definition does not affect the applicability of the detectors presented in this chapter. Furthermore, the existence of a ν_{eq} as in Equation (2.4) is only required for the detector presented in Section 2.2. ◁

2.1.3 Detector Applicability Conditions

Based on the error dynamics in Equation (2.3), Propositions 2.1 to 2.3 provide a sufficient condition for the presented detectors to be applicable. As an exemplification, in Section 2.1.4 we will prove that the propositions hold for the SMOs from [204] and [9].

Proposition 2.1. *In Equation* (2.3)*, A_{11} is Hurwitz, $K_\nu \succ 0$ is a diagonal matrix*[1]*, C_2 is invertible, and $K_2 \neq 0$.*

Proposition 2.2. *The following conditions on e_2 hold.*[2]

$$\begin{cases} |e_2| \preceq \tilde{e}_2 \preceq \bar{e}_2 \\ sign(\dot{e}_2) = -sign(Pe_y) \\ if\ \dot{e}_2 \succ 0 : \quad \underline{\dot{e}}_2^+ \preceq \underline{\dot{e}}_2^+ \preceq |\dot{e}_2| \preceq \tilde{\dot{e}}_2^+ \preceq \bar{\dot{e}}_2^+ \\ if\ \dot{e}_2 \prec 0 : \quad \underline{\dot{e}}_2^- \preceq \underline{\dot{e}}_2^- \preceq |\dot{e}_2| \preceq \tilde{\dot{e}}_2^- \preceq \bar{\dot{e}}_2^- \end{cases} \tag{2.5}$$

[1] K_ν only affects the detector in Section 2.2, as such the part of the proposition relating to K_ν is only required for the detector in Section 2.2.

[2] For the detector in Section 2.3 only bounds on the nominal system are required. This because detection occurs by comparing two thresholds without comparing them to the behaviour of a residual.

where $\tilde{e}_2$, $\tilde{\dot{e}}_2^+$, $\underset{\sim}{\dot{e}}_2^+$, $\tilde{\dot{e}}_2^-$, *and* $\underset{\sim}{\dot{e}}_2^-$ *are the unknown true bounds, and* $\bar{e}_2$, $\bar{\dot{e}}_2^+$, $\underline{\dot{e}}_2^+$, $\bar{\dot{e}}_2^-$, *and* $\underline{\dot{e}}_2^-$, *are the known bounds on* e_2. *Furthermore, equivalent bounds for the nominal system can be obtained, denoted with superscript* ^{0}.

Remark 2.3. The unknown bounds on e_2 introduced in Proposition 2.2 may not admit an algebraic closed form, albeit they can still be computed numerically from Equation (2.3) and the true bounds on the anomaly and uncertainties introduced in Assumption 2.1. The known bounds, instead, need only to satisfy Equation (2.5) and can be freely defined by the user in any form. ◁

The relation between true-anomalous and known-nominal bounds can be conveniently written as

$$\begin{aligned} &\tilde{e}_2 + \delta_e = \bar{e}_2^0 + \delta_f(f), \\ &\tilde{\dot{e}}_2^+ + \delta_{\dot{e}} = \bar{\dot{e}}_2^{0,+} + \delta_{\dot{f}}^+(f)\ ;\ \underset{\sim}{\dot{e}}_2^+ - \delta_{\dot{e}} = \underline{\dot{e}}_2^{0,+} + \delta_{\dot{f}}^+(f), \\ &\tilde{\dot{e}}_2^- + \delta_{\dot{e}} = \bar{\dot{e}}_2^{0,-} - \delta_{\dot{f}}^-(f)\ ;\ \underset{\sim}{\dot{e}}_2^- - \delta_{\dot{e}} = \underline{\dot{e}}_2^{0,-} - \delta_{\dot{f}}^-(f), \end{aligned} \tag{2.6}$$

where $\delta_e > 0$ and $\delta_{\dot{e}} > 0$ represent the difference between the true and known bound, and $\delta_f : \mathbb{R}^r \longrightarrow \mathbb{R}^p$, $\delta_{\dot{f}}^+ : \mathbb{R}^r \longrightarrow \mathbb{R}^p$, and $\delta_{\dot{f}}^- : \mathbb{R}^r \longrightarrow \mathbb{R}^p$ represent the effect of an anomaly. Here, and in the following, the superscripts $^+$ and $^-$ denote a variable relates to time periods during which the sign of $\dot{e}_2$ is, respectively, positive or negative.

Proposition 2.3. *For any* j *and* d_f *such that* $|f_{(j)}| \geq d_f$, *there exists a* $\gamma > 0$ *and an index* i *such that either of the following holds.*

1. $\delta_{f,(i)}(f) \geq 0$, $\delta_{\dot{f},(i)}^+(f) \leq -\gamma d_f$ *and* $\delta_{\dot{f},(i)}^-(f) \leq 0$.
2. $\delta_{f,(i)}(f) \geq 0$, $\delta_{\dot{f},(i)}^+(f) \geq 0$ *and* $\delta_{\dot{f},(i)}^-(f) \geq \gamma d_f$.

Remark 2.4. Proposition 2.1 presents some requirements on the observer matrices which are common for SMOs. Furthermore, Proposition 2.2 bounds the area around the ideal sliding surface to which the observer error is attracted. These conditions will form the basis of the detector design. Lastly, Proposition 2.3 requires the anomaly to affect the system, which is needed for the anomaly to be detected. ◁

2.1.4 Proof of Applicability to Existing sliding mode observers

In this section, Propositions 2.1 to 2.3 from Section 2.1.3 are proven to hold for the SMOs proposed in [204] and [9]. Similar proofs exist for many other existing SMOs such as [101, 200, 203, 251, 259]. The proofs presented here serve as an exemplification.

SMO From (Tan and Edwards, 2003)[204]

The SMO design by Tan and Edwards considers a system with model uncertainty and allows for estimation of both actuator and sensor anomalies. The work, however, does not consider measurement noise and requires the matching condition to hold. Here, this SMO is applied on a system with measurement noise $F\zeta_2$, such that the observer error dynamics

from equations (23) and (24) in [204] can be written in the general form Equation (2.3) as

$$\begin{cases} \dot{e}_1 = A_{11}e_1 + A_{12}e_2 + E_{11}\zeta_1 + E_{12}\zeta_2 \\ \dot{e}_2 = A_{21}e_1 + A_{22}e_2 + E_{21}\zeta_1 + E_{22}\zeta_2 + N_2 f + K_2 \nu \\ e_y \triangleq \hat{y} - y = e_2 - F\zeta_2 \\ \nu = -\text{sign}(Pe_y) \end{cases}$$

where ζ_1, ζ_2 and f are bounded (see Equation (3) and below in [204]), such that Assumption 2.1 holds. Below we will prove that Propositions 2.1 to 2.3 hold for the SMO from [204].

Proof. **(Proposition 2.1)** The proof can be found in Equation (19), the Remark below Equation (21), and Equation (24) of [204]. ∎

Proof. **(Proposition 2.2)** We extend Proposition 1 in [204]. Here statement (26) in [204] depends on $e_2^\top P\nu < 0$[3], which is true trivially for a system without measurement noise. For a system with measurement noise, however, this can be untrue if $-F\zeta_2 < e_2 < F\zeta_2$. Therefore, only practical convergence to an area $|e_2| \leq \max_t(F\zeta_2) = \tilde{e}_2$ can be proven. This allows to define $\bar{e}_2 = |F|\bar{\zeta}_2$. By substituting ρ in the right hand side of Equation (24) in [204] it can be proven that $\text{sign}(\dot{e}_2) = -\text{sign}(Pe_y)$. Furthermore, bounds on $\dot{e}_2$ can be obtained by bounding the right hand side of Equation (24) in [204]. ∎

Proof. **(Proposition 2.3)** From the bounds on e_2 in Proposition 2.2 it can directly be found that $\delta_f = 0$ and $\delta_f^- = \delta_f^+ = N_2 f$, where N_2 is full column rank. ∎

SMO From (Keijzer et al., 2019)[9]

The work by Keijzer et al. is one of the few which relaxes the matching condition for anomaly estimation while still only using a single FOSMO. By doing so, however, the state partition x_1 cannot be estimated. Furthermore, [9] already considers system uncertainties and measurement noise, such that the detectors are applicable without any change to the observer. The SMO error dynamics in [9] can be written as

$$\begin{cases} \dot{e}_1 = A_{11}e_1 + A_{12}e_2 + E_{11}\zeta_1 + E_{12}\zeta_2 + N_1 f \\ \dot{e}_2 = A_{21}e_1 + A_{22}e_2 + E_{21}\zeta_1 + E_{22}\zeta_2 + N_2 f + K_2 \nu \\ e_y \triangleq \hat{y} - y = e_2 - \zeta_2 \\ \nu = -\text{sign}(e_y) \end{cases} \tag{2.7}$$

where ζ_1, ζ_2 and f are bounded (see Assumptions 2 and 3 in [9]), such that Assumption 2.1 holds. Below we present proofs of Propositions 2.1 to 2.3 from Section 2.1.3.

Proof. **(Proposition 2.1)** Proof of these statements can be found in Assumption 4 and Proposition 1 of [9]. ∎

Proof. **(Proposition 2.2)** Proof of these statements can be found in Proposition 1 in [9], where known bounds $\bar{e}_2$, $\bar{\dot{e}}_2^+$, $\underline{\dot{e}}_2^+$, $\bar{\dot{e}}_2^-$, and $\underline{\dot{e}}_2^-$ have been derived directly. The true bounds $\tilde{e}_2$, $\tilde{\dot{e}}_2^+$, $\underline{\tilde{e}}_2^+$, $\tilde{\dot{e}}_2^-$, and $\underline{\tilde{e}}_2^-$ can be found following the same methodology. ∎

[3] [204] uses e_y to denote e_2.

Proof. **(Proposition 2.3)** From Proposition 1 in [9] it can be seen that $\delta_f = 0$ and $\delta_{\bar{f}}^- = \delta_{\bar{f}}^+ = r(f)$, where in steady state $r(f) = (F_2 - A_{21}A_{11}^{\dagger}F_1)f$. By Assumption 5 in [9] $F_2 - A_{21}A_{11}^{\dagger}F_1$ is full column-rank. ∎

2.1.5 Detector Design Problem

In this chapter two SMO based anomaly detectors are designed. The detector presented in Section 2.2 consists of a lower and upper threshold on the so-called equivalent output injection (EOI), which functions as a residual. The detector in Section 2.3 does not use a residual, but compares two bounds on the observer error to perform detection. While both detectors use different modes of detection they are both based on the same SMO and have also been designed using the same design criteria:

1. The detector is applicable to a general class of systems and SMOs which fit the general error dynamics of Equation (2.3) and for which Propositions 2.1 to 2.3 hold.
2. The detector is deterministic and robust to uncertainties, i.e. there are no false positives.
3. If $\delta_e = 0$ and $\delta_{\dot{e}} = 0$, for any non-zero anomaly there exists a realisation of the uncertainty and noise such that detection occurs.
4. Any anomaly of sufficient magnitude, which is sustained for a sufficient duration, is detected for all realisations of the uncertainty and noise. Here the sufficient magnitude and duration are specified in Theorems 2.3 and 2.5.

2.2 Detection using Equivalent Output Injection based Residual

The detection logic which will be presented in this section is based on comparing the EOI, ν_{eq} defined in Equation (2.4), to a set of robust detection thresholds. To this end, first the EOI dynamics will be written in time domain in Section 2.2.1. The threshold designs will then be presented in Section 2.2.2. Proofs of robusteness and detectablity are presented in Section 2.2.3.

2.2.1 Equivalent Output Injection Dynamics

Recall the definition of the EOI in Equation (2.4). As ν is piece-wise constant, the time response of each element of the EOI, $\nu_{\text{eq},(i)}$, can be written in closed form. To simplify notation, for each element $\nu_{\text{eq},(i)}$, let us denote $k_i = K_{\nu,(i,i)}$. Furthermore, we define the so-called *switching times*, $\{t_j\}_i$, as the sequence of times at which $\nu_{(i)}$ changes sign. Note that the switching times are not equally spaced, but depend on the system dynamics. In the following, wherever possible, derivations will be shown for one element $\nu_{\text{eq},(i)}$ and the subscript i will be dropped to ease notation. Furthermore, without loss of generality, it is assumed that $\nu_{(i)}$ is positive during each period $[t_{2j} \;\; t_{2j+1}]$, and $\nu_{(i)}$ is negative during each period $[t_{2j+1} \;\; t_{2j+2}]$. With this, the EOI response over any period $[t_{2j} \;\; t]$ where $t_{2j} \le t \le t_{2j+1}$, can be written as

$$\nu_{\text{eq},(i)}(t) = e^{-k(t-t_{2j})}\nu_{\text{eq},(i)}(t_{2j}) + \left(1 - e^{-k(t-t_{2j})}\right). \tag{2.8}$$

During the next period $[t_{2j+1} \;\; t_{2j+2}]$, $\nu_{(i)} = -1$, so the EOI response over any period $[t_{2j+1} \;\; t]$, where $t_{2j+1} \leq t \leq t_{2j+2}$, can be written as

$$\nu_{\text{eq},(i)}(t) = e^{-k(t-t_{2j+1})}\nu_{\text{eq},(i)}(t_{2j+1}) - (1 - e^{-k(t-t_{2j+1})}). \tag{2.9}$$

Substituting Equation (2.8), with $t = t_{2j+1}$, into Equation (2.9) gives

$$\nu_{\text{eq},(i)}(t) = e^{-k(t-t_{2j})}\nu_{\text{eq},(i)}(t_{2j}) - e^{-k(t-t_{2j})} + 2e^{-k(t-t_{2j+1})} - 1, \tag{2.10}$$

for $t_{2j+1} \leq t \leq t_{2j+2}$. Substituting Equation (2.10), with $t = t_{2j+2}$, into itself, and repeating this process N times, the EOI at t_{2N} for any $N \in \mathbb{Z}^+$ can be calculated as

$$\nu_{\text{eq},(i)}(t_{2N}) = e^{-k(t_{2N}-t_0)}\nu_{\text{eq},(i)}(t_0) - e^{-k(t_{2N}-t_0)} + 2\sum_{j=1}^{2N-1}\left((-1)^{j+1}e^{-k(t_{2N}-t_j)}\right) - 1. \tag{2.11}$$

An example of a nominal EOI response, with the corresponding behaviour of e_2 is shown on the left in Figure 2.1.

2.2.2 Equivalent Output Injection Threshold

In this section the detection threshold is designed as an upper bound on the nominal EOI response. This way, by construction, the threshold is guaranteed to have no false positives, i.e. design criterion 2 from Section 2.1.5 is satisfied. The resulting threshold consists of two parts. First, a threshold is designed which bounds the EOI response considering only one period between switches. This threshold is called the *peak threshold*. However, it will be proven that there exist no sufficient conditions guaranteeing detection using only this threshold. Therefore, a so-called *sustained condition* is designed to serve as an initial condition for the peak threshold. The resulting threshold will be called the *combined threshold*. Sufficient conditions for anomaly detection using the combined threshold are presented in Section 2.2.3.

Because the threshold is modelled as a bound on the nominal EOI, first recall the EOI responses in Equations (2.8) and (2.11). From these EOI responses, a particular observation can be made, which will form the basis of the whole threshold design:

The EOI is fully determined by its initial value and the duration of the periods between switches.

These periods between switches, $t_j - t_{j-1}$, can be bound based on the known limits on e_2 from Proposition 2.2. Bounding the duration of these periods in nominal conditions will thus form the core of the threshold design.

Remark 2.5. In the following the design procedure will only be shown for the upper threshold. The lower one can be derived similarly and only the end result will be stated.◁

Peak Threshold

The peak threshold considers the worst-case behaviour of the nominal EOI over one period between switches. As can be seen in Equation (2.8), this occurs for the maximum duration of a period between switches, which we will denote $\tilde{t}^{0,u}$. $\tilde{t}^{0,u}$ occurs for the hypothetical

behaviour of e_2 where e_2 moves from its minimum, $-\tilde{e}_2^0$, to its maximum, $\tilde{e}_2^0$, with the minimum rate, $\dot{e}_2^{0,+} = \underline{\dot{e}}_2^{0,+}$, as illustrated in the right part of Figure 2.1. This leads to the definition below. Similarly also the known bound $\bar{t}_i^{0,u}$ is defined below, based on the known bounds on e_2.

$$\tilde{t}_i^{0,u} \triangleq \frac{2\tilde{e}_{2,(i)}^0}{\underline{\dot{e}}_{2,(i)}^{0,+}} \; ; \; \bar{t}_i^{0,u} \triangleq \frac{2\bar{e}_{2,(i)}^0}{\underline{\dot{e}}_{2,(i)}^{0,+}},$$

where we will drop the subscript i to ease notation. With these definitions, by Proposition 2.2, $\bar{t}^{0,u} \geq \tilde{t}^{0,u}$. Substituting $t - t_{2j} = \bar{t}^{0,u}$ in Equation (2.8), gives the bound on the nominal EOI

$$\bar{v}_{\mathrm{eq},(i)}^{\mathrm{peak}}(t_{2j}) \triangleq e^{-k\bar{t}^{0,u}} v_{\mathrm{eq},(i)}(t_{2j}) + 1 - e^{-k\bar{t}^{0,u}},$$

which is the so-called peak threshold. Here the argument t_{2j} denotes the time at which the threshold is calculated, or reset, based on the current value of the EOI, $v_{\mathrm{eq},(i)}(t_{2j})$. The resulting threshold is constant until a new peak threshold is calculated at $t_{2(j+1)}$. This process of constructing the peak threshold is illustrated on the right side of Figure 2.1. The threshold is used with the anomaly detection logic

$$\exists\, i,j \text{ s.t. } v_{\mathrm{eq},(i)}(t) > \bar{v}_{\mathrm{eq},(i)}^{\mathrm{peak}}(t_{2j}) \text{ for } t \in [t_{2j}\ t_{2(j+1)}]. \tag{2.12}$$

A lower peak threshold can be designed similarly as

$$\underline{v}_{\mathrm{eq},(i)}^{\mathrm{peak}}(t_{2j+1}) = e^{-k\bar{t}^{0,l}} v_{\mathrm{eq},(i)}(t_{2j+1}) - 1 + e^{-k\bar{t}^{0,l}},$$
$$\bar{t}^{0,l} \triangleq \frac{2\bar{e}_{2,(i)}^0}{\underline{\dot{e}}_{2,(i)}^{0,-}},$$

for which the anomaly detection logic is defined as

$$\exists\, i,j \text{ s.t. } v_{\mathrm{eq},(i)}(t) < \underline{v}_{\mathrm{eq},(i)}^{\mathrm{peak}}(t_{2j+1}) \text{ for } t \in [t_{2j+1}\ t_{2j+3}]. \tag{2.13}$$

This lower threshold has to be calculated, or reset, at every t_{2j+1} based on the current value of the EOI $v_{\mathrm{eq},(i)}(t_{2j+1})$, and holds until t_{2j+3}.

The peak thresholds, as presented above, are applicable to the considered SMOs, and are deterministic, i.e. design criteria 1 and 2 hold[4]. However, its detection capabilities are not consistent, thus failing to meet criterion 4. This issue is formalised by the following theorem.

Theorem 2.1. *If $\tilde{e}_2 \leq \max_t(C_2^{-1}F\zeta_2(t))$, no sufficient condition on f exists guaranteeing anomaly detection using the peak thresholds. That is, regardless of f, there always exists a realization of $\zeta_2(t)$ such that neither of the detection conditions are satisfied.*

Proof. From Proposition 2.2 and the hypothesis one can derive, $|e_2| \leq \tilde{e}_2 \leq \max_t(C_2^{-1}F\zeta_2)$, which implies there always exists a ζ_2 such that $C_2^{-1}F\zeta_2 = e_2$. Substituting this ζ_2 in the definition of e_y from Equation (2.3) gives $e_y = 0$. Thus there always exists a ζ_2 such that $e_y = 0$. By the definition of the sign function, a switch occurs when $e_y = 0$, thus there always

[4]Design criterion 3 also holds for the peak thresholds, however the proof will not be provided.

exists a realisation of ζ_2 that makes the time between switches arbitrarily small. Detection with the peak threshold occurs only if the time between two switches is sufficiently large, specifically if $t_{2j+1} - t_{2j} > \min(\bar{t}^{0,u}, \bar{t}^{0,l})$. Therefore, detection with the peak threshold can never be guaranteed. ■

2

Remark 2.6. In Section 2.1.4 it has been proven that the hypothesis of Theorem 2.1, $\tilde{e}_2 \leq \max_t(C_2^{-1}F\zeta_2)$, holds for the SMOs from [204] and [9]. ◁

To satisfy design criterion 4, the threshold design needs to be changed. In particular, we no longer want to use $v_{\text{eq}}(t_{2j})$ and $v_{\text{eq}}(t_{2j+1})$ as reset conditions for the upper and lower peak thresholds, respectively. This will allow us to decouple the detection performances from the actual trajectory of v_{eq}, which depends on the uncertainty realization and not only on the anomaly f. To achieve this, in the following global bounds on $v_{\text{eq}}(t_{2j})$ and $v_{\text{eq}}(t_{2j+1})$ will be designed.

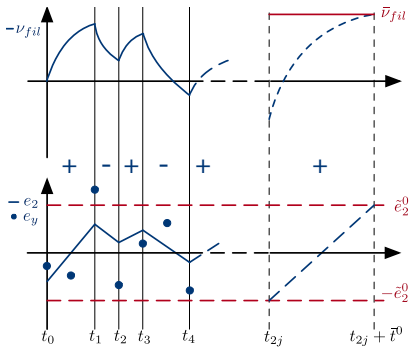


Figure 2.1: Example nominal EOI response - $t_0 \leq t \leq t_4$; Worst-case EOI response for the peak threshold design - $t_{2j} \leq t \leq t_{2j} + \bar{t}^0$. Both with corresponding e_2 behaviour.

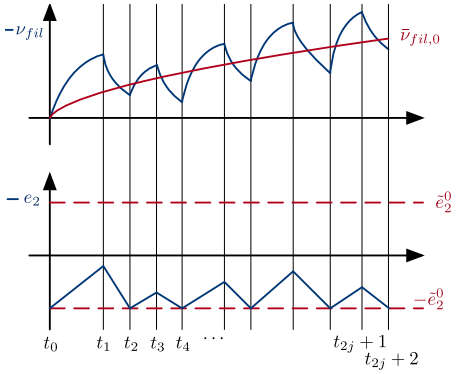


Figure 2.2: Worst-case EOI response for the sustained condition design with corresponding hypothetical e_2 behaviour.

Sustained Condition & Combined Threshold

In this section the so-called sustained condition, denoted by $\bar{\nu}_{\text{eq},0,(i)}$, is introduced as an initial condition for the peak threshold. The sustained condition replaces the reset to $v_{\text{eq},(i)}(t_{2j})$, which was used for the upper peak threshold. Using the sustained condition as initial condition for the peak threshold gives the so-called combined threshold as

$$\bar{\nu}_{\text{eq},(i)}(t_{2j}) = e^{-k\bar{t}^{0,u}}\,\bar{\nu}_{\text{eq},0,(i)}(t_{2j}) + 1 - e^{-k\bar{t}^{0,u}}\,. \tag{2.14}$$

To guarantee that the combined threshold does not result in any false detection, for nominal behaviour the sustained condition should globally upper-bound $v_{\text{eq},(i)}(t_{2j})$. By doing so the combined threshold can globally bound the nominal EOI without requiring the resets previously needed for the peak threshold. Furthermore, $\bar{\nu}_{\text{eq},0}$ should be an initial condition for the peak threshold. Therefore, the hypothetical behaviour of e_2 leading to $\bar{\nu}_{\text{eq},0}$ should also be an initial condition for the behaviour of e_2 leading to the peak threshold. Therefore, as the hypothetical behaviour leading to the peak threshold starts at $e_2 = -\tilde{e}_2$, for the design

of $\bar{\nu}_{\text{eq},0}$, e_2 needs to be constrained as $e_2(t_{2j}) = -\tilde{e}_2 \ \forall j$. How this is achieved can be seen in Figure 2.2. Now we will use the bounds on e_2 from Proposition 2.2, together with the newly found constraint $e_2(t_{2j}) = -\tilde{e}_2 \ \forall j$ to bound the time between switches as

$$\begin{aligned} e_{2,(i)}(t_{2j+2}) - e_{2,(i)}(t_{2j}) &= \int_{t_{2j}}^{t_{2j+2}} \dot{e}^0_{2,(i)} \mathrm{d}t = \int_{t_{2j}}^{t_{2j+1}} |\dot{e}^0_{2,(i)}| \mathrm{d}t - \int_{t_{2j+1}}^{t_{2j+2}} |\dot{e}^0_{2,(i)}| \mathrm{d}t \\ &= \dot{e}^{0,+}_{2,(i)}(t_{2j+1} - t_{2j}) - \dot{e}^{0,-}_{2,(i)}(t_{2j+2} - t_{2j+1}) = 0 \\ \longrightarrow \frac{t_{2j+1} - t_{2j}}{t_{2j+2} - t_{2j+1}} &= \frac{\dot{e}^{0,-}_{2,(i)}}{\dot{e}^{0,+}_{2,(i)}}, \end{aligned} \tag{2.15}$$

where $\dot{e}^{0,+}_{2,(i)}$ and $\dot{e}^{0,-}_{2,(i)}$ are the average of $|\dot{e}^0_{2,(i)}|$ over periods $[t_{2j}\ t_{2j+1}]$, and $[t_{2j+1}\ t_{2j+2}]$, respectively. These averages, $\dot{e}^{0,+}_{2,(i)}$ and $\dot{e}^{0,-}_{2,(i)}$, can be bound in the same way as $|\dot{e}^0_2|$ is bound by Proposition 2.2. Using these bounds, the ratio between switching times defined in Equation (2.15) can be bound for nominal behaviour as

$$\frac{t_{2j+1} - t_{2j}}{t_{2j+2} - t_{2j+1}} \leq \frac{\bar{\dot{e}}^{0,-}_{2,(i)}}{\underline{\dot{e}}^{0,+}_{2,(i)}}.$$

Now we will use this bound on the duration between switches to bound the EOI. Let us define $r_e^{0,u} = \frac{\bar{\dot{e}}^{0,-}_{2,(i)}}{\underline{\dot{e}}^{0,+}_{2,(i)}}$, and $t_{j^-} = t_{2j} - t_{2j-1}$, such that we can write $t_{2j+1} - t_{2j} \leq r_e^{0,u} t_{j^-}$. Using this bound in the EOI response from Equation (2.11) gives the upper sustained condition as

$$\begin{aligned} \bar{\nu}_{\text{eq},0,(i)}(t_{2j}) =& e^{-k(1+r_e^{0,u})\sum_{\ell=0}^{j} t_{\ell^-}} \nu_{\text{eq},(i)}(t_0) \\ &- e^{-k(1+r_e^{0,u})\sum_{\ell=0}^{j} t_{\ell^-}} - 1 + 2\sum_{\ell=1}^{2j-1} \left((-1)^{\ell+1} e^{-k(\sum_{q=1}^{\lceil \frac{\ell}{2} \rceil} t_{q^-} + r_e^{0,u} \sum_{q=1}^{\lfloor \frac{\ell}{2} \rfloor} t_{q^-})} \right). \end{aligned} \tag{2.16}$$

which can be calculated at time t_{2j}, for each $j \in \mathbb{Z}^+$, and is valid over the period $[t_{2j}\ t_{2j+2}]$. Substituting this sustained condition in Equation (2.14) gives the combined threshold. Note that, by construction, this combined threshold satisfies design criteria 1 and 2. The corresponding detection logic is given by

$$\exists\ i,j \text{ s.t. } \nu_{\text{eq},(i)}(t) > \bar{\nu}_{\text{eq},(i)}(t_{2j}) \text{ for } t \in [t_{2j}\ t_{2(j+1)}]. \tag{2.17}$$

Remark 2.7. The sustained condition from Equation (2.16) can be calculated recursively as

$$\bar{\nu}_{\text{eq},0,(i)}(t_{2j}) = e^{-k(1+r_e^{0,u})t_{j^-}} \nu_{\text{eq},0,(i)}(t_{2j-2}) - e^{-k(1+r_e^{0,u})t_{j^-}} + 2e^{-kt_{j^-}} - 1, \tag{2.18}$$

to reduce the computational load. ◁

A lower combined threshold can be designed similarly as

$$\underline{\nu}_{\text{eq},(i)}(t_{2j+1}) = e^{-k\bar{t}^{0,l}} \underline{\nu}_{\text{eq},0,(i)}(t_{2j+1}) - 1 + e^{-k\bar{t}^{0,l}}, \tag{2.19}$$

$$\begin{aligned} \underline{\nu}_{\text{eq},0,(i)}(t_{2j+1}) =& e^{-k(1+r_e^{0,l})\sum_{\ell=0}^{j} t_{\ell^+}} \nu_{\text{eq},(i)}(t_1) \\ &+ e^{-k(1+r_e^{0,l})\sum_{\ell=0}^{j} t_{\ell^+}} + 1 - 2\sum_{\ell=1}^{2j-1} \left((-1)^{\ell+1} e^{-k(\sum_{q=1}^{\lceil \frac{\ell}{2} \rceil} t_{q^+} + r_e^{0,l} \sum_{q=1}^{\lfloor \frac{\ell}{2} \rfloor} t_{q^+})} \right), \end{aligned}$$

with $t_{j+} = t_{2j+1} - t_{2j}$, $t_{2j+2} - t_{2j+1} \leq r_e^{0,l} t_{j+}$, $r_e^{0,l} = \frac{\bar{\dot{e}}^{0,+}_{2,(i)}}{\underline{\dot{e}}^{0,-}_{2,(i)}}$, and the detection logic

$$\exists\ i,j \text{ s.t. } \nu_{\text{eq},(i)}(t) < \underline{\nu}_{\text{eq},(i)}(t_{2j+1}) \text{ for } t \in [t_{2j}\ t_{2(j+1)}]. \tag{2.20}$$

Even though this combined threshold is not reset at every switch, like the peak threshold was, it still requires to be recalculated at every switch, as t_{j-} and t_{j+} are actual durations between switches. Furthermore, as t_{j-} and t_{j+} are also influenced by the system uncertainty and measurement noise, the combined threshold is different for each uncertainty realisation. Therefore, in the following a constant upper-bound to the combined threshold will be designed, which can be calculated off-line.

Constant Combined Threshold

In this section a constant upper-bound to the combined threshold is designed. This threshold will be called the *constant combined threshold*. A constant threshold reduces the computational burden to a single off-line calculation. To calculate the constant threshold, without loss of generality, assume $t_{j-} = t_-$ for all j. This allows us to rewrite Equation (2.16) as

$$\bar{\nu}_{\text{eq},0,(i)} = e^{-k(1+r_e^{0,u})Nt_-}\nu_{\text{eq},(i)}(t_0) + 2(e^{-kt_-} - 1)\sum_{i=0}^{N} e^{-ki(1+r_e^{0,u})t_-} + 1 + e^{-kN(1+r_e^{0,u})t_-} - 2e^{-k(N+1+Nr_e^{0,u})t_-}.$$

Considering the effect of N alone, this bound will always increase for increasing N. Therefore, take $N \longrightarrow \infty$ to get a simplified constant threshold as

$$\lim_{N\to\infty} \bar{\nu}_{\text{eq},0,(i)} = 1 + 2(e^{-kt_-} - 1)\lim_{N\to\infty}\sum_{i=0}^{N} e^{-ki(1+r_e^{0,u})t_-} = 1 - 2\frac{e^{-kt_-} - 1}{e^{-k(1+r_e^{0,u})t_-} - 1}.$$

Only considering the effect of t_-, this expression is maximized for minimal t_-. So, by taking the limit for $t_- \longrightarrow 0$, once again a simplified upper-bound on the time-varying threshold is obtained. Using L'Hospital's rule this gives

$$\bar{\nu}^{\text{const}}_{\text{eq},0,(i)} = 1 - 2\frac{-k}{-k(1+r_e^{0,u})} = \frac{r_e^{0,u} - 1}{1 + r_e^{0,u}}.$$

Substituting the definition of $r_e^{0,u}$ this gives

$$\bar{\nu}^{\text{const}}_{\text{eq},0,(i)} = \frac{\bar{\dot{e}}^{0,-}_{2,(i)} - \underline{\dot{e}}^{0,+}_{2,(i)}}{\bar{\dot{e}}^{0,-}_{2,(i)} + \underline{\dot{e}}^{0,+}_{-2,(i)}}. \tag{2.21}$$

Substituting this expression in Equation (2.14) gives the constant combined threshold as

$$\bar{\nu}^{\text{const}}_{\text{eq},(i)} = e^{-k\bar{t}^{0,u}}\bar{\nu}^{\text{const}}_{\text{eq},0,(i)} + 1 - e^{-k\bar{t}^{0,u}}, \tag{2.22}$$

which is used with detection logic

$$\exists\ i \text{ s.t. } \nu_{\text{eq},(i)}(t) > \bar{\nu}^{\text{const}}_{\text{eq},(i)}\ \forall t. \tag{2.23}$$

A lower combined constant threshold can be designed similarly, resulting in

$$\underline{\nu}^{\text{const}}_{\text{eq},(i)} = e^{-k\bar{t}^{0,l}}\,\underline{\nu}^{\text{const}}_{\text{eq},0,(i)} - 1 + e^{-k\bar{t}^{0,l}},$$
$$\underline{\nu}^{\text{const}}_{\text{eq},0,(i)} = -\frac{\bar{e}^{0,+}_{2,(i)} - \underline{\dot{e}}^{0,-}_{2,(i)}}{\bar{e}^{0,+}_{2,(i)} + \underline{\dot{e}}^{0,-}_{2,(i)}},$$

with detection logic

$$\exists\ i \text{ s.t. } \nu_{\text{eq},(i)}(t) < \underline{\nu}^{\text{const}}_{\text{eq},(i)}\ \forall\ t\,. \tag{2.24}$$

To summarize, in this section, first the so-called *peak threshold* $\bar{\nu}^{\text{peak}}_{\text{eq}}$ has been designed. This threshold does allow for anomaly detection, but, detection can never be guaranteed. To address this sensitivity to measurement noise, the *sustained condition*, $\bar{\nu}_{\text{eq},0}$, was introduced as a global initial condition from which the combined threshold, $\bar{\nu}_{\text{eq}}$, can be calculated. For this combined threshold anomaly detection can be guaranteed, as will be proven in Section 2.2.3. However, it still has to be recalculated online at every switch of ν. To reduce the computational burden a *constant combined threshold* $\bar{\nu}^{\text{const}}_{\text{eq}}$ has been designed which over-bounds the combined threshold.

Remark 2.8. The derived detection thresholds are based on a novel approach to bound ν_{eq}. As such, a full analytical derivation and a suitable notation were required. However, this does not lead to a high computational cost. $\bar{\nu}_{\text{eq}}$ can be obtained online by Equations (2.14) and (2.18); $\bar{\nu}^{\text{const}}_{\text{eq}}$ can be obtained offline by Equations (2.21) and (2.22). ◁

2.2.3 Detectability Analysis

In this section the performance of the combined threshold is analysed. In doing so it will be proven that the threshold satisfies design criteria 3 and 4. First, in Theorem 2.2 a condition will be presented for which there exists a realisation of the noise and uncertainty such that detection occurs. Then, in Corollary 2.1, it will be proven that with δ_e and $\delta_{\dot{e}}$ the condition from Theorem 2.2 reduces to $f \neq 0$, proving design criterion 3 is satisfied.

Theorem 2.2. *If* $|\delta^+_{\dot{f}}(f)\underline{\dot{e}}^{0,-}_2 + \delta^-_{\dot{f}}(f)\bar{\dot{e}}^{0,+}_2| > \delta_{\dot{e}}(\bar{\dot{e}}^{0,+}_2 + \underline{\dot{e}}^{0,-}_2)$*, and* $\delta_f(f)\underline{\dot{e}}^{0,+}_2 - \delta^+_{\dot{f}}(f)\bar{e}^0_2 > \delta_{\dot{e}}\bar{e}^0_2 + \delta_e\underline{\dot{e}}^{0,+}_2$ *or* $\delta_f(f)\underline{\dot{e}}^{0,-}_2 + \delta^-_{\dot{f}}(f)\bar{e}^0_2 > \delta_{\dot{e}}\bar{e}^0_2 + \delta_e\underline{\dot{e}}^{0,+}_2$ *there exists a realisation of the uncertainty* ζ_1 *and noise* ζ_2 *such that detection occurs with the combined threshold.*

Proof. In order to prove that there exists a realisations of ζ_1 and ζ_2 such that detection occurs (using the upper threshold), we first design a function $\tilde{\nu}_{\text{eq}}$ such that $\exists t, \zeta_1, \zeta_2$ s.t. $\nu_{\text{eq}}(t) \geq \tilde{\nu}_{\text{eq}}$. Then, based on this function

$$\tilde{\nu}_{\text{eq}} > \bar{\nu}_{\text{eq}} \tag{2.25}$$

needs to hold to prove the theorem. The behaviour leading to the upper combined threshold is based on the realisations of ζ_1 and ζ_2 that maximize ν_{eq}. Therefore, with the same methodology, but using the true-anomalous bounds instead of the known-nominal bounds, $\tilde{\nu}_{\text{eq}}$ is defined as

$$\tilde{\nu}_{\text{eq},(i)} = e^{-k\tilde{t}^u}\,\tilde{\nu}_{\text{eq},0,(i)} + \left(1 - e^{-k\tilde{t}^u}\right),$$

where $\tilde{t}^u = \frac{2\tilde{e}_2}{\dot{\tilde{e}}_2^+}$ and $\tilde{v}_{\text{eq},0,(i)}$ is defined as in Equation (2.16) where we replace $r_e^{0,u} \leftarrow \tilde{r}_e^u = \frac{\tilde{\dot{e}}_2^-}{\dot{\tilde{e}}_2^+}$. Satisfying Equation (2.25) is now implied by $\tilde{t}^u > \bar{t}^{0,u}$ and $\tilde{r}_e^u > r_e^{0,u}$. Using Equation (2.6) $\tilde{t}^u > \bar{t}^{0,u}$ can be written as

$$\delta_f(f)\underline{\dot{e}}_2^{0,+} - \delta_{\dot{f}}^+(f)\bar{e}_2^0 > \delta_{\dot{e}}\bar{e}_2^0 + \delta_e\underline{\dot{e}}_2^{0,+},$$

and $\tilde{r}_e^u > r_e^{0,u}$ can be written as

$$\delta_{\dot{f}}^+(f)\underline{\dot{e}}_2^{0,-} + \delta_{\dot{f}}^-(f)\bar{\dot{e}}_2^{0,+} < -\delta_{\dot{e}}(\bar{\dot{e}}_2^{0,+} + \underline{\dot{e}}_2^{0,-}).$$

Similarly, using the lower peak threshold, we obtain

$$\delta_f(f)\underline{\dot{e}}_2^{0,-} + \delta_{\dot{f}}^-(f)\bar{e}_2^0 > \delta_{\dot{e}}\bar{e}_2^0 + \delta_e\underline{\dot{e}}_2^{0,+},$$
$$\delta_{\dot{f}}^+(f)\underline{\dot{e}}_2^{0,-} + \delta_{\dot{f}}^-(f)\bar{\dot{e}}_2^{0,+} > \delta_{\dot{e}}(\bar{\dot{e}}_2^{0,+} + \underline{\dot{e}}_2^{0,-}).$$

These conditions can be rewritten to those in the theorem statement. ∎

Corollary 2.1. *Assume $\delta_e = 0$ and $\delta_{\dot{e}} = 0$. If $f \neq 0$ there exists a realisation ζ_2 and ζ_1 for which detection occurs.*

Proof. Using the equalities in the theorem statement, the conditions on f in Theorem 2.2 reduce to $|\delta_{\dot{f}}^+(f)\underline{\dot{e}}_2^{0,-} + \delta_{\dot{f}}^-(f)\bar{\dot{e}}_2^{0,+}| > 0$, and $\delta_f(f)\underline{\dot{e}}_2^{0,+} - \delta_{\dot{f}}^+(f)\bar{e}_2^0 > 0$ or $\delta_f(f)\underline{\dot{e}}_2^{0,-} + \delta_{\dot{f}}^-(f)\bar{e}_2^0 > 0$. By Proposition 2.3 these conditions are implied by $f \neq 0$. ∎

In the following, a sufficient condition will be presented guaranteeing anomaly detection in terms of a minimum anomaly magnitude, i.e. all anomalies continuously larger than this magnitude are guaranteed to be detected in finite time. This proves design condition 4 holds.

Theorem 2.3. *If*

$$\delta_{\dot{f}}^+(f) + \delta_{\dot{f}}^-(f) + (\delta_{\dot{f}}^+(f) - \delta_{\dot{f}}^-(f))\bar{v}_{eq} < -(\underline{\dot{e}}_2^{0,-} + \bar{\dot{e}}_2^{0,+})\bar{v}_{eq} + (\underline{\dot{e}}_2^{0,-} - \bar{\dot{e}}_2^{0,+} + 2\delta_{\dot{e}})$$
or
$$-\delta_{\dot{f}}^+(f) - \delta_{\dot{f}}^-(f) - (\delta_{\dot{f}}^+(f) - \delta_{\dot{f}}^-(f))\underline{v}_{eq} < (\underline{\dot{e}}_2^{0,-} + \bar{\dot{e}}_2^{0,+})\underline{v}_{eq} + (\underline{\dot{e}}_2^{0,+} - \bar{\dot{e}}_2^{0,-} + 2\delta_{\dot{e}}),$$

an anomaly is guaranteed to be detected within finite time.

Proof. To prove that detection is guaranteed for all realisations of ζ_1 and ζ_2, first define a function such that $\exists t$ s.t. $v_{\text{eq}}(t) \geq \underline{v}_{\text{eq}}\ \forall \zeta_1, \zeta_2$. Then, based on this function, the relation

$$\underline{v}_{\text{eq}} > \bar{v}_{\text{eq}} \tag{2.26}$$

needs to hold to prove the theorem statement. For the design of $\underline{v}_{\text{eq}}$, consider the behaviour leading to the lower sustained condition, $\underline{v}_{\text{eq},0}$. The lower sustained condition is designed such that for all realisations of ζ_1 and ζ_2, $v_{\text{eq}}(t_{2j+1}) \geq \underline{v}_{\text{eq},0}(t_{2j+1})$ if $e_2(t_{2j+3}) \geq e_2(t_{2j+1})$. Furthermore, as e_2 is bounded, $\exists t$ s.t. $e_2(t_{2j+3}) \geq e_2(t_{2j+1})$. Therefore, with the same methodology,

but using the true-anomalous bounds instead of the known-nominal bounds, $\underline{\nu}_{\text{eq}}$ can be defined as

$$\underline{\nu}_{\text{eq},(i)} = -\frac{\tilde{\dot{e}}^{+}_{2,(i)} - \underset{\sim}{\dot{e}}^{-}_{2,(i)}}{\tilde{\dot{e}}^{+}_{2,(i)} + \underset{\sim}{\dot{e}}^{-}_{2,(i)}}.$$

With this, detection can be guaranteed, according to Equation (2.26), if

$$-\frac{\tilde{\dot{e}}^{+}_{2,(i)} - \underset{\sim}{\dot{e}}^{-}_{2,(i)}}{\tilde{\dot{e}}^{+}_{2,(i)} + \underset{\sim}{\dot{e}}^{-}_{2,(i)}} > \bar{\nu}_{\text{eq},(i)}$$

which can be simplified to

$$\delta^{+}_{\dot{f}}(f) + \delta^{-}_{\dot{f}}(f) + (\delta^{+}_{\dot{f}}(f) - \delta^{-}_{\dot{f}}(f))\bar{\nu}_{\text{eq}} < -(\underline{\dot{e}}^{0,-}_{2} + \bar{\dot{e}}^{0,+}_{2})\bar{\nu}_{\text{eq}} + (\underline{\dot{e}}^{0,-}_{2} - \bar{\dot{e}}^{0,+}_{2} + 2\delta_{\dot{e}}).$$

where subscript (i) is dropped to ease notation. Similarly considering detection by the lower threshold we obtain

$$-\delta^{+}_{\dot{f}}(f) - \delta^{-}_{\dot{f}}(f) - (\delta^{+}_{\dot{f}}(f) - \delta^{-}_{\dot{f}}(f))\underline{\nu}_{\text{eq}} < (\underline{\dot{e}}^{0,+}_{2} + \bar{\dot{e}}^{0,-}_{2})\underline{\nu}_{\text{eq}} + (\underline{\dot{e}}^{0,+}_{2} - \bar{\dot{e}}^{0,-}_{2} + 2\delta_{\dot{e}}).$$

■

Corollary 2.2. *If $\tilde{f}$ is sufficiently large there always exists an f such that the conditions in Theorem 2.3 hold .*

Proof. By Assumption 2.1 and Proposition 2.3 there exists an f such that $\delta^{+}_{\dot{f},(i)}(f) < -\gamma d_f$ for any $0 < d_f < \tilde{f}$ and $\delta^{-}_{\dot{f},(i)}(f) \leq 0$. Substituting this in the first condition of Theorem 2.3 -for detection with the upper threshold- gives

$$\tilde{f} > \frac{(\underline{\dot{e}}^{0,-}_{2} + \bar{\dot{e}}^{0,+}_{2})\bar{\nu}_{\text{eq}} - (\underline{\dot{e}}^{0,-}_{2} - \bar{\dot{e}}^{0,+}_{2} + 2\delta_{\dot{e}})}{\gamma(1 + \bar{\nu}_{\text{eq}})}. \tag{2.27}$$

Similarly for detection with the lower threshold we get

$$\tilde{f} > \frac{(\underline{\dot{e}}^{0,-}_{2} + \bar{\dot{e}}^{0,+}_{2})\underline{\nu}_{\text{eq}} + (\underline{\dot{e}}^{0,+}_{2} - \bar{\dot{e}}^{0,-}_{2} + 2\delta_{\dot{e}})}{\gamma(\underline{\nu}_{\text{eq}} - 1)}. \tag{2.28}$$

Therefore, if $\tilde{f}$ satisfies Equation (2.27) or (2.28), there exists an f s.t. one of the conditions in Theorem 2.3 holds. ■

2.3 Detection using Thresholds on Observer Error

In this section, an anomaly detector will be presented which is based on direct analyses of the behaviour of observer error e_2. The detection logic uses thresholds on the observer error e_2 based on the bounds in Proposition 2.2, and the relation $e_y = C_2 e_2 - F\zeta_2$ from Equation (2.3). The resulting thresholds, $\underline{\epsilon}_{2,(i)} \leq e^{0}_{2,(i)} \leq \bar{\epsilon}_{2,(i)}$, will be used for anomaly detection. Preferably one would directly monitor this condition, and detect an anomaly

when it is violated. However, as e_2 is not known to the observer, this is not possible. Alternatively, the condition

$$\underline{\epsilon}_{2,(i)} > \bar{\epsilon}_{2,(i)} \tag{2.29}$$

can be monitored. Satisfying this condition implies violation of $\underline{\epsilon}_{2,(i)} \leq e_{2,(i)} \leq \bar{\epsilon}_{2,(i)}$, and can thus serve as detection condition. In the following, first the thresholds $\underline{\epsilon}_{2,(i)}$ and $\bar{\epsilon}_{2,(i)}$ will be defined in Section 2.3.1. Then it will be shown that detection using these thresholds conforms to the design criteria from Section 2.1.5.

2.3.1 Observer Error Thresholds

The thresholds $\underline{\epsilon}_{2,(i)}$ and $\bar{\epsilon}_{2,(i)}$ will be constructed using three bounds on the observer error dynamics in Equation (2.3). Of these bounds, two are bounds which hold only in nominal conditions and one bound also holds when the anomaly occurs. The combination of these bounds will be used to detect the occurrence of an anomaly. In the following these bounds will be formally introduced and combined to form the thresholds. Subscripts (i) will be omitted for ease of notation.

The first bound on e_2 is taken directly from Proposition 2.2, saying that at all time

$$|e_2^0| < \bar{e}_2^0 .$$

The second bound depends on e_y and can thus only be checked at times the observer receives a measurement y, which will be denoted by the sequence of times $\{t_{m_j}\}$. Note that the system dynamics are continuous time and the measurement times are only defined to obtain an explicit expression for the bounds. Using this, from Equation (2.3) and Assumption 2.1 it can be derived that

$$\begin{aligned} &\underline{e}_2^y(t_{m_j}) \leq e_2(t_{m_j}) \leq \bar{e}_2^y(t_{m_j}) \; \forall j \;\; \text{where} \\ &\underline{e}_2^y(t_{m_j}) = C_2^{-1}(e_y(t_{m_j}) - F\bar{\zeta}_2(t_{m_j})) \;\&\; \bar{e}_2^y(t_{m_j}) = C_2^{-1}(e_y(t_{m_j}) + F\bar{\zeta}_2(t_{m_j})) . \end{aligned} \tag{2.30}$$

Then, at any time t_{m_j}, e_2^0 can be bounded as

$$\max(\underline{e}_2^y(t_{m_j}), -\bar{e}_2^0) \leq e_2^0(t_{m_j}) \leq \min(\bar{e}_2^y(t_{m_j}), \bar{e}_2^0) \tag{2.31}$$

Furthermore, bounds on $\dot{e}_2^0$ are known from Proposition 2.2. With these, e_2 can be bound for each t in the period $[t_{m_{j-1}}\ t_{m_j}]$ as

$$\begin{aligned} &\text{If } \dot{e}_2(t_{m_{j-1}}) > 0 \\ &\qquad \int_{t_{m_{j-1}}}^{t} \underline{\dot{e}}_2^{0,+}(T)\mathrm{d}T \leq e_2^0(t_{m_j}) - e_2^0(t_{m_{j-1}}) \leq \int_{t_{m_{j-1}}}^{t} \bar{\dot{e}}_2^{0,+}(T)\mathrm{d}T . \\ &\text{If } \dot{e}_2(t_{m_{j-1}}) < 0 \\ &\qquad -\int_{t_{m_{j-1}}}^{t} \bar{\dot{e}}_2^{0,-}(T)\mathrm{d}T \leq e_2^0(t_{m_j}) - e_2^0(t_{m_{j-1}}) \leq -\int_{t_{m_{j-1}}}^{t} \underline{\dot{e}}_2^{0,-}(T)\mathrm{d}T . \end{aligned} \tag{2.32}$$

The above bounds require further inspection. At first sight they seem to depend only on the modeled healthy system behaviour through $\bar{\dot{e}}_2^0$ and $\underline{\dot{e}}_2^0$. However, the bounds also depend on the real behaviour. As can be seen, the integration duration is dictated by the sign of $\dot{e}_2$,

which through Proposition 2.2 is related to e_y. The bounds in Equations (2.31) and (2.32) can be combined for $j \geq 1$ to form the thresholds as

$$\begin{aligned}
&\text{If } \dot{e}_2(t_{m_{j-1}}) > 0 \\
&\quad \bar{\epsilon}_2(t_{m_j}) = \min(\bar{\epsilon}_2(t_{m_{j-1}}) + \int_{t_{m_{j-1}}}^{t_{m_j}} \bar{\dot{e}}_2^{0,+}(T)\mathrm{d}T, \bar{e}_2^y(t_{m_j}), \bar{e}_2^0) \\
&\quad \underline{\epsilon}_2(t_{m_j}) = \max(\underline{\epsilon}_2(t_{m_{j-1}}) + \int_{t_{m_{j-1}}}^{t_{m_j}} \underline{\dot{e}}_2^{0,+}(T)\mathrm{d}T, \underline{e}_2^y(t_{m_j}), -\bar{e}_2^0) \\
&\text{If } \dot{e}_2(t_{m_{j-1}}) < 0 \\
&\quad \bar{\epsilon}_2(t_{m_j}) = \min(\bar{\epsilon}_2(t_{m_{j-1}}) - \int_{t_{m_{j-1}}}^{t_{m_j}} \underline{\dot{e}}_2^{0,-}(T)\mathrm{d}T, \bar{e}_2^y(t_{m_j}), \bar{e}_2^0) \\
&\quad \underline{\epsilon}_2(t_{m_j}) = \max(\underline{\epsilon}_2(t_{m_{j-1}}) - \int_{t_{m_{j-1}}}^{t_{m_j}} \bar{\dot{e}}_2^{0,-}(T)\mathrm{d}T, \underline{e}_2^y(t_{m_j}), -\bar{e}_2^0)
\end{aligned} \tag{2.33}$$

The bounds in Equation (2.31) can be used to obtain the initial thresholds $\bar{\epsilon}_2(t_{m_0})$ and $\underline{\epsilon}_2(t_{m_0})$. Based on these thresholds, the detection condition

$$\underline{\epsilon}_2(t_{m_j}) > \bar{\epsilon}_2(t_{m_j}) \tag{2.34}$$

can be monitored at every measurement time t_{m_j} to detect anomalies.

2.3.2 Detectability Analysis

In this section it will be proven that the proposed detector conforms to the design criteria from Section 2.1.5. In Theorem 2.4 it will be proven that design criterion 2 holds. Design criteria 3 and 4 will be proven to hold in Corollary 2.3 and Theorem 2.5 respectively.

Theorem 2.4 (Robustness). *Consider the system in Equation* (2.1)*, observer in Equation* (2.2) *and detection criterion in Equation* (2.34)*. In nominal conditions the detection criterion will never be satisfied, i.e. there will be no false detection.*

Proof. Recall from Section 2.2 the sequence $\{t_j\}$ which denotes the times at which $\dot{e}_2$ changes sign and that $\dot{e}_2 > 0$ during periods $[t_{2j}\ t_{2j+1}]$. Furthermore recall $\dot{e}_2^+$ denoting the average $|\dot{e}_2|$ while $\dot{e}_2 > 0$, and $\dot{e}_2^-$ the average $|\dot{e}_2|$ while $\dot{e}_2 < 0$.

For analysis purposes, the continuous dynamics of e_2 from Equation (2.3) can be rewritten in a discrete form based on the switching times as

$$e_2(t_{2j+2N}) = e_2(t_{2j}) + \sum_{\ell=0}^{N} c_{j+\ell} \ \forall N \in \mathbb{Z}, \text{ where } c_j = t_j^+ \dot{e}_2^+ - t_j^- \dot{e}_2^-, \tag{2.35}$$

with $t_j^+ = t_{2j+1} - t_{2j}$ and $t_j^- = t_{2j+2} - t_{2j+1}$. Based on Equation (2.35), in nominal conditions c_j can be bounded as

$$-e_2(t_{2j}) + \max(\underline{e}_2^y(t_{2j}), -\bar{e}_2^0) \leq \sum_{\ell=0}^{N} c_{j+\ell} \leq -e_2(t_{2j}) + \min(\bar{e}_2^y(t_{2j}), \bar{e}_2^0) \ \forall N \in \mathbb{Z}, \tag{2.36}$$

using the bounds on e_2 from Proposition 2.2 and $e_y = C_2 e_2 - F\zeta_2$ from Equation (2.3).

Now, we will have a look at the dynamics of the thresholds $\bar{\epsilon}_2$ and $\underline{\epsilon}_2$ from Equation (2.33). Without loss of generality, we will analyse the dynamics of $\bar{\epsilon}_2$ and $\underline{\epsilon}_2$ using three scenarios, that together describe all possible behaviour.

1. $\bar{\epsilon}_2$ and $\underline{\epsilon}_2$ are fully determined by the integral bounds from Equation (2.32).
2. $\bar{\epsilon}_2$ or $\underline{\epsilon}_2$ is affected by the instantaneous bounds from Equation (2.31).
3. $\bar{\epsilon}_2$ and $\underline{\epsilon}_2$ are both affected by the instantaneous bounds from Equation (2.31).

If we prove that $\bar{\epsilon}_2 > \underline{\epsilon}_2$ in nominal conditions for each scenario, then robustness is guaranteed. Starting with scenario 1, we can write the discrete form of the dynamics of $\bar{\epsilon}_2$ over a period $[t_{2j}\ t_{2j+2}]$ as

$$\bar{\epsilon}_2(t_{2j+2}) = \bar{\epsilon}_2(t_{2j}) + \bar{\dot{e}}_2^{0,+} t_j^+ - \underline{\dot{e}}_2^{0,-} t_j^- ,$$

where the definition of c_j can be used to eliminate t_j+ as

$$\bar{\epsilon}_2(t_{2j+2}) = \bar{\epsilon}_2(t_{2j}) + \frac{t_j^-}{\dot{e}_2^+}(\bar{\dot{e}}_2^{0,+}\dot{e}_2^- - \underline{\dot{e}}_2^{0,-}\dot{e}_2^+) + \frac{\bar{\dot{e}}_2^{0,+}}{\dot{e}_2^+} c_j . \tag{2.37}$$

The latter can be extended for $\bar{\epsilon}_2(t_{2j+2N})$ for any $N \in \mathbb{Z}$ as

$$\bar{\epsilon}_2(t_{2j+2N}) = \bar{\epsilon}_2(t_{2j}) + \sum_{\ell=0}^{N-1}\left(\frac{t_{j+\ell}^-}{\dot{e}_2^+}(\bar{\dot{e}}_2^{0,+}\dot{e}_2^- - \underline{\dot{e}}_2^{0,-}\dot{e}_2^+) + \frac{\bar{\dot{e}}_2^{0,+}}{\dot{e}_2^+} c_{j+\ell} \right) . \tag{2.38}$$

Similarly for $\underline{\epsilon}_2(t_{2j+2N})$ we can derive

$$\underline{\epsilon}_2(t_{2j+2N}) = \underline{\epsilon}_2(t_{2j}) + \sum_{\ell=0}^{N-1}\left(\frac{t_{j+\ell}^-}{\dot{e}_2^+}(\underline{\dot{e}}_2^{0,+}\dot{e}_2^- - \bar{\dot{e}}_2^{0,-}\dot{e}_2^+) + \frac{\underline{\dot{e}}_2^{0,+}}{\dot{e}_2^+} c_{j+\ell} \right) . \tag{2.39}$$

It can be seen that in nominal conditions, when $\underline{\dot{e}}_2^{0,-} \le \dot{e}_2^- \le \bar{\dot{e}}_2^{0,-}$ and $\underline{\dot{e}}_2^{0,+} \le \dot{e}_2^+ \le \bar{\dot{e}}_2^{0,+}$,

$$\begin{aligned} \bar{\epsilon}_2(t_{2j+2N}) - \bar{\epsilon}_2(t_{2j}) &\ge \sum_{\ell=0}^{N-1} \frac{\bar{\dot{e}}_2^{0,+}}{\dot{e}_2^+} c_{j+\ell} \ge \sum_{\ell=0}^{N-1} c_{j+\ell} \\ \underline{\epsilon}_2(t_{2j+2N}) - \underline{\epsilon}_2(t_{2j}) &\le \sum_{\ell=0}^{N-1} \frac{\underline{\dot{e}}_2^{0,+}}{\dot{e}_2^+} c_{j+\ell} \le \sum_{\ell=0}^{N-1} c_{j+\ell} \end{aligned} \tag{2.40}$$

By subtracting these inequalities it can be found that $\bar{\epsilon}_2(t_{2j+2N}) - \underline{\epsilon}_2(t_{2j+2N}) \ge \bar{\epsilon}_2(t_{2j}) - \underline{\epsilon}_2(t_{2j})$, i.e. considering scenario 1, the difference between $\bar{e}_2$ and $\underline{e}_2$ is non-decreasing. This leaves to prove that no detection occurs in scenarios 2 and 3. In scenario 2, if the lower bound is affected by Equation (2.31), $\underline{\epsilon}_2(t_{2j+2N}) = \max(\underline{e}_2^y(t_{2j+2N}), -\bar{e}_2^0)$. Then use Equations (2.36) and (2.40) to derive that in nominal conditions

$$\bar{\epsilon}_2(t_{2j+2N}) \ge \bar{\epsilon}_2(t_{2j}) - e_2(t_{2j}) + \max(\underline{e}_2^y(t_{2j}), -\bar{e}_2^0) \ge \max(\underline{e}_2^y(t_{2j}), -\bar{e}_2^0) = \underline{\epsilon}_2(t_{2j+2N}),$$

which implies there is no detection. Lastly, in scenario 3

$$\bar{\epsilon}_2(t_{2j}) - \underline{\epsilon}_2(t_{2j}) = \min(\bar{e}_2^y(t_{2j}), \bar{e}_2^0) - \max(\underline{e}_2^y(t_{2j}), -\bar{e}_2^0) \ge 0 .$$

■

Theorem 2.5 (Detectability). *An anomaly starting at t_{2j} is guaranteed to be detected before t_{2j+2N} if $N > \frac{4\bar{e}_2^0}{\phi\delta}$ where $\phi \leq \frac{\underline{t}_{j+\ell}^-}{\dot{e}_2^+}$ $\forall\, 0 < \ell \leq N$ and $\delta > 0$ is such that $\delta_{\dot{f}}^-(f) \geq \frac{\bar{e}_2^{0,+}\bar{e}_2^{0,-} - \underline{\dot{e}}_2^{0,+}\underline{\dot{e}}_2^{0,-} + \delta}{\bar{e}_2^{0,+}}$ or $\delta_{\dot{f}}^+(f) \leq -\frac{\bar{e}_2^{0,+}\bar{e}_2^{0,-} - \underline{\dot{e}}_2^{0,+}\underline{\dot{e}}_2^{0,-} + \delta}{\bar{e}_2^{0,-}}$ during the period $[t_{2j}\ t_{2j+2N}]$.*

Proof. First, note that the bounds on e_2 in Equations (2.31) and (2.32) are always more conservative when used separately, than when used together in the threshold in Equation (2.33). Therefore, for the purpose of proving a sufficient condition for detectability, they can be used interchangeably. In this proof we can, therefore, without loss of generality consider the following behaviour of $\bar{\epsilon}_2$ and $\underline{\epsilon}_2$. At t_{2j} both $\bar{\epsilon}_2$ and $\underline{\epsilon}_2$ are determined by Equation (2.31) and at any subsequent time $\bar{\epsilon}_2$ is governed by Equation (2.32) and $\underline{\epsilon}_2$ by Equation (2.31).[5]

By case 2 in Proposition 2.3 there exists a sufficiently large f such that $\delta_{\dot{f}}^+(f) \geq 0$ and $\delta_{\dot{f}}^-(f) \geq \frac{\bar{e}_2^{0,+}\bar{e}_2^{0,-} - \underline{\dot{e}}_2^{0,+}\underline{\dot{e}}_2^{0,-} + \delta}{\bar{e}_2^{0,+}}$. With such f, using Equation (2.6) we can derive

$$\dot{e}_2^- \leq \bar{\dot{e}}_2^- = \bar{\dot{e}}_2^{0,-} - \delta_{\dot{f}}^-(f) \leq \frac{\underline{\dot{e}}_2^{0,+}\underline{\dot{e}}_2^{0,-} - \delta}{\bar{e}_2^{0,+}},$$
$$\dot{e}_2^+ \geq \underline{\dot{e}}_2^{0,+},$$

such that

$$\bar{\dot{e}}_2^{0,+}\dot{e}_2^- - \underline{\dot{e}}_2^{0,-}\dot{e}_2^+ \leq \underline{\dot{e}}_2^{0,+}\underline{\dot{e}}_2^{0,-} - \delta - \underline{\dot{e}}_2^{0,-}\underline{\dot{e}}_2^{0,+} = -\delta. \tag{2.41}$$

Substituting this inequality and $\phi \leq \frac{\underline{t}_{j+\ell}^-}{\dot{e}_2^+}$ $\forall j, \ell$ in Equation (2.38)[6] gives

$$\bar{\epsilon}_2(t_{2j+2N}) - \bar{\epsilon}_2(t_{2j}) \leq -N\phi\delta + \sum_{\ell=0}^{N-1} c_{j+\ell}.$$

Using the bound on c_j from Equation (2.36) gives

$$\bar{\epsilon}_2(t_{2j+2N}) \leq \bar{\epsilon}_2(t_{2j}) - N\phi\delta - e_2(t_{2j}) + \min(\bar{e}_2^y(t_{2j}), \bar{e}_2^0),$$

where by Equation (2.31), $\bar{\epsilon}_2(t_{2j}) - e_2(t_{2j}) \leq \min(\bar{e}_2^y(t_{2j}), \bar{e}_2^0) - \max(\underline{e}_2^y(t_{2j}), -\bar{e}_2^0)$ such that

$$\bar{\epsilon}_2(t_{2j+2N}) \leq -N\phi\delta + 2\min(\bar{e}_2^y(t_{2j}), \bar{e}_2^0) - \max(\underline{e}_2^y(t_{2j}), -\bar{e}_2^0).$$

Meanwhile, always $\underline{\epsilon}_2(t_{2j+2N}) \geq \max(\underline{e}_2^y(t_{2j+2N}), -\bar{e}_2^0)$, such that

$$\bar{\epsilon}_2(t_{2j+2N}) - \underline{\epsilon}_2(t_{2j+2N}) \leq -N\phi\delta + 2\min(\bar{e}_2^y(t_{2j}), \bar{e}_2^0) - \max(\underline{e}_2^y(t_{2j}), -\bar{e}_2^0) - \max(\underline{e}_2^y(t_{2j+2N}), -\bar{e}_2^0),$$

which can be simplified to

$$\bar{\epsilon}_2(t_{2j+2N}) - \underline{\epsilon}_2(t_{2j+2N}) \leq -N\phi\delta + 4\bar{e}_2^0. \tag{2.42}$$

[5] Although it is not proven here, the considered behaviour does correspond to the typical behaviour leading to detection of an anomaly

[6] Recall from the proof of Theorem 2.4 that Equation (2.38) is a discrete equivalent of Equation (2.32).

Detection occurs when $\bar{\epsilon}_2(t_{2j+2N}) < \underline{\epsilon}_2(t_{2j+2N})$ which, by Equation (2.42), is guaranteed for $N > \frac{4\bar{e}_2^0}{\phi\delta}$.

Alternatively, by case 1 in Proposition 2.3 there exists a sufficiently large f such that $\delta_f^+(f) \leq -\frac{\bar{e}_2^{0,+}\bar{e}_2^{0,-} - \underline{e}_2^{0,+}\underline{e}_2^{0,-} + \delta}{\bar{e}_2^{0,-}}$ and $\delta_f^+(f) \leq 0$ for which the same result can be obtained using the same procedure. ∎

Corollary 2.3 (Detectability)**.** *Assume $\delta_{\dot{e}} = 0$. If $f \neq 0$ there exists a realisation of the uncertainty ζ_1 and ζ_2 for which detection occurs.*

Proof. Consider case 2 in Proposition 2.3 such that a nonzero f implies $\delta_f^+(f) \geq 0$ and $\delta_f^-(f) \geq \delta_0$ for any $\delta_0 > 0$. Furthermore, by definition there exists a realisation of ζ_1 and ζ_2 such that $\dot{e}_2^- = \underline{e}_2^-$ and $\dot{e}_2^+ = \tilde{e}_2^-$. Using these relations in Equation (2.6) one can then write

$$\begin{aligned} \bar{e}_2^{0,+} &\leq \dot{e}_2^+ + \delta_{\dot{e}}\,, \\ \underline{e}_2^{0,-} &\geq \underline{e}_2^- + \delta_0 - \delta_{\dot{e}}\,, \end{aligned} \tag{2.43}$$

which means

$$\bar{e}_2^{0,+}\dot{e}_2^- - \underline{e}_2^{0,-}\dot{e}_2^+ \leq (\dot{e}_2^+ + \delta_{\dot{e}})\dot{e}_2^- - (\dot{e}_2^- + \delta_0 - \delta_{\dot{e}})\dot{e}_2^+ = (\dot{e}_2^+ + \dot{e}_2^-)\delta_{\dot{e}} - \delta_0\dot{e}_2^+\,. \tag{2.44}$$

Using $\delta_{\dot{e}} = 0$ and $\delta = \delta_0\dot{e}_2^+ > 0$ this becomes

$$\bar{e}_2^{0,+}\dot{e}_2^- - \underline{e}_2^{0,-}\dot{e}_2^+ \leq -\delta\,, \tag{2.45}$$

which is equivalent to Equation (2.41) which is proven to be a sufficient condition for detection in Theorem 2.5. ∎

2.4 Results of Application to a CVP under MITM Attack

In this section the SMO based anomaly detectors presented in Sections 2.2 and 2.3 will be applied to a collaborative vehicle platoon (CVP) subject to man-in-the-middle (MITM) attacks as introduced in Section 1.1.1. The detectors will be compared based on their performance in this application. First, the CVP model will be defined, and then rewritten such that it can be used for implementation of the detectors. Then, through simulations, the effect of the SMO tuning parameters is demonstrated in Section 2.4.1 using a simple step attack. Detection results are shown for a more elaborate attack scenario in Section 2.4.2.

The results presented in this section are obtained using a CVP of a lead vehicle followed by one follower. The platoon is connected in a predecessor-follower communication topology as illustrated in Figure 1.2 and is used to communicate the control input. Each vehicle in the CVP is modeled as

$$\begin{cases} \dot{p}_i = v_i \\ \dot{v}_i = a_i \\ \dot{a}_i = \dfrac{1}{\tau_i}(u_i - a_i) \end{cases} \tag{2.46}$$

where i denotes the vehicle number, p_i, v_i and a_i denote its position, velocity and acceleration, u_i is the applied control input, and τ_i is the engine time constant.

Assumption 2.2. The input of each vehicle i is constrained by $u_{\min} \le u_i \le u_{\max}$. ◁

Each vehicle aims to keep a reference distance, $d_{r,i}$ from its predecessor, defined as

$$d_{r,i} \triangleq r + h v_i, \tag{2.47}$$

where h denotes the reference time headway and r is the reference distance at standstill. Based on this vehicle dynamics, the model of the platoon, from the perspective of the follower, can be written as

$$\begin{cases} \begin{bmatrix} \dot{p}_l \\ \dot{v}_l \\ \dot{a}_l \\ \dot{p}_f \\ \dot{v}_f \\ \dot{a}_f \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{\tau_l} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{\tau_f} \end{bmatrix} \begin{bmatrix} p_l \\ v_l \\ a_l \\ p_f \\ v_f \\ a_f \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1}{\tau_l} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \frac{1}{\tau_f} \end{bmatrix} \begin{bmatrix} u_l \\ u_f \end{bmatrix} \\ y = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_l \\ v_l \\ a_l \\ p_f \\ v_f \\ a_f \end{bmatrix} + \zeta_2 \end{cases} \tag{2.48}$$

It is considered that the follower might not know exactly the dynamics of the lead vehicle, i.e. there is uncertainty in τ_l. To account for this, define $\hat{\tau}_l$ as the estimate of τ_l used by the follower vehicle such that, with some bounded r_τ, we can write $\tau_l = r_\tau \hat{\tau}_l$. Furthermore, u_l is only known to the follower vehicle through communication, which is affected by the MITM attack. Therefore define a_u as the additive MITM attack and denote $u_l^a = u_l + a_u$ as the attacked input known to the follower vehicle. This allows us to write

$$\begin{cases} \begin{bmatrix} \dot{p}_l \\ \dot{v}_l \\ \dot{a}_l \\ \dot{p}_f \\ \dot{v}_f \\ \dot{a}_f \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{\hat{\tau}_l} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{\tau_f} \end{bmatrix} \begin{bmatrix} p_l \\ v_l \\ a_l \\ p_f \\ v_f \\ a_f \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1}{\hat{\tau}_l} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \frac{1}{\tau_f} \end{bmatrix} \begin{bmatrix} u_l^a \\ u_f \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\hat{\tau}_l} \\ 0 \\ 0 \\ 0 \end{bmatrix} \zeta_1 + \begin{bmatrix} 0 \\ 0 \\ -\frac{1}{\hat{\tau}_l} \\ 0 \\ 0 \\ 0 \end{bmatrix} f \\ y = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_l \\ v_l \\ a_l \\ p_f \\ v_f \\ a_f \end{bmatrix} + \zeta_2 \end{cases} \tag{2.49}$$

where $\zeta_1 = (r_\tau - 1)(u_l - a_l)$ such that all uncertainty is concentrated in ζ_1 and ζ_2, and $f = a_u$ such that it contains all anomalies. This model can be cast in the form of Equation (2.1) as

$$\begin{cases} \dot{x}_1 = A_{11}x_1 + A^s_{12}x_2 + E_{11}\zeta_1 + B_1 u + N_1 f \\ \dot{x}_2 = A_{21}x_1 + A^s_{22}x_2 + B_2 u \\ y = x_2 + \zeta_2 \end{cases}$$

where $x_1 = [p_l,\ a_l]^\top$ and $x_2 = [p_l - p_f,\ v_l - v_f,\ v_f,\ a_f]^\top$, allowing it to be used for the SMO based detectors. Of the SMOs considered in Section 2.1.4, only the one from [9] can be applied to this model as $N_1 \neq 0$. The noise is bounded as $\bar{\zeta}_2 = [15\ 30\ 3\ 15]^\top \cdot 10^{-2}$. Other model and observer parameters used in this section are presented in Table 2.1.

Table 2.1: Parameters used in simulation

Param.	Value	Param.	Value	Param.	Value
τ_f	$0.1\ [s]$	τ_l	$0.11\ [s]$	$\hat{\tau}_l$	$0.1\ [s]$
$\bar{\zeta}_1$	$1\ \left[\frac{m}{s^2}\right]$	$\bar{f}$	$10\ \left[\frac{m}{s^2}\right]$	A_{21}	0
A_{22}	$-0.1I_4$	P	I_4	K_1	0

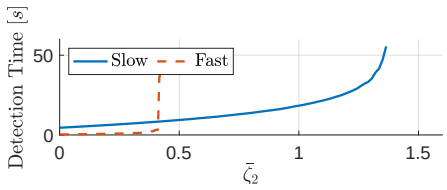
2.4.1 Parameter Study

To investigate the effect of the SMO tuning parameters on detection performance we introduce two sets of design parameters which will be referred to as the *slow* and *fast* parameter sets. The slow parameter set is $K_2 = [2.35\ 3.3\ 2.2\ 3.6]$, $K_v = 0.1 \cdot I_4$; and the fast parameter set is $K_2 = [10.35\ 11.3\ 10.2\ 11.6]$, $K_v = 2 \cdot I_4$. A step attack with magnitude 2.8 $[m/s^2]$ will then be applied to CVPs with different bounds on the measurement uncertainty $\bar{\zeta}_2$. Figures 2.3a and 2.3b show the detection times for the EOI and observer error based detectors respectively. Note that for this parameter study the measurement noise bounds on each measurement are taken to be equal.

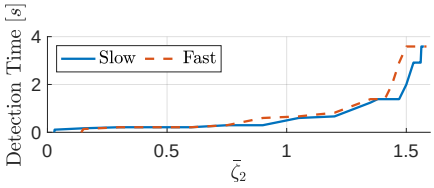
For the EOI based detector one can see in Figure 2.3a that for low noise bounds faster detection is obtained with the fast parameter set. However, for larger noise bounds the same attack is no longer detected. This because for the same noise bound the threshold corresponding to the fast parameter set is higher than for the slow parameter set. Based on this result, the optimal parameter set for any application of the presented detection threshold depends on the system uncertainty, including measurement noise, and the expected anomaly magnitude. However, note that only step anomalies are considered in this comparison. Different anomalies, like ramp or oscillatory anomalies may lead to different conclusions. As the detector is guaranteed to have no false detections, it is possible to simultaneously use multiple detectors, without loss in accuracy. Each detector can then be designed for a specific type of fault.

For the observer error based detector one can see in Figure 2.3b that there is much less difference in performance between the slow and fast parameter sets. This can partially be attributed to the fact that K_v is not a tuning parameter for this method. On a smaller scale however, one can still see that detection with the fast parameter set is faster for small uncertainties and slower for large uncertainties.

Comparing the EOI and observer error based detectors, most importantly one can see that the observer error based detector is much faster than the EOI based method.[7] Furthermore, one can see that with the observer error based method the same anomaly can be detected upto even larger uncertainties than the slow parameter set with the EOI based detector. Both effects can be attributed to the removal of the filter that generates ν_{eq} (see Equation (2.4)). The use of the filter introduces a trade-off between detection speed and threshold magnitude, which is completely removed using the observer error based detection.



(a) Detection time using the EOI based detector from Section 2.2



(b) Detection time using the observer error based detector from Section 2.3

Figure 2.3: Detection time of a step attack of 2.8 m/s^2 for different measurement noise bounds $\bar{\zeta}_2$. $\bar{\zeta}_1 = 1$ is kept constant. Note the difference in time-scale between the figures.

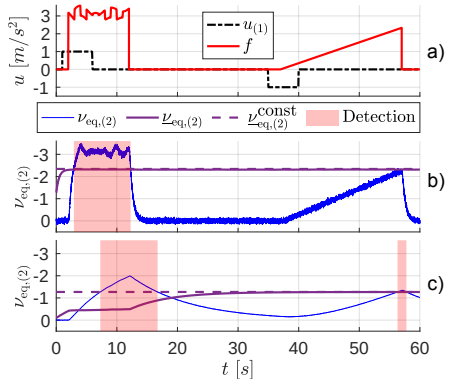


Figure 2.4: a) Input of lead vehicle and cyber-attack. b),c) Detection of the attack by the EOI based detector. Second element of EOI with its lower threshold. Vertical axes are inverted to highlight the estimation capability of the SMO. b) Fast parameters; c) Slow parameters.

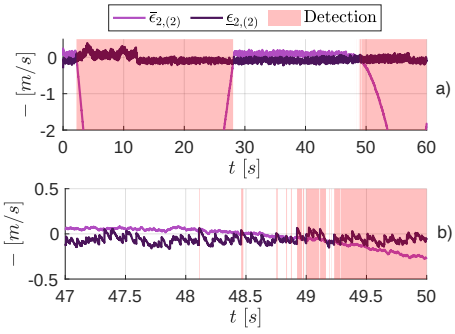


Figure 2.5: Detection of the attack by the observer error based detector. Second element of the observer error bounds. a) whole time range. b) around time of first detection of slope attack.

2.4.2 Simulation Scenario

In this section the anomaly detectors are applied to the scenario shown in Figure 2.4a. The true input of the leader vehicle is shown with the dashed black line and two attacks are introduced on inter-vehicle communication, which are depicted in red. First, at 2 seconds a varying step-like attack is introduced. Secondly, at 37 seconds a ramp attack is introduced.

[7] Note that the detection time scale in Figures 2.3a and 2.3b are different.

For this scenario the EOI based detector has been applied with the *slow* and *fast* parameter sets and the observer error based method only with the *slow* parameter set.[8]

Detection performance of the EOI based detector is shown in Figures 2.4b and 2.4c for the fast and slow parameter set respectively. Here, only $\nu_{\text{eq},(2)}$ with its corresponding thresholds is shown as $\hat{f} = -\nu_{\text{eq},(2)}$ and as such only this element of the EOI is relevant for detection. $\nu_{\text{eq},(2)}$ is depicted in Figures 2.4b and 2.4c by the blue line, the solid purple line is the corresponding lower combined threshold, the dashed purple line is the lower constant combined threshold, and the red areas indicate the cyber-attack is detected by the constant combined threshold.

As shown in Figures 2.4b and 2.4c, the threshold for the slow parameter set is closer to zero than for the fast parameter set. In general, the threshold is lower for lower values of K_2 and K_ν. Therefore, with the slow parameter set smaller cyber-attacks can be detected. This can also be seen in the presented scenario where the ramp-shaped attack is detected at 55.8 s with the slow parameter set but not with the fast parameter set. Conversely, if the attack is sufficiently large, detection with the fast parameter set is faster as illustrated by detection of the first step-like attack. Here the attack is detected at 3 s with the fast parameter set and at 7.1 s with the slow parameter set. In the considered platooning scenario, the step-like attack causes a crash between the vehicles at 6.2 s, meaning only detection with the fast parameter set is sufficiently fast. For the ramp attack a crash occurs at 56.2 s, meaning detection with the slow parameter set at 55.8 s is sufficiently fast. Therefore, both parameter sets need to be used simultaneously to provide sufficiently fast detection for all attacks in this simulation example.

Furthermore, as $\hat{f} = -\nu_{\text{eq},(2)}$, the estimation capability of the SMO can also directly be seen from Figures 2.4b and 2.4c. One can see that especially with the fast parameter set a good, albeit a bit noisy, estimate of the attack is obtained.

Detection performance of the observer error based detector is shown in Figure 2.5 for the scenario from Figure 2.4a. The light and dark purple lines indicate the upper and lower thresholds, $\bar{\epsilon}_{2,(2)}$ and $\underline{\epsilon}_{2,(2)}$. One can see that first detection of both the step-like and the ramp attack occurs faster than with the EOI based detector using either parameter set. In Figure 2.5b a detail is shown around the first detection of the ramp attack. One can see that the first detection is caused by a jump in the lower threshold which is the result of a high value of the measurement noise $\zeta_{2,(2)}$. Thus, between this first detection at 48.1 [s] and consistent detection around 49.3 [s] detection can occur or not depending on the noise distribution. One can see that after 49.3 [s] detection will occur regardless of the noise distribution. Therefore, still detection guarantees can be given and this dependence on the noise is not of concern. Furthermore, it should be noted that for the observer error based detection the detection decision is maintained until long after the attack has stopped. If required, this effect can be mitigated by adding a saturation on the detection thresholds that serves as an anti-windup for the integrator in the threshold design.

2.5 Results of Application to an aircraft under OFC

The aircraft servo loop application introduced in Section 1.1.2 is a nonlinear single-input single-output (SISO) system, which does not conform to the standard form in Equation (2.1)

[8] This has been done as the observer error based detection is not very sensitive to the design parameters.

which is a linear multiple-input multiple-output (MIMO) system with known nonlinearities. The methods presented are however also applicable to a class of nonlinear SISO systems, including the aircraft servo loop. Simulations as well as flight test data have been used to show the performance of the error based detector from Section 2.3 in detecting an oscillatory failure case (OFC).

In Section 2.5.1, first, a short overview of other work on OFC detection is presented. Then, from Section 2.5.2 onward the developed SMO error based detector will be implemented. First, sufficient conditions on nonlinear SISO systems are presented for which the SMO based detectors from Sections 2.2 and 2.3 are applicable without change. Secondly, in Section 2.5.3 it is shown how the observer error based detector is applied to detect the OFC. Then results of a Monte Carlo (MC) simulation show its detection performance in Section 2.5.4 and its robustness is validated using flight test data in Section 2.5.5.

2.5.1 Existing Work on Detection of OFCs

The first research works on OFC detection were conducted in the 90's [12] and describe an oscillatory failure identification system that uses several combinations of linear methods and signal processing techniques where each method is designed for a different fault scenario. Since these first works, the OFC detection problem has gained significant interest and a wide range of approaches have been tested and published. Part of the current industrial state of practice has been published by Airbus in 2010 [13]. Non-linear filtering techniques have been widely used and assessed in the industrial environment [260, 261] and they are now fully part of the most recent industrial state of practice [262] with a certified solution embedded and flying on the Airbus A350 long range aircraft.

Pons et al. applied a learning approach based on interval analysis to OFC identification [263]. Varga and Ossmann [264] developed a linear parameter-varying (LPV) based identification approach for oscillatory failure cases. Due to the oscillatory nature of the fault to detect, differentiator approaches have been successfully applied and tested on real data [265, 266]. Sifi et al. [267] uses an H_∞ observer for OFC detection on new generation Electro-Hydraulic Actuators. Sun et al. [268] proposed a linear time-invariant model based robust fast adaptive fault estimator with unknown input decoupling for oscillatory fault detection. Alwi and Edwards [269] used an adaptive sliding mode differentiator to reconstruct OFC signals for the purpose of detection. More recently, Goupil et al. [270, 271] developed and industrially tested a pure data-driven approach for OFC detection based on similarity index computation.

One way to improve a model-based approach to OFC detection is to enhance the residual evaluation step. Varga and Ossmann [272] used the Narendra criteria [273] as an adaptive way to evaluate the residual using a forgetting factor, as opposed to simply thresholding the residual. Trinh et al. [274] performed a quantitative analysis of a bank of residuals through a correlation test. Lavigne et al. [275] investigated the Wald test by exploiting the different statistical nature of the residual in the fault-free and faulty case.

All of the aforementioned works generally concern a classical hydraulic actuator. Some research has also been performed on OFCs in new generations of actuators such as Electro-Hydraulic Actuators [276]. Oscillatory behavior detection for other kinds of systems can also be found in academic literature. For example, Loutridis [277] investigated damage detection in gear systems using empirical model decomposition.

2.5.2 Applicability Conditions for Nonlinear SISO systems

Consider a nonlinear system of the form

$$\begin{cases} \dot{x} = f(x,u), \\ y = x + \zeta_2, \end{cases} \tag{2.50}$$

where $f(\cdot)$ denotes the true system dynamics, and $x \in \mathbb{R}$, $u \in \mathbb{R}$, $\zeta_2 \in \mathbb{R}$ are the state, input and measurement noise respectively.[9] This can be rewritten as

$$\begin{cases} \dot{x} = \hat{f}_0(\hat{x},u) + \underbrace{(f_0(x,u) - \hat{f}_0(\hat{x},u))}_{\theta} + \underbrace{(f(x,u) - f_0(x,u))}_{\Phi}, \\ y = x + \zeta_2, \end{cases} \tag{2.51}$$

where $\hat{f}(\cdot)$ denotes a known dynamics estimate, θ is the model uncertainty and Φ is the effect of the anomaly. Using an SMO of the form

$$\begin{cases} \dot{\hat{x}} = \hat{f}_0(\hat{x},u) - K_L(y - \hat{y}) + K_2 \text{sgn}(y - \hat{y}), \\ \hat{y} = \hat{x}, \end{cases} \tag{2.52}$$

gives error dynamics

$$\begin{cases} \dot{e} = K_L e + \theta + K_L\zeta_2 + \Phi - K_2 \text{sgn}(e_y), \\ e_y = e + \zeta_2, \end{cases} \tag{2.53}$$

where $e = x - \hat{x}$ and $e_y = y - \hat{y}$. This resembles Equation (2.3) where $e = e_2$ and e_1 does not exist as the problem is scalar. Furthermore, $A_{22} = E_{22} = K_L$, $E_{21}\zeta_1 = \theta$, $N_2 f = \Phi$ and $P = C_2 = F = 1$. From this error dynamics an equivalent of Proposition 2.1 is achieved with $K_L < 0$ and $|\theta|$ bounded by known bound $\bar{\theta}$. Below it will be proven that Propositions 2.2 and 2.3 can always be made to hold.

Proof. **(Proposition 2.2)** Pick $K_2 > \bar{\theta} + |K_L|\bar{\zeta}_2$ such that $\text{sgn}(\dot{e}^0) = -\text{sgn}(e_y^0)$. Then define $V = \frac{e^{0^2}}{2}$ such that

$$\dot{V} = \dot{e}^0 e^0 = K_L e^{0^2} + \theta + K_L\zeta_2 - K_2 \text{sgn}(e^0 + \zeta_2)e^0 \tag{2.54}$$

which can be simplified if $|e^0| > \bar{\zeta}_2$ as

$$\dot{V} < K_L e^{0^2} + \theta + K_L\zeta_2 - (\bar{\theta} + |K_L|\bar{\zeta}_2) \leq K_L e^{0^2} \leq 0 \tag{2.55}$$

such that $V = e^{0^2}$ is a lyapunov function if $|e^0| > \bar{\zeta}_2$. This means that e^0 will converge to a region around the origin $|e^0| \leq \bar{\zeta}_2 = \bar{e}^0$.

Furthermore, bounds on $\dot{e}$ can be derived directly from Equation (2.53) as

$$\begin{aligned} \bar{\dot{e}}^+ &= K_L e_y + 2\bar{\theta} + \Phi + |K_L|\bar{\zeta}_2, \\ \underline{\dot{e}}^+ &= K_L e_y + \Phi + |K_L|\bar{\zeta}_2, \\ \bar{\dot{e}}^- &= K_L e_y - \Phi + |K_L|\bar{\zeta}_2, \\ \underline{\dot{e}}^+ &= K_L e_y + 2\bar{\theta} - \Phi + |K_L|\bar{\zeta}_2, \end{aligned} \tag{2.56}$$

[9] There is research on SMOs for much more general classes of nonlinear systems. It is likely the detectors could be adapted to be used for such systems too, but this has not been investigated.

The existence of known bounds $\bar{\dot{e}}^+$, $\bar{\dot{e}}^-$, $\underline{\dot{e}}^+$, and $\underline{\dot{e}}^+$ implies the existence of true bounds $\tilde{\dot{e}}^+$, $\tilde{\dot{e}}^-$, $\underline{\dot{e}}^+$, and $\underline{\dot{e}}^+$. Thereby proving the proposition. ∎

Proof. **(Proposition 2.3)** From the bounds in Equation (2.56) one can see

$$\begin{aligned}
\bar{\dot{e}}^+ &= \bar{\dot{e}}^{0,+} + \Phi, \\
\underline{\dot{e}}^+ &= \underline{\dot{e}}^{0,+} + \Phi, \\
\bar{\dot{e}}^- &= \bar{\dot{e}}^{0,-} - \Phi, \\
\underline{\dot{e}}^- &= \underline{\dot{e}}^{0,-} - \Phi,
\end{aligned} \tag{2.57}$$

such that $\delta_f(F) = 0$ and $\delta_{\dot{f}^+} = \delta_{\dot{f}^-} = \Phi$. ∎

2.5.3 Application of Observer Error based Detection to the OFC

The servo-loop control system of an actuator of a civil aircraft can be modelled as

$$\begin{aligned}
\dot{p} &= V_c \sqrt{\frac{\Delta P + \frac{\delta k_{aero} \operatorname{sgn}(V_c)}{S}}{\Delta P_{ref} + \frac{k_d V_c^2}{S}}} + \Gamma, \\
V_c &= k_c(K(p_{\text{ref}} - p_{\text{meas}}) + i_f), \\
\delta &= k_p(p), \quad p_{\text{ref}} = k_\delta(\delta_{\text{des}}), \\
\delta_{\text{meas}} &= \delta + \xi_\delta, \quad p_{\text{meas}} = p + \xi_p + p_f,
\end{aligned} \tag{2.58}$$

where p is the servo rod position, δ is the control surface deflection, and V_c is the commanded voltage to the servo. p_{ref} and p_{meas} are the desired and measured servo rod position, which are only used in the internal servo-loop control. δ_{des} and δ_{meas} are the desired and measured control surface deflections, which are, respectively, the only input and output of the system. ξ_p and ξ_δ are the measurement noises for the rod position and control surface deflection sensors, respectively.

Furthermore, ΔP_{ref}, S, and K are typically known parameters and ΔP, k_d, and k_{aero} are unknown parameters. Additionally, Γ represents unmodelled, but bounded, behaviour of the real servo. k_p and k_δ are non-increasing known functions and k_c is a non-decreasing known function, all of which are defined as lookup tables. Lastly, we consider anomalies p_f and i_f that can occur in the servo rod position measurement and the commanded current, respectively. These anomalies can occur in *solid* or *liquid* form. If the anomaly occurs in *solid* form the anomalous oscillatory signal replaces the nominal signal. If it occurs in the *liquid* form oscillatory signal is added to the nominal signal. As such, these anomalies can be modeled as

$$\text{Liquid:} \begin{cases} p_f = p_f^{\text{osc}} \\ i_f = i_f^{\text{osc}} \end{cases}$$

$$\text{Solid:} \begin{cases} p_f = p_{\text{ref}} - p_{\text{meas}} + p_f^{\text{osc}} \\ i_f = -K(p_{\text{ref}} - p_{\text{meas}}) + i_f^{\text{osc}} \end{cases}$$

where p_f^{osc} and i_f^{osc} are sinusoidal with an unknown, but constant frequency and amplitude.

The models presented above are part of an aerospace industrial benchmark on fault detection, dedicated to fault detection in the flight control system of a civil commercial aircraft, which was developed by Airbus and Stellenbosch University. Detection of OFCs within this industrial benchmark was posed as one of three competitions organized in the context of the 2020 IFAC World Congress. Such competitions are organized to enhance Industry participation in IFAC events and to bridge the gap with Academia. Furthermore, it gives the opportunity for participants to compete against other international teams.

Next, we will introduce a few assumptions on this model that will be used in the remainder of this section.

Assumption 2.3. Unknown parameters ΔP, k_d, and k_{aero} can be expressed as the summation of known nominal values, ΔP_N, k_{d_N} and k_{aero_N}, and unknown variations $\Delta\tilde{P}$, $\tilde{k}_d$ and $\tilde{k}_{\text{aero}}$ with known bounds. Furthermore, the umodelled dynamics Γ is bounded as $|\Gamma| < \gamma$. ◁

Assumption 2.4. The sensor noises ξ_p and ξ_δ are zero-mean and can be bounded for all time as $|\xi_p| \leq \bar{\xi}_p$ and $|\xi_\delta| \leq \bar{\xi}_\delta$, respectively. ◁

Assumption 2.5. The faults p_f and i_f can be bounded for all time as $|p_f| \leq \bar{p}_f$ and $|i_f| \leq \bar{i}_f$, respectively. ◁

Following the notation introduced in Section 2.5.2, Equation (2.58) defines $f(p, p_{\text{ref}}) = \dot{p}$. Excluding the effect of the faults gives the nominal dynamics as

$$\begin{aligned} f_0(p, p_{\text{ref}}) = \dot{p}^0 = V_c^0 \sqrt{\frac{\Delta P + \frac{\delta k_{aero}\text{sgn}(V_c^0)}{S}}{\Delta P_{ref} + \frac{k_d V_c^{0\,2}}{S}}} + \Gamma, \\ V_c^0 = k_c(K(p_{\text{ref}} - p_{\text{meas}})). \end{aligned} \tag{2.59}$$

Furthermore, using Assumption 2.3 a known model of the nominal dynamics can be obtained as

$$\begin{aligned} \hat{f}_0(\hat{p}, \delta_{\text{des}}) = \dot{p}_{\text{model}} = \hat{V}_c \sqrt{\frac{\Delta P_N}{\Delta P_{ref} + \frac{k_{d_N} \hat{V}_c^2}{S}}}, \\ \hat{V}_c = k_c(K(k_\delta(\delta_{\text{des}}) - k_p^{-1}(\delta_{\text{meas}}))). \end{aligned} \tag{2.60}$$

Here $k_\delta(\delta_{\text{des}}) = p_{\text{ref}}$ is just a change of notation representing that the inputs of the detector are δ_{des} and δ_{meas}. Based on these equations we can derive the bounds on $|\zeta_2|$ and $|\theta|$ required for implementation of the OFC detection.

Firstly, for the actuator servo loop, the measured signal is δ_{meas}, which is then transformed according to the relations in Equation (2.58) as $y = k_p^{-1}(\delta_{\text{meas}})$ to form an approximation of $x = p$. The relation $y = x + \zeta_2$ then gives

$$\begin{aligned} \zeta_2 = k_p^{-1}(k_p(p) + \xi_\delta) - p, \\ \bar{\zeta}_2 = \max_{p, \xi_\delta} |k_p^{-1}(k_p(p) + \xi_\delta) - p|, \end{aligned} \tag{2.61}$$

where $\delta_{\text{meas}} = k_p(p) + \xi_\delta$ is just a change of notation to make explicit that the bound only depends on p and ξ_δ.

Secondly, the bound on $|\theta|$ can be derived. This process is a bit more elaborate, but in principle finds $\bar{\theta}$ as

$$\bar{\theta} = \max_{\tilde{k}_d, \Delta\tilde{P}, \tilde{k}_{\text{aero}}, \Gamma, \xi_\delta, \xi_p} (|\dot{p}^0 - \dot{p}_{\text{model}}|). \tag{2.62}$$

This can be expanded to

$$\bar{\theta} = \max(|\max_{\tilde{k}_d, \Delta\tilde{P}, \tilde{k}_{\text{aero}}, \Gamma, \xi_\delta, \xi_p} (\dot{p}^0) - \dot{p}_{\text{model}}|, |\min_{\tilde{k}_d, \Delta\tilde{P}, \tilde{k}_{\text{aero}}, \Gamma, \xi_\delta, \xi_p} (\dot{p}^0) - \dot{p}_{\text{model}}|),$$

such that, to obtain $\bar{\theta}$ we only require to derive the bounds on $\dot{p}^0$ over all uncertainties. Based on Equation (2.59) we will first derive bounds on V_c^0 and δ based on the uncertainty in the sensor noise. We will then use these bounds and the bounds on the uncertain model parameters to bound $\dot{p}^0$.

First note that the detector does not know p_{meas}, but needs to derive it from δ_{meas} using the relations in Equation (2.58) as $p_{\text{meas}} = k_p^{-1}(\delta_{\text{meas}} - \xi_\delta) + \xi_p$. Using this relation and the definition of V_c^0 from Equation (2.59) we can write $\underline{V}_c^0 \leq V_c^0 \leq \bar{V}_c^0$ where

$$\begin{aligned} \underline{V}_c^0 &= k_c(K(p_{\text{ref}} - (k_p^{-1}(\delta_{\text{meas}} + \bar{\xi}_\delta) - \bar{\xi}_p))), \\ \bar{V}_c^0 &= k_c(K(p_{\text{ref}} - (k_p^{-1}(\delta_{\text{meas}} - \bar{\xi}_\delta) + \bar{\xi}_p))). \end{aligned} \tag{2.63}$$

Furthermore, δ can be bound as

$$\delta_{\text{meas}} - \bar{\xi}_\delta = \underline{\delta} \leq \delta \leq \bar{\delta} = \delta_{\text{meas}} + \bar{\xi}_\delta. \tag{2.64}$$

Now we will bound $\dot{p}^0$, where we set all instances of V_c^0 appearing in Equation (2.59) independently to achieve the extremes. This results in

$$\begin{cases} \max(\dot{p}^0) = \bar{V}_c^0 \sqrt{\dfrac{\max\limits_{\Delta\tilde{P}}(\Delta P) + \max\limits_{\tilde{k}_{\text{aero}}}(k_{\text{aero}}) \max\limits_{\delta, V_c^0}(\frac{\delta \text{sgn}(V_c^0)}{S})}{\Delta P_{\text{ref}} + \min\limits_{\tilde{k}_d}(k_d) \min\limits_{V_c^0}(\frac{V_c^{0\,2}}{S})}} + \max\limits_{\Gamma}(\Gamma) \text{ if } \bar{V}_c^0 > 0 \\ \max(\dot{p}^0) = \bar{V}_c^0 \sqrt{\dfrac{\min\limits_{\Delta\tilde{P}}(\Delta P) + \min\limits_{\tilde{k}_{\text{aero}}}(k_{\text{aero}}) \min\limits_{\delta, V_c^0}(\frac{\delta \text{sgn}(V_c^0)}{S})}{\Delta P_{\text{ref}} + \max\limits_{\tilde{k}_d}(k_d) \max\limits_{V_c^0}(\frac{V_c^{0\,2}}{S})}} + \max\limits_{\Gamma}(\Gamma) \text{ if } \bar{V}_c^0 \leq 0 \end{cases} \tag{2.65}$$

where $\max\limits_{\delta, V_c^0}(\delta \text{sgn}(V_c^0))$ and $\min\limits_{\delta, V_c^0}(\delta \text{sgn}(V_c^0))$ can be obtained by calculating the expression for all four combinations of the extremes of δ and V_c^0.

An observer of the form Equation (2.52) has been applied to this model together with the observer error based detector from Section 2.3. The results are shown in the following sections.

2.5.4 Monte Carlo Study

In this section, the robustness and detection performance of the observer error based detection scheme from Section 2.3, with the modifications presented above, is demonstrated

through exhaustive Monte Carlo (MC) simulations within the industrial benchmark for OFC detection developed by Airbus and Stellenbosch University. To this end, first the benchmark simulation will be introduced, followed by the results obtained through MC simulations, and analysis thereof.

Industrial OFC Benchmark

The OFC detection scheme was tested in simulation using a Simulink model that was developed by Airbus and Stellenbosch University for the Aerospace Industrial Benchmark on Fault Detection competition at the 21st IFAC World Congress. The Simulink model is a high fidelity model of a control loop for a specific flight phase, namely the cruise phase, and for a given flight point. It allows generating the control surface position corresponding to a given flight path angle or load factor command. It also allows the injection of OFCs originating from the rod position measurement or the current commanded by the flight control computer (FCC), at different frequencies and amplitudes. A top-level diagram of the Simulink model is shown in Figure 2.6.

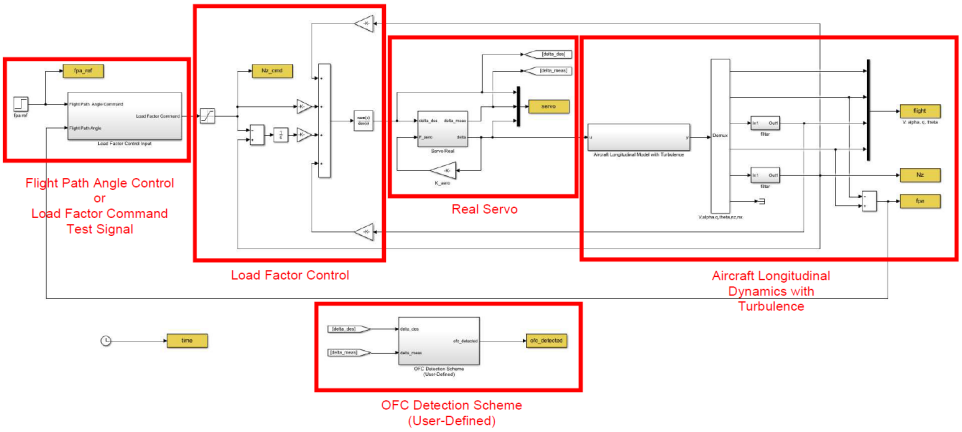


Figure 2.6: Simulink Model for OFC Benchmark.

The Simulink model consists of the following functional blocks: The real servo actuator with the option to inject an OFC, the aircraft longitudinal dynamics with turbulence, the load factor controller, the flight path angle controller, and the OFC detection scheme.

The real servo actuator block implements the model in Equation (2.58) that describes the motion of the servo actuator in response to a control surface deflection command, and also allows an OFC to be injected. The aircraft longitudinal dynamics block models the longitudinal motion of the aircraft in response to the elevator control surface deflection, including the effect of external wind turbulence. The wind turbulence is modelled using a Von Karman Wind Turbulence Model from the Simulink Aerospace Blockset. The load factor controller block controls the load factor of the aircraft by commanding the elevator control surface deflection using the measured load factor and measured pitch rate of the aircraft as feedback. The flight path angle controller block controls the flight path angle of the aircraft to maintain level flight by commanding the load factor using the measured

flight path angle as feedback. The OFC detection scheme block implements the SMO-based approach described in Section 2.3.

Detector Performance Analysis through Monte Carlo Simulation
Extensive MC simulations were performed using this benchmark model using the parameters in Table 2.2. These parameters have been identified based on extensive flight test data from the considered actuator. Furthermore the observer gains have been chosen as in Table 2.3. Simulations have been performed while injecting faults in the commanded current and rod position sensor as described in Section 2.5.3. Detection performance is shown for the full range of considered fault frequencies, amplitudes, and fault types.

Table 2.2: Actuator model parameters

Parameter	Value [Unit]	Parameter	Range [Unit]
ΔP_{ref}	21 $[N/mm^2]$	ΔP	$[15, 29]$ $[N/mm^2]$
K	0.4 $[mA/mm]$	k_d	$[3, 6.2]$ $[N \cdot s^2/mm^2]$
S	5000 $[mm^2]$	k_{aero}	$[435, 975]$ $[N/\text{deg}]$
		Γ	$[-0.07, 0.07] \frac{\mathrm{d}^2 p_{\text{ref}}}{\mathrm{d}t^2}$

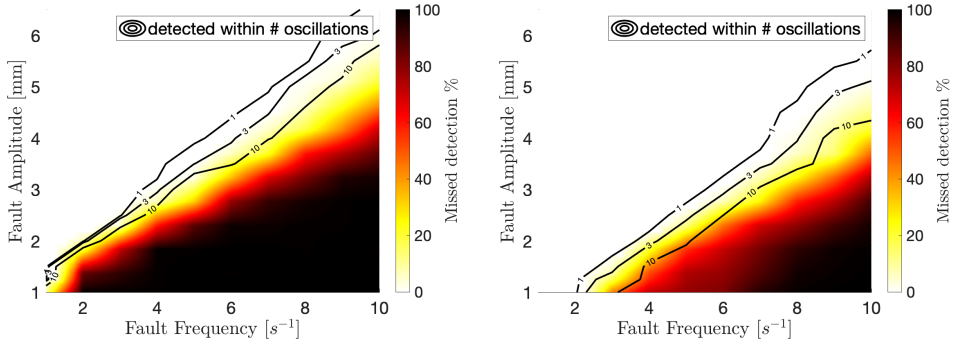
Table 2.3: Observer gains and nominal model parameters

Parameter	Value [Unit]	Parameter	Value [Unit]
K_2	$\tilde{\theta} + \lvert K_L \rvert \tilde{\zeta}_2 + \eta$	ΔP_N	28 $[N/mm^2]$
K_L	-1 $[s^{-1}]$	k_{d_N}	5.5 $[n \cdot s^2/mm^2]$
η	0.1 $[-]$	k_{aero_N}	650 $[N/\text{deg}]$

The results of the Monte Carlo simulations are shown in Figures 2.7 and 2.8 for all considered fault types. These results are obtained by performing 200 simulations for each combination of fault frequency and amplitude, with different uncertainties. For each simulation instance the uncertain parameters ΔP, k_d, k_{aero}, and Γ are drawn from a uniform distribution within the possible set defined in Assumption 2.3. Furthermore, excitation of the system is obtained through p_{ref}, which is calculated by using the load factor control in flight path mode. The controller is tasked with performing a stabilization task under four different turbulence conditions.

Recall that the objective of the detection scheme is to detect on OFC within a specified maximum number of oscillations, while having no false alarms. Therefore in Figures 2.7 and 2.8 the contours show the regions for which detection always occurs within 1, 3, and 10 oscillations. Furthermore, the background colour shows the percentage of *missed detections*, which is defined as no detection within 3 oscillations of the fault. Note that the choice to show results for detection within 1, 3, and 10 oscillations has been made as an example as they span a realistic range of potential detection requirements, but they do not reflect the actual requirement for the considered actuator.

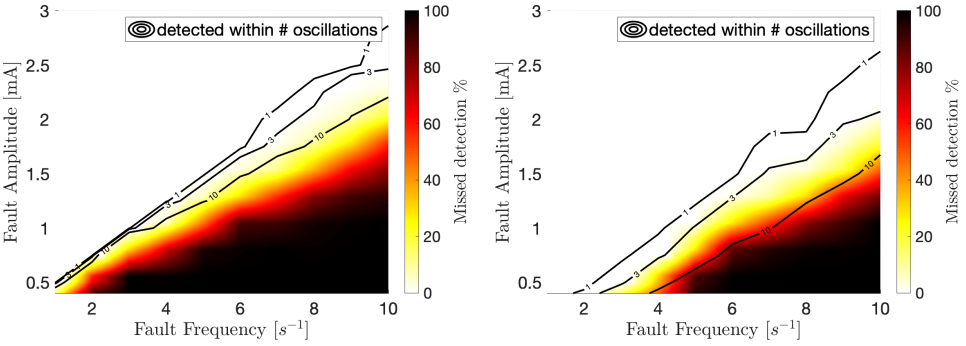
From the contours in Figures 2.7 and 2.8 it can be seen that for all fault types, faults with a sufficiently large amplitude can be consistently detected within any specified maximum number of oscillations. This shows that the detection objective can be achieved for all fault



(a) Liquid fault in rod position sensor

(b) Solid fault in rod position sensor

Figure 2.7: Detection performance for oscillatory faults in the rod position sensor with varying amplitude and frequency



(a) Liquid fault in commanded current

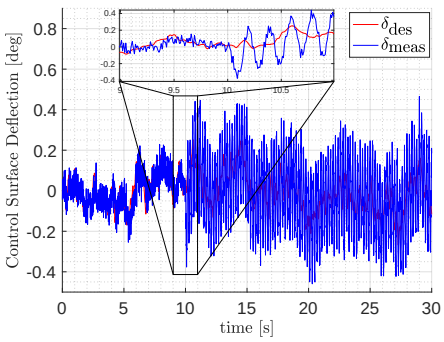
(b) Solid fault in commanded current

Figure 2.8: Detection performance for oscillatory faults in the commanded current with varying amplitude and frequency

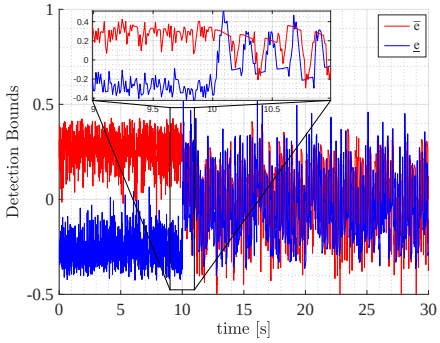
types and frequencies. Furthermore, during the 96,000 simulations performed to obtain the Monte Carlo results, no false alarms were recorded, demonstrating the robustness of the detection scheme.

Furthermore, it can be seen that the fault amplitude required for consistent detection increases approximately linearly with fault frequency. This finding is supported by theory through Theorem 2.5, where it is proven that detection is guaranteed for sufficiently positive (or negative) faults which persist longer than $t_{2j+2N} - t_{2j}$. For the considered zero-mean liquid faults, this means detection guarantees for higher frequency faults demand a larger amplitude. Lastly, it can be seen that the detector consistently shows better detection performance for solid faults than for liquid faults. Unlike liquid faults, solid faults are not zero-mean. Therefore, we can once again invoke Theorem 2.5 to explain the improved detection performance. The nonzero mean of the solid fault will always cause an increase of either the duration for which the oscillatory fault is positive or negative.

To get a feeling for the type of data from which the extensive Monte Carlo results presented above are obtained, Figure 2.9 shows simulation results for a single realisation of the uncertainty. Here, a liquid fault in the commanded current with frequency 5 Hz and amplitude 1.5 mA is introduced at 10 seconds under light turbulence conditions. Figure 2.9a shows the control surface deflection for the performed stabilisation maneuver, and the behaviour of the fault detection bounds $\bar{\epsilon}$ and $\underline{\epsilon}$ is shown in Figure 2.9b. One can see that for this realisation of the uncertainty fault detection occurs well within 0.1 $[s]$.

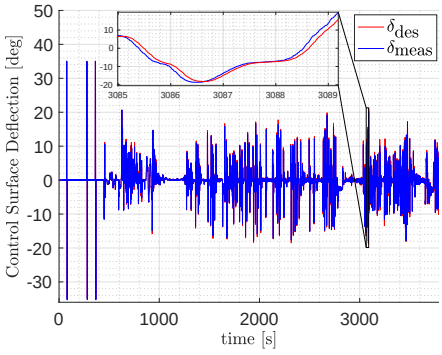


(a) Measured and commanded actuator deflection for a stabilization task under light turbulence.

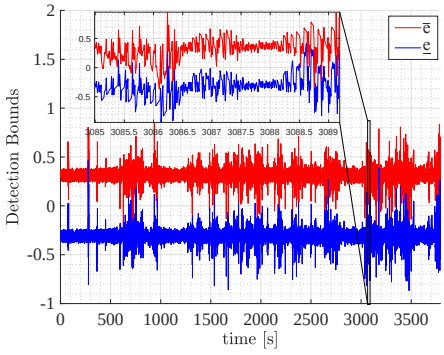


(b) Fault detection bounds $\bar{\epsilon}$ and $\underline{\epsilon}$ leading to detection of the fault within one oscillation.

Figure 2.9: Example of behaviour of the actuator and fault detector under a fault in the commanded current. A fault with frequency of 5 Hz and amplitude of 1.5 mA occurs at 10 seconds.



(a) Maneuvers performed to excite the system



(b) Detection bounds indicating no false alarm

Figure 2.10: Maneuvers and detector response during healthy behaviour of the performed flight test.

2.5.5 Application on Flight-Test Data

To bridge the gap between academic research and real practice, measurements of the actuator behaviour have been obtained from flight tests performed by Airbus. These measurements show the healthy actuator behaviour and are utilized here to validate Theorem 2.4, where it is proven the detector is free of false alarms. The verification of the

Table 2.4: Metrics describing the flight test data and corresponding detection bounds.

Data set	Duration [min]	δ_{meas} [deg]		$\bar{e} - \underline{e}$ [mm]		
		Variance	Range	Range	Median	< 0
1	63	16.6	$[-35.3, 35.1]$	$[0.11, 0.74]$	0.62	No
2	22	10.6	$[-33.5, 24.3]$	$[0.08, 0.74]$	0.62	No
3	24	23.7	$[-30.6, 26.1]$	$[0.07, 0.74]$	0.62	No

robustness is a key step in industrial acceptance of the developed solution. In particular OFCs are very rare events so the assessment of the probability to degrade the so-called *mean time between failure* of flight control equipment is of primary interest. The missed detection rate has been verified through extensive simulations as presented in Section 2.5.4.

The obtained data corresponds to typical in-service sensor measurements of a real commercial aircraft, with the sampling rate of typical flight control computers. The provided data sets include the desired control surface position (the command), generated by the flight control laws, as well as the measured control surface position. Three different data sets have been obtained from flight tests with the same aircraft type. The first data set is of a flight lasting over one hour and starting with flight control checks on the ground, followed by many dynamic maneuvers until the cruise phase. The second data set is a complete but short flight, containing take-off and landing phases, in which steady maneuvers as well as dynamic maneuvers are performed. The third data set covers a highly dynamic flight phase. The evolution of the control surface deflection for data set 1 is shown in Figure 2.10a. The other data sets are quantified in Table 2.4. One can see that all data sets comprise a large range of possible deflections and contain several highly dynamic flight phases. In Figure 2.10b the time-varying bounds on e used for detection are shown when applied to data set 1. An excerpt of a dynamic phase of the flight is highlighted, where it can be seen that the bounds more closely approach each other, but still no false detection occurs. To give some more insight into the evolution of the detection bounds, Table 2.4 presents some properties of $\bar{\epsilon} - \underline{\epsilon}$ also for the other data-sets. One can see that for all data sets $\bar{\epsilon} - \underline{\epsilon}$ comes close to 0, which is as expected considering the dynamic flight phases in each data set. However, no false alarm is triggered in any data set, which validates the robustness of the detection scheme.

Computational Complexity

Computational complexity is important to determine the real-time applicability of the detection scheme on a commercial aircraft. Therefore the number of scalar operations, such as lookup tables, addition, multiplication, and logic operations, for each update of the detection scheme are counted. One update of the detector requires 120 scalar operations, of which 20 are used to update the SMO and the remaining 100 are used to construct the bounds on e and perform detection.

2.6 Discussion & Conclusion

sliding mode observers (SMOs) have been used extensively for anomaly estimation, allowing for exact anomaly estimation under ideal assumptions such as the absence of measurement

noise. In this chapter the anomaly detection problem has been addressed when these SMOs are applied to systems with unmatched uncertainties and measurement noise. To this end two robust detectors are presented which are applicable to a large class of SMOs.

The applicability of the designed detectors can be evaluated based on three propositions relating the structure of the SMO error dynamics, boundedness of the nominal SMO uncertainty, and the influence of the anomaly. Based on this, it can be concluded the threshold is applicable to a large class of SMOs for linear MIMO systems. Furthermore, it has been shown that the detectors are also applicable to a smaller class of SMOs for nonlinear SISO systems. Further research into the applicability of these detectors to larger classes of SMOs for nonlinear systems is promising and should be pursued.

The first SMO based detector uses the so-called equivalent output injection (EOI), which is also traditionally used for anomaly estimation, as a residual for anomaly detection. Robust detection thresholds on this EOI are derived based on the nominal observer error dynamics. The second SMO-based detector directly constructs two thresholds on the observer error. As the true observer error is not available, these thresholds are compared directly for anomaly detection.

Strong guarantees on detectability of anomalies are presented for both methods. Furthermore both methods, by design, guarantee there are no false alarms. The main advantage of the observer error based method over the EOI-based method is that it generally performs detection faster and can consistently detect smaller anomalies. This is caused by the the low-pass filter that generates EOI, which is removed with the observer error based method. Therefore, the trade-off between detection speed and minimal detectable anomaly that comes with this filter does not apply to the error based detector.

Both detectors have been applied to a collaborative vehicle platoon (CVP) for detection of man-in-the-middle (MITM) cyber-attacks on the communication between vehicles. Here it has been shown that in the presented scenario both methods provide detection before safety is lost. Furhtermore, the advantages of the observer error based detector over the EOI based method have been demonstrated.

The error based detector has also been applied to an aircraft servo loop for detection of a so-called oscillatory failure case (OFC). It has been shown through Monte Carlo simulations, on a benchmark developed by Airbus and Stellenbosch University, that any sufficiently large OFC fault can be detected within a predefined number of oscillations of the fault, although the detection performance is better for low frequency faults. Robustness of the detector has been validated on real nominal flight test data, where no false alarms were recorded in almost 2 hours of flight test data.

3

A Topology-Switching Approach to Anomaly Accommodation in CVPs

The wireless communication used by vehicles in collaborative vehicle platoons (CVPs) is vulnerable to cyber-attacks, which threaten their safe operation. To address this issue, in this chapter a safety preserving controller is proposed. The proposed controller is based on topology-switching coalitional model predictive control (MPC), which utilises a reduced unknown input observer (R-UIO) to detect and isolate the cyber-attacks. Attacked communication links are then disabled to accommodate the attack. Furthermore, the MPC controller is designed to be resilient against undetected attacks and the uncertainty derived from disabling communication links. The proposed control method conforms to a relaxed string stability condition and is guaranteed to be safe from crashes. The tracking performance of the proposed topology-switching controller is illustrated on a simulated CVP of four vehicles. It is shown that the proposed topology-switching coalitional controller has better performance than controllers using other communication topologies.

This chapter is based on
Twan Keijzer, Paula Chanfreut, José María Maestre, and Riccardo M.G. Ferrari. Collaborative vehicle platoons with guaranteed safety against cyber-attacks. Transactions on Intelligent Transportation Systems, *under review.*

Inter-vehicle communication is an integral part of collaborative vehicle platoons (CVPs), allowing them to achieve good tracking performance at low inter-vehicle distances. Therefore, this inter-vehicle communication is the topic of extensive research. This has lead to many control approaches using different communication protocols [50, 51], communication topologies, and communication signals. A common choice is to communicate the intended acceleration in a one-directional precedessor-follower topology [8, 27, 52, 96]. But many other topologies have been proposed, such as [28, 37], which consider communication in a coalitional communication topology, forming disjoint coalitions of cooperative agents that are fully connected.

A promising direction in this field is the introduction of topology-switching coalitional control, in which the communication topology is changed on-line to trade-off performance with communication and computation cost [34, 279, 280]. See [37, 281, 282] for examples of its application in irrigation canals, traffic systems, and solar parabolic plants. In this regard, [39, 283–286] deal with CVPs where the communication topology switches due to vehicles joining and leaving the CVP, the possible inter-vehicle communication failures, and the existence of a maximum distance over which vehicles can communicate. These works stress the relevance of flexible controllers able to accommodate these dynamic communication constraints while providing performance and stability guarantees. In particular, by using the results of [27], the work of [39] presents distributed model predictive control (DMPC) for CVPs with switching topologies and guarantees convergence of the predicted terminal states. [286] proposes a switching control law to achieve string stability in heterogeneous CVPs with communication losses, and [287] studies the influence of the communication topology on the stability and scalability of CVPs considering linear feedback controllers. Additionally, the literature includes other control strategies that similarly handle switching communication topologies and/or clustering of local agents outside the field of CVP, such as the *reconfiguration-based DMPC* proposed in [288], the *plug and play* controller in [35, 289], and the *sparsity-promoting* DMPCs in [290, 291].

The communication within CVPs is, however, not only beneficial. The exchange of data in CVPs can also be subject to cyber-attacks, which threatens its safe operation [83, 87, 292]. Therefore, controllers able to mitigate these attacks are required. To this end e.g., [63] uses a combination of state and time delay observers and [293] implements a modified DMPC resilient against denial of service (DoS) attacks. Closely related, [96] designs a controller for CVP robust against faults causing loss of communication. The literature dealing with other attack types such as injection attacks seems more scarce, e.g., [294] deals with various malicious threats and proposes a robust consensus strategy relying on the availability of sufficient uncorrupted communication links. A larger body of work deals with additive faults such as [221], where an integrated fault tolerant control based on a reduced unknown input observer (R-UIO) is presented, and others like [162, 164, 295]. These approaches can in some cases also be employed for robustness against cyber-attacks.

In this chapter, following works as [96], which note that CVPs can also operate safely with less communication, albeit with degraded performance, we use a topology-switching control law to guarantee safety and maintain performance in CVPs under cyber-attacks. In particular, the chosen approach integrates a coalitional model predictive control (MPC) controller for nominal CVP control with an R-UIO based method for cyber-attack detection and a topology-switching law to accommodate the attacks. In the design of this integrated

safety preserving controller, two main contributions are given. Firstly, the topology-switching control framework is adapted to be robust to involuntary topology changes for cyber-attack mitigation. Secondly, the design of an MPC controller with constraints for safety and string stability of a CVP is presented, which is recursively feasible under involuntary topology changes due to cyber-attacks.

The remainder of the chapter is organized as follows. Section 3.1 presents the design requirements for each component of the proposed safety preserving controller. Section 3.2 introduces the design of the R-UIO used for cyber-attack detection. Section 3.3 presents the topology-switching rule and the formulation of the MPC problem to be solved by each vehicle. Section 3.4 provides the theoretical guarantees of safety and string stability for the proposed control scheme. Section 3.5 presents numerical results on a CVP of 4 vehicles following a leader. Finally, Section 3.6 provides conclusions and future research directions.

Notation

$x(n|k)$ denotes the predicted value of variable x at time instant n computed at time instant $k \leq n$. $\text{conv}(\mathcal{X})$ denotes the convex hull of set $\mathcal{X}$. For a set $\mathcal{X}$, $Q_{\mathcal{X}} = [Q_i]_{i\in\mathcal{X}}$ denotes a block diagonal matrix with $|\mathcal{X}|$ blocks Q_i, where $|\mathcal{X}|$ is the cardinality of $\mathcal{X}$.

3.1 Problem Formulation

In this chapter we consider a CVP as described in Section 1.1.1. The considered problem is that of developing a control law, which can provide safety from cyber-attacks on the communication within the CVP. At the basis of this control law is a dynamic communication topology in which the vehicles can assemble into cooperative groups, hereafter referred to as *coalitions*. By terminating affected communication channels once a cyber-attack is detected, this topology-switching can be used to accommodate the attacks. The goal of this chapter is to present an integrated safety preserving controller based on this principle.

To this end, in the remainder of this section, we will present a short introduction to topology-switching coalitional control followed by a model of the attack and the coalitional CVP. Finally, a list of design requirements for the resulting control law will be presented.

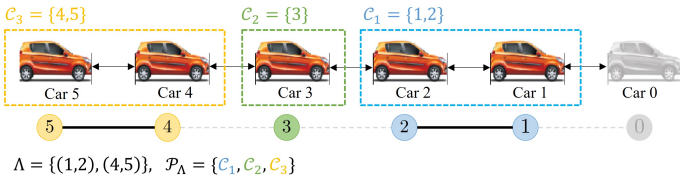


Figure 3.1: Topology and resulting coalitions in an example with 5 vehicles and a leader vehicle which is not part of any coalition.

3.1.1 Topology-Switching communication

Following the *coalitional* control approach of [279, 296], we assume that vehicles are interconnected by a set of wireless communication links that allow each vehicle i to exchange local measurement y_i. These links are considered to be bidirectional, i.e., any pair of connected vehicles can both send and receive information to/from the other. Furthermore,

we also consider multi-hop communication, i.e., vehicles connected by a path of enabled links can exchange data.

Communication links can be dynamically enabled and disabled, leading to different communication topologies. Any communication topology induces a partition of the set of vehicles into coalitions. Considering this, let us introduce the following notation:

- The set $\mathcal{C} \subseteq \mathcal{N}$ denotes a coalition of vehicles, i.e., a group of vehicles that exchange data and coordinate their actions for their joint benefit.
- Λ denotes the topology of the communication network.
- Set $\mathcal{P}_\Lambda$ denotes the partition into coalitions induced by communication topology Λ,

$$\mathcal{P}_\Lambda = \left\{ C_1, C_2, ..., C_{|\mathcal{P}_\Lambda|} \right\}, \tag{3.1}$$

where $\cup_{C_i \in \mathcal{P}_\Lambda} C_i = \mathcal{N}$ and $C_i \cap C_j = \emptyset$, for all $C_i, C_j \in \mathcal{P}_\Lambda$. Note that the number of coalitions in the system will be an integer number ranging from 1, if all the vehicles cooperate, to N, in case the vehicles operate in a decentralized fashion. See Figure 3.1 for an illustration of these concepts.

3.1.2 Unreliable data exchange

The communication of local measurements between vehicles in a coalition may be affected by man-in-the-middle (MITM) cyber-attacks. The performed attacks are considered additive and independent for each communication channel. To simplify notation let Λ be the chosen communication topology and $\mathcal{C} \in \mathcal{P}_\Lambda$ any of the resulting coalitions. Then, at each time instant k, vehicle $i \in \mathcal{C}$ receives the signals

$$y_j^i(k) = y_j(k) + a_{y_j}^i(k)$$

from each vehicle $j \in \mathcal{C} \setminus \{i\}$. Here $a_{y_j}^i(k)$ is the attack on the measurement vector sent from vehicle j to vehicle i. Note that nominally $a_{y_j}^i(k) = 0$.

Remark 3.1. The additive formulation of the attack does not lack any generality, as any received measurement can be generated using this formulation. Furthermore, the attack definition used is also applicable if the attacks stem from other attack vectors such as a malicious agent. ◁

Remark 3.2. Vehicles in different coalitions cannot attack each other because there is no inter-coalition communication. Therefore, working in a decentralized manner, i.e., when all coalitions are singletons, avoids the possibility of being attacked. A decentralized platoon, however, has lower performance due to the lack of coordination. ◁

3.1.3 Coalition model

In this chapter, let us consider a CVP formed by a set $\mathcal{N} = \{1, ..., N\}$ of locally controlled vehicles (see Figure 3.1). This CVP consists of vehicles that can be modeled as

$$\begin{cases} \dot{p}_i = v_i \\ \dot{v}_i = a_i \\ \dot{a}_i = \dfrac{1}{\tau_i}(u_i - a_i) \end{cases} \tag{3.2}$$

where $i \in \mathcal{N}$ denotes the vehicle number, p_i, v_i and a_i denote its position, velocity and acceleration, u_i is the applied control input, and τ_i is the engine time constant.

Assumption 3.1. The input of each vehicle $i \in \mathcal{N}$ is constrained by $u_{\min} \leq u_i \leq u_{\max}$. ◁

Each vehicle aims to keep a reference distance, $d_{r,i}$ from its predecessor, defined as

$$d_{r,i} \triangleq r + hv_i, \tag{3.3}$$

where h denotes the reference time headway and r is the reference distance at standstill. Additionally, CVPs require a form of string stability [297], which ensures that disturbances that occur in the CVP do not grow further down the CVP. Here we define two variants of string stability in time-domain as

Definition 3.1 (Strict string stability)**.** A vehicle CVP is strictly string stable if for a given t_0 if

$$\left| \frac{v_i(t_1) - v_i(t_0)}{v_{i-1}(t_1) - v_{i-1}(t_0)} \right| < 1 \; \forall t_1 > t_0, i.$$

◁

Definition 3.2 (Relaxed string stability)**.** A vehicle CVP is relaxed string stable if for a given t_0 $d_i(t) > 0 \; \forall \; i, t$ and

$$\exists \, j, l < j \text{ s.t. } \left| \frac{v_j(t_1) - v_j(t_0)}{v_l(t_1) - v_l(t_0)} \right| < 1 \; \forall t_1 > t_0 \, .$$

◁

Remark 3.3. Strict string stability assures that the impact of disturbances decreases between any two vehicles moving further away from the source of the disturbance. Relaxed string stability allows for bounded violations of strict string stability between any two vehicles, as long as after some number of vehicles the string stability property is regained. ◁

To achieve these goals each vehicle measures its own velocity and acceleration v_i and a_i, and the distance and relative velocity to its predecessor $d_i = p_i - p_{i-1}$ and $\Delta v_i = v_i - v_{i-1}$. Furthermore, each vehicle communicates measurements. The system in Equation (3.2) can, to this end, be written in discrete time space form as

$$\begin{cases} x_i(k+1) &= A_{i,i} x_i(k) + B_{i,i} u_i(k) + A_{i,i-1} x_{i-1}(k), \\ y_i(k) &= x_i(k), \end{cases} \tag{3.4}$$

with $x_i = [e_{\mathrm{d},i} \; d_i \; v_i \; a_i \; \Delta v_i]^\top$ and $A_{i,i}$, $B_{i,i}$ and $A_{i,i-1}$ are derived from the discretization of Equation (3.2) with sample time T. Note that, given Equation (3.4), at any instant k, the state of any vehicle i in the CVP is only coupled with that of its predecessor. The model of a vehicle i in coalition $\mathcal{C}$[1] then is

$$\begin{cases} x_C(k+1) = A_C x_C(k) + B_C u_C^i(k) + w_C(k), \\ \quad w_C(k) = A_C^w x_{p_C}(k), \\ x_{p_C}(k+1) = A_{p_C} x_{p_C}(k) + B_{p_C} u_{p_C}(k), \\ \quad y_C^i(k) = C_C x_C(k) + C_{\mathrm{a}}^i a_{y_C}^i(k), \end{cases} \tag{3.5}$$

[1] For the sake of clarity, hereafter we use $\mathcal{C}$ to refer to a coalition in general, but note that there may be a number of different coalitions in the system simultaneously as indicated in Equation (3.1)

where $x_C = [x_j]_{j\in C} \in \mathbb{R}^{|C|n}$ is the aggregation of the states of all vehicles in C, $u_C^i = [u_j^i]_{j\in C} \in \mathbb{R}^{|C|m}$ and $y_C^i = [y_j^i]_{j\in C} \in \mathbb{R}^{|C|p}$ are respectively the coalitional input and output as known by vehicle $i \in C$, $a_{y_C}^i = [a_{y_j}^i]_{j\in C} \in \mathbb{R}^{|C|p}$ are the aggregated attacks on vehicle i from all vehicles in the coalition and w_C represents the coupling of coalition C with its predecessor vehicle p_C. Here p_C is defined as $p_C = \min_{j\in C}(j-1)$. For example, in Figure 3.1, the predecessor of coalition 3, formed by vehicles 4 and 5, is vehicle 3, i.e. $p_{C_3} = 3$. Futhermore, matrices A_C, A_{p_C}, A_C^w, B_C, B_{p_C}, and C_C are built according to the model in Equation (3.4), and C_a^i is a matrix that maps the attacks in $a_{y_C}^i$ into the corresponding components of y_C^i.

3

As shown in Figure 3.1, the overall system can be seen as a sequence of cooperative substrings which respectively follow a vehicle whose actions are uncertain, yet bounded as $u_{\min} \le u_{p_C} \le u_{\max}$ (recall Assumption 3.1). Furthermore, due to the possibility of cyber-attacks, uncertainty also exists in the data communicated among vehicles. As all detected cyber-attacks are accomodated, also this uncertainty is bounded. Specifically, using the cyber-attack detector and topology switching rule, as in Sections 3.2 and 3.3.1, the effect of the attack can be bounded in a convex set $\mathcal{A}_C$, which will be defined later.

Remark 3.4. Since the cyber-attack $a_{y_C}^i(k)$ affects the measurement vector y_C^i, it can affect the computation of $u_C^i(k)$, i.e., the vehicle behaviour. ◁

3.1.4 Controller Design Requirements

In this chapter a safety preserving control scheme is proposed which comprises three main components, as shown in Figure 3.2. The requirements for each component are defined below.

1. The local cyber-attack detector in each vehicle should robustly detect cyber-attacks on the communication between all vehicles within its coalition. Furthermore, guarantees on detectability of the attacks should be available.
2. The topology-switching law should form coalitions such that
   (a) all active communication links provide a significant performance improvement.
   (b) coalitions with active communication links are not affected by detected cyber-attacks.
   (c) the CVP is sufficiently connected to be relaxed string stable.
3. The coalitional controller should generate control input which
   (a) provides optimal reference tracking control robust against all undetected attacks and uncertain actions of preceding vehicle p_C.
   (b) avoids crashes for all vehicles within the CVP, even when communication links are attacked, i.e., $d_i(k) > 0,\ \forall i \in \mathcal{N},\ k \ge 0$.
   (c) guarantees that in healthy conditions, strict string stability holds within each coalition.

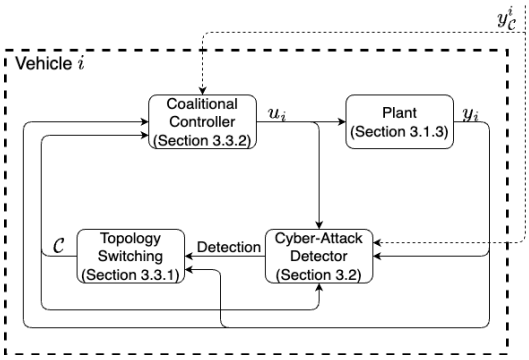


Figure 3.2: Block diagram of the control solution used in each vehicle. The dotted arrows indicate signals communicated from other vehicles in the coalition.

3.2 Cyber-Attack Detector Design

Each vehicle in the CVP is required to implement a detector which allows for robust detection of cyber-attacks. It has been chosen to use a cyber-attack detection method based on the R-UIO from [221] for this purpose. An R-UIO uses the fact that the state is measured to reduce the dimension of the problem and only estimate the unknown input, i.e. the cyber-attack. This is convenient for the topology-switching control as this allows for low computational complexity, even when applied to larger coalitions. Furthermore, the R-UIO design can be easily automated for different vehicles and communication topologies.

A discretized version of this R-UIO is presented in this section along with a detection threshold and guarantees on its performance. As can be seen in Figure 3.2, when an attack is detected, a signal is sent to the topology switching module, which will then disable the corresponding communication link.

To aid the implementation of the R-UIO, the system in Equation (3.5) can be augmented by aggregating the state and attack as

$$\begin{cases} \bar{x}_i(k+1) = \bar{A}_C \bar{x}_i(k) + \bar{B}_C u_C^i(k) + \bar{d}_i(k), \\ y_C^i(k) = \bar{C}_i \bar{x}_i(k), \end{cases} \tag{3.6}$$

where

$$\bar{x}_i(k) = \begin{bmatrix} x_C(k) \\ a_{y_C}^i(k) \end{bmatrix}, \ \bar{d}_i(k) = \begin{bmatrix} w_C(k) \\ a_{y_C}^i(k+1) \end{bmatrix}, \ \bar{B}_C = \begin{bmatrix} B_C \\ 0 \end{bmatrix}, \quad \bar{A}_C = \begin{bmatrix} A_C & 0 \\ 0 & 0 \end{bmatrix}, \ \bar{C}_i = \begin{bmatrix} C_C & C_{\mathrm{a}}^i \end{bmatrix}.$$

Here, we define the sets $\mathcal{D}_i = \{\bar{d}_i | w_C \in \mathcal{W}_C, \Delta a_{y_C}^{i,\mathrm{ind}} \in \mathcal{A}_C\}$ and $\mathcal{D}_i^0 = \{\bar{d}_i | w_C \in \mathcal{W}_C, \Delta a_{y_C}^i = 0\}$ as the possible disturbances $\bar{d}_i$ in, respectively, attacked and healthy conditions.

For a system of the form in Equation (3.6), the R-UIO can estimate the attack signals $a_{y_C}^i(k) = L\bar{x}_i(k)$ with estimator $\hat{\alpha}_i$ as

$$\begin{cases} \hat{x}_{\alpha_i}(k+1) = M_i \hat{x}_{\alpha_i}(k) + G_i u_C^i(k) + R_i y_C^i(k), \\ \hat{\alpha}_i(k) = \hat{x}_{\alpha_i}(k) + H_i y_C^i(k), \end{cases} \tag{3.7}$$

where

$$\begin{aligned} \|M_i\| &< 1\,, \\ M_i T_i + R_i \bar{C}_i - T_i \bar{A}_C &= 0\,, \\ T_i + H_i \bar{C}_i - L &= 0\,, \\ G_i - T_i \bar{B}_C &= 0\,. \end{aligned} \tag{3.8}$$

Here, the matrices in Equation (3.7) are designed based on Equation (3.6) to satisfy the constraints in Equation (3.8). A procedure for this design is presented in [221]. To achieve stability of the dicrete-time observer, only the condition on M_i is changed with respect to [221]. Furthermore, T_i appears in the full derivations of the R-UIO as a transformation matrix, and can be freely designed such that the conditions in Equation (3.8) hold. Then, if Equation (3.8) holds, the observer error dynamics are reduced to

$$e_i(k+1) = M_i e_i(k) + (H_i \bar{C}_i - L)\bar{d}_i(k), \tag{3.9}$$

where $e_i = \hat{\alpha}_i - L\bar{x}_i(k)$. Let us now define a threshold T_α for attack detection as

$$\begin{aligned} T_\alpha(k) &\triangleq \|(I - M_i)^{-1}\| D_i^0, \\ D_i^0 &= \max_{\bar{d}_i \in D_i^0} \|(H_i \bar{C}_i - L)\bar{d}_i\|, \end{aligned}$$

such that detection is triggered when

$$\|\hat{\alpha}_i(k)\| > T_\alpha(k)\,. \tag{3.10}$$

Then the following robustness and detectability results can be presented, which prove the detector conforms to the requirements of Section 3.1.4.

Lemma 3.1 (Robustness)**.** *The threshold T_α is robust to uncertainties and does not lead to false detection.*

Proof. First, under healthy conditions Equation (3.9) can be simplified as

$$\hat{\alpha}_i(k+1) = M_i \hat{\alpha}_i(k) + (H_i \bar{C}_i - L)\bar{d}_i(k)\,.$$

Now, if we initialize $\hat{\alpha}_i(0) = 0$, $\hat{\alpha}_i(k)$ can be written as

$$\hat{\alpha}_i(k) = \sum_{j=0}^{k-1} M_i^j (H_i \bar{C}_i - L)\bar{d}_i(j) \tag{3.11}$$

so that

$$\|\hat{\alpha}_i(k)\| \le \sum_{j=0}^{k-1} \|M_i^j\| D_i^0 \le \|(I - M_i)^{-1}\| D_i^0 = T_\alpha(k)\,,$$

which concludes the proof. ∎

To analyze detectability under attacks, let us define

$$D_i \triangleq \max_{\bar{d}_i \in D_i} \|(H_i \bar{C}_i - L)\bar{d}_i\|\,. \tag{3.12}$$

Theorem 3.1 (Detectability)**.** *A sufficient condition for attack detection by the R-UIO is* $a^i_{yc}(k) \notin \mathcal{A}_C$*, with*

$$\mathcal{A}_C \triangleq \{ a^i_{yc} \,:\, \|a^i_{yc}\| \leq \|(I - M_i)^{-1}\|(D_i + D^0_i)\}$$

being the set of attacks not guaranteed to be detected.

Proof. Using Equation (3.9), the disturbance bound in Equation (3.12), and assuming the R-UIO is initialized in healthy conditions, i.e., $e_i(0) = 0$, we can derive

$$\|e_i(k)\| \leq \|(I - M_i)^{-1}\|D_i\,.$$

This implies

$$\|\hat{\alpha}_i(k)\| \geq \|a^i_{yc}(k)\| - \|(I - M_i)^{-1}\|D_i.$$

Using this in combination with the detection condition from Equation (3.10), detection is guaranteed if

$$\|a^i_{yc}(k)\| - \|(I - M_i)^{-1}\|D_i > T_\alpha(k),$$

which, by definition of T_α, proves the theorem. ∎

Corollary 3.1. *The set of attacks for which detection is not guaranteed can be over-bounded by a convex polytope* $\bar{\mathcal{A}}_C \triangleq conv(\{\alpha_0, \ldots, \alpha_\eta\}) \supset \mathcal{A}_C$*, where all* $\alpha_\bullet \in \mathbb{R}^{|C|p}$*.*

3.3 Topology Switching Controller

In this section, the coalitional controller is presented together with the topology switching rule. It has been chosen to implement MPC as the MPC framework allows for handling of requirements in the form of constraints and handling of uncertainty in the form of scenario based MPC. Additionally, MPC has already been implemented frequently with switching communication topologies. In the remainder of this section, first, the topology switching rule will be defined. Then, the MPC problem to be solved by each vehicle is formulated.

3.3.1 Topology Switching Rule

The topology switching rule determines the communication topology, i.e. which vehicles form coalitions at each time instant. During normal operation its main goal is to balance tracking performance and coordination efforts. In particular, during normal operation the following switching-rule is employed: a communication link between vehicles $i \in \mathcal{C}$ and $i-1 \notin C$ is established if $|\Delta v_i| > T_\mathrm{v}$ or $|e_{\mathrm{d},i}| > T_\mathrm{d}$. Here T_v and T_d are design parameters.

However, when the communication is subject to cyber-attacks, the main goal shifts to preserving safety. Communication links on which a cyber-attack is performed are dangerous to maintain and, therefore, they are disabled upon detection of a cyber-attack. By doing so, Theorem 3.1 implies that only attacks $a^i_{yc} \in \mathcal{A}_C$ can affect the CVP. These ideas are formalized for a CVP of N vehicles in Algorithm 1.

3.3.2 MPC Problem

As can be seen in Equation (3.5), vehicles within each coalition are affected by uncertainty through the actions of the vehicle preceding the coalition and undetected cyber-attacks on the communicated signals. In particular, attacks in set $\mathcal{A}_C$, as defined in Theorem 3.1,

Algorithm 1 Topology Switching Rule

Initialize: $\mathcal{C}_1 = \{1\}$ and $j = 1$.
for all vehicles $i = 2 \dots N$ **do**
  if $(|\Delta v_i| > T_\mathrm{v}$ **or** $|e_{\mathrm{d},i}| > T_\mathrm{d})$ **and** $\|\hat{\alpha}_i(k)\| \leq T_\alpha(k)$
    Set $\mathcal{C}_j = \{\mathcal{C}_j, i\}$.
  else
    Set $j = j + 1$ and $\mathcal{C}_j = \{i\}$.
  end if
end for

may not be detected. Resilience to these uncertainties is guaranteed using scenario-based MPC. A detected attack, on the other hand, will lead to direct accommodation of the attack though these topology switching rule, which will disable the communication link.

Scenario-based approaches consider a set of realizations of the uncertainties affecting the system. The MPC problem is formulated so that the implemented inputs satisfy the system constraints in these scenarios, while optimizing an objective function that typically weights the performance costs in all of these scenarios. Although the scenario-based approach usually provides stochastic guarantees on constraints satisfaction, here the extreme realizations of the vehicles' behaviour and undetected attacks are considered, such that safety guarantees in all cases can be obtained.

In the remainder of this section, first the considered uncertainty scenarios are presented. Then an *ideal* MPC problem is introduced that serves to, in simple terms, show the goal of the MPC problem. As this *ideal* MPC problem cannot be solved in real-time, the *practical* MPC problem is introduced, which is an adapted version of the *ideal* MPC problem that can be solved.

Uncertainty Scenarios

At each time instant k, each vehicle $i \in \mathcal{C}$ considers a set of S realizations of the uncertainty. In particular, each scenario $s \in \mathcal{S} = \{1, ..., S\}$ defines a possible trajectory of the coalitions' predecessor input, i.e.,

$$\hat{\mathbf{u}}_{p_\mathcal{C},s} = [\hat{u}_{p_\mathcal{C},s}(k|k), ..., \hat{u}_{p_\mathcal{C},s}(k + N_\mathrm{p} - 1|k)],$$

where N_p is the length of the prediction horizon, and possible undetected attacks on the measurement vector, $\hat{a}^i_{y_\mathcal{C},s}$. As used above, in what follows, let subscript s indicate the scenarios, e.g., $x_{\mathcal{C},s}(n|k)$ will denote the prediction made at time instant k for the state of coalition $\mathcal{C}$ at time instant n in scenario s.

The set of scenarios $\mathcal{S}$ comprises of three different classes of scenarios. First, define $\mathcal{S}_\mathrm{e} \subset \mathcal{S}$ as the subset of *extreme* scenarios, which imply that the predecessor input and the undetected cyber-attacks take their extreme values, i.e.,

$$\begin{aligned}
\mathcal{S}_\mathrm{e} = \{ s \in \mathcal{S} \ : \ \hat{u}_{p_\mathcal{C},s}(n|k) \in & \begin{cases} \{u_\mathrm{min}, u_\mathrm{max}\} & \text{if } v_{p_\mathcal{C}}(n|k) \in [0, v_\mathrm{max}], \\ 0 & \text{otherwise,} \end{cases} \\
& \hat{a}^i_{y_\mathcal{C},s}(k|k) \in \{a_0, \dots, a_\eta\}, \\
& n = k, ..., k + N_\mathrm{p} - 1 \}.
\end{aligned}$$

These scenarios are used to guarantee safety of the CVP, even when the communication is attacked. Secondly, define the set of *healthy extreme* scenarios, $\mathcal{S}_0 \subset \mathcal{S}$, to involve the extreme inputs, while there is no cyber-attack. That is,

$$\mathcal{S}_0 = \{ s \in \mathcal{S} \ : \ \hat{u}_{p_C,s}(n|k) \in \begin{cases} \{u_{\min}, u_{\max}\} & \text{if } v_{p_C}(n|k) \in [0, v_{\max}], \\ 0 & \text{otherwise,} \end{cases}$$
$$\hat{a}^i_{y_C,s}(k|k) = 0,$$
$$n = k, ..., k + N_{\mathrm{p}} - 1 \Big\}.$$

These healthy extreme scenarios are used to provide string stability in healthy conditions. Lastly, a set of *design* scenarios, $\mathcal{S}_{\mathrm{d}}$, can be chosen freely to include other hypotheses on the possible scenarios, i.e.

$$\mathcal{S}_{\mathrm{d}} = \{ s \in \mathcal{S} \ : \ u_{\min} \le \hat{u}_{p_C,s}(n|k) \le u_{\max},$$
$$\hat{a}^i_{y_C,s}(k|k) \in \bar{\mathcal{A}}_C,$$
$$n = k, ..., k + N_{\mathrm{p}} - 1 \Big\}.$$

These design scenarios will be used to weigh the cost function. The user can add any finite number of *design* scenarios at the expense of an increase in computational burden. Finally, $\mathcal{S}$ is defined as $\mathcal{S} = \mathcal{S}_{\mathrm{e}} \cup \mathcal{S}_0 \cup \mathcal{S}_{\mathrm{d}}$.

Ideal MPC Problem

The ideal MPC problem satisfies the safety and string-stability conditions that are required, but, as will be shown, it cannot be used directly for real-time control. Modifications to make this possible will lead to the *practical* MPC problem of Section 3.3.2. The ideal MPC problem can be formulated as

$$\min_{\mathbf{u}^i_C} \ J_C(y^i_C(k), \Delta \mathbf{u}^i_C) \tag{3.13}$$

subject to:

Prediction model

$$x_{C,s}(k|k) = y^i_C(k) - C^i_a \hat{a}^i_{y_C,s}(k|k), \tag{3.14a}$$
$$x_{C,s}(n+1|k) = A_C x_{C,s}(n|k) + B_C u^i_C(n|k) + w_{C,s}(n|k), \tag{3.14b}$$
$$x_{p_C,s}(k|k) = x_{p_C,s}(k|k-1), \tag{3.14c}$$
$$x_{p_C,s}(n+1|k) = A_{p_C} x_{p_C,s}(n|k) + B_{p_C} \hat{u}_{p_C,s}(n|k), \tag{3.14d}$$
$$w_{C,s}(n|k) = A^w_C x_{p_C,s}(n|k), \tag{3.14e}$$
$$u^i_C(n|k) \in [u_{\min} \ \ u_{\max}]^{|C|}, \tag{3.14f}$$
$$\forall s \in \mathcal{S},$$

Safety

$$d_{j,s}(n|k) \ge 0 \ \forall j \in \mathcal{C}, \forall s \in \mathcal{S}_{\mathrm{e}}, \tag{3.15}$$

String stability

$$\mathrm{sgn}(\Delta v_{j,s}(n|k)) = \mathrm{sgn}(d v_{j,s}(k|k)) \ \forall j \in \mathcal{C}_{\setminus p_C+1}, \forall s \in \mathcal{S}_0, \tag{3.16}$$
$$\forall n = k, \dots, k + N_{\mathrm{p}} - 1,$$

where cost function $J_C(y_C^i(k), \mathbf{u}_C^i)$ is of the form

$$J_C(y_C^i(k), \Delta\mathbf{u}_C^i) = \sum_{n=k}^{k+N_\mathrm{p}-1} \left(\sum_{s\in S_\mathrm{d}} p_s x_{C,s}(n+1|k)^\mathrm{T} Q_C x_{C,s}(n+1|k) \; + \Delta u_C^i(n|k)^\mathrm{T} R_C \Delta u_C^i(n|k) \right).$$

Here $Q_C = [Q_i]_{i\in C}$ and $R_C = [R_i]_{i\in C}$ are positive definite weighting matrices, and $p_s > 0$ represents the probability assigned to scenario s. Furthermore, $\mathbf{u}_C^i$ is the sequence vector $\mathbf{u}_C^i = [{u_C^i}^\top(k|k), \ldots, {u_C^i}^\top(k+N_\mathrm{p}-1|k)]^\top$, and $\Delta u_C^i(n|k)$ is defined as $\Delta u_C^i(n|k) = u_C^i(n|k) - u_C^i(n-1|k)$.[2] Finally, $dv_{i,s}(k|k) = v_{i,s}(k+N_\mathrm{p}|k) - v_{i,s}(k|k)$ denotes the predicted change of velocity of vehicle i over the prediction horizon. Note that unlike a min-max approach, here the deterministic worst case scenarios $\mathcal{S}_\mathrm{e}$ and $\mathcal{S}_0$ are used to guarantee safety using constraint satisfaction, but the minimization is performed based on the design scenarios $\mathcal{S}_d$.

In this ideal MPC problem, Constraints from Equations (3.14a) to (3.14f) predict the coalition behaviour over the prediction horizon for all considered scenarios. Note that in Equation (3.14c), if the topology changes between time steps $k-1$ and k, subscript p_C denotes the vehicle preceding the new coalition at time step k. Furthermore, the constraints in Equations (3.15) and (3.16) ensure safety in all conditions and string stability in all healthy conditions, respectively. This is proven in the following lemma

Lemma 3.2. *For a fixed communication topology, if the constraints in Equations* (3.15) *and* (3.16) *hold*

- *No crashes occur in the CVP, even when the system is under attack.*
- *Strict string stability is guaranteed within each coalition for the healthy system.*

Proof. If the constraint in Equation (3.15) holds, then for all extreme scenarios it holds $d_{i,s} > 0$. This implies there are no crashes for all possible uncertainties, including those from undetected cyber-attacks.

The constraint in Equation (3.16) guarantees strict string stability according to Definition 3.1 within each coalition for the healthy system. This can easily be derived by noting that, starting from $\Delta v_j(k) = 0$, the constraint in Equation (3.16) implies that if $dv_{j,s}(k|k) > 0$, then $v_{j-1,s}(n|k) > v_{j,s}(n|k)$ and thus $dv_{j-1,s}(k|k) > dv_{j,s}(k|k)$. Conversely if $dv_{j,s}(k|k) < 0$ then $dv_{j-1,s}(k|k) < dv_{j,s}(k|k)$. ∎

Unfortunately, the ideal MPC problem in Equation (3.13) cannot be readily implemented to find the vehicles' inputs in real-time. Firstly, the string stability constraint in Equation (3.16) is non-linear, complicating the solution of the ideal MPC problem in real-time. Secondly, both the safety and string stability constraints in constraint in Equations (3.15) and (3.16) are not recursively feasible for all scenarios. For these reasons, a *modification* of the ideal MPC problem in Equation (3.13) is proposed, which, at the expense of a certain loss of optimality, results in a recursively feasible quadratic optimization with linear constraints.

Practical MPC Problem

To obtain a recursively feasible quadratic MPC problem with linear constraints, the safety and string stability constraints need to be reformulated. This will be done in the following.

[2] Note that to obtain $\Delta u_C^i(k|k)$, it is considered that $u_C^i(k-1|k) = u_C^i(k-1|k-1)$.

Safety constraints To make the ideal safety constraint in Equation (3.15) recursively feasible, it needs to be extended so that a feasible solution exists in all scenarios, including emergency braking of the car preceding the coalition and any undetected attack on the communication. To achieve the required robustness to uncertainty, the distance between vehicles is bounded based on the relative velocity and acceleration between vehicles. By doing so, the vehicles' speed of approach to its predecessor becomes more limited at smaller inter-vehicle distances. The exact relation of the practical safety constraint is derived such that the inter-vehicle distance is always safe, as

$$d_{j,s}(m|k) \geq 0, \tag{3.17a}$$

$$d_{j,s}(m|k) \geq -\Delta v_{j,s}(m|k)\delta(m|k), \tag{3.17b}$$

$$d_{j,s}(m|k) \geq -\left(\Delta v_{j,s}(m|k) + \tau \Delta a_{j,s}(m|k)\right)\delta(m|k), \tag{3.17c}$$

for all scenarios $s \in \mathcal{S}_e$, for all $j \in \mathcal{C}$, and for $m = k+1$. Here, $\Delta a_{j,s} = a_{j-1,s} - a_{j,s}$ and

$$\delta(m|k) = \gamma(k|k) - (m-k)T,$$

with $\gamma(k|k)$ the time to standstill as defined below.

Definition 3.3. $\gamma(k|k)$ is an upper bound on the time to standstill of vehicle j, when $u_j(\kappa) = u_{\min}\ \forall \kappa \geq k$. ◁

Remark 3.5. $\gamma(k|k)$ can be implicitly calculated through the model in Equation (3.4) given initial conditions $v_j(k)$ and $a_j(k)$. ◁

The constraints in Equation (3.17) are all based on the idea that

$$d_j(n) > d_j(k) + \min_{k \leq \kappa < n} (\Delta v_j(\kappa))\gamma(k|k) > 0\,,$$

i.e. the change in distance can be bounded by a product of bounds on the relative velocity and the time to standstill. This relation is expanded for three situations. In boundary case $d_{j,s}(n|k) = 0$, when the constraint in Equation (3.17a) is active, the relative velocity can only be positive. In the other cases sufficient distance must be held to guarantee recursive feasibility. The constraint in Equation (3.17b) is active only if the acceleration is positive, and the constraint in Equation (3.17c) is active only if both the relative velocity and acceleration are negative. The full proof of recursive feasibility of this constraint is deferred to that of Theorem 3.2 in the next section.

String stability constraints The ideal string stability constraint in Equation (3.16) is both non-linear and there are no guarantees that it can be recursively satisfied. Therefore a major reformulation of the constraint is required for it to be applicable in the practical MPC problem. First, define the positive and negative components of $dv_{j,s}$ as

$$v_{j,s}(k+N_p|k) - v_{j,s}(k|k) = dv_{j,s}^{\text{pos}} + dv_{j,s}^{\text{neg}}, \tag{3.18a}$$

$$dv_{j,s}^{\text{pos}} \geq 0,\ dv_{j,s}^{\text{neg}} \leq 0\,. \tag{3.18b}$$

Furthermore, to assure that $dv_{j,s}^{\text{pos}}$ and $dv_{j,s}^{\text{neg}}$ will not unnecessarily cancel each other, a term $\beta_1(dv_{j,s}^{\text{pos}} - dv_{j,s}^{\text{neg}})$ is added to the cost function. With this, a linear constraint equivalent to the constraint in Equation (3.16) could be obtained as

$$\gamma\, dv_{j,s}^{\text{neg}}(k|k) \le \Delta v_{j,s}(n|k) \le \gamma\, dv_{j,s}^{\text{pos}}(k|k). \tag{3.18c}$$

for all $n = k \dots k + N_{\text{p}} - 1$, where γ is a sufficiently large constant. This constraint is however still not recursively feasible, as implicitly it does not allow for sign changes of $\Delta v_{j,s}$ and $dv_{j,s}$. This because $dv_{j,s}(k|k)$ is required to have the same sign as all relative velocities over the prediction horizon $\Delta v_{j,s}(n|k)$, including the current relative velocity $\Delta v_{j,s}(k|k)$. The current relative velocity is fixed, and therefore, also the sign of $dv_{j,s}(k|k)$ and $\Delta v_{j,s}(n|k)\ \forall n = k \dots k + N_{\text{p}} - 1$ cannot be changed. This repeats for each next prediction horizon such that the sign of $\Delta v_{j,s}$ and $dv_{j,s}$ can never change. To this end, the constraint is changed to

$$N_{\text{p}} T \gamma\, dv_{j,s}^{\text{neg}} \le d_{j,s}(k + N_{\text{p}}|k) - d_{j,s}(k|k) \le N_{\text{p}} T \gamma\, dv_{j,s}^{\text{pos}}.$$

Here, $d_{j,s}(k + N_{\text{p}}|k) - d_{j,s}(k|k) = \sum_{n=k}^{k+N_{\text{p}}-1} \Delta v_{j,s}(n|k) T$ so that this constraint is equivalent to Equation (3.18c) if the sign of $\Delta v_{j,s}$ is constant over the prediction horizon. During normal operation, the sign of $\Delta v_{j,s}$ is constant except when the CVP transitions between accelerating and decelerating or vice versa. A proof that string stability can be achieved, even when such a transition occurs, is presented in Theorem 3.3.

Lastly, a sensible value for the design constant γ is chosen. From the distance reference defined in Equation (3.3), it can be seen that with any change in velocity dv_j, the reference distance changes with $h\, dv_j$. Therefore, set γ such that $N_{\text{p}} T \gamma = h$ and the distance between vehicles never changes more than required to track the reference. This means the controller will not overshoot the distance reference, and thus the relative velocity will not change sign during a continuous acceleration/deceleration maneuver. This gives the final constraint

$$h\, dv_{j,s}^{\text{neg}} - \epsilon_s \le d_{j,s}(k + N_{\text{p}}|k) - d_{j,s}(k|k) \le h\, dv_{j,s}^{\text{pos}} + \epsilon_s \tag{3.18d}$$

for scenarios $s \in \mathcal{S}_0$ and for vehicles $j \in C_{\setminus p_C+1}$. Here ϵ_s is a slack variable.

Using the constraints defined above, at each time instant k, each vehicle $i \in C$ solves an MPC optimization problem formulated as follows:

$$\begin{aligned} \min_{\mathbf{u}_C^i, dv_{j,\text{pos}}, dv_{j,\text{neg}}, \epsilon_s} \quad & J_C(y_C^i(k), \mathbf{u}_C^i) + \sum_s \sum_{j \in C} (\beta_1(dv_{j,s}^{\text{pos}} - dv_{j,s}^{\text{neg}}) + \beta_2 \epsilon_s) \\ \text{s.t.} \quad & (3.14),\ \forall s \in \mathcal{S}, \\ & (3.17),\ \forall j \in C, \forall s \in \mathcal{S}_{\text{e}}, \\ & (3.18),\ \forall j \in C, \forall s \in \mathcal{S}_0, \\ & \forall n = k, \dots, k + N_{\text{p}} - 1, \end{aligned} \tag{3.19}$$

where β_1 and β_2 weigh the slack variables used in the constraint in Equation (3.18).

Remark 3.6. Note that the MPC problem in Equation (3.19) can be solved locally by each vehicle $i \in C$ once all vectors y_j^i, for $j \in C/\{i\}$, are received. ◁

3.4 Control Scheme Properties

In this section, it is proven that the controller design presented in Sections 3.2 and 3.3 and visualised in Figure 3.2 complies with the requirements of Section 3.1.4. First, it is guaranteed that no crash occurs at all time, even when the CVP is subject to cyber-attack. Secondly, it is proven that, in healthy condition, there exists an input sequence such that the CVP conforms to the relaxed string stability as defined in Definition 3.2. For readability, proofs of the presented Lemmas can be found in Section 3.7.

3.4.1 Safety Properties

To prove that no crashes occur at any time, it will be proven that the safety condition in Equation (3.17) always holds, even when the topology changes.

Lemma 3.3. *Consider that at time instant k, a feasible solution of the MPC problem in Equation* (3.19) *can be found by all vehicles $i \in \mathcal{N}$ for $m = k+1$. Then, an input $u_C^i(k+1|k)$ exists such that the constraint in Equation* (3.17) *is also satisfied by all vehicles for $m = k+2$.*

Proof. For readability, proofs of the Lemmas in this section can be found in Section 3.7. ∎

Lemma 3.4. *Given Lemma 3.3, the constraint in Equation* (3.17) *with $m = k+2$ is also satisfied by all vehicles $i \in \mathcal{N}$ at time instant $k+1$, even when the communication topology changes.*

Theorem 3.2. *It is guaranteed that no crashes occur at any time, i.e. $d_{i,s}(t) \geq 0 \; \forall i, t, s$.*

Proof. In Lemmas 3.3 and 3.4 it has been shown that the safety constraints in the MPC problem in Equation (3.19) are recursively feasible both for a constant topology and over topology switches. Furthermore, all safety constraints imply $d_{i,s}(t) \geq 0$, which proves the theorem statement. ∎

3.4.2 String Stability Properties

In this section, it will be proven that in healthy conditions there always exists an input for which relaxed string stability (Definition 3.2) is achieved in the CVP. To this end, it will first be proven that strict string stability can be achieved within each coalition. Then, it will be shown that the violation of the string stability between coalitions is bounded when using the proposed topology switching law, so that the whole CVP is relaxed string stable.

Lemma 3.5. *If the soft constraint in Equation* (3.18) *holds for each vehicle $j \in C$ with $\epsilon_s = 0$, there exist an input sequence $\mathbf{u}_C^i$ for each vehicle $i \in C$ such that the coalition is strictly string stable.*

Theorem 3.3. *There exist an input sequence $\mathbf{u}_C^i$ for each vehicle $i \in C$ such that the healthy CVP is relaxed string stability.*

Proof. Lemma 3.5 proves strict string stability within a coalition. This only leaves to prove that the violation of string stability is always upper-bounded between coalitions, i.e. that $v_j(k_2) - v_j(k_1) - (v_{j-1}(k_2) - v_{j-1}(k_1))$ is bounded for all $k_1 > 0, k_2 > k_1, j \in C, j-1 \notin C$. By the switching law from Algorithm 1

$$v_j(k_2) - v_j(k_1) - (v_{j-1}(k_2) - v_{j-1}(k_1)) = \Delta v_j(k_1) - \Delta v_j(k_2) \leq 2T_v\,,$$

for all $k_1 > 0, k_2 > k_1, j \in \mathcal{N}$. ∎

Table 3.1: Parameters used in the simulation example

Parameter	Value	Parameter	Value	Parameter	Value
τ	0.1 $[s^{-1}]$	Q_i	diag(100,0,0,0,10)	β_1	0.1
r	10 $[m]$	R_i	50	β_2	1e5
h	0.5 $[s]$	$u_{\min}, u_{\max}$	$-10, 10$ $[ms^{-2}]$	T_v	0.2 $[ms^{-1}]$
N_p	10	S_d	$\{s \mid \hat{u}_{pc} = 0, \hat{a}^i_{yc} = 0\}$	T_d	0.2 $[m]$

3

3.5 Simulation for Vehicle Platoon Control

In this section, the proposed control method is applied to a CVP of 4 vehicles following a leader vehicle. The vehicles are modeled according to Equation (3.4) with a sampling time of $T = 0.05$ $[s]$. The input of the leader vehicle, which defines the CVP maneuvers is shown as the dashed line in Figure 3.3. The attacks injected in the communication are shown as dashed lines in Figure 3.5. The simulation parameters are given in Table 2.1.

Figure 3.3 shows the evolution of the states of all vehicles in this scenario. Figure 3.4 shows the evolution of the communication topology. Overall, the behaviour of the CVP is smooth and the tracking error over the whole scenario is at most 0.4m. Furthermore, note that when the tracking error is low, the CVP tends to operate in a decentralized manner as intended, thus saving coordination efforts. It is, however, important to shed some more light on a few noteworthy points.

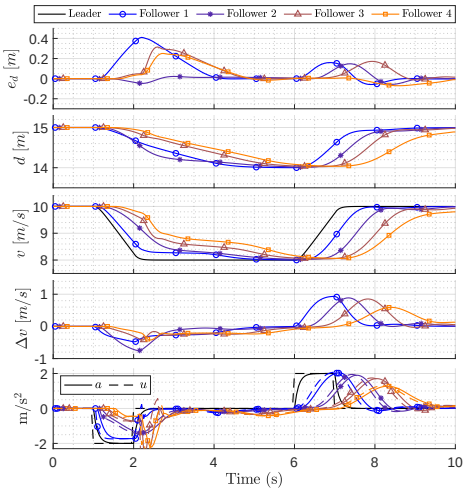


Figure 3.3: Evolution of the states and input of all vehicles.

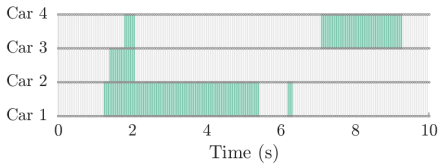


Figure 3.4: Evolution of the communication topologies.

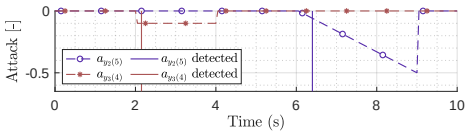


Figure 3.5: Evolution of the attacks on the communication and the times of first detection.

First, in the period between 1 and 2.2 seconds in Figure 3.3, the tracking error of vehicle 1 grows more than that of all other vehicles. Since the CVP is decelerating, all vehicles need negative relative velocity for reference tracking, causing the safety constraint to become more restrictive and forcing the vehicles to brake more than desired for reference tracking. Vehicles 2, 3, and 4 can enable communication with the preceding vehicle, as can be seen in

Figure 3.4. Thereby, the uncertainty is reduced and thus the effect of the safety constraint can be limited without compromising safety. Vehicle 1 cannot initiate communication because the preceding vehicle is the lead vehicle. Therefore, its tracking error increases during deceleration maneuvers.

Notice also that vehicles 1 to 4 start operating with full communication from approximately 2s. Nevertheless, the detection of the attack on the signals that vehicle 4 transmits forces isolation of vehicle 4 at around 2.2s as the corresponding link is deactivated. The latter causes the spike in the acceleration of vehicle 4 that can be seen in Figure 3.3. At the time the communication with vehicle 4 is disabled, it is still decelerating and its relative velocity is negative. Therefore, to keep fulfilling the safety constraint after disabling communication, the relative velocity needs to be suddenly increased. Similarly, the communication with vehicle 2 is disabled when the attack on its communication is detected (see Figures 3.4 and 3.5).

Furthermore, the performance of the proposed safety preserving topology-switching control law has been compared with other possible communication topologies. Table 3.2 shows the cumulative costs obtained with different communication topologies and attacks scenarios. Here *switching* is used to refer to the situation in which the vehicles change their communication topology only according to the tracking error and relative velocity, not considering any attack detection. Furthermore, *full communication* and *no communication* indicate, respectively, the situation in which all communication links are permanently enabled and disabled. The cumulative costs are provided for a case without attack and for three attack scenarios of varying severity. In particular, the output attacks shown in Figure 3.5 are considered and scaled according to the factors indicated in Table 3.2. Throughout all scenarios the lead vehicle follows the acceleration profile given in Figure 3.3.

It can be seen in Table 3.2 that the proposed safety preserving method has a cost equal to the *switching* topology in healthy conditions and a higher, but constant, cost for all attack conditions. One can see that the lack of attack detection in the *switching* and *full communication* topologies cause losses of feasibility, which for the *switching* topology already happens at the smallest considered attack. Moreover, the constant cost of the *no communication* approach always exceeds the cost of the proposed switching approach with attack detection. This shows that the proposed approach provides the best results over all attack scenarios.

Notice, however, that for smaller attacks, the *full communication* topology still has better performance than the proposed safety preserving approach. This extra cost of the proposed approach is caused by the conservativeness involved in the protection of the system against possible attackers. The CVP is decentralised once an attack is detected, without checking whether the effect of the attack is larger than the effect of the decentralisation. It should, however, be remarked that, even in these conditions, the switching approach allows reducing the communication and computation burden, thus providing benefits with respect to the *full communication* approach.

Table 3.2: Value of the cost function for various communication topologies and attack scenarios.

Attack scenario → / Communication topology ↓	None	Small (x2)	Medium (x3)	Large (x4)
Full communication	9.60e3	9.40e3	1.18e4	Infeasible
No communication	1.37e4	1.37e4	1.37e4	1.37e4
Switching	1.14e4	Infeasible	Infeasible	Infeasible
Switching with Attack Detection	1.14e4	1.27e4	1.27e4	1.27e4

3.6 Conclusions

A topology-switching coalitional MPC controller for CVPs is proposed, which allows for guaranteeing safety even when communication between vehicles is subject to cyber-attacks. To this end, the topology-switching coalitional MPC is integrated with an R-UIO for cyber-attack detection.

The proposed MPC law optimizes a cost function weighing performance and control effort to determine the control action subject to constraints that guarantee strict predecessor-follower string stability within each coalition. Likewise, a topology switching law enables/disables communication links when the tracking error or relative velocity between any two vehicles exceeds/falls below a chosen threshold. This significantly reduces the cooperation costs with respect to a *full communication* approach. Furthermore, it allows the topology-switching coalitional MPC to obtain relaxed string stability over the full CVP.

In case of attack, the focus of the control law shifts towards guaranteeing safety. Specifically, when a cyber-attack is detected on a certain communication link, all communication on this communication link is disabled. MPC constraints are in place to avoid crashes when such a forced topology switch occurs. Furthermore, even if the cyber-attacks are undetected safety is still guaranteed.

In summary, the designed control law provides a trade-off between performance and control effort while reducing the cooperation costs. Furthermore, relaxed string stability of the CVP can be obtained in nominal conditions, and safety is guaranteed even when the CVP is under attack. These properties have been shown theoretically and are illustrated using a CVP of 4 vehicles following a lead vehicle.

3.7 Proofs of Lemmas 3.3-3.5

In what follows, the proofs of Lemmas 3.3 to 3.5 are provided. For further use in these proofs, we first introduce the following propositions:

Proposition 3.1. *Variable $\delta(n|k)$ is positive if $u_j(\kappa) = u_{\min}\ \forall k \le \kappa < n$ and $\Delta v_{j,s}(n|k) < 0$.*

Proof. By definition, $\gamma(k|k)$ represents an upper bound on the time to standstill when $u_j(\kappa) = u_{\min}$, $\forall \kappa \ge k$. Therefore, $\delta(n|k) = \gamma(k|k) - (n-k)T$ is an upper bound on the time left to standstill at instant $n > k$ if $u_j(\kappa) = u_{\min}$, $\forall k \le \kappa < n$. Furthermore, $\Delta v_{j,s}(n|k) < 0$ implies that $v_{j,s}(n|k) > 0$, i.e. vehicle j is not at standstill at time n. This proves that, if $\Delta v_{j,s}(n|k) < 0$, then $\delta(n|k) > 0$. ∎

Proposition 3.2. *The set $\{x_{j,s}(n|k)\}_{\forall s\in\mathcal{S}_e}$, which contains the predicted states of vehicle j at instant $(n|k)$ in the extreme scenarios, is bounded and form a convex set such that $x_{j,s}(n|k)\in\mathcal{X}_j^e(n|k)\triangleq conv(\{x_{j,s}(n|k)\}_{\forall s\in\mathcal{S}_e})\ \forall s\in\mathcal{S}$.*

Proof. Separately address the two sources of uncertainty. Firstly, the undetected cyber-attacks $\hat{a}^j_{y_C,s}$ are bounded and $\hat{a}^j_{y_C,s}\in\text{conv}(\{\hat{a}^j_{y_C,s}\}_{\forall s\in\mathcal{S}_e})\ \forall s\in\mathcal{S}$. Furthermore, physics dictates that the realized state of each vehicle $j\in\mathcal{N}$ is always bounded. Therefore, $x_{C,s}(k|k)$ in Equation (3.14a) is also bounded and $x_{C,s}(k|k)\in\text{conv}(\{x_{C,s}(k|k)\}_{\forall s\in\mathcal{S}_e})\ \forall s\in\mathcal{S}$.

Secondly, the unknown input of the vehicle p_C, $\hat{u}_{p_C,s}$ is also bounded and $\hat{u}_{p_C,s}\in\text{conv}(\{\hat{u}_{p_C,s}\}_{\forall s\in\mathcal{S}_e})$, $\forall s\in\mathcal{S}$. As $w_{C,s}\sim a_{p_C,s}$ and $\text{conv}(\{a_{p_C,s}\}_{\forall s\in\mathcal{S}_e})\subseteq\text{conv}(\{\hat{u}_{p_C,s}\}_{\forall s\in\mathcal{S}_e})$, it can also be stated that $w_{C,s}$ is bounded and $w_{C,s}\in\text{conv}(\{w_{C,s}\}_{\forall s\in\mathcal{S}_e})$, $\forall s\in\mathcal{S}$.

Thus, Equation (3.14b) is a linear update equation for which $\mathcal{S}_e$ defines convex bounds on both the input and the initial condition. As $x_{j,s}(n|k)$ is considered only for $n<k+N_p$, this is sufficient to prove $x_{j,s}(n|k)$ is bounded and $x_{j,s}(n|k)\in\text{conv}(\{x_{j,s}(n|k)\}_{\forall s\in\mathcal{S}_e})\ \forall s\in\mathcal{S}$. ∎

Proof. **(Lemma 3.3)** All constraints in Equation (3.17) are lower bounds on $d_{j,s}(k+1|k)$ and each of the constraints is only active in a subset of the state-space, as discussed in Section 3.3.2. Therefore, each constraint can be considered separately. The lemma will be proven for each constraint by showing that if the constraint holds for $m=k+1$, using input $u_j(k+1|k)=u_{\min}$, it still holds for $m=k+2$. To this end, use the following relations:

$$d_{j,s}(k+2|k)=d_{j,s}(k+1|k)+T\Delta v_{j,s}(k+1|k), \tag{3.20a}$$

$$\Delta v_{j,s}(k+2|k)=\Delta v_{j,s}(k+1|k)+T\Delta a_{j,s}(k+1|k), \tag{3.20b}$$

$$\tau a_{j,s}(k+2|k)=(\tau-T)a_{j,s}(k+1|k)+Tu_j(k+1|k). \tag{3.20c}$$

Now, firstly, consider the constraint in Equation (3.17a) for $d_{j,s}(k+1|k)$, which is active only if $\Delta v_j(k+1|k)\geq 0$.[3] Then, if the constraint in Equation (3.17a) is satisfied, i.e., $d_{j,s}(k+1|k)\geq 0$, the following holds:

$$d_{j,s}(k+2|k)=d_{j,s}(k+1|k)+T\Delta v_{j,s}(k+1|k)\geq 0,$$

which proves the lemma for the constraint in Equation (3.17a). Secondly, consider the constraint in Equation (3.17b), which is active only if $\Delta v_{j,s}(k+1|k)\leq 0$ and $\Delta a_{j,s}(k+1|k)\geq 0$, and recall Proposition 3.1. Then, if the constraint in Equation (3.17b) is satisfied, i.e. $d_{j,s}(k+1|k)\geq -\Delta v_{j,s}(k+1|k)\delta(k+1|k)$, the following holds:

$$\begin{aligned}
d_{j,s}(k+2|k)&=d_{j,s}(k+1|k)+T\Delta v_{j,s}(k+1|k)\\
&\geq -\Delta v_{j,s}(k+1|k)\delta(k+1|k)+T\Delta v_{j,s}(k+1|k)\\
&=-\Delta v_{j,s}(k+1|k)\delta(k+2|k)\\
&=-(\Delta v_{j,s}(k+2|k)-\Delta a_{j,s}(k+1|k)T)\delta(k+2|k)\\
&\geq -\Delta v_{j,s}(k+2|k)\delta(k+2|k),
\end{aligned}$$

where Equations (3.20a) and (3.20b) have been used, and the fact that $\Delta a_{j,s}(k+1|k)\geq 0$ in the last inequality. This proves the lemma for the constraint in Equation (3.17b). Lastly,

[3] Note that this is a necessary condition for Equation (3.17a) to be active. If $\Delta v_{j,s}(k+1|k)<0$, the constraint in Equation (3.17b) is more restrictive. However, if $\tau\Delta a_j(k+1|k)<-\Delta v_j(k+1|k)$ the constraint in Equation (3.17c) can be more restrictive even if $\Delta v_j(k+1|k)\geq 0$.

consider the constraint in Equation (3.17c), which is active only if $\Delta a_{j,s}(k+1|k) \leq 0$. Then, if the constraint in Equation (3.17c) is satisfied, use Equation (3.20) to derive

$$\begin{aligned}
&d_{j,s}(k+2|k) = d_{j,s}(k+1|k) + T\Delta v_{j,s}(k+1|k)\\
&\geq -(\Delta v_{j,s}(k+1|k) + \tau\Delta a_{j,s}(k+1|k))\delta(k+1|k) + T\Delta v_{j,s}(k+1|k)\\
&\geq -(\Delta v_{j,s}(k+1|k) + \tau\Delta a_{j,s}(k+1|k))\delta(k+1|k) + T(\Delta v_{j,s}(k+1|k) + \tau\Delta a_{j,s}(k+1|k))\\
&= -(\Delta v_{j,s}(k+1|k) + \tau\Delta a_{j,s}(k+1|k))\delta(k+2|k)\\
&\geq -(\Delta v_{j,s}(k+2|k) - T\Delta a_{j,s}(k+1|k))\delta(k+2|k) - (\tau\Delta a_{j,s}(k+2|k)\\
&\quad + T\Delta a_{j,s}(k+1|k))\delta(k+2|k) + T\Delta u_{j,s}(k+1|k)\delta(k+2|k),
\end{aligned}$$

where $\Delta u_{j,s}(k+1|k) \triangleq u_{j-1,s}(k+1|k) - u_{j,s}(k+1|k)$, such that with the chosen input, $\Delta u_j(k+1|k) = u_{j-1}(k+1|k) - u_{\min} \geq 0$. Then,

$$d_{j,s}(k+2|k) \geq -(\Delta v_{j,s}(k+2|k) + \tau\Delta a_{j,s}(k+2|k))\delta(k+2|k),$$

which proves the lemma for the constraint in Equation (3.17c).[4] ∎

Proof. **(Lemma 3.4)** Without loss of generality, consider only the relation between vehicles $j, j-1 \in \mathcal{N}$ and the following changes to the coalitions:

(a) At instant k, the vehicles form a coalition $\mathcal{C} = \{j-1, j\}$, and it *breaks up* into $\mathcal{C}_1 = \{j-1\}$ and $\mathcal{C}_2 = \{j\}$ at $k+1$.

(b) At instant k, the vehicles are in different coalitions, say $\mathcal{C}_1 = \{j-1\}$ and $\mathcal{C}_2 = \{j\}$, and they *join* into a single $\mathcal{C} = \{j-1, j\}$ at $k+1$.

(c) Coalition $\mathcal{C} = \{j-1, j\}$ remains constant.

In Lemma 3.3, it is proven that if Equation (3.17) holds for $x_{j,s}(k+1|k)$, there exists an input sequence such that it also holds for $x_{j,s}(k+2|k)$, $\forall s \in \mathcal{S}$. Therefore, a sufficient condition for the constraint in Equation (3.17) to hold also for $x_{j,s}(k+2|k+1)$ is

$$x_{j,s}(k+2|k+1) \in \mathcal{X}_j^{\mathrm{e}}(k+2|k),\ \forall k \geq 0,\ \forall s \in \mathcal{S}. \tag{3.21}$$

The existence of set $\mathcal{X}_j^{\mathrm{e}}(k+2|k)$ is proven by Proposition 3.2. Using the prediction model in Equation (3.14), obtain

$$\begin{aligned}
x_{j,s}(k+2|k+1) =& A_j x_{j,s}(k+1|k+1) + B_j u_j(k+1|k+1)\\
&+ A_j^w x_{j-1,s}(k+1|k+1),\\
x_{j,s}(k+2|k) =& A_j x_{j,s}(k+1|k) + B_j u_j(k+1|k)\\
&+ A_j^w x_{j-1,s}(k+1|k),
\end{aligned} \tag{3.22}$$

where we chose the input $u_j(k+1|k+1) = u_j(k+1|k)$. Furthermore, by Proposition 3.2,

$$x_{j,s}(k+1|k) \in \mathcal{X}_j^{\mathrm{e}}(k+1|k),\ \forall i \in \mathcal{C},\ \forall s \in \mathcal{S},\ \forall k \geq 0,$$

[4] Above the lemma is proved if the active constraint is fixed. It is, however, possible that the active constraint changes between time instants $k+1$ and $k+2$. Similar approaches can be used to prove the lemma for each of these cases. These full proofs are however omitted here.

and as the realised state $x_{j,s}(k+1|k+1)$ is the outcome of one of the possible scenarios in $\mathcal{S}$, also

$$x_{j,s}(k+1|k+1) \in \mathcal{X}_j^{\mathrm{e}}(k+1|k),\ \forall i \in \mathcal{C},\ \forall s \in \mathcal{S},\ \forall k \geq 0. \tag{3.23}$$

That is, the new initial condition $x_{j,s}(k+1|k+1)$ is bounded by the prediction based on the extreme scenario at time k.

Consider case (a), where using Equation (3.14c), $x_{j-1,s}(k+1|k+1) = x_{j-1,s}(k+1|k) \in \mathcal{X}_{j-1}^{\mathrm{e}}(k+1|k)$ for all $s \in \mathcal{S}$. In cases (b) and (c), as vehicles j and $j-1$ are in the same coalition at time $k+1$, Equation (3.23) can be directly applied for vehicle $j-1$ too, such that $x_{j-1,s}(k+1|k+1) \in \mathcal{X}_{j-1}^{\mathrm{e}}(k+1|k)$ for all $s \in \mathcal{S}$.

Substituting the results above into Equation (3.22) implies Equation (3.21) holds, proving the lemma. ∎

Proof. **(Lemma 3.5)** Define $dv_j^{n_2}(n_1) \triangleq v_j(n_1+n_2) - v_j(n_1)$ as the change in velocity of vehicle j over a period of n_2 time-steps, and recall $dv_{j,s}(k) = dv_{j,s}^{\mathrm{pos}}(k) + dv_{j,s}^{\mathrm{neg}}(k) = v_{j,s}(k+N_{\mathrm{p}}|k) - v_{j,s}(k|k)$ is the predicted change in velocity over the length of the prediction horizon.

Now, without loss of generality, consider that, dictated by the considered maneuver of the lead vehicle, $dv_{j,s}(k) \geq 0$, $\forall k \in [0, N)$ and $dv_{j,s}(k) \leq 0$, $\forall k \in [N, N_2)$. Furthermore, consider initially $\Delta v_j(0) \leq 0$. Starting from this initial condition, by the constraint in Equation (3.18d)

$$\begin{aligned} &\Delta v_{j,s}(k) \leq 0,\ \forall k \in [0, n_s), \\ &\exists\ n_s < N_{\mathrm{p}} \text{ s.t. } \Delta v_{j,s}(n_s) \geq 0, \end{aligned}$$

for all $s \in \mathcal{S}_0$. Then, by the definition of the scenarios $\mathcal{S}_0$, also

$$\begin{aligned} &\Delta v_j(k) \leq 0,\ \forall k \in [0, n), \\ &\exists\ n < N_{\mathrm{p}} \text{ s.t. } \Delta v_j(n) \geq 0. \end{aligned}$$

Furthermore,

$$\exists\ \mathbf{u}_j \text{ s.t. } \Delta v_j(k) \leq 0,\ \forall k \in (n, N).$$

Note that $u_j(k) = u_{j-1}(k)\ \forall k \in (n, N)$ is one of the input sequences that guarantees this. With this, derive

$$dv_j^{\ell-k}(k) - dv_{j-1}^{\ell-k}(k) = \Delta v_j(k) - \Delta v_j(\ell) \leq 0,\ \forall k \in [0, \ell),$$

where $\ell < N$. This implies strict string stability according to Definition 3.1 for time steps 0 to N.

At time N, $\Delta v_j(N) \geq 0$ and $dv_{j,s}(N) \leq 0$, which is a similar situation to the the initial one. Therefore, following the same line of reasoning, $\exists\ n_2 < N + N_{\mathrm{p}}$ such that

$$\begin{aligned} \Delta v_j(k) \geq 0,\ \forall k \in [N, N+n_2), \\ \Delta v_j(n_2) \leq 0, \text{and} \\ \exists\ \mathbf{u}_j \text{ s.t. } \Delta v_j(k) \geq 0,\ \forall k \in (N+n_2, N_2), \end{aligned}$$

such that

$$dv_j^{\ell-k}(k) - dv_{j-1}^{\ell-k}(k) = \Delta v_j(k) - \Delta v_j(\ell) \geq 0,$$

for all $k \geq N$ and $k < N + n_2 \leq \ell < N_2$, which implies strict string stability for time steps N to N_2. At time step N_2 the situation is then as it was initially, such that the proof can be repeated for all time. ∎

4

Anomaly Prevention through Fully Homomorphic Encryption

The current industrial state of practice of encryption in cyber-physical systems (CPSs) is based on end-to-end encryption techniques preventing intrusion in remote connections between the plant, sensors and controllers. To close the loop of encryption in control, recently fully homomorphic encryption (FHE) has been proposed, which makes it possible to perform addition and multiplication of encrypted data. Using FHE in control thus allows to develop controllers which can operate on encrypted data, preventing loss of confidentiality in the controller. FHE in its current form is however not practically applicable for real-time control at high update rates due to its increased computational load. In this chapter a reformulation of the GSW FHE scheme is proposed which reduces the computation load. This reformulated FHE scheme is then implemented on a field-programmable gate array (FPGA) for stabilisation of an inverted double pendulum. It will be shown that with this implementation the unstable plant can be stabilised in real-time by fully utilising the computational advantages of the reformulated scheme on the FPGA.

This chapter is based on

Pieter Stobbe, Twan Keijzer, and Riccardo M.G. Ferrari. A fully homomorphic encryption scheme for real-time safe control. In Conference on Decision and Control, *2022.*

With the development of large scale cyber-physical systems (CPSs), cryptography, which was originally developed for information technology (IT) systems, is increasingly utilised in operation technology (OT). Large scale CPSs must be securely monitored and controlled over long distances using remote connections between the plant, sensors and the controller. Loss of confidentiality of data transmitted through such remote connections can only feasibly be prevented via encryption.

The industrial state of practice is to use so-called *end-to-end encryption*, which provides confidentiality in the remote connections, but requires decryption at the receiver [106, 299, 300]. This is ideal for IT systems but has several drawbacks when used in control. Specifically, measurements that are sent over an end-to-end encrypted connection need to be decrypted at the controller to be processed. For automatic control, performing this decryption and re-encryption at the controller is not necessary and even forms a risk to security as loss of confidentiality in the controller is no longer prevented. At the same time, these decryption and re-encryption steps create computational overhead, limiting the controller update time, which reduces the stability margin and can possibly destabilise the plant.

Homomorphic encryption (HME) schemes present an alternative that can solve these problems by allowing for multiplication and/or addition of encrypted numbers. With this, the decryption and re-encryption at the controller side are no longer required, providing confidentiality in the controller and eliminating the related computational overhead. There are two main types of HME: partially homomorphic encryption (PHE) and fully homomorphic encryption (FHE). PHE schemes support only multiplication or addition, whereas FHE schemes support both.

The first HME scheme was the Rivest–Shamir–Adleman (RSA) scheme [106], followed by PHE schemes such as EL Gamal [108] and Paillier [109]. More recently lattice-based FHE schemes have been introduced in [117, 118, 301]. These schemes allow for the implementation of a broad range of feedback control, however, their high computational complexity prevents them from being used in real-time on conventional hardware.

PHE schemes have also been proposed for control, such as in [110] which proposes a combination of the El Gamal [108] and RSA [106] schemes. This control scheme, however, requires the controller state to be sent to the plant for decryption and re-encryption at each time step, adding additional overhead. More recently, [302] has demonstrated direct feedback control with the PHE scheme from [109]. Due to the limited homomorphic properties of the PHE scheme, the controller uses plain-text controller gains, posing a risk to security.

Recently, more attention has been directed to FHE schemes for control, such as in [119, 124–126, 303]. These schemes however still suffer from two problems. Firstly, the representation of encrypted signals requires orders of magnitude more storage than the original plain-text. This means that, due to limited computation and bandwidth resources, real-time control with FHE is limited in complexity and update rate. In [124], a two-state linear controller is implemented with an update rate of 2 Hz while in [120] a direct feedback controller for high-level control of a drone reaches an update rate of 10 Hz.

Secondly, both PHE and FHE only allow for encryption of unsigned integers, whereas typically rational numbers are required in control. For this purpose rational numbers can be represented as unsigned integers through the Q format. The transformation to get Q format

numbers retains all properties of the rational numbers, such that any calculations can equivalently be performed on Q format numbers. Using the Q format, however, introduces a limitation when used in combination with PHE or FHE. By the structure of Q format numbers, the result of a multiplication of two Q format numbers requires more memory than the original numbers. Specifically, the number of bits that correspond to the fraction in the rational number increase. With limited memory dedicated to represent each number, this eventually leads to insufficient bits to represent the whole numbers, i.e. leading to a overflow. Under plain-text operation, right hand bit-shifts can be used to round of to the desired precision and prevent this overflow. However, no FHE schemes currently support homomorphic right hand bit-shifts without excessive penalties on *multiplicative depth*, which is defined as the maximum allowed number of consecutive multiplications. Several solutions to this problem have been proposed, such as periodic reset [119] and scaling of the state space matrices [303]. These methods, however, affect stability and performance of the control schemes to which they are applied. Therefore, they cannot be directly applied to existing control designs.

The problems of computational complexity and fixed precision have hindered the acceptance of FHE for real-time control. In this chapter we propose an FHE scheme for real-time secure control implemented on a field-programmable gate array (FPGA) which address these issues. The contributions presented in this chapter are:

- The GSW FHE scheme [117] has been reformulated to reduce computational complexity by introducing the so-called *reduced cipher*. The reduced cipher is a change of representation of the original cipher which allows for more intuitive manipulation and interpretation of the scheme.
- The reformulated FHE scheme is implemented on an FPGA for real-time control of an unstable plant to demonstrate the benefits of the novel *reduced cipher*.

The resulting FHE scheme can be used in combination with a large class of existing control schemes. Furthermore, an FPGA was chosen for implementation to fully utilise the reduction in computational complexity caused by the reduced cipher. The scheme can, however, also be implemented on any conventional hardware.

In the following, Section 4.1 introduces the considered control setup and the GSW FHE scheme. In Section 4.2 the *reduced cipher* is introduced and it is proven it allows for reduced computational complexity. Section 4.3 shows the benefit of the reformulated FHE by implementing it on an FPGA for control of a inverted double pendulum. Lastly, some concluding remarks are presented in Section 4.4.

Notation

For a positive scalar x, we denote individual digits of its binary representation as $x^{[i]}$. That is, $x = \sum_{i=0}^{\infty} 2^i x^{[i]}$. For any $x \in \mathbb{N}$ we define $(x)^l = \sum_{i=0}^{l-1} 2^i x^{[i]}$ which are the l least significant binary digits of x, such that if $x \leq q$ where $q = 2^l - 1$, then $(x)^l = x$ and if $x > q$, then $(x)^l \neq x$. We denote $[x]^l = [x^{[0]}, \dots, x^{[l]}]$ as a vector whose elements are the binary digits of $(x)^l$; $g = [2^0, \dots, 2^{l-1}]^\top$ and the set $\mathbb{Z}_q = \{0, \dots, q-1\}$, where $q \in \mathbb{N}$. We denote bit-shifts of a x by i bits as $x \ll i = 2^i x$ and $x \gg i = 2^{-i} x$. These concepts can be extended to matrices $X \in \mathbb{N}^{n_1 \times n_2}$, where $(X)^l$, $[X]^l$, and bitshifts are applied element-wise. $G_n = I_n \otimes g$, while the encrypted version of a variable x is denoted as $\mathbf{E}(x)$.

4.1 Problem Statement

In this chapter a reformulation of the GSW FHE scheme will be presented which reduces its computational load, increasing its applicability in safe and secure control. To this end, in Section 4.1.1, we will first introduce a general control scenario to which the reformulated FHE scheme is applicable. Then the GSW FHE scheme will be presented in its original form in Section 4.1.2. In this section the scheme is also directly written in a novel, simplified notation, which will aid in reformulating the FHE scheme in Section 4.2. Lastly, Section 4.1.3 discusses current challenges in applying FHE for control to motivate the presented research.

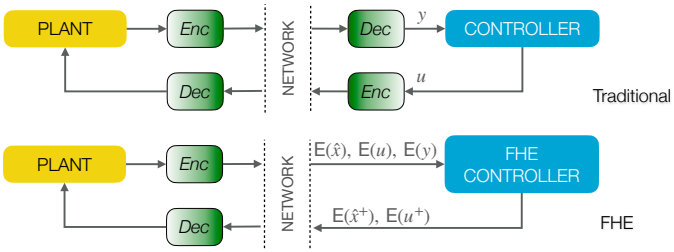


Figure 4.1: Top: A plant controlled remotely using traditional encryption. Bottom: A plant controlled remotely using FHE.

4.1.1 Control Scenario

In this chapter we consider a general nonlinear plant of the form

$$\begin{cases} \dot{x} = \mathfrak{f}(x,u) + \zeta_1, \\ y = \mathfrak{h}(x,u) + \zeta_2, \end{cases} \tag{4.1}$$

and a discrete time, dynamical linear controller of the form

$$\begin{cases} \hat{x}^+ = \mathfrak{g}(\hat{x},u,y,L), \\ u^+ = \mathfrak{v}(\hat{x}^+,K), \end{cases} \tag{4.2}$$

where superscript + denotes the next time-step. Furthermore, $x \in \mathbb{R}^\rho$ is the state, $\hat{x} \in \mathbb{R}^\rho$ the state estimate, $u \in \mathbb{R}^\gamma$ the input and $y \in \mathbb{R}^\nu$ the output. $\mathfrak{f}(\cdot)$ and $\mathfrak{h}(\cdot)$ are the known state transition and output functions, and ζ_1 and ζ_2 represent system and measurement uncertainties. The controller consists of two parts: $\mathfrak{g}(\cdot)$ is a state estimator generating $\hat{x}$; and $\mathfrak{v}(\cdot)$ is a state feedback controller generating control input u. In this chapter we consider the plant is controlled by an encrypted version of this controller, which using notation introduced before, is denoted by

$$\begin{cases} \mathbf{E}(\hat{x}^+) = \tilde{\mathfrak{g}}(\mathbf{E}(\hat{x}),\mathbf{E}(u),\mathbf{E}(y),\mathbf{E}(L)), \\ \mathbf{E}(u^+) = \tilde{\mathfrak{v}}(\mathbf{E}(\hat{x}^+),\mathbf{E}(K)). \end{cases} \tag{4.3}$$

An illustration of the proposed encrypted control loop is shown in Figure 4.1 together with an illustration of a traditional encrypted control loop. One can see that FHE has two

advantages with respect to the traditional encryption. Firstly, with FHE the communicated signals remain encrypted, and therefore secure, while they are processed in the controller. Secondly, with FHE only one pair of encryption and decryption is required, reducing the added delay. The specifics of the proposed encrypted control loop will be discussed in Section 4.3 and illustrated in Figure 4.2.

This work is not concerned with the design of the nominal unencrypted controller itself, but rather focuses on the implementation of FHE to secure any new or existing control systems. Therefore, we introduce some basic assumptions on the unencrypted controller and the closed loop behaviour of the plant and controller before continuing.

Assumption 4.1. The control law in Equation (4.2) can be constructed with addition, subtraction and multiplication operations only. This holds true for all linear control methods such as proportional-integral-derivative (PID) and state-feedback control as well as linear-quadratic regulators (LQRs) [304]. ◁

Assumption 4.2. The plant in Equation (4.1) is stabilised by the unencrypted controller in Equation (4.2). ◁

4.1.2 Fully Homomorphic Encryption Scheme by Gentry et al.

The GSW FHE scheme was introduced in [118] as a great improvement in simplicity and efficiency of FHE schemes based on learning with errors (LWE). It eliminated the computationally heavy relinearisation step and reduced the scheme to just four procedures: Key generation, encryption, homomorphic operations, and decryption. Key generation is performed at the plant at initialization to allow for encryption and decryption, which are also performed at the plant, as visualised in Figure 4.1. The homomorphic operations are then performed remotely in the FHE based encrypted controller. Gentry et al. introduced four basis functions needed to perform these procedures.

Definition 4.1. For any matrix $a \in \mathbb{N}^{N\times(n+1)}$, $b \in \mathbb{N}^{N\times N}$, and $c \in \mathbb{Z}_q^{n+1\times 1}$

$$\textbf{BitDecomp}(a) = [a]^l \tag{4.4}$$

$$\textbf{BitDecomp}^{-1}(b) = b \cdot G_{n+1} \tag{4.5}$$

$$\textbf{Flatten}(b) = [b \cdot G_{n+1}]^l \tag{4.6}$$

$$\textbf{PowersOf2}(c) = G_{n+1} \cdot c \tag{4.7}$$

◁

Note that these functions are directly cast in the novel notation allowing them to be easily understood and manipulated. We can now use these functions to also define the procedures of the GSW FHE scheme in this new notation.

Key Generation

A public-private key pair is generated as follows: Pick parameters $m \in \mathbb{N}$, $n \in \mathbb{N}$ and $q \in \mathbb{N}$ based on the required security and precision respectively. The private key is $s = [1, -t]^\top$ where $t \in \mathbb{Z}_q^{1\times n}$ is sampled uniformly on the interval $[0, q-1]$. The public key is $A = [b, B]$ where $b = B \cdot t^\top + e$, each element of $B \in \mathbb{Z}_q^{m\times n}$ is sampled uniformly on the interval $[0, q-1]$, and each element of $e \in \mathbb{Z}_q^m$ is sampled from the χ_q distribution [116].

Encryption

A message $\mu \in \mathbb{Z}_q$ can be encrypted as a cipher $C \in \mathbb{Z}_2^{N\times N}$ via the following relation

$$C = \mathbf{Enc}(\mu) = \mathbf{Flatten}(\mu \cdot I_N + \mathbf{BitDecomp}(R \cdot A)) = [(\mu \cdot I_N + [R \cdot A]^l) \cdot G_{n+1}]^l, \tag{4.8}$$

where $N = l(n+1)$ depends on the message size through $l = \lfloor \log_2(q) \rfloor$ and n relates to the security of the encryption scheme. Furthermore, each element of $R \in \mathbb{Z}_2^{N\times m}$ is sampled uniformly on the interval $[0,1]$.

Decryption

Ciphers are decrypted using the **MPDec** algorithm as proposed in [305]:

$$\mu = \mathbf{MPDec}((C\mathbf{PowersOf2}(s))^l) = \mathbf{MPDec}((CG_{n+1}s)^l) \tag{4.9}$$

The **MPDec** algorithm [305] uses the first l elements of its input to retrieve μ. Proof that the correct message is retrieved in this way can be found in [118].

Homomorphic Operations

The homomorphic operations for ciphers $C_1 = \mathbf{Enc}(\mu_1)$, $C_2 = \mathbf{Enc}(\mu_2)$ and scalar α are

$$\begin{aligned}
\text{Sum: } C_3 &= \mathbf{Flatten}(C_1 + C_2) = [(C_1 + C_2) \cdot G_{n+1}]^l, \\
\text{Product: } C_4 &= \mathbf{Flatten}(C_1 \cdot C_2) = [(C_1 \cdot C_2) \cdot G_{n+1}]^l, \\
\text{Scalar product: } C_5 &= \mathbf{Flatten}(\mathbf{Flatten}(\alpha I_N) \cdot C_2) = [([(\alpha I_N) \cdot G_{n+1}]^l C_2) \cdot G_{n+1}]^l, \\
\text{Scalar sum: } C_6 &= \mathbf{Flatten}(\alpha I_N + C_2) = [(\alpha I_N + C_2) \cdot G_{n+1}]^l.
\end{aligned} \tag{4.10}$$

For these homomorphic operations it is proven that they are equivalent to the corresponding plain-text operations, i.e.

$$\begin{aligned}
\mu_3 &= \mu_1 + \mu_2 \quad &\Longleftrightarrow \mu_3 &= \mathbf{MPDec}((C_3 G_{n+1} s)^l), \\
\mu_4 &= \mu_1 \mu_2 \quad &\Longleftrightarrow \mu_4 &= \mathbf{MPDec}((C_4 G_{n+1} s)^l), \\
\mu_5 &= \alpha \mu_2 \quad &\Longleftrightarrow \mu_5 &= \mathbf{MPDec}((C_5 G_{n+1} s)^l), \\
\mu_6 &= \alpha + \mu_2 \quad &\Longleftrightarrow \mu_6 &= \mathbf{MPDec}((C_6 G_{n+1} s)^l).
\end{aligned} \tag{4.11}$$

4.1.3 FHE in Control

The GSW FHE scheme [118] has excellent theoretical properties, but there are two obstacles which, until now, have prevented implementation of the scheme in control. Firstly, any message $\mu \in \mathbb{Z}_q$ containing l bits of information, when encrypted, becomes a cipher $C \in \mathbb{Z}_2^{N\times N}$ containing $N^2 = (n+1)^2 l^2$ bits of information. Therefore storage and transfer of ciphers requires more memory than unencrypted equivalents. The problem of size becomes even more pronounced when performing homomorphic operations. Direct implementation of homomorphic operations requires multiple steps in which intermediate ciphers can become as large as $\mathbb{Z}_N^{N\times N}$ containing $N^2(\lfloor \log_2(N) \rfloor + 1)$ bits of information.

Even more important than the strain on storage and communication, is the strain on the computational resources. For direct implementation of homomorphic addition,

$N^2(n+2)+N(n+1)$ addition operations and $N^2(n+1)$ multiplication operations are needed, whereas its unencrypted equivalent requires only a single addition. In this chapter a so-called *reduced cipher* is introduced to reduce the computational load of FHE, allowing for faster update rates of control laws.

The second obstacle is the representation of real numbers with unsigned integers. To this end we employ the commonly used fixed precision representation called Q format [302].[1] Alternatives using floating point numbers are currently being researched [306] but are not yet sufficiently mature. Q format allows for representing a fixed accuracy number β with an integer message $\mu \in \mathbb{Z}_p$ where $\lfloor \log_2(p) \rfloor + 1 = m_q + n_q$ as

$$\begin{aligned} \beta &= -2^{m_q-1}\mu^{[m_q+n_q-1]} + \sum_{i=0}^{m_q+n_q-2} 2^{i-n_q}\mu^{[i]} \\ \mu &= \begin{cases} 2^{n_q}\beta & \text{if } \beta \geq 0 \\ -2^{m_q+n_q} + |\beta|2^{n_q} & \text{if } \beta < 0 \end{cases} \end{aligned} \tag{4.12}$$

such that β can be any value in $[-2^{m_q-1}, 2^{m_q-1})$ with an accuracy of 2^{-n_q}. When performing multiplication of two messages $\mu_3 = \mu_1 \cdot \mu_2$, where μ_1 and μ_2 are obtained from Equation (4.12), the result has to fit a $m_q + 2n_q$ sized register to yield an exact result. The available storage for each message is limited and so after a certain number of consecutive multiplications overflow would occur.

Therefore, conventionally, a right-bitshift by n_q bits is performed after each multiplication such that the $m_q + n_q$ least significant bits of μ_3 can be used to retrieve $\beta_1\beta_2$[2]. However, no HME scheme supports such operation on ciphers without penalty on multiplicative depth. Thus, consecutive multiplications have formed a great obstacle in HME. This problem is important for application to control systems, which often have internal states in the controller that are updated at each timestep without being decrypted. Until now this obstacle has been dealt with using a periodic reset [302] or by transforming the state space variables [303]. These methods, however, affect the stability and performance of the controller such that direct implementation of FHE with existing control schemes is not guaranteed to work.

In this work we use a solution based on [110], where the encrypted controller state is sent to the plant for decryption at each controller update to overcome the lack of homomorphic addition in the ElGamal PHE scheme [108]. To this end, in [110], the encrypted controller state is sent to the plant in parts, where they need to be decrypted, added together and re-encrypted to a single cipher. In this work, similarly, the encrypted controller state is sent to the plant where it is decrypted, the necessary right-hand bitshifts are performed, and re-encrypted. This also prevents any issues with multiplicative depth.

4.2 Reduced Ciphers for Fast FHE Implementation

In this section the so-called *reduced cipher* will be presented that allows us to reformulate the GSW FHE scheme for computationally efficient implementation of encrypted control.

[1] We will be using the Q-notation as introduced by Texas-Instruments, which is used in code libraries such as the TMS320C64x+ IQmath.

[2] rounded down to the nearest 2^{-n_q}, due to truncation during the right hand bitshift.

It will be shown that with the *reduced cipher*, encryption, homomorphic operations, and decryption can be made orders of magnitude more computationally efficient, enabling real-time implementation of FHE for control.

Given a cipher $C \in \mathbb{Z}_2^{N\times N}$, the so-called *reduced cipher* will be denoted by $\tilde{C} \in \mathbb{Z}_q^{N\times(n+1)}$ and is defined as

$$\tilde{C} \triangleq \mathbf{BitDecomp}^{-1}(C) = CG_{n+1}. \tag{4.13}$$

Note that the reduced cipher contains exactly the same information as the original cipher. However, by using the reduced cipher for encryption, decryption and homomorphic operations the computational complexity of these procedures can be greatly reduced. In Theorem 4.1 it will be shown how the reduced cipher can be used to perform homomorphic operations. Then in Corollary 4.1 it is shown how these same principles can be used to perform encryption and decryption with the reduced cipher.

4

Lemma 4.1. *For any matrix $\Lambda \in \mathbb{N}^{n_1\times n_2}$, we have $[\Lambda]^l G_{n_2} = (\Lambda)^l$.*

Proof. Consider $\alpha \in \mathbb{N}$. for any α it holds $(\alpha)^l = \sum_{i=0}^{l-1} 2^i \alpha^{[i]} = \left[\alpha^{[0]}, \dots, \alpha^{[l-1]}\right] \cdot g = [\alpha]^l \cdot g$. Then apply this relation on each element of Λ, giving $(\Lambda)^l = [\Lambda]^l \cdot I_{n_2} \otimes g = [\Lambda]^l \cdot G_{n_2}$. ∎

Theorem 4.1. *Given ciphers $C_1, C_2 \in \mathbb{Z}_2^{N\times N}$ and scalar $\alpha \in \mathbb{Z}^q$ the existing homomorphic operations as in Equation* (4.10) *can equivalently be written using the reduced cipher as*

$$C_3 = [(C_1 + C_2)G_{n+1}]^l \leftrightarrow \tilde{C}_3 = \left(\tilde{C}_1 + \tilde{C}_2\right)^l \tag{4.14}$$

$$C_4 = [(C_1 \cdot C_2)G_{n+1}]^l \leftrightarrow \tilde{C}_4 = \left(C_1 \cdot \tilde{C}_2\right)^l \tag{4.15}$$

$$C_5 = [[\alpha G_{n+1}]^l \cdot C_1 G_{n+1}]^l \leftrightarrow \tilde{C}_5 = ([\alpha G_{n+1}]^l \tilde{C}_1)^l \tag{4.16}$$

$$C_6 = [(\alpha I_N + C_1)G_{n+1}]^l \leftrightarrow \tilde{C}_6 = (\alpha G_{n+1} + \tilde{C}_1)^l \tag{4.17}$$

Proof. Use the definition of the reduced cipher in Equation (4.13), Definition 4.1, and Lemma 4.1 to derive

$$\begin{aligned}
\tilde{C}_3 &= [(C_1 + C_2)G_{n+1}]^l G_{n+1} = (C_1 G_{n+1} + C_2 G_{n+1})^l = \left(\tilde{C}_1 + \tilde{C}_2\right)^l \\
\tilde{C}_4 &= [(C_1 \cdot C_2)G_{n+1}]^l G_{n+1} = \left(C_1 \cdot \tilde{C}_2\right)^l \\
\tilde{C}_5 &= \left[[\alpha I_N G_{n+1}]^l \cdot C_1 G_{n+1}\right]^l G_{n+1} = \left([\alpha G_{n+1}]^l \cdot \tilde{C}_1\right)^l \\
\tilde{C}_6 &= [(\alpha I_N + C_1)G_{n+1}]^l G_{n+1} = (\alpha G_{n+1} + C_1 G_{n+1})^l = \left(\alpha G_{n+1} + \tilde{C}_1\right)^l
\end{aligned}$$

∎

Corollary 4.1. *Encryption and decryption can be reformulated in terms of reduced ciphers.*

Proof. Encryption is performed using Equation (4.8), where $\mathbf{BitDecomp}(RA) = [RA]^l \in \mathbb{Z}_2^{N\times N}$ is of the same form as a cipher. Encryption is thus a special case of the homomorphic scalar sum as defined in Equation (4.10). Applying the results in Equation (4.17) and Lemma 4.1 to encryption thus yields

$$\tilde{C} = (\mu G_{n+1} + [R \cdot A]^l G_{n+1})^l = (\mu G_{n+1} + R \cdot A)^l. \tag{4.18}$$

Decryption, as defined in Equation (4.9), is reformulated in terms of the reduced cipher as

$$\mu = \mathbf{MPDec}((CG_{n+1}s)^{l}) = \mathbf{MPDec}((\tilde{C}s)^{l}) \tag{4.19}$$

■

Theorem 4.1 and Corollary 4.1 have shown that all operations needed for an encrypted control scheme can be performed based on the reduced cipher. In the following we will show that using the reduced cipher also reduces the computational complexity of these operations. Most importantly, it will be shown that the reduced cipher completely eliminates the need for performing hardware multiplications to perform homomorphic operations. To this end Lemmas 4.2 and 4.3 will show how the multiplication operations remaining in Equations (4.15) to (4.17) can be performed using bit-operations on the hardware level.

Table 4.1: Number of hardware operations required for homomorphic operations.

	sum		product	
	Cipher	Red. Cipher	Cipher	Red. Cipher
Bit Operation	0	0	$\mathcal{O}(n^3l^3)$	$\mathcal{O}(n^3l^2)$
Addition	$\mathcal{O}(n^3l^2)$	$\mathcal{O}(n^2l)$	$\mathcal{O}(n^3l^3)$	$\mathcal{O}(n^3l^2)$
Multiplication	$\mathcal{O}(n^3l^2)$	0	$\mathcal{O}(n^3l^2)$	0
Memory	$\mathcal{O}(n^2l^2)$	$\mathcal{O}(n^2l^2)$	$\mathcal{O}(n^2l^2\log(nl))$	$\mathcal{O}(n^2l^2)$
	scalar sum		scalar product	
	Cipher	Red. Cipher	Cipher	Red. Cipher
Bit Operation	$\mathcal{O}(n^3l^2)$	$\mathcal{O}(l)$	$\mathcal{O}(n^2l^3)$	$\mathcal{O}(n^2l^2)$
Addition	$\mathcal{O}(n^3l^2)$	$\mathcal{O}(nl)$	$\mathcal{O}(n^3l^3)$	$\mathcal{O}(n^2l^2)$
Multiplication	$\mathcal{O}(n^2l)$	0	$\mathcal{O}(n^3l^2)$	0
Memory	$\mathcal{O}(n^2l^2)$	$\mathcal{O}(n^2l^2)$	$\mathcal{O}(n^2l^2\log(l))$	$\mathcal{O}(n^2l^2)$

Lemma 4.2. *Given any $\alpha \in \mathbb{Z}^q$, αG_{n+1} can be generated by using only $l-1$ bit-shifts on the hardware level.*

Proof. $\alpha G_{n+1} = I_{n+1} \otimes \alpha g$, i.e. αG_{n+1} consist only of $(n+1)$ identical instances of αg. Now, recall $x \ll n$ denotes n left bit-shifts of scalar x, which is equivalent to a multiplication by 2^n. Using this notation αg can be generated as $\alpha g = [\alpha, \alpha \ll 1, \ldots, \alpha \ll l-1]^\top$ using only $l-1$ bit-shifts. ■

Lemma 4.3. *Given any $C_1 \in \mathbb{Z}_2^{N\times N}$ and $\tilde{C}_2 \in \mathbb{Z}_q^{N\times(n+1)}$, $C_1 \cdot \tilde{C}_2$ can be generated by using $\mathcal{O}(n^3l^2)$ bit-masks and $\mathcal{O}(n^3l^2)$ additions on the hardware level.*

Proof. As the entries of C_1 are binary, multiplications between entries of C_1 and $\tilde{C}_2$ can be performed as bit-masks. Specifically, the bit-masks can be performed by overlaying a row of C_1 on a column of $\tilde{C}_2$. The outcomes are then added to obtain one entry of the resulting reduced cipher matrix. This operation requires N bit-masks and $N-1$ additions and needs to be repeated for each entry of the resulting reduced cipher matrix, i.e. $N(n+1)$ times. Therefore, in total $\mathcal{O}(n^3l^2)$ bit-masks and $\mathcal{O}(n^3l^2)$ additions are required. ■

The total reduction of computational complexity obtained by using the *reduced ciphers* for the homomorphic operations is summarised in Table 4.1. The table shows the hardware level computations and memory utilisation of the operations that are involved in evaluating the homomorphic operations from Equation (4.10), both with and without the use of *reduced ciphers*. It can be seen that the overall number of hardware operations is reduced and the homomorphic operations no longer require multiplication at the hardware level. Note, however, that the *Reduced ciphers* contain the same amount of data as original ciphers and therefore the communication bandwidth required to transfer the ciphers is unchanged.

4.3 Results on a Simulated Plant

In this section, the reformulated FHE scheme is applied to the control of an inverted double pendulum. It will be shown that it is possible to perform stabilising control of the unstable plant in real-time with the encrypted controller. To this end the reformulated FHE based control system is applied on two FPGAs as shown in Figure 4.2. The use of FPGAs allows for an efficient implementation of the reformulated FHE scheme. In the remainder of this section, it will first be shown how the properties of an FPGA are beneficial to the implementation of the reformulated FHE scheme. Then, the used control setup will be discussed. Lastly, the results obtained with this setup will be presented and analysed.

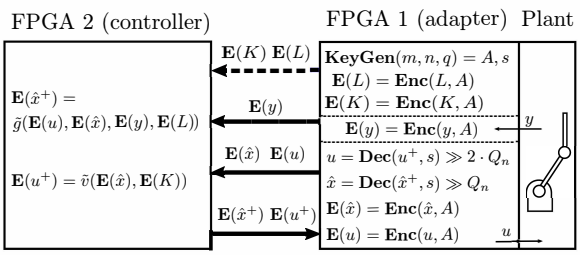


Figure 4.2: The experimental setup. FPGA 1 performs encryption and decryption. FPGA 2 performs the encrypted control. Key exchange, indicated with the dashed arrow, is only required at initialization.

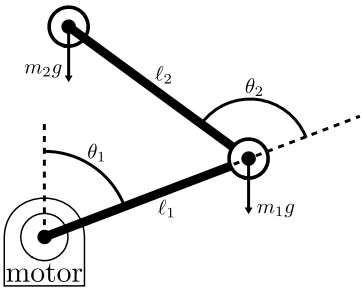


Figure 4.3: Double pendulum as modeled by Equation (4.20).

4.3.1 Hardware Resources of an FPGA

An FPGA contains generic logic cells and memory components of differing sizes and configurability. The most common type of logic cell is called an adaptive logic module (ALM). These can be configured to perform any operation. However, though ALMs can be configured to perform multiplication, this leads to an inefficient use of resources. Therefore, FPGAs are generally equipped with digital signal processing (DSP) slices which are specifically designed to perform multiplication. Unfortunately, due to their heavy requirements on the die space, there are far fewer DSP slices than ALMs. To illustrate, ALMs are usually available in the order of tens of thousands, whereas there are usually only in the order of ten DSP slices. Therefore, if a design's speed relies on multiplication, the limited number of DSP-slices is often a bottle-neck to the computational speed.

As shown in Section 4.2, the reformulated FHE scheme with the *reduced cipher* allows for implementation without any multiplications on the hardware level. This is beneficial

to the computational efficiency of the scheme on any platform, however due to the limited number of DSP-slices it is particularly beneficial on an FPGA. By eliminating this bottleneck, many more operations can be implemented in parallel. FPGAs can be programmed to operate without the need for a software layer, dedicating all hardware resources to the encrypted control scheme. Therefore, implementation of encrypted control on an FPGA with the reformulated FHE scheme allows for higher update rates than an implementation on conventional hardware, or with the original FHE scheme.

4.3.2 Experimental Setup

The encrypted control scheme has been implemented on two Nexys 4 FPGAs in the configuration as shown in Figure 4.2. The control loop in Figure 4.2 works as follows: At initialisation FPGA 1, the adapter, generates an encryption key pair and uses this to encrypt the controller gain matrices. These are then shared with the remote encrypted controller in FPGA 2. Then, for each control update, first, the adapter encrypts measurement vector y and sends it to the controller, which computes the control input $\mathbf{E}(u^+)$ and state estimate $\mathbf{E}(\hat{x}^+)$. These signals are then sent back to the adapter for decryption, where the control input is applied to the plant. Additionally, u^+ and $\hat{x}^+$ are bit-shifted, re-encrypted and sent back to the controller along with the new measurements $\mathbf{E}(y)$. The additional communication of $\mathbf{E}(\hat{x}^+)$ and $\mathbf{E}(u)$ is required to extend multiplicative depth of the homomorphic operations. This obstacle of FHE is explained in more detail in Section 4.1.3. This solution to extend multiplicative depth of the homomorphic operations is inspired by [110] where the same principle is used to overcome the lack of homomorphic in the ElGamal PHE scheme [108].

The chosen plant is the inverted double pendulum depicted in Figure 4.3. The dynamics of the double pendulum's state $\theta = [\theta_1 \ \theta_2]^\top$ is modeled as

$$\begin{cases} M(\theta)\ddot{\theta} + C(\theta,\dot{\theta})\dot{\theta} + G(\theta) = T, \\ T + \tau_e \dot{T} = k_m u \end{cases} \tag{4.20}$$

$$M(\theta) = \begin{bmatrix} P_1 + P_2 + 2P_3\cos\theta_2 & P_2 + P_3\cos\theta_2 \\ P_2 + P_3\cos\theta_2 & P_2 \end{bmatrix}$$

$$C(\theta,\dot{\theta}) = \begin{bmatrix} b_1 - P_3\dot{\theta}_2\sin\theta_2 & -P_3(\dot{\theta}_1+\dot{\theta}_2)\sin\theta_2 \\ P_3\dot{\theta}_1\sin\theta_2 & b_2 \end{bmatrix} \quad G(\theta) = \begin{bmatrix} -g_1\sin\theta_1 - g_2\sin(\theta_1+\theta_2) \\ -g_2\sin(\theta_1+\theta_2) \end{bmatrix}$$

$$P_1 = m_1c_1^2 + m_2l_1^2 + I_1,\ P_2 = m_2c_2^2 + I_2,\ P_3 = m_2l_1c_2,\ g_1 = (m_1c_1 + m_2l_1)g,\ g_2 = m_2c_2g$$

where θ_1 and θ_2 denote the angles of the pendulum links as shown in Figure 4.3. Furthermore, m_1,m_2 are the masses of the links; l_1, l_2 are their lengths; c_1, c_2 are the centers of mass; I_1, I_2 are the mass moments of inertia; b_1,b_2 are the damping coefficients of the joints; k_m,τ_e are the electrical motor gain and time constant, and g is the gravitational acceleration. All these model parameters can be found in Table 4.2.

Next we will present the design of a controller that can attain reference tracking of the double inverted pendulum around $\theta = \dot{\theta} = [0\ 0]^\top$, i.e. around the unstable equilibrium where both pendulums are upright. To this end, a discrete time linearization of the model

in Equation (4.20) around $\theta = \dot{\theta} = [0\ 0]^\top$ is made as

$$\begin{cases} x(k+1) = A_d x(k) + B_d u(k), \\ y(k) = C_d x(k), \end{cases} \tag{4.21}$$

where $x(k) = [\theta_1(k)\ \dot{\theta}_1(k)\ \theta_2(k)\ \dot{\theta}_2(k)\ T]^\top$, $y(k) = [\theta_1(k)\ \theta_2(k)]^\top$, and A_d, B_d, and C_d are matrices of appropriate size. This linearized model is used to implement an observer and state feedback controller as

$$\begin{cases} \hat{x}(k+1) = A_d \hat{x}(k) + B_d u(k) + L(y(k) - C_d \hat{x}(k)), \\ u(k+1) = K\hat{x}(k+1), \end{cases} \tag{4.22}$$

where L is the observer gain and K is the state feedback gain. The controller is updated at a rate of $f = 100\ Hz$. L has been obtained by placing the observer poles at [0.7 0.5 0.8 0.6 0.85] and $K = [-12.6\ -1.8\ -9.8\ -0.95\ 0.015]$. The input and the state estimate are initialized at $u(1) = 0$ and $\hat{x}(0) = 0$. The controller in Equation (4.22) is encrypted to obtain the equivalent controllers $\tilde{\mathfrak{g}}(\cdot)$ and $\tilde{\mathfrak{v}}(\cdot)$ of the form in Equation (4.3). The encryption parameters used can be found in Table 4.2.

Table 4.2: Model and encryption parameters

Parameter	Value	Parameter	Value	Parameter	Value
m_1	$0.125\ kg$	m_2	$0.05\ kg$	g	$9.81 ms^{-2}$
l_1	$0.1\ m$	l_2	$0.1\ m$	n	7
c_1	$-0.04\ m$	c_2	$0.06\ m$	ℓ	64
I_1	$0.074\ kgm^2$	I_2	$0.00012\ kgm^2$	m	7
b_1	$4.8\ kgs^{-1}$	b_2	$0.0002\ kgs^{-1}$	m_q	10
k_m	$50\ Nm$	τ_e	$0.03\ s$	n_q	22

4.3.3 Performance

Figure 4.4 shows the results of the system in Equation (4.20) being controlled according to $\tilde{\mathfrak{g}}(\cdot), \tilde{\mathfrak{v}}(\cdot)$. To obtain these results the double pendulum is initialized at an initial state $\theta = \theta_0^\top = [0.0289, 0.1156]^\top$, $\dot{\theta} = \dot{\theta}_0{}^\top = [0.0669, 0.0049]^\top$ and $T = T_0 = 0$, and is controlled such that both pendulums point upwards, i.e. $\theta = [0\ 0]^\top$.

The results shown in Figure 4.4 are obtained using a hardware simulation of the FPGAs coupled to a high resolution simulation of the double pendulum. With this implementation of the encrypted control on the FPGAs a new control input can be generated every $0.8\ ms$, which is more than sufficient for the needed update rate of $100\ Hz$. Achieving such update rate would not have been possible using the original cipher or on conventional hardware.

One can see in Figure 4.4 that the encrypted observer estimates the states correctly and the plant is stabilized by the encrypted controller. Controlling the plant without encryption, i.e. according to Equation (4.22), yields identical results. This illustrates that the encrypted plain-text controllers are indeed equivalent. The experimental setup serves to highlight the contributions made to FHE. One can see that the plant is controlled towards an unstable equilibrium which requires a fast update rate of the encrypted controller. Due to the use of the *reduced cipher*, this has become possible on the chosen hardware (two Nexys 4 FPGAs).

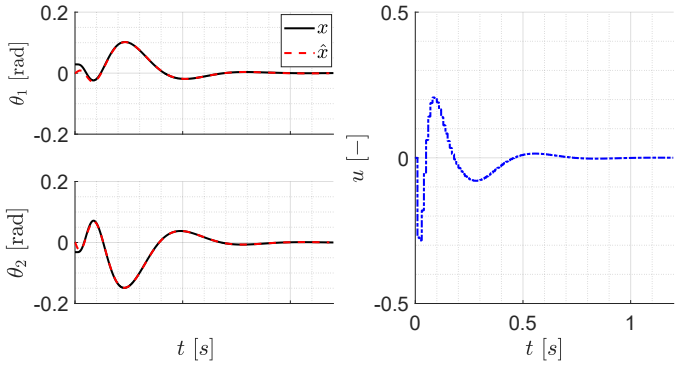


Figure 4.4: Simulation results, θ_1, θ_2 and control effort u

4

4.4 Conclusion

With the development of large scale CPSs, encryption, which was originally developed for IT systems, is increasingly utilised in control systems to prevent intrusion of remote connections. The current industrial state of practice is to use so-called *end-to-end encryption*, which requires decryption, and re-encryption, of the signals that are processed at the controller. This forms a risk to security, as intrusion in the controller no longer prevented. Furthermore, these decryption and re-encryption steps create computational overhead, limiting the controller update rate and reducing the stability margin.

FHE has been developed such that operations can be performed on encrypted signals. Therefore, it has the potential to close the loop of encryption for secure control. The main obstacle to widespread implementation of FHE in control is the high computational complexity. In this chapter, the GSW FHE scheme has been reformulated using the so-called *reduced cipher*. This allows for reducing the computational complexity of the scheme, while preserving its theoretical properties. Specifically, the need for hardware multiplications to perform homomorphic operations is completely eliminated and the total number of hardware operations is reduced by an order of magnitude. Furthermore the *reduced cipher* allows for more intuitive interpretation and manipulation of the scheme, opening the door to further improvements.

This reformulated FHE scheme has been implemented on an FPGA for real-time control of an inverted double pendulum. A standard digital control law has been encrypted with FHE and is able to stabilize the plant around its unstable equilibrium. The controller computation time is 0.8 ms on a Nexys 4 FPGA, which is more than sufficient for the chosen 100 Hz update rate.

The presented FHE scheme is, to the best of the authors knowledge, the first FHE scheme that has been implemented for real-time control of an unstable plant. In future work it is interesting to extend these results to larger and more complex plants to show the full capability of the scheme. Furthermore, it is interesting to use the reformulation of the scheme to explore how to perform right hand bit shifts and other operations like division on encrypted data. This would be an elegant solution to the problem of overflow by consecutive multiplications of encrypted fixed precision numbers.

5

Conclusion & Discussion

In the last two decades cyber-physical systems (CPSs) have increasingly become a part of industrial processes and consumer products. CPSs are extremely useful as they provide the combined advantages of their physical and cyber components, allowing for coordination of the actions of the physical components. However, this interaction between physical and cyber components also causes CPSs to be burdened with both safety and security risks. Therefore, research on safety and security risks is especially interesting for CPSs.

In this dissertation three advances within the field of safety and security of CPS have been presented. These advances each focus on one or more methods to perform mitigation of safety and security risks. Namely, Chapter 2 addresses sliding mode observer (SMO)-based anomaly detection, Chapter 3 presents a guaranteed safe control law which integrates resilience, detection and accommodation of anomalies in a collaborative vehicle platoon (CVP). Finally, Chapter 4 introduces an encrypted control approach based on fully homomorphic encryption (FHE) in order to prevent cyber anomalies. The contributions of this dissertation within each of these topics will be summarised in Section 5.1. Then, these contributions are placed in the context of the broader field of safety and security of CPS in Section 5.2. Lastly, some recommendations for future work are presented in Section 5.3.

5.1 Contributions to Safety & Security of CPS

In Chapter 2, two anomaly detection methods are presented which provide detection capabilities for a large class of existing SMOs, while preserving their inherent state and anomaly estimation capabilities. The first method, presented in Section 2.2, performs detection using thresholds on the equivalent output injection (EOI), which is related to the anomaly estimate. The second method, presented in Section 2.3, performs detection by comparing upper and lower thresholds on the state estimation error. By design, both detection methods are free of false detection. Furthermore, strong guarantees on detectability are provided.

In Chapter 3 a solution for safe control of CVPs under man-in-the-middle (MITM) attacks is presented. The control law is based on coalitional model predictive control (MPC), where optimal coalitions of vehicles are determined online based on performance and safety criteria. Nominally, the CVP is controlled to trade-off tracking performance with communication and computational load, while also maintaining safety and string stability

under uncertainty. When the CVP is under attack, however, the priorities of the control law shift to primarily maintaining safety. A reduced unknown input observer (R-UIO) based detection method is implemented to detect any sufficiently large MITM attack. The communication topology is then changed to disable the attacked communication link. Meanwhile the MPC is designed to be resilient to undetected attacks and the uncertainty caused by disabled communication links. It is shown that this control approach is guaranteed to be safe and even retains good tracking performance while the CVP is under attack.

In Chapter 4 the GSW FHE scheme has been reformulated to increase its computational efficiency. The reformulation allows for reduction of the total number of hardware operations by orders of magnitude and completely eliminates the need for multiplications at the hardware level. The reformulated FHE scheme has then been implemented on two field-programmable gate arrays (FPGAs) for real-time remote encrypted control of an unstable system. Such implementation would not have been possible using the original formulation or on conventional hardware.

5.2 Significance and Limitations of the Contributions

The contributions of this dissertation all address a specific topic within safety and security of CPS, and have been discussed as such in detail throughout the dissertation. In this section, however, we will zoom out and discuss the significance and limitations of the contributions in view of the broader field of safety and security of CPS.

SMO based Anomaly Detection

The presented SMO based detection methods, depending on how they are configured, allow for detection of physical anomalies and cyber anomalies affecting data integrity. This is not limiting within the field of model based detection. This because cyber anomalies affecting data availability are trivially detected and those affecting data confidentiality cannot be detected by any model based detection scheme. Yet, within the broader field, model based detection is only a small part of a solution providing safety and security of CPS.

The presented SMO based detectors are, however, applicable to a general class of linear systems and first order sliding mode observers (FOSMOs). Furthermore, an initial extension to a class of non-linear single-input single-output (SISO) systems is presented. Additionally, literature on FOSMOs for general non-linear systems suggests further extensions are possible too. Next to the broad applicability, the detectors are computationally efficient and can provide for fast detection. This combination of general applicability, computational efficiency and fast detection makes the detectors suitable in many scenarios.

Despite their advantages, the SMO based detectors are not suitable to detect covert, zero-dynamics, or replay attacks. This is a major limitation inherent to many model-based detection methods [36, 61, 64, 68, 213]. There are, however, recent works that aim to address this problem. For example, (dynamic) watermarking has been proposed to allow for detecting replay attacks [66, 307] and stealthy attacks [91]. And in [35], two complementary model-based detectors are used to detect covert attacks.

Topology-Switching Control for Anomaly Accommodation in CVPs

The presented topology-switching coalitional MPC method with R-UIO based anomaly detection provides resilience, detection and accommodation of anomalies in a CVP. The scheme mainly considers malicious physical anomalies and cyber-anomalies affecting integrity and availability of communicated measurements, but is also applicable to accidental physical anomalies and cyber anomalies affecting integrity and availability in the controller. This means only cyber anomalies affecting confidentiality are not addressed by this method.

The presented method, thus comes quite close to providing an integrated solution to safety and security of CVPs. This does, however, come with some limitations on its applicability. Most importantly, the scheme assumes that all vehicles in the CVP use the same control law, such that vehicles within the same coalition know each others control input without communicating it.[1] Furthermore, for all vehicles within a coalition to know each others control input, each vehicle needs to solve the MPC problem for the whole coalition. This increases computational complexity of the MPC problem, which is already inherently computationally expensive due to the required optimisation.

Additionally, disabling communication as a means of anomaly accommodation, has some limitations too. Not so for any anomaly that directly affects the communication, e.g. a MITM or denial of service (DoS) attack, or packet-loss. Here the accommodation addresses the source of the problem and provides safety and security from it. However, for anomalies that affect the vehicles themselves the accommodation is less suitable. Even though the accommodation still provides safety for these anomalies, the accommodation does not address their source. For malicious anomalies to vehicles one might argue that the communication from that vehicle then also cannot be trusted. However, for accidental anomalies to a vehicle, such as engine failure, it is clear that it is in the interest of the whole platoon to aid the affected vehicle instead of ignoring it.

Anomaly Prevention through FHE

The presented FHE based encrypted controller allows for prevention of cyber-anomalies affecting confidentiality of data during communication as well as in the controller. Additionally, any anomaly affecting the integrity of the cipher will have an unpredictable and typically large impact on the corresponding plain-text, allowing for trivial detection of cyber-anomalies affecting integrity. Therefore, although typically encryption is classified as a prevention method, for anomalies affecting integrity it is better classified as a detection method. This means that the presented FHE scheme, when not combined with a form of anomaly accommodation, can only address cyber-anomalies affecting confidentiality.

However, unlike traditional encryption schemes, FHE does provide its prevention and detection capabilities against anomalies on the communication as well as in the controller. Unfortunately, FHE does still have quite some limitations on its applicability. Firstly, FHE only allows for encrypting controllers consisting of addition and multiplication operations. Secondly, FHE has a high computational load, even after the reformulation of the scheme that is presented in this dissertation. Therefore, in its current form, encrypted control using FHE can only be implemented in real-time on dedicated hardware such as an FPGA.

[1] This limitation is not trivially resolved by adding communication of the control input as the communication channels are possibly subject to anomalies.

Additionally, a problem that is not addressed in this dissertation is that of communication of the ciphers between the plant and the controller. The reformulation of FHE presented in this dissertation reduces its computational load, but the ciphers still contain the same amount of information and their load on communication is still high. Furthermore, the solution to extend multiplicative depth implemented in this dissertation requires additional communication, adding to the problem. As the reduced computational load allows for higher update rates, while the load on the communication is unchanged, this will likely become the new bottleneck limiting the performance.

An Overview of Contributions

		Anomaly		Applicability		Comp. load
		Cyber	Physical	Plant	Controller	
SMO (Chapter 2)	Prevention			Linear (+ nonlinear SISO)	Any	Low
	Resilience					
	Detection	I	x			
	Accommodation					
Topology-Switching MPC (Chapter 3)	Prevention			CVP	Same MPC in all vehicles	Medium
	Resilience	IA	x			
	Detection	IA	x			
	Accommodation[2]	IA	x			
FHE (Chapter 4)	Prevention	C		Any	Only addition and products	High
	Resilience					
	Detection	I				
	Accommodation					

Table 5.1: A high level overview of the contributions in this dissertation. Here C, I, and A refer to Confidentiality, Integrity and Availability as defined in Definitions 1.5 to 1.7.

The contributions presented in this dissertation each contribute to safety and security in CPS by addressing different anomalies through prevention, resilience, detection, accommodation or a combination thereof. In Table 5.1 an overview of the contributions of this dissertation is given, showing the addressed anomalies and type of mitigation as well as the applicability and computational load.

One can see that each contribution addresses safety and security in a different way. Especially, the SMO based detector addresses only detection for a large class of systems, while the topology-switching MPC addresses resilience, detection and accommodation for a large class of anomalies, but only for a single system. Furthermore, one can see that none of the methods address all anomalies. Specifically, confidentiality is the only anomaly not addressed by the topology-switching MPC. This anomaly is, however, addressed by the FHE based encrypted control. Unfortunately, it is currently not possible to implement FHE-based encrypted MPC. This is mainly due to the limited applicability of the FHE

[2]The accommodation method is well suited for e.g. a MITM attack. However, for physical failures such as an engine failure another method, where the faulty vehicle is assisted instead of disconnected, might be preferable.

encryption to different types of control. Specifically it relates to the inability to evaluate *if* statements. There is however research on partially homomorphic encryption (PHE) based encrypted explicit MPC, where only the part of the controller that can be encrypted is performed remotely [113]. The remaining computations, among which the *if* statements, are performed in plain-text at the plant. A similar solution might allow for FHE based encrypted MPC in the future.

5.3 Recommendations for Future Work

Following the contributions in this dissertation, in this section some recommendations for future work are presented.

SMO based Anomaly Detection

- ***Extend the presented detectors to be applicable to higher order SMOs and a larger class of nonlinear systems:*** The broad applicability of the presented detectors is a great benefit. However, it has not yet reached its full potential. Therefore, it is recommended to investigate the applicability of the detectors to a larger class of nonlinear systems. Furthermore, in the last decade higher order SMOs are being introduced to increase applicability and estimation performance compared to FOSMOs. Therefore, also the detectors stand to greatly benefit from application to such higher order SMOs. It is thus very interesting to explore applicability of the detectors to such higher order SMOs.

- ***Reformulate the detector designs to allow for stochastic analysis of robustness and detectability:*** It is common for anomaly detectors to make a trade-off between robustness and detectability. The detectors presented in this dissertation don't allow for such a trade-off as they are designed based on bounds on the uncertainty, providing deterministic guarantees on robustness and detectability. By considering the uncertainties in a stochastic framework and propagating the stochastic variables through the SMO based detector, also stochastic guarantees on robustness and detectability can be obtained. This will allow for making the common trade-off between robustness and detectability that is not possible in the deterministic setting.

- ***Formalise the performance comparison between the two presented SMO based anomaly detectors:*** Two detectors based on SMOs have been presented in this dissertation. Their detection performance is currently only compared though a simulation of a CVP subject to a MITM attack. A formal theoretical comparison can reveal whether the findings of this single simulation example extend in general.

- ***Utilise the SMO anomaly estimation capabilities to design a safety preserving control law that integrates resilience, detection and accommodation:*** One of the main advantages of using SMOs for anomaly detection is that they can concurrently also estimate state and anomaly. This additional information can be used to perform accommodation of detected anomalies. Furthermore, the detectors have strong detectability guarantees, which can serve as input to the design of a controller that is resilient against undetected anomalies. These properties, however, have not yet been utilised to design a safety preserving controller.

Topology-Switching Control for Anomaly Accommodation in CVPs

- ***Increase the flexibility of the definition of coalitions to allow for non-consecutive coalitions and incorporate this in the controller design:*** The presented scheme strongly relies on the ability to change the communication topology to increase nominal performance and guarantee safety and security from anomalies. Therefore, it is important to allow for full flexibility in defining the coalitions. The proposed topology-switching coalitional controller disables affected communication links. With the currently used definition of coalitions, this means communication is also no longer possible between the vehicles before and after the affected vehicle. Such communication might, however, still be beneficial to the performance of the CVP. By making non-consecutive coalitions possible, this potential can be utilised.

- ***Reduce computational load of the scheme by minimizing the amount of overlapping calculation in each vehicle:*** In the proposed topology-switching coalitional controller each vehicle within a coalition calculates the control action of the whole coalition. Reducing the overlap in calculations between vehicles, e.g. by using a consensus based approach, can aid in decreasing the computational complexity.

- ***Reduce the amount of R-UIO gains to be pre-calculated and saved:*** In the proposed controller an R-UIO based anomaly detection is used. The R-UIO gains for all possible communication topologies are pre-calculated and saved in each vehicle. Especially for larger platoons the computational load and memory usage of this approach becomes problematic as the amount of possible communication topologies grows exponentially with the number of vehicles. It is therefore recommended to reduce the amount of R-UIO gains to be pre-calculated and saved. Possible directions that can be investigated to achieve this are making the R-UIO gains modular or by parameterised based on coalition size and vehicle location.

Anomaly Prevention through FHE

- ***Find solutions to use FHE to partially encrypt more complex control methods while not compromising security:*** In the presented design the encrypted controller is strictly constrained to only use addition and multiplication. There are however methods to only partially encrypt a control law without compromising security. An example of this for MPC using PHE is presented in [113]. It is interesting to study how a similar approach can be used in combination with FHE.

- ***Find solutions to use FHE based encrypted control in distributed control systems:*** In the presented design of the encrypted controller the adapter at the plant is a trusted agent that knows the private key for decryption. Secure control occurs at another location, where the private key is not known. In general, using FHE, always at least one trusted agent is required to know the private key. Therefore, this approach is not directly applicable to distributed control systems, where typically all local systems perform similar computations and actions. It is thus interesting to find extensions of FHE based encrypted control to distributed control systems. Here, one can take inspiration from secure multi-party computation (sMPC)[128–130].

Bibliography

References

[1] James P. Farwell and Rafal Rohozinski. Stuxnet and the future of cyber war. *Survival*, 53(1):23–40, 2011.

[2] Ralph Langner. Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security & Privacy*, 9(3):49–51, 2011.

[3] Chris Valasek and Charlie Miller. Remote exploitation of an unaltered passenger vehicle. In *Blackhat*, 2015.

[4] Robert M. Lee, Michael J. Assante, and Tim Conway. Analysis of the cyber attack on the ukrainian power grid. Technical report, Electricity Information Sharing and Analysis Center (EISAC), 2016.

[5] Julia Nilsson, Mattias Brännström, Erik Coelingh, and Jonas Fredriksson. Lane change maneuvers for automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 18(5):1087–1096, 2017.

[6] Dale Richards and Alex Stedmon. To delegate or not to delegate: A review of control frameworks for autonomous cars. *Applied Ergonomics*, 53:383–388, 2016. Transport in the 21st Century: The Application of Human Factors to Future User Needs.

[7] Tom van der Sande and Henk Nijmeijer. *From Cooperative to Autonomous Vehicles*, pages 435–452. Springer International Publishing, Cham, 2017.

[8] Jeroen Ploeg, Bart M. T. Scheepers, Ellen van Nunen, Nathan van de Wouw, and Henk Nijmeijer. Design and experimental evaluation of cooperative adaptive cruise control. In *Conference on Intelligent Transportation Systems*, pages 260–265, 2011.

[9] Twan Keijzer and Riccardo M.G. Ferrari. A sliding mode observer approach for attack detection and estimation in autonomous vehicle platoons using event triggered communication. In *Conference on Decision and Control (CDC)*, pages 5742–5747, 2019.

[10] Karl Koscher, Alexei Czeskis, Franziska Roesner, Shwetak Patel, Tadayoshi Kohno, Stephen Checkoway, Damon McCoy, Brian Kantor, Danny Anderson, Hovav Shacham, and Stefan Savage. Experimental security analysis of a modern automobile. In *2010 IEEE Symposium on Security and Privacy*, pages 447–462, 2010.

[11] Aengus Collins. The global risks report 2018. Technical report, World Economic Forum, 2018.

[12] H.M. Besch, H.G. Giesseler, and J. Schuller. Impact of electronic flight control system (efcs) failure cases on structural design loads. In *83rd Meeting of the AGARD Structures and Materials Panel*, page 14, 1996.

[13] Philippe Goupil. Oscillatory failure case detection in the a380 electrical flight control system by analytical redundancy. *Control Engineering Practice*, 18(9):1110–1119, 2010.

[14] Ludovic Piètre-Cambacédès and Claude Chaudet. The sema referential framework: Avoiding ambiguities in the terms "security" and "safety". *International Journal of Critical Infrastructure Protection*, 3(2):55–66, 2010.

[15] Maria B. Line, Odd Nordland, Lillian Røstad, and Inger Anne Tøndel. Safety vs. Security? In *Proceedings of the Eighth International Conference on Probabilistic Safety Assessment & Management (PSAM)*. ASME Press, 01 2006.

[16] Michelle S. Chong, Henrik Sandberg, and André M.H. Teixeira. A tutorial introduction to security and privacy for cyber-physical systems. In *2019 18th European Control Conference (ECC)*, pages 968–978, 2019.

[17] Seyed Mehran Dibaji, Mohammad Pirani, David Bezalel Flamholz, Anuradha M. Annaswamy, Karl Henrik Johansson, and Aranya Chakrabortty. A systems and control perspective of cps security. *Annual Reviews in Control*, 47:394–411, 2019.

[18] Artem A. Nazarenko and Ghazanfar Ali Safdar. Survey on security and privacy issues in cyber physical systems. *AIMS Electronics and Electrical Engineering*, 3(2):111–143, 2019.

[19] Johannes Geismann and Eric Bodden. A systematic literature review of model-driven security engineering for cyber–physical systems. *Journal of Systems and Software*, 169:110697, 2020.

[20] George K. Fourlas and George C. Karras. A survey on fault diagnosis and fault-tolerant control methods for unmanned aerial vehicles. *Machines*, 9(9), 2021.

[21] Derui Ding, Qing-Long Han, Xiaohua Ge, and Jun Wang. Secure state estimation and control of cyber-physical systems: A survey. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(1):176–190, 2021.

[22] Chen Gao, Xiao He, Hongli Dong, Hongjian Liu, and Guangran Lyu. A survey on fault-tolerant consensus control of multi-agent systems: trends, methodologies and prospects. *International Journal of Systems Science*, 0(0):1–14, 2022.

[23] Brett T. Stewart, Aswin N. Venkat, James B. Rawlings, Stephen J. Wright, and Gabriele Pannocchia. Cooperative distributed model predictive control. *Systems and Control Letters*, 59:460–469, 8 2010.

[24] Andong Liu, Wen An Zhang, Li Yu, Huaicheng Yan, and Rongchao Zhang. Formation control of multiple mobile robots incorporating an extended state observer and distributed model predictive approach. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50:4587–4597, 11 2020.

[25] Qin Liu, Hossam Seddik Abbas, and Javad Mohammadpour Velni. An lmi-based approach to distributed model predictive control design for spatially-interconnected systems. *Automatica*, 95:481–487, 9 2018.

[26] Paul A. Trodden and J. M. Maestre. Distributed predictive control with minimization of mutual disturbances. *Automatica*, 77:31–43, 3 2017.

[27] Yang Zheng, Shengbo Eben Li, Keqiang Li, Francesco Borrelli, and J. Karl Hedrick. Distributed model predictive control for heterogeneous vehicle platoons under unidirectional topologies. *Transactions on Control Systems Technology*, 25:899–910, 5 2017.

[28] Anca Maxim and Constantin Florin Caruntu. Coalitional distributed model predictive control strategy for vehicle platooning applications. *Sensors*, 22, 2 2022.

[29] Tieshan Li, Rong Zhao, C. L. Philip Chen, Liyou Fang, and Cheng Liu. Finite-time formation control of under-actuated ships using nonlinear sliding mode control. *IEEE Transactions on Cybernetics*, 48(11):3243–3253, 2018.

[30] Francesco Borrelli and Tamás Keviczky. Distributed lqr design for identical dynamically decoupled systems. *Transactions on Automatic Control*, 53(8):1901–1912, 2008.

[31] Dimos V. Dimarogonas and Emilio Frazzoli. Distributed event-triggered control strategies for multi-agent systems. In *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 906–910, 2009.

[32] Anders Rantzer. Dynamic dual decomposition for distributed control. In *2009 American Control Conference*, pages 884–888, 2009.

[33] Martin Jilg and Olaf Stursberg. Optimized distributed control and topology design for hierarchically interconnected systems. In *2013 European Control Conference (ECC)*, pages 4340–4346, 2013.

[34] Paula Chanfreut, José M. Maestre, and Eduardo F. Camacho. A survey on clustering methods for distributed and networked control systems. *Annual Reviews in Control*, 52:75–90, 2021.

[35] Alexander J. Gallo, Mustafa Sahin Turan, Francesca Boem, Thomas Parisini, and Giancarlo Ferrari-Trecate. A distributed cyber-attack detection scheme with application to dc microgrids. *Transactions on Automatic Control*, 65:3800–3815, 9 2020.

[36] Alexander J. Gallo, Angelo Barboni, and Thomas Parisini. On detectability of cyber-attacks for large-scale interconnected systems. In *IFAC-PapersOnLine*, volume 53, pages 3521–3526. Elsevier B.V., 2020.

[37] Paula Chanfreut, José María Maestre, and Eduardo F. Camacho. Coalitional model predictive control on freeways traffic networks. *IEEE Transactions on Intelligent Transportation Systems*, 22(11):6772–6783, 2021.

[38] Paula Chanfreut, Twan Keijzer, Riccardo M.G. Ferrari, and Jose Maria Maestre. A topology-switching coalitional control and observation scheme with stability guarantees. *IFAC-PapersOnLine*, 53(2):6477–6482, 2020. 21st IFAC World Congress.

[39] Keqiang Li, Yougang Bian, Shengbo Eben Li, Biao Xu, and Jianqiang Wang. Distributed model predictive control of multi-vehicle systems with switching communication topologies. *Transportation Research Part C: Emerging Technologies*, 118, 9 2020.

[40] Marcello Farina, Xinglong Zhang, and Riccardo Scattolini. A hierarchical multi-rate mpc scheme for interconnected systems. *Automatica*, 90:38–46, 4 2018.

[41] Lakshmi Dhevi Baskar, Bart De Schutter, and Hans Hellendoorn. Model predictive control for intelligent speed adaptation in intelligent vehicle highway systems. In *2008 IEEE International Conference on Control Applications*, pages 468–473, 2008.

[42] Brett T. Stewart, James B. Rawlings, and Stephen J. Wright. Hierarchical cooperative distributed model predictive control. In *Proceedings of the 2010 American Control Conference*, pages 3963–3968, 2010.

[43] Riccardo Scattolini. Architectures for distributed and hierarchical model predictive control – a review. *Journal of Process Control*, 19(5):723–731, 2009.

[44] Gerrit J.L. Naus, René P.A. Vugts, Jeroen Ploeg, Marinus J.G. Van De Molengraft, and Maarten Steinbuch. String-stable cacc design and experimental validation: A frequency-domain approach. *IEEE Transactions on Vehicular Technology*, 59:4268–4279, 11 2010.

[45] Yuanheng Zhu, Haibo He, and Dongbin Zhao. Lmi-based synthesis of string-stable controller for cooperative adaptive cruise control. *IEEE Transactions on Intelligent Transportation Systems*, 21:4516–4525, 11 2020.

[46] Roozbeh Kianfar, Paolo Falcone, and Jonas Fredriksson. A control matching model predictive control approach to string stable vehicle platooning. *Control Engineering Practice*, 45:163–173, 12 2015.

[47] William B. Dunbar and Derek S. Caveney. Distributed receding horizon control of vehicle platoons: Stability and string stability. *IEEE Transactions on Automatic Control*, 57:620–633, 3 2012.

[48] Elham Semsar-Kazerooni, Jan Verhaegh, Jeroen Ploeg, and Mohsen Alirezaei. Cooperative adaptive cruise control: An artificial potential field approach. In *Intelligent Vehicles Symposium*, pages 361–367, 2016.

[49] E. Semsar-Kazerooni, K. Elferink, J. Ploeg, and H. Nijmeijer. Multi-objective platoon maneuvering using artificial potential fields. *IFAC-PapersOnLine*, 50(1):15006–15011, 2017.

[50] Theodore Willke, Patcharinee Tientrakool, and Nicholas Maxemchuk. A survey of inter-vehicle communication protocols and their applications. *IEEE Communications Surveys and Tutorials*, 11:3–20, 2009.

[51] Ejaz Ahmed and Hamid Gharavi. Cooperative vehicular networking: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 19(3):996–1014, 2018.

[52] Roberto Merco, Francesco Ferrante, and Pierluigi Pisu. Network aware control design for string stabilization in vehicle platoons: An lmi approach. In *Proceedings of the American Control Conference*, pages 539–544. IEEE, 2019.

[53] Cedric Levy-Bencheton and Eleni Darra. Cyber security and resilience of intelligent public transport. Technical report, European Union Agency for Cyber-Security, 2015.

[54] J. Harding, G.R. Powell, R. Yoon, J. Fikentscher, C. Doyle, D. Sade, M. Lukuc, J. Simons, and J. Wang. Vehicle-to-vehicle communications: Readiness of v2v technology for application. Technical report, NHTSA, 8 2014.

[55] Kim Hartmann and Christoph Steup. The vulnerability of uavs to cyber attacks - an approach to the risk assessment. In *2013 5th International Conference on Cyber Conflict (CYCON 2013)*, pages 1–23, 2013.

[56] Chaitanya Rani, Hamidreza Modares, Raghavendra Sriram, Dariusz Mikulski, and Frank L Lewis. Security of unmanned aerial vehicle systems against cyber-physical attacks. *The Journal of Defense Modeling and Simulation*, 13(3):331–342, 2016.

[57] Samuel Woo, Hyo Jin Jo, and Dong Hoon Lee. A practical wireless attack on the connected car and security protocol for in-vehicle can. *IEEE Transactions on Intelligent Transportation Systems*, 16(2):993–1006, 2015.

[58] Nicolas Falliere, Liam O. Murchu, and Eric Chien. W32.stuxnet dossier. Technical report, Symantec, 2011.

[59] Jill Slay and Michael Miller. Lessons learned from the maroochy water breach. In Eric Goetz and Sujeet Shenoi, editors, *Critical Infrastructure Protection*, pages 73–82, Boston, MA, 2008. Springer US.

[60] Thomas Miller, Alexander Staves, Sam Maesschalck, Miriam Sturdee, and Benjamin Green. Looking back to look forward: Lessons learnt from cyber-attacks on industrial control systems. *International Journal of Critical Infrastructure Protection*, 35:100464, 2021.

[61] André Teixeira, Iman Shames, Henrik Sandberg, and Karl Henrik Johansson. A secure control framework for resource-limited adversaries. *Automatica*, 51:135–148, 2015.

[62] Abhishek Gupta, Cédric Langbort, and Tamer Başar. Optimal control in the presence of an intelligent jammer with limited actions. In *49th IEEE Conference on Decision and Control (CDC)*, pages 1096–1101, 2010.

[63] Zoliekha Abdollahi Biron, Satadru Dey, and Pierluigi Pisu. Resilient control strategy under denial of service in connected vehicles. In *Proceedings of the American Control Conference*, pages 4971–4976. Institute of Electrical and Electronics Engineers Inc., 6 2017.

[64] Yilin Mo, Sean Weerakkody, and Bruno Sinopoli. Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Systems Magazine*, 35(1):93–109, 2015.

[65] Takashi Irita and Toru Namerikawa. Detection of replay attack on smart grid with code signal and bargaining game. In *2017 American Control Conference (ACC)*, pages 2112–2117, 2017.

[66] Riccardo M.G. Ferrari and André M.H. Teixeira. Detection and isolation of replay attacks through sensor watermarking. *IFAC-PapersOnLine*, 50(1):7363–7368, 2017.

[67] Mohammad Deghat, Valery Ugrinovskii, Iman Shames, and Cédric Langbort. Detection and mitigation of biasing attacks on distributed estimation networks. *Automatica*, 99:369–381, 1 2019.

[68] Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, 2013.

[69] Henrik Sandberg and André M.H. Teixeira. From control system security indices to attack identifiability. In *2016 Science of Security for Cyber-Physical Systems Workshop (SOSCYPS)*, pages 1–6, 2016.

[70] Michelle S. Chong and Margreta Kuijper. Characterising the vulnerability of linear control systems under sensor attacks using a system's security index. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 5906–5911, 2016.

[71] Zhanghan Tang, Margreta Kuijper, Michelle S. Chong, Iven Mareels, and Christopher Leckie. Linear system security—detection and correction of adversarial sensor attacks in the noise-free case. *Automatica*, 101:53–59, 2019.

[72] André M. H. Teixeira. *Security Metrics for Control Systems*, pages 99–121. Springer International Publishing, Cham, 2021.

[73] Sribalaji C. Anand, André M. H. Teixeira, and Anders Ahlén. Risk assessment of stealthy attacks on uncertain control systems, 2021.

[74] Roy S. Smith. A decoupled feedback structure for covertly appropriating networked control systems. *IFAC Proceedings Volumes*, 44(1):90–95, 2011. 18th IFAC World Congress.

[75] Henrik Sandberg, André Teixeira, and K. H. Johansson. On security indices for state estimators in power networks. In *Preprints of the First Workshop on Secure Control Systems, CPSWEEK 2010, Stockholm, Sweden*, 2010.

[76] Hilary E. Brown and Christopher L. Demarco. Risk of cyber-physical attack via load with emulated inertia control. *IEEE Transactions on Smart Grid*, 9(6):5854–5866, 2018.

[77] Peyman Mohajerin Esfahani, Maria Vrakopoulou, Kostas Margellos, John Lygeros, and Göran Andersson. Cyber attack in a two-area power system: Impact identification using reachability. In *Proceedings of the 2010 American Control Conference*, pages 962–967, 2010.

[78] Yilin Mo, Tiffany Hyun-Jin Kim, Kenneth Brancik, Dona Dickinson, Heejo Lee, Adrian Perrig, and Bruno Sinopoli. Cyber–physical security of a smart grid infrastructure. *Proceedings of the IEEE*, 100(1):195–209, 2012.

[79] Quanyan Zhu, Linda Bushnell, and Tamer Başar. Game-theoretic analysis of node capture and cloning attack with multiple attackers in wireless sensor networks. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 3404–3411, 2012.

[80] Paula Chanfreut, J.M. Maestre, and H. Ishii. Vulnerabilities in distributed model predictive control based on jacobi-gauss decomposition. In *European Control Conference*, pages 2587–2592, 2018.

[81] Pablo Velarde, José M. Maestre, Hideaki Ishii, and Rudy R. Negenborn. Vulnerabilities in lagrange-based distributed model predictive control. *Optimal Control Applications and Methods*, 39:601–621, 3 2018.

[82] Abdullahi Chowdhury, Gour Karmakar, Joarder Kamruzzaman, Alireza Jolfaei, and Rajkumar Das. Attacks on self-driving cars and their countermeasures: A survey. *IEEE Access*, 8:207308–207342, 2020.

[83] Mani Amoozadeh, Arun Raghuramu, Chen-nee Chuah, Dipak Ghosal, H. Michael Zhang, Jeff Rowe, and Karl Levitt. Security vulnerabilities of connected vehicle streams and their impact on cooperative driving. *IEEE Communications Magazine*, 53(6):126–132, 2015.

[84] Jonathan Petit and Steven E. Shladover. Potential cyberattacks on automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 16(2):546–556, 2015.

[85] Simon Parkinson, Paul Ward, Kyle Wilson, and Jonathan Miller. Cyber threats facing autonomous and connected vehicles: Future challenges. *IEEE Transactions on Intelligent Transportation Systems*, 18(11):2898–2915, 2017.

[86] Shantanu Sardesai, Denis Ulybyshev, Lotfi ben Othmane, and Bharat Bhargava. Impacts of security attacks on the effectiveness of collaborative adaptive cruise control mechanism. In *IEEE International Smart Cities Conference (ISC2)*, 2018.

[87] Rens van der Heijden, Thomas Lukaseder, and Frank Kargl. Analyzing attacks on cooperative adaptive cruise control (cacc). In *2017 IEEE Vehicular Networking Conference (VNC)*, pages 45–52. IEEE, 2017.

[88] Muhammad Awais Javed and Elyes Ben Hamida. On the interrelation of security, qos, and safety in cooperative its. *IEEE Transactions on Intelligent Transportation Systems*, 18(7):1943–1957, 2017.

[89] James Weimer, Soummya Kar, and Karl Henrik Johansson. Distributed detection and isolation of topology attacks in power networks. In *Proceedings of the 1st International Conference on High Confidence Networked Systems*, page 65–72, New York, NY, USA, 2012. Association for Computing Machinery.

[90] Riccardo M.G. Ferrari and André M.H. Teixeira. Detection and isolation of routing attacks through sensor watermarking. In *2017 American Control Conference (ACC)*, pages 5436–5442, 2017.

[91] Riccardo M. G. Ferrari and André M. H. Teixeira. A switching multiplicative watermarking scheme for detection of stealthy cyber-attacks. *IEEE Transactions on Automatic Control*, 66(6):2558–2573, 2021.

[92] James Weimer, Seyed Alireza Ahmadi, José Araujo, Francesca Madia Mele, Dario Papale, Iman Shames, Henrik Sandberg, and Karl Henrik Johansson. Active actuator fault detection and diagnostics in hvac systems. In *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, page 107–114, New York, NY, USA, 2012. Association for Computing Machinery.

[93] Jim Marcicki, Simona Onori, and Giorgio Rizzoni. Nonlinear fault detection and isolation for a lithium-ion battery management system. In *Proceedings of the ASME 2010 Dynamic Systems and Control Conference*, 2010.

[94] Vanessa Smet, Francois Forest, Jean-Jacques Huselstein, Frédéric Richardeau, Zoubir Khatir, Stéphane Lefebvre, and Mounira Berkani. Ageing and failure modes of igbt modules in high-temperature power cycling. *IEEE Transactions on Industrial Electronics*, 58(10):4931–4941, 2011.

[95] Christopher Edwards and Sarah K. Spurgeon. A sliding mode observer based fdi scheme for the ship benchmrk. *European Journal of Control*, 6:341–355, 2000.

[96] Ellen Van Nunen, Jeroen Ploeg, Alejandro Morales Medina, and Henk Nijmeijer. Fault tolerancy in cooperative adaptive cruise control. In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pages 1184–1189, 2013.

[97] Ellen van Nunen, Jan Verhaegh, Emilia Silvas, Elham Semsar-Kazerooni, and Nathan van de Wouw. Robust model predictive cooperative adaptive cruise control subject to v2v impairments. In *International Conference on Intelligent Transportation Systems*, 2017.

[98] Jeroen Ploeg, Elham Semsar-Kazerooni, Guido Lijster, Nathan Van De Wouw, and Henk Nijmeijer. Graceful degradation of cooperative adaptive cruise control. *IEEE Transactions on Intelligent Transportation Systems*, 16:488–497, 2 2015.

[99] Tijmen Pollack, Gertjan Looye, and Frans Van der Linden. Design and flight testing of flight control laws integrating incremental nonlinear dynamic inversion and servo current control. In *AIAA Scitech 2019 Forum*, 2019.

[100] Twan Keijzer, Gertjan Looye, Q Ping Chu, and Erik-Jan Van Kampen. Design and flight testing of incremental backstepping based control laws with angular accelerometer feedback. In *AIAA Scitech 2019 Forum*, 2019.

[101] Halim Alwi and Christopher Edwards. Fault detection and fault-tolerant control of a civil aircraft using a sliding-mode-based scheme. *IEEE Transactions on Control Systems Technology*, 16:499–510, 5 2008.

[102] Rolf Isermann. *Process Models and Fault Modelling*, pages 61–82. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[103] Weiting Zhang, Dong Yang, and Hongchao Wang. Data-driven methods for predictive maintenance of industrial equipment: A survey. *IEEE Systems Journal*, 13(3):2213–2227, 2019.

[104] W. Diffie and M. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22(6):644–654, 1976.

[105] Imran Makhdoom, Mehran Abolhasan, and Justin Lipman. A comprehensive survey of covert communication techniques, limitations and future challenges. *Computers & Security*, 120:102784, 2022.

[106] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.

[107] R. L. Rivest, L. Adleman, and M. L. Dertouzos. On data banks and privacy homomorphisms. *Nato.Adv.Sci.I.F-Com.*, page 169, 1978.

[108] Taher El Gamal. A public key cryptosystem and a signature scheme based on discrete logarithms. In George Robert Blakley and David Chaum, editors, *Advances in Cryptology*, pages 10–18, Berlin, Heidelberg, 1985. Springer Berlin Heidelberg.

[109] Pascal Paillier. Public-key cryptosystems based on composite degree residuosity classes. *Advances in Cryptology-EUROCRYPT'99*, page 223–238, 1999.

[110] Kiminao Kogiso and Takahiro Fujita. Cyber-security enhancement of networked control systems using homomorphic encryption. In *Conference on Decision and Control*, 12 2015.

[111] Farhad Farokhi, Iman Shames, and Karl H. Johansson. Private and secure coordination of match-making for heavy-duty vehicle platooning. *IFAC-PapersOnLine*, 50(1):7345–7350, 2017. 20th IFAC World Congress.

[112] Farhad Farokhi, Iman Shames, and Nathan Batterham. Secure and private control using semi-homomorphic encryption. *Control Engineering Practice*, 67:13–20, 2017.

[113] Moritz Schulze Darup, Adrian Redder, Iman Shames, Farhad Farokhi, and Daniel Quevedo. Towards encrypted mpc for linear constrained systems. *IEEE Control Systems Letters*, 2(2):195–200, 2018.

[114] Moritz Schulze Darup, Adrian Redder, and Daniel E. Quevedo. Encrypted cooperative control based on structured feedback. *IEEE Control Systems Letters*, 3(1):37–42, 2019.

[115] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. In *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '05, page 84–93, New York, NY, USA, 2005. Association for Computing Machinery.

[116] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, 56(6):34:1–34:40, 2009.

[117] Craig Gentry. *A fully homomorphic encryption scheme.* PhD thesis, Stanford University, 2009.

[118] Craig Gentry, Amit Sahai, and Brent Waters. Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based. *Advances in Cryptology*, page 75–92, 2013.

[119] Carlos Murguia, Farhad Farokhi, and Iman Shames. Secure and private implementation of dynamic controllers using semihomomorphic encryption. *IEEE Transactions on Automatic Control*, 65(9):3950–3957, 2020.

[120] J.H. Cheon, K. Han, S.-M. Hong, H.J. Kim, J. Kim, S. Kim, H. Seo, H. Shim, and Y. Song. Toward a secure drone system: Flying with real-time homomorphic authenticated encryption. *IEEE Access*, 6:24325–24339, 2018.

[121] Nils Schlüter and Moritz Schulzedarup. On the stability of linear dynamic controllers with integer coefficients. *IEEE Transactions on Automatic Control*, pages 1–1, 2021.

[122] Nils Schlüter, Matthias Neuhaus, and Moritz Schulze Darup. Encrypted dynamic control with unlimited operating time via fir filters. In *2021 European Control Conference (ECC)*, pages 952–957, 2021.

[123] Moritz Schulze Darup, Andreea B. Alexandru, Daniel E. Quevedo, and George J. Pappas. Encrypted control for networked systems: An illustrative introduction and current challenges. *IEEE Control Systems Magazine*, 41(3):58–78, 2021.

[124] Junsoo Kim, Chanhwa Lee, Hyungbo Shim, Jung Hee Cheon, Andrey Kim, Miran Kim, and Yongsoo Song. Encrypting controller using fully homomorphic encryption for security of cyber-physical systems. *IFAC-PapersOnLine*, 49(22):175–180, 2016. 6th IFAC Workshop on Distributed Estimation and Control in Networked Systems NECSYS 2016.

[125] Junsoo Kim, Hyungbo Shim, Henrik Sandberg, and Karl H Johansson. Method for Running Dynamic Systems over Encrypted Data for Infinite Time Horizon without Bootstrapping and Re-encryption. In *60th IEEE Conference on Decision and Control*, pages 5614–5619, 2021.

[126] Mariano Perez Chaher, Bayu Jayawardhana, and Junsoo Kim. Homomorphic Encryption-Enabled Distance-Based Distributed Formation Control with Distance Mismatch Estimators. In *60th IEEE Conference on Decision and Control*, pages 4915–4922, 2021.

[127] Junsoo Kim, Hyungbo Shim, and Kyoohyung Han. *Comprehensive Introduction to Fully Homomorphic Encryption for Dynamic Feedback Controller via LWE-Based Cryptosystem*, pages 209–230. Springer Singapore, Singapore, 2020.

[128] Andrew C. Yao. Protocols for secure computations. In *23rd Annual Symposium on Foundations of Computer Science (sfcs 1982)*, pages 160–164, 1982.

[129] Andreea B. Alexandru and George J. Pappas. *Secure Multi-party Computation for Cloud-Based Control*, pages 179–207. Springer Singapore, Singapore, 2020.

[130] Katrine Tjell, Nils Schlüter, Philipp Binfet, and Moritz Schulze Darup. Secure learning-based mpc via garbled circuit. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 4907–4914, 2021.

[131] Nirupam Gupta, Jonathan Katz, and Nikhil Chopra. Privacy in distributed average consensus. *IFAC-PapersOnLine*, 50(1):9515–9520, 2017.

[132] Jorge Cortés, Geir E. Dullerud, Shuo Han, Jerome Le Ny, Sayan Mitra, and George J. Pappas. Differential privacy in control and network systems. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 4252–4272, 2016.

[133] Yilin Mo and Richard M. Murray. Privacy preserving average consensus. In *53rd IEEE Conference on Decision and Control*, pages 2154–2159, 2014.

[134] Erfan Nozari, Pavankumar Tallapragada, and Jorge Cortés. Differentially private average consensus: Obstructions, trade-offs, and optimal algorithm design. *Automatica*, 81:221–231, 2017.

[135] Apostolos I. Rikos, Themistoklis Charalambous, Karl H. Johansson, and Christoforos N. Hadjicostis. Distributed event-triggered algorithms for finite-time privacy-preserving quantized average consensus. *IEEE Transactions on Control of Network Systems*, pages 1–12, 2022.

[136] Ahmet Cetinkaya, Hideaki Ishii, and Tomohisa Hayakawa. An overview on denial-of-service attacks in control systems: Attack models and security analyses. *Entropy*, 21, February 2019.

[137] Liang Zhao and Guang-Hong Yang. Adaptive sliding mode fault tolerant control for nonlinearly chaotic systems against dos attack and network faults. *Journal of the Franklin Institute*, 354(15):6520–6535, 2017.

[138] Claudio De Persis and Pietro Tesi. Input-to-state stabilizing control under denial-of-service. *IEEE Transactions on Automatic Control*, 60(11):2930–2944, 2015.

[139] Zhicheng Li, Bin Hu, and Zaiyue Yang. Co-design of distributed event-triggered controller for string stability of vehicle platooning under periodic jamming attacks. *IEEE Transactions on Vehicular Technology*, 70(12):13115–13128, 2021.

[140] Saurabh Amin, Alvaro A. Cárdenas, and S. Shankar Sastry. Safe and secure networked control systems under denial-of-service attacks. In Rupak Majumdar and Paulo Tabuada, editors, *Hybrid Systems: Computation and Control*, pages 31–45, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[141] Kyriakos G. Vamvoudakis and João P. Hespanha. Cooperative q-learning for rejection of persistent adversarial inputs in networked linear quadratic systems. *IEEE Transactions on Automatic Control*, 63(4):1018–1031, 2018.

[142] S Longhi, A Monteriù, and M Vaccarini. Cooperative control of underwater glider fleets by fault tolerant decentralized mpc. In *Proceedings of the 17th IFAC World Congress*, 2008.

[143] J.M. Maestre, Paul A. Trodden, and Hideaki Ishii. A distributed model predictive control scheme with robustness against non-compliant controllers. In *IEEE Conference on Decision and Control (CDC)*, pages 3704–3709, 2018.

[144] P. Velarde, J.M. Maestre, H. Ishii, and R.R. Negenborn. Scenario-based defense mechanism fordistributed model predictive control. In *Conference on Decision and Control*, pages 6171–6176, 2017.

[145] Wicak Ananduta, José María Maestre, Carlos Ocampo-Martinez, and Hideaki Ishii. Resilient distributed energy management for systems of interconnected microgrids, 9 2018.

[146] James Usevitch and Dimitra Panagou. Resilient leader-follower consensus to arbitrary reference values. In *2018 Annual American Control Conference (ACC)*, pages 1292–1298, 2018.

[147] Hamza Fawzi, Paulo Tabuada, and Suhas Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, 2014.

[148] Shreyas Sundaram and Christoforos N. Hadjicostis. Distributed function calculation via linear iterative strategies in the presence of malicious agents. *IEEE Transactions on Automatic Control*, 56(7):1495–1508, 2011.

[149] Michelle S. Chong, Henrik Sandberg, and João P. Hespanha. A secure state estimation algorithm for nonlinear systems under sensor attacks. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 5743–5748, 2020.

[150] Seyed Mehran Dibaji and Hideaki Ishii. Resilient consensus of second-order agent networks: Asynchronous update rules with delays. *Automatica*, 81:123–132, 2017.

[151] Yasser Shoukry, Pierluigi Nuzzo, Alberto Puggelli, Alberto L. Sangiovanni-Vincentelli, Sanjit A. Seshia, and Paulo Tabuada. Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach. *IEEE Transactions on Automatic Control*, 62:4917–4932, 10 2017.

[152] Mohammad Hossein Basiri, Mohammad Pirani, Nasser L. Azad, and Sebastian Fischmeister. Security-aware optimal actuator placement in vehicle platooning. *Asian Journal of Control*, 7 2021.

[153] Leslie Lamport, Robert Shostak, and Marshall Pease. The byzantine generals problem. *ACM Trans. Program. Lang. Syst.*, 4(3):382–401, jul 1982.

[154] Waseem Abbas, Aron Laszka, and Xenofon Koutsoukos. Improving network connectivity and robustness using trusted nodes with application to resilient consensus. *IEEE Transactions on Control of Network Systems*, 5(4):2036–2048, 2018.

[155] Mohammad Pirani, Simone Baldi, and Karl Henrik Johansson. Impact of Network Topology on the Resilience of Vehicle Platoons. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–12, 2022.

[156] Hojjat A. Izadi, Youmin Zhang, and Brandon W. Gordon. Fault tolerant model predictive control of quad-rotor helicopters with actuator fault estimation. In *IFAC Proceedings Volumes (IFAC-PapersOnline)*, volume 44, pages 6343–6348, 2011.

[157] Ning Zhou, Yu Kawano, and Ming Cao. Adaptive failure-tolerant control for spacecraft attitude tracking. In *Proceedings of the 15th IFAC Symposium on Large Scale Complex Systems: Theory and Applications,*, pages 67–72, 2019.

[158] Gonzalo Garcia and Shahriar Keshmiri. Nonlinear model predictive controller for navigation, guidance and control of a fixed-wing uav. In *AIAA Guidance, Navigation, and Control Conference*, 2011.

[159] Alain Yetendje, Maria Seron, and José De Doná. Robust multisensor fault tolerant model-following mpc design for constrained systems. *International Journal of Applied Mathematics and Computer Science*, 22:211–223, 3 2012.

[160] Abdel-Razzak Merheb, Hassan Noura, and François Bateman. Passive fault tolerant control of quadrotor uav using regular and cascaded sliding mode control. In *2013 Conference on Control and Fault-Tolerant Systems (SysTol)*, pages 330–335, 2013.

[161] Gianmario Rinaldi, Prathyush P. Menon, Christopher Edwards, and Antonella Ferrara. Distributed super-twisting sliding mode observers for fault reconstruction and mitigation in power networks. In *Proceedings of the IEEE Conference on Decision and Control*, pages 5550–5555. Institute of Electrical and Electronics Engineers Inc., 7 2018.

[162] Riadh Hmidi, Ali Ben Brahim, Fayçal Ben Hmida, and Anis Sellami. Robust fault tolerant control design for nonlinear systems not satisfying matching and minimum phase conditions. *International Journal of Control, Automation and Systems*, 18:2206–2219, 9 2020.

[163] T. Espinoza, A. E. Dzul, R. Lozano, and P. Parada. Backstepping - sliding mode controllers applied to a fixed-wing uav. *Journal of Intelligent & Robotic Systems*, 73(1):67–79, 2014.

[164] Shaocheng Tong, Baoyu Huo, and Yongming Li. Observer-based adaptive decentralized fuzzy fault-tolerant control of nonlinear large-scale systems with actuator failures. *IEEE Transactions on Fuzzy Systems*, 22, 2 2014.

[165] Tijmen Pollack and Erik-Jan Van Kampen. Robust stability and performance analysis of incremental dynamic inversion-based flight control laws. In *AIAA Scitech 2021 Forum*, 2021.

[166] Bart Helder, Erik-Jan Van Kampen, and Marilena Pavel. Online adaptive helicopter control using incremental dual heuristic programming. In *AIAA Scitech 2021 Forum*, 2021.

[167] Xuerui Wang, Erik-Jan van Kampen, Qiping Chu, and Peng Lu. Incremental sliding-mode fault-tolerant flight control. *Journal of Guidance, Control, and Dynamics*, 42(2):244–259, 2019.

[168] Xuerui Wang and Sihao Sun. Incremental fault-tolerant control for a hybrid quadplane uav subjected to a complete rotor loss. *Aerospace Science and Technology*, 125:107105, 2022.

[169] Alan S. Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12(6):601–611, 1976.

[170] Paul M. Frank. Enhancement of robustness in observer-based fault detection. *International Journal of Control*, 59:955–981, 1994.

[171] P M Frank and X Ding. Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *J. Proc. Cont.*, 7:403–424, 1997.

[172] A. Emami-Naeini, M.M. Akhter, and S.M. Rock. Effect of model uncertainty on failure detection: the threshold selector. *IEEE Transactions on Automatic Control*, 33(12):1106–1115, 1988.

[173] Mogens Blanke, Michel Kinnaert, Jan Lunze, and Marcel Staroswiecki. *Diagnosis and Fault-Tolerant Control*. Springer Verlag, 2 edition, 2006.

[174] Steven X. Ding. *Model-based fault diagnosis techniques: Design schemes, algorithms, and tools*. Springer Verlag, 2008.

[175] Andre Teixeira, Iman Shames, Henrik Sandberg, and Karl H. Johansson. Distributed fault detection and isolation resilient to network model uncertainties. *IEEE Transactions on Cybernetics*, 44:2024–2037, 2014.

[176] Riccardo M.G. Ferrari, Thomas Parisini, and Marios M. Polycarpou. Distributed fault detection and isolation of large-scale discrete-time nonlinear systems: An adaptive approximation approach. *IEEE Transactions on Automatic Control*, 57:275–290, 2 2012.

[177] Raul Quinonez, Jairo Giraldo, Luis Salazar, Erick Bauman, Alvaro Cardenas, and Zhiqiang Lin. Savior: Securing autonomous vehicles with robust physical invariants. In *Proceedings of the 29th USENIX Security Symposium*, pages 895–912, 2020.

[178] S. Joe Qin. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36:220–234, 2012.

[179] Liguo Qin, Xiao He, and D. H. Zhou. A survey of fault diagnosis for swarm systems. *Systems Science and Control Engineering*, 2:13–23, 2014.

[180] Chih Che Sun, Adam Hahn, and Chen Ching Liu. Cyber security of a power grid: State-of-the-art. *International Journal of Electrical Power and Energy Systems*, 99:45–56, 7 2018.

[181] Xiaodong Zhang, Marios M. Polycarpou, and Thomas Parisini. Fault diagnosis of a class of nonlinear uncertain systems with lipschitz nonlinearities using adaptive estimation. *Automatica*, 46:290–299, 2 2010.

[182] Yimeng Dong, Nirupam Gupta, and Nikhil Chopra. False data injection attacks in bilateral teleoperation systems. *IEEE Transactions on Control Systems Technology*, 28:1168–1176, 5 2020.

[183] M. F. Hassan, M. A. Sultan, and M. S. Attia. Fault detection in large-scale stochastic dynamic systems. *IEE Proceedings Control Theory and Applications*, 139:119–124, 1992.

[184] Walter H. Chung, Jason L. Speyer, and Robert H. Chen. A decentralized fault detection filter. *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, 123:237–247, 2001.

[185] Shamanth Shankar, Swaroop Darbha, and Aniruddha Datta. Design of a decentralized detection filter for a large collection of interacting lti systems. *Mathematical Problems in Engineering*, 8:233–248, 2002.

[186] Adriano Fagiolini, Gianni Valenti, Lucia Pallottino, Gianluca Dini, and Antonio Bicchi. Decentralized intrusion detection for secure cooperative multi-agent systems. In *2007 46th IEEE Conference on Decision and Control*, pages 1553–1558, 2007.

[187] Shreyas Sundaram, Miroslav Pajic, Christoforos N. Hadjicostis, Rahul Mangharam, and George J. Pappas. The wireless control network: Monitoring for malicious behavior. In *49th IEEE Conference on Decision and Control (CDC)*, pages 5979–5984, 2010.

[188] André Teixeira, Henrik Sandberg, and Karl H. Johansson. Networked control systems under cyber attacks with applications to power networks. In *Proceedings of the 2010 American Control Conference, ACC 2010*, pages 3690–3696. IEEE Computer Society, 2010.

[189] Sridhar Adepu and Aditya Mathur. Distributed attack detection in a water treatment plant: Method and case study. *IEEE Transactions on Dependable and Secure Computing*, 18:86–99, 1 2021.

[190] Joseph K. Scott, Rolf Findeisen, Richard D. Braatz, and Davide M. Raimondo. Input design for guaranteed fault diagnosis using zonotopes. *Automatica*, 50:1580–1589, 2014.

[191] Joseph K. Scott, Davide M. Raimondo, Giuseppe Roberto Marseglia, and Richard D. Braatz. Constrained zonotopes: A new tool for set-based estimation and fault detection. *Automatica*, 69:126–136, 7 2016.

[192] Davide M. Raimondo, Giuseppe Roberto Marseglia, Richard D. Braatz, and Joseph K. Scott. Closed-loop input design for guaranteed fault diagnosis using set-valued observers. *Automatica*, 74:107–117, 12 2016.

[193] Amir Khazraei, Hamed Kebriaei, and Farzad Rajaei Salmasi. An optimal linear dynamic detection method for replay attack in cyber-physical systems, 2019.

[194] Christopher Edwards and Sarah K. Spurgeon. On the development of discontinuous observers. *International Journal of Control*, 59(5):1211–1229, 1994.

[195] T. Floquet, J. P. Barbot, W. Perruquetti, and M. Djemai. On the robust fault detection via a sliding mode disturbance observer. *International Journal of Control*, 77:622–629, 5 2004.

[196] C. P. Tan, F. Crusca, and M. Aldeen. Extended results on robust state estimation and fault detection. *Automatica*, 44:2027–2033, 8 2008.

[197] Luca Massimiliano Capisani, Antonella Ferrara, Alejandra Ferreira De Loza, and Leonid M. Fridman. Manipulator fault diagnosis via higher order sliding-mode observers. *IEEE Transactions on Industrial Electronics*, 59:3979–3986, 2012.

[198] Xianghua Wang, Chee Pin Tan, and Donghua Zhou. A novel sliding mode observer for state and fault estimation in systems not satisfying matching and minimum phase conditions. *Automatica*, 79:290–295, 5 2017.

[199] Alexey Zhirabok, Alexander Zuev, and Alexey Shumsky. Fault identification via sliding mode observers in nonlinear systems not satisfying matching and minimum phase conditions. In *Proceedings of the European Control Conference*, 2021.

[200] Chee Pin Tan and Christopher Edwards. Sliding mode observers for robust fault detection and reconstruction. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 15:347–352, 2002.

[201] Weitian Chen and Mehrdad Saif. Sliding mode output estimator based fault detection, isolation and estimation for systems with unmatched unknown inputs. In *Proceedings of the IEEE International Conference on Control Applications*, pages 1753–1758, 2006.

[202] Christopher Edwards, Sarah K. Spurgeon, Ron J. Patton, and Petra Klotzek. Sliding mode observers for fault detection. *IFAC Proceedings Volumes*, 30:507–512, 8 1997.

[203] Christopher Edwards, Sarah K Spurgeon, and Ron J Patton. Sliding mode observers for fault detection and isolation. *Automatica*, 36:541–553, 2000.

[204] Chee Pin Tan and Christopher Edwards. Sliding mode observers for robust detection and reconstruction of actuator and sensor faults. *International Journal of Robust and Nonlinear Control*, 13:443–463, 4 2003.

[205] Junqi Yang and Fanglai Zhu. Fdi design for uncertain nonlinear systems with both actuator and sensor faults. *Asian Journal of Control*, 17:213–224, 2015.

[206] Junqi Yang, Fanglai Zhu, Xin Wang, and Xuhui Bu. Robust sliding-mode observer-based sensor fault estimation, actuator fault detection and isolation for uncertain nonlinear systems. *International Journal of Control, Automation and Systems*, 13:1037–1046, 10 2015.

[207] Xing Gang Yan and Christopher Edwards. Robust decentralized actuator fault detection and estimation for large-scale systems using a sliding mode observer. *International Journal of Control*, 81:591–606, 4 2008.

[208] Prathyush P. Menon and Christopher Edwards. Sliding mode observers for fault detection in a network of linear dynamical systems. In *IFAC Proceedings Volumes (IFAC-PapersOnline)*, pages 150–155, 2009.

[209] Mi Lv, Wenwu Yu, Yuezu Lv, Jinde Cao, and Wei Huang. An integral sliding mode observer for cps cyber security attack detection. *Chaos*, 29, 4 2019.

[210] Christopher Edwards, Halim Alwi, and Prathyush P. Menon. *Applications of Sliding Observers for FDI in Aerospace Systems*, volume 440, pages 341–360. Springer Verlag, 2013.

[211] Niloofar Jahanshahi and Riccardo M.G. Ferrari. Attack detection and estimation in cooperative vehicles platoons: A sliding mode observer approach. *IFAC-PapersOnLine*, 51(23):212–217, 2018.

[212] Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo. Cyber-physical security via geometric control: Distributed monitoring and malicious attacks. In *IEEE Conference on Decision and Control (CDC)*, pages 3418–3425, 2012.

[213] Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo. Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pages 2195–2201, 2011.

[214] Cheolhyeon Kwon, Weiyi Liu, and Inseok Hwang. Security analysis for cyber-physical systems against stealthy deception attacks. In *Proceedings of the American Control Conference*, pages 3344–3349, 2013.

[215] Roberto Merco, Zoleikha Abdollahi Biron, and Pierluigi Pisu. Replay attack detection in a platoon of connected vehicles with cooperative adaptive cruise control. In *American Control Conference*, pages 5582–8887. IEEE, 2018.

[216] Abdelrahman Khalil, Mohammad Al Janaideh, Khaled F Aljanaideh, and Deepa Kundur. Output-only fault detection and mitigation of networks of autonomous vehicles. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2257–2264, 2020.

[217] Feng Xu, Sorin Olaru, Vicenc Puig, Carlos Ocampo-Martinez, and Silviu Iulian Niculescu. Sensor-fault tolerance using robust mpc with set-based state estimation and active fault isolation. In *Proceedings of the IEEE Conference on Decision and Control*, volume 2015-February, pages 4953–4958. Institute of Electrical and Electronics Engineers Inc., 2014.

[218] Mohammadreza Davoodi, Nader Meskin, and Khashayar Khorasani. Simultaneous fault detection and consensus control design for a network of multi-agent systems. *Automatica*, 66:185–194, 2016.

[219] Mohammad Pirani, Ehsan Hashemi, Amir Khajepour, Baris Fidan, Bakhtiar Litkouhi, Shih Ken Chen, and Shreyas Sundaram. Cooperative vehicle speed fault diagnosis and correction. *IEEE Transactions on Intelligent Transportation Systems*, 20:783–789, 2 2019.

[220] Alvaro A. Cárdenas, Saurabh Amin, Zong-Syun Lin, Yu-Lun Huang, Chi-Yen Huang, and Shankar Sastry. Attacks against process control systems: Risk assessment, detection, and response. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, page 355–366, New York, NY, USA, 2011. Association for Computing Machinery.

[221] Jianglin Lan and Ron J. Patton. A new strategy for integration of fault estimation within fault-tolerant control. *Automatica*, 69:48–59, 2016.

[222] R. J. Patton, C. Kambhampati, A. Casavola, P. Zhang, S. Ding, and D. Sauter. A generic strategy for fault-tolerance in control systems distributed over a network. *European Journal of Control*, 13:280–296, 2007.

[223] Jeroen Ligthart, Elham Semsar-Kazerooni, Jeroen Ploeg, Mohsen Alirezaei, and Henk Nijmeijer. Controller design for cooperative driving with guaranteed safe behavior. In *2018 IEEE Conference on Control Technology and Applications (CCTA)*, pages 1460–1465, 2018.

[224] 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.

[225] Mohammadreza Davoodi, Nader Meskin, and Khashayar Khorasani. *Integrated Fault Diagnosis and Control Design of Linear Complex Systems.* Institution of Engineering and Technology (The IET), 2018.

[226] Twan Keijzer and Riccardo M.G. Ferrari. Threshold design for fault detection with first order sliding mode observers. *Automatica*, 146:110600, 2022.

[227] Twan Keijzer, Japie A.A. Engelbrecht, Phillipe Goupil, and Riccardo M.G. Ferrari. A sliding mode observer approach to oscillatory fault detection in commercial aircraft. *Control Engineering Practice*, under review.

[228] Twan Keijzer, Fabian Jarmolowitz, and Riccardo M.G. Ferrari. Detection of cyber-attacks in collaborative intersection control. In *European Control Conference*, 2021.

[229] Jie Chen and Hong-Yue Zhang. Robust detection of faulty actuators via unknown input observers. *International Journal of Systems Science*, 22(10):1829–1839, 1991.

[230] Jie Chen, Ron J. Patton, and Hongyue Zhang. Design of unknown input observers and robust fault detection filters. *International Journal of Control*, 63(1):85–105, 1996.

[231] Damien Koenig. Unknown input proportional multiple-integral observer design for linear descriptor systems: Application to state and fault estimation. *IEEE Transactions on Automatic Control*, 50(2):212–217, 2005.

[232] Peter Fogh Odgaard and Jakob Stoustrup. Fault tolerant control of wind turbines using unknown input observers. *IFAC Proceedings Volumes*, 45(20):313–318, 2012. 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes.

[233] Jianglin Lan and Ron J. Patton. Integrated fault estimation and fault-tolerant control for uncertain lipschitz nonlinear systems. *International Journal of Robust and Nonlinear Control*, 27(5):761–780, 2017.

[234] Mehrdad Saif and Yuping Guan. A New Approach to Robust Fault Detection and Identification. *Transactions on Aerospace and Electronic Systems*, 29(3):685–695, 1993.

[235] M. Darouach. On the novel approach to the design of unknown input observers. *IEEE Transactions on Automatic Control*, 39(3):698–699, 1994.

[236] F J J Hermans· and M B Zarrop·. Sliding mode observers for robust sensor monitoring. In *13th IFAC World Congress*, pages 6530–6535, 1996.

[237] Leonid Fridman, Arie Levant, and Jorge Davila. High-order sliding-mode observation and identification for linear systems with unknown inputs. In *IEEE Conference on Decision and Control*, pages 5567–5572, 2006.

[238] Thierry Floquet, Christopher Edwards, and Sarah K. Spurgeon. On sliding mode observers for systems with unknown inputs. *International Journal of Adaptive Control and Signal Processing*, pages 938–956, 2007.

[239] Leonid Fridman, Jorge Davila, and Arie Levant. High-order sliding-mode observation of linear systems with unknown inputs. *IFAC Proceedings Volumes*, 41(2):4779 – 4790, 2008.

[240] Christopher Edwards. A comparison of sliding mode and unknown input observers for fault reconstruction. In *Conference on Decision and Control*, pages 5279–5284, 2004.

[241] Christopher Edwards and Chee Pin Tan. A comparison of sliding mode and unknown input observers for fault reconstruction. *European J. of Control*, 12(3):245–260, 2006.

[242] Leonid Fridman, Arie Levant, and Jorge Davila. Observation of linear systems with unknown inputs via high-order sliding-modes. *International Journal of Systems Science*, 38(10):773–791, 2007.

[243] Leonid Fridman, Jorge Davila, and Arie Levant. High-order sliding-mode observation and fault detection. In *Conference on Decision and Control*, pages 4317–4322, 2007.

[244] H. Ríos, Jorge Davila, Leonid Fridman, and Christopher Edwards. Fault detection and isolation for nonlinear systems via high-order-sliding-mode multiple-observer. *International Journal of Robust and Nonlinear Control*, 25:2871–2893, 2015.

[245] Alejandra Ferreira de Loza, David Henry, Jérôme Cieslak, Ali Zolghadri, and Jorge Davila. Sensor fault diagnosis using a non-homogeneous high-order sliding mode observer with application to a transport aircraft. *IET Control Theory & Applications*, 9(4):598–607, 2015.

[246] Chee Pin Tan and Christopher Edwards. Robust Fault Reconstruction in Uncertain Linear Systems Using Multiple Sliding Mode Observers in Cascade. *IEEE Transactions on Automatic Control*, 55(4):855–867, 2010.

[247] Reza Raoufi, H. J. Marquez, and Alan S.I. Zinober. H∞ sliding mode observers for uncertain nonlinear lipschitz systems with fault estimation synthesis. *International Journal of Robust and Nonlinear Control*, 20:1785–1801, 2010.

[248] Francisco J. Bejarano. Partial unknown input reconstruction for linear systems. *Automatica*, 47(8):1751–1756, 2011.

[249] Jian Zhang, Akshya Kumar Swain, and Sing Kiong Nguang. Robust sensor fault estimation scheme for satellite attitude control systems. *Journal of the Franklin Institute*, 350(9):2581–2604, 2013.

[250] A. N. Zhirabok, A. E. Shumsky, and A. V. Zuev. Fault diagnosis in linear systems via sliding mode observers. *International Journal of Control*, 94(2):327–335, 2021.

[251] Xianghua Wang, Chee Pin Tan, and Donghua Zhou. A novel sliding mode observer for state and fault estimation in systems not satisfying matching and minimum phase conditions. *Automatica*, 79:290–295, 2017.

[252] Arie Levant. higher-order sliding modes differentiation and output-feedback control. *International Journal of Control*, 76(9-10):924–941, 2003.

[253] A. S. Poznyak. Stochastic output noise effects in sliding mode state estimation. *International Journal of Control*, 76(9-10):986–999, 2003.

[254] Satadru Dey, Sara Mohon, Pierluigi Pisu, and Beshah Ayalew. Sensor fault detection, isolation, and estimation in lithium-ion batteries. *IEEE Transactions on Control Systems Technology*, 24:2141–2149, 11 2016.

[255] Junqi Yang, Fanglai Zhu, and Wei Zhang. Sliding-mode observers for nonlinear systems with unknown inputs and measurement noise. *International Journal of Control, Automation and Systems*, 11(5):903–910, 2013.

[256] Indira Nagesh and Christopher Edwards. A sliding mode observer based fdi scheme for a nonlinear satellite systems. In *IEEE International Conference on Control Applications*, pages 159–164, 2011.

[257] Alessandro Pilloni, Alessandro Pisano, Elio Usai, and Ruben Puche-Panadero. Detection of rotor broken bar and eccentricity faults in induction motors via second order sliding mode observer. In *51st IEEE Conference on Decision and Control*, pages 7614–7619, 2012.

[258] Wei Ao, Yongdong Song, and Changyun Wen. Adaptive cyber-physical system attack detection and reconstruction with application to power systems. *IET Control Theory and Applications*, 10:1458–1468, 8 2016.

[259] Chee Pin Tan and Christopher Edwards. An LMI approach for designing sliding mode observers. *International Journal of Control*, 74(16):1559–1568, 2001.

[260] E. Alcorta-Garcia, A. Zolghadri, and P. Goupil. A nonlinear observer-based strategy for aircraft oscillatory failure detection: A380 case study. *IEEE Transactions Aerospace and Electronic Systems*, 47(4):2792–2806, 2011.

[261] L. Lavigne, A. Zolghadri, P. Goupil, and P. Simon. A model-based technique for early and robust detection of oscillatory failure case in a380 actuators. *International Journal of Control, Automation and Systems*, 9(1):42–49, 2011.

[262] Ali Zolghadri, David Henry, Jérome Cieslak, Denis Efimov, and Philippe Goupil. *Fault Diagnosis and Fault-Tolerant Control and Guidance for Aerospace Vehicles: From theory to application.* Advances in Industrial Control. Springer London Ltd, 2013.

[263] R. Pons, C. Jauberthie, L. Travé-Massuyès, and P. Goupil. Interval analysis based learning for fault model identification. application to control surfaces oscillatory failures. In *International Workshop on Qualitative Reasoning*, pages 115–122, 2008.

[264] Varga A. and Ossmann D. Lpv-model based identification approach of oscillatory failure cases. In *IFAC International Symposium on Fault Detection, Supervision and Safety of Technical Processes*, pages 1347–1352, Mexico City, Mexico, 2012.

[265] Jérome Cieslak, Denis Efimov, Ali Zolghadri, David Henry, and Philippe Goupil. Oscillatory failure case detection for aircraft using non-homogeneous differentiator in noisy environment. In *2nd CEAS Specialist Conference on Guidance, Navigation & Control*, pages 394–413, April 2013.

[266] Denis Efimov, Jérôme Cieslak, Ali Zolghadri, and David Henry. Actuator fault detection in aircraft systems: Oscillatory failure case study. *Annual Reviews in Control*, 37(1):180–190, 2013.

[267] Mohcine Sifi, Loïc Lavigne, Franck Cazaurang, and Philippe Goupil. Oscillatory failure detection in flight control system of civil aircraft: Eha actuator servo loop case study. In *R3ASC'12*, pages 55–64, 2012.

[268] Xiaoyu Sun, Ron J Patton, and Philippe Goupil. Robust adaptive fault estimation for a commercial aircraft oscillatory fault scenario. In *Proceedings of 2012 UKACC International Conference on Control*, pages 595–600. IEEE, 2012.

[269] Halim Alwi and Christopher Edwards. An adaptive sliding mode differentiator for actuator oscillatory failure case reconstruction. *Automatica*, 49(2):642–651, 2013.

[270] Philippe Goupil, S. Urbano, and J.Y. Tourneret. A data-driven approach to detect faults in the airbus flight control system. *IFAC-PapersOnLine*, 49(17):52–57, 2016.

[271] S. Urbano, E. Chaumette, P. Goupil, and J.Y. Tourneret. A data-driven approach for actuator servo loop failure detection. *IFAC-PapersOnLine*, 50(1):13544–13549, 2017.

[272] Andreas Varga and Daniel Ossmann. Lpv model-based robust diagnosis of flight actuator faults. *Control Engineering Practice*, 31:135–147, 2014.

[273] Kumpati S Narendra and Jeyendran Balakrishnan. Adaptive control using multiple models. *IEEE transactions on automatic control*, 42(2):171–187, 1997.

[274] Do Hieu Trinh, Benoît Marx, Philippe Goupil, and José Ragot. Oscillatory failure detection in the flight control system of a civil aircraft using soft sensors. *New Sensors and Processing Chain*, pages 85–105, 2014.

[275] Loic Lavigne, Franck Cazaurang, Luciano Fadiga, and Philippe Goupil. New sequential probability ratio test: Validation on a380 flight data. *Control Engineering Practice*, 22:1–9, 2014.

[276] H. Sachs, U. B. Carl, and F. Thielecke. An approach to the investigation of oscillatory failure cases in electro-hydraulic actuation systems. In *Recent advances in aerospace actuation systems and components*, pages 99–104, Toulouse, France, 2007.

[277] S.J. Loutridis. Damage detection in gear systems using empirical mode decomposition. *Engineering Structures*, 26(12):1833–1841, 2004.

[278] Twan Keijzer, Paula Chanfreut, José María Maestre, and Riccardo M.G. Ferrari. Collaborative vehicle platoons with guaranteed safety against cyber-attacks. *Transactions on Intelligent Transportation Systems*, under review.

[279] Filiberto Fele, Jose M Maestre, and Eduardo F Camacho. Coalitional control: Cooperative game theory and control. *IEEE Control Systems*, 37(1):53–69, 2017.

[280] Pablo R. Baldivieso-Monasterios and Paul A. Trodden. Coalitional predictive control: Consensus-based coalition forming with robust regulation. *Automatica*, 125:109380, 2021.

[281] Filiberto Fele, José M Maestre, S Hashemy, David Muñoz de la Peña, and Eduardo F Camacho. Coalitional model predictive control of an irrigation canal. *Journal of Process Control*, 24(4):314–325, 2014.

[282] Eva Masero, José Ramón D Frejo, José M Maestre, and Eduardo F Camacho. A light clustering model predictive control approach to maximize thermal power in solar parabolic-trough plants. *Solar Energy*, 214:531–541, 2021.

[283] Baocang Ding, Liang Ge, Hongguang Pan, and Peng Wang. Distributed mpc for tracking and formation of homogeneous multi-agent system with time-varying communication topology. *Asian Journal of Control*, 18(3):1030–1041, 2016.

[284] Pangwei Wang, Hui Deng, Juan Zhang, Li Wang, Mingfang Zhang, and Yongfu Li. Model predictive control for connected vehicle platoon under switching communication topology. *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[285] Hui Zhao, Xuewu Dai, Qi Zhang, and Jinliang Ding. Robust event-triggered model predictive control for multiple high-speed trains with switching topologies. *IEEE Transactions on Vehicular Technology*, 69(5):4700–4710, 2020.

[286] Youssef Abou Harfouch, Shuai Yuan, and Simone Baldi. An adaptive switched control approach to heterogeneous platooning with intervehicle communication losses. *IEEE Transactions on Control of Network Systems*, 5(3):1434–1444, 2017.

[287] Yang Zheng, Shengbo E. Li, Jianqiang Wang, Dongpu Cao, and Keqiang Li. Stability and scalability of homogeneous vehicular platoon:study on the influence of information flow topologies. *T. on intelligent transportation systems*, 17(1):14–26, 2015.

[288] Yongsong Wei, Shaoyuan Li, and Yi Zheng. Enhanced information reconfiguration for distributed model predictive control for cyber-physical networked systems. *International Journal of Robust and Nonlinear Control*, 30(1):198–221, 2020.

[289] Stefano Riverso, Francesca Boem, Giancarlo Ferrari-Trecate, and T. Parisini. Plug-and-play fault detection and control-reconfiguration for a class of nonlinear large-scale constrained systems. *Transactions on Automatic Control*, 61(12):3963–3978, 2016.

[290] Abhishek Jain, Aranya Chakrabortty, and Emrah Biyik. Distributed wide-area control of power system oscillations under communication and actuation constraints. *Control Engineering Practice*, 74:132–143, 2018.

[291] Guannan Lou, Wei Gu, Yinliang Xu, Ming Cheng, and Wei Liu. Distributed mpc-based secondary voltage control scheme for autonomous droop-controlled microgrids. *IEEE transactions on sustainable energy*, 8(2):792–804, 2016.

[292] Jason J. Haas. The effects of wireless jamming on vehicle platooning. Technical report, University of Illinois, 2009.

[293] Mohammad H. Basiri, Nasser L. Azad, and Sebastian Fischmeister. Attack resilient heterogeneous vehicle platooning using secure distributed nonlinear model predictive control. In *Mediterranean conf. on control and automation*, pages 307–312, 2020.

[294] Alberto Petrillo, Antonio Pescapé, and Stefania Santini. A collaborative approach for improving the security of vehicular scenarios: The case of platooning. *Computer Communications*, 122:59–75, 2018.

[295] Xiaoran Feng and Ron Patton. Active fault tolerant control of a wind turbine via fuzzy MPC and moving horizon estimation. In *Proceedings of the 19th IFAC World Congress*, pages 3633–3638. IFAC, 2014.

[296] José M Maestre, Rudy R Negenborn, et al. *Distributed model predictive control made easy*, volume 69. Springer, 2014.

[297] Jeroen Ploeg, Dipan P Shukla, Nathan van de Wouw, and Henk Nijmeijer. Controller Synthesis for String Stability of Vehicle Platoons. *IEEE Transactions on Intelligent Transportation Systems*, 15(2):854–865, 2014.

[298] Pieter Stobbe, Twan Keijzer, and Riccardo M.G. Ferrari. A fully homomorphic encryption scheme for real-time safe control. In *Conference on Decision and Control*, 2022.

[299] Morris Dworkin, Elaine Barker, James Nechvatal, James Foti, Lawrence Bassham, E. Roback, and James Dray. Advanced encryption standard (aes), November 2001.

[300] Rhett Smith. *Cryptography Concepts and Effects on Control System Communications*, pages 1–9. Schweitzer Engineering Laboratories, Inc., 2018.

[301] Jung Hee Cheon and Damien Stehlé. Fully homomophic encryption over the integers revisited. In Elisabeth Oswald and Marc Fischlin, editors, *Advances in Cryptology – EUROCRYPT*, pages 513–536, 2015.

[302] Julian Tran, Farhad Farokhi, Michael Cantoni, and Iman Shames. Implementing homomorphic encryption based secure feedback control. *Control Engineering Practice*, 97:104350, 2020.

[303] Junsoo Kim, Hyungbo Shim, and Kyoohyung Han. Dynamic controller that operates over homomorphically encrypted data for infinite time horizon. *IEEE Transactions on Automatic Control*, pages 1–1, 2022.

[304] Karl J. Åström and Richard M. Murray. *Feedback systems: An introduction for scientists and Engineers*. Princeton University Press, 2021.

[305] Daniele Micciancio and Chris Peikert. Trapdoors for lattices: Simpler, tighter, faster, smaller. In *Advances in Cryptology – EUROCRYPT*, pages 700–718, 2012.

[306] Subin Moon and Younho Lee. An efficient encrypted floating-point representation using HEAAN and TFHE. *Security and Communication Networks*, 2020:1–18, 03 2020.

[307] Alexander J. Gallo, Mustafa S. Turan, Francesca Boem, Giancarlo Ferrari-Trecate, and Thomas Parisini. Distributed watermarking for secure control of microgrids under replay attacks. *IFAC-PapersOnLine*, 51(23):182–187, 2018. 7th IFAC Workshop on Distributed Estimation and Control in Networked Systems NECSYS 2018.

Glossary

ACC adaptive cruise control.

ALM adaptive logic module.

AV autonomous vehicle.

CACC collaborative adaptive cruise control.

CC cruise control.

CIA confidentiality, integrity and availability.

CPS cyber-physical system.

CVP collaborative vehicle platoon.

DMPC distributed model predictive control.

DoS denial of service.

DSP digital signal processing.

EOI equivalent output injection.

FBW fly-by-wire.

FCC flight control computer.

FHE fully homomorphic encryption.

FHSS frequency hopping spread spectrum.

FOSMO first order sliding mode observer.

FPGA field-programmable gate array.

FTC fault tolerant control.

HME homomorphic encryption.

HVAC heating, ventilation and air-conditioning.

I/O input/output.

ICS industrial control system.

IT information technology.

LMI linear matrix inequality.

LPV linear parameter-varying.

LQR linear-quadratic regulator.

LWE learning with errors.

MC Monte Carlo.

MIMO multiple-input multiple-output.

MITM man-in-the-middle.

MPC model predictive control.

OFC oscillatory failure case.

OT operation technology.

PHE partially homomorphic encryption.

PID proportional-integral-derivative.

QoS quality of service.

R-UIO reduced unknown input observer.

RSA Rivest–Shamir–Adleman.

SISO single-input single-output.

SMC sliding mode controller.

SMO sliding mode observer.

sMPC secure multi-party computation.

UAV unmanned aerial vehicle.

UIO unknown input observer.

VHDL VHSIC hardware description language.

Curriculum Vitæ

Twan Keijzer

1994/09/19 Born in Amsterdam, The Netherlands

Education

2019-2023 PhD in Systems & Control
Delft University of Technology
Thesis: Advances in Safety and Security of Cyber Physical Systems
Promotor: Prof. Dr. Ir. J.W. van Wingerden
Copromotor: Dr. R.M.G. Ferrari

2015-2018 MSc. in Aerospace Engineering
Delft University of Technology
Honours: Artificial Inteligence at University of Amsterdam
Thesis: Design and Flight Testing of Incremental Backstepping based Control Laws with Angular Accelerometer Feedback
Supervisors: Dr. Ir. G. Looye, Dr. Ir. Q.P. Chu

2012-2015 BSc. in Aerospace Engineering (Cum Laude)
Delft University of Technology
Minor: Robotics at Nanyang Technological University
Thesis: Meet the Martians: Design of a Controllable Inflatable Aeroshell
Supervisor: Dr. Ir. H. Damveld

Awards

2020 Winner of the *Aerospace Industrial Benchmark on Fault Detection* competition at the 21st IFAC World Congress

List of Publications

1. *T. Keijzer*, P. Chanfreut, J.M. Maestre, and R.M.G. Ferrari, "Collaborative Vehicle Platoons with guaranteed Safety against Cyber-Attacks," Transactions on Inteligent Transportation Systems, under review

2. *T. Keijzer*, J.A.A. Engelbrecht, P. Goupil, and R.M.G. Ferrari, "A sliding mode observer approach to oscillatory fault detection in commercial aircraft," Control Engineering Practice, under review

3. *T. Keijzer* and R. M. G. Ferrari, "Threshold Design for Fault Detection with First Order Sliding Mode Observers," Automatica, 146:110600, 2022.

4. *T. Keijzer*, A.J. Gallo and R.M.G. Ferrari, "Hierarchical Cyber-Attack Detection in Large-Scale Interconnected Systems," 2022 IEEE 61st Conference on Decision and Control (CDC), 2022.

5. P. Stobbe, *T. Keijzer* and R.M.G. Ferrari, "A Fully Homomorphic Encryption Scheme for Real-Time Safe Control," 2022 IEEE 61st Conference on Decision and Control (CDC), 2022.

6. *T. Keijzer*, F. Jarmolowitz and R. M. G. Ferrari, "Detection of Cyber-Attacks in Collaborative Intersection Control," 2021 European Control Conference (ECC), 2021, pp. 62-67.

7. *T. Keijzer* and R.M.G. Ferrari, "Detection of Network and Sensor Cyber-Attacks in Platoons of Cooperative Autonomous Vehicles: a Sliding-Mode Observer Approach," 2021 European Control Conference (ECC), 2021, pp. 515-520.

8. *T. Keijzer* and R.M.G. Ferrari, "A Sliding Mode Observer Approach to the Aerospace Industrial Benchmark on Fault Detection", 2020. Winner of the "Aerospace Industrial Benchmark on Fault Detection" competition at the 21st IFAC World Congress

9. P. Chanfreut, *T. Keijzer*, R.M.G. Ferrari, J.M. Maestre, "A Topology-Switching Coalitional Control and Observation Scheme with Stability Guarantees," IFAC-PapersOnLine, Volume 53, Issue 2, 2020, Pages 6477-6482.

10. *T. Keijzer* and R.M.G. Ferrari, "A Sliding Mode Observer Approach for Attack Detection and Estimation in Autonomous Vehicle Platoons using Event Triggered Communication," 2019 IEEE 58th Conference on Decision and Control (CDC), 2019, pp. 5742-5747.

11. *T. Keijzer*, G. Looye, Q.P. Chu and E.J. van Kampen. "Design and Flight Testing of Incremental Backstepping based Control Laws with Angular Accelerometer Feedback," AIAA 2019-0129. AIAA Scitech 2019 Forum. January 2019.

Included in this thesis.
Won a best paper, tool demonstration, or proposal award.

Propositions

accompanying the dissertation

Advances in Safety and Security of Cyber-Physical Systems

by

Twan Keijzer

1. Assuming it is possible to provide safety and security from all possible anomalies, this can only be achieved by complementary prevention, resilience, detection and accommodation techniques. [This Thesis]
2. If the risk that an anomaly poses to the safety and security of a system is sufficiently small, it does not need to be considered. [This Thesis]
3. Scientific papers should include simulation results of systems that fall outside the theoretical framework of the paper to demonstrate a certain flexibility of the proposed method that could not be shown in theory. [This Thesis]
4. Journalism often serves as a tool to detect societal problems, but it relies on others to provide adequate accommodation of these problems, as well as prevention of, and resilience to, future problems. Therefore, it is essential for journalism to be taken seriously.
5. The ultimate goal of automation should be a world in which everyone can freely pursue their passion.
6. Traditions often exist for a reason, but should never be an excuse to reject change.
7. By travelling the world one can explore the sense and nonsense of all traditions and use this to independently develop habits and plant the seeds for new traditions.
8. The beauty of nature is derived from the fact that it is forever temporary.
9. Everyone is, to some extent, a hypocrite. People should remind themselves of this before judging others.

These propositions are regarded as opposable and defendable, and have been approved as such by the promotors prof.dr.ir. J.W. van Wingerden and dr. R. Ferrari