Cyber-Security of Over-Actuated Cyber-Physical Systems: A Study on Dynamic Positioning Systems

Sina Ghodsi

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Electrical and Computer Engineering) at

Concordia University

Montréal, Québec, Canada

August 2024

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Sina Ghodsi**

Entitled: **Cyber-Security of Over-Actuated Cyber-Physical Systems: A Study on Dynamic Positioning Systems**

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

Master of Applied Science (Electrical and Computer Engineering)

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

Signed by the Final Examining Committee:

______________________________ Chair
Dr. Rastko R. Selmic

______________________________ External Examiner
Dr. Walter Lucia

______________________________ Examiner
Dr. Rastko R. Selmic

______________________________ Supervisor
Dr. Khashayar Khorasani

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

____________ 2024 ______________________________
Dr. Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Cyber-Security of Over-Actuated Cyber-Physical Systems: A Study on Dynamic Positioning Systems

Sina Ghodsi

Over-actuated systems, are crucial for advanced control functionalities in various industries, with Dynamic Positioning (DP) systems serving as a prime example. DP systems, commonly found in marine vessels and offshore platforms, maintain fixed positions using multiple thrusters. As Cyber-Physical Systems (CPS), DP systems integrate computation, communication, and physical components to achieve control objectives. However, their reliance on communication networks makes them vulnerable to False Data Injection (FDI) cyber-attacks, where adversaries can compromise signals sent from the controller to the thrusters. This thesis presents methods for secure estimation, attack reconstruction, isolation, and compensation within a centralized thrust allocation framework. Key contributions include achieving these goals without relying on strict input-output matrix conditions, and allowing thrusters to remain operational even when affected by FDI attacks. Furthermore, the thesis addresses the challenge of reducing and shifting the attack surface, particularly in over-actuated systems, where numerous communication links increase vulnerability. To mitigate this, the allocation scheme is transformed from a centralized to a decentralized framework. In this approach, control signals are sent to a randomly chosen thruster, periodically switching to prevent successful attacks. Thrusters then coordinate through a consensus network to realize the control commands, while the consensus protocol is resilient under attacks on the communication channels between the thrusters. The decentralized protocol is effective in both closed-loop and open-loop operations. Finally, the estimation and compensation techniques developed earlier in the thesis are also applied to enhance the resilience of the decentralized architecture.

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Professor Khorasani, for his guidance, patience, and support throughout my program. I learned a lot from him during my master's and I will forever be indebted to him.

I am also profoundly grateful to my aunt and my parents for their support and encouragement, which have been crucial to my success.

Additionally, I would like to thank my friends Shahab Chehraghi, Kasra Khorsand and Mohammadreza Nematollahi for their constant encouragement and compassion during this journey.

I dedicate this thesis to my aunt, Rana Samadfam, whose unwavering support and encouragement have been instrumental in this achievement. Thank you for everything.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Preliminaries

1.1.1 Marine Vessels and Cyber-physical Systems

Over-actuated systems play an essential role in various engineering applications due to their enhanced capabilities in maneuvering and control. These systems are prevalent across industries ranging from aerospace to marine, offering flexibility and robustness in handling complex operations. Among these applications, surface vessels stand out as a notable example of over-actuated systems. Ships are crucial in global trade. Export and import of goods and commodities across the world heavily rely on ships. Furthermore, ships are also used for oil drilling, scientific expeditions, leisure and travelling, and maritime defense. Ships also have significant environmental impacts, that range from sea pollution, greenhouse gas emissions, and habitat destruction. Thus, it is of great importance to study the operations of ships considering the mentioned factors.

For the stated reasons, maritime safety is a crucial topic. The safety of marine vessels and the associated protocols for vessels' safe operations are vital to preventing collisions and accidents, preventing environmental catastrophes, warding against piracy and hijacking, and ensuring coordinated response.

Ships are complex systems that consist of various subsystems. These subsystems include propulsion system, power management system, position, navigation and timing (PNT), course keeping and steering system, etc. Traditional and legacy marine vessels employed standalone subsystems that were controlled separately, and were generally isolated from the the rest of the subsystems, or at least the communication between them was minimal. Furthermore, these vessels heavily relied on a human operator for seamless operation. Therefore, these conventional ways of controlling and operating the marine vessels suffered from lack of integration and automation and a centralized monitoring.

In order to overcome the above shortcomings, industrial settings, including marine vessels, have taken advantage of communication and digitization and resorted to networked control systems (NCS). Since hardware devices for networks and network nodes have become cheaper, closed control loops over a communication network have become increasingly prevalent. Control systems communicating with sensors and actuators over a network are called NCSs [3]. In addition to enabling remote data transfers, networks also reduce wire connection complexity and media costs, and they facilitate maintenance [4].

Having touched upon the advantages of NCS in optimizing maritime control and operations, the discourse naturally progresses towards the emerging framework of Cyber-Physical Systems (CPS). Cyber-physical systems employ interconnected communication networks for the supervision and control of operations, facilitated by distributed sensors, actuators, and embedded computers [5]. Recently, marine vessels have adopted the concept of Integrated Platform Management System (IPMS) as an NSC. IPMS is a set of hardware and software elements that enable the automation, control and supervision of almost all shipboard equipment. Given that the IPMS of a ship integrates with physical processes, enables real-time monitoring and control, employs distributed sensors and actuators, and resorts to communication networks for the transmission of data, it certainly qualifies as a CPS. The concept of integrated automation of surface vessels is not new. It was previously the responsibility of the Integrated Machinery Control System (IMCS) or Machinery Control and Surveillance System (MCAS) to integrate the supervisory control of various subsystems on board a vessel, such as machinery control, dynamic positioning system, propulsion control, power management, auxiliary systems, etc. IMCS has gradually been replaced by IPMS in recent years, with more and more subsystems and components of vessels being incorporated into IPMS. There are some instances where IPMS and IMCS, or MCS (Machine Control System) can be used interchangeably. A variety of IPMS solutions and configurations are available according to the vessel's needs [6].

1.1.2 Dynamic Positioning System

One of the most important subsystems of IPMS onboard most of the marine vessels, is dynamic positioning (DP) system. DP is a system responsible for simultaneous control of three horizontal motions (surge, sway and yaw motion). The Norwegian classification society, defines DP as following:
”A dynamically positioned vessel, is a vessel which maintain its position (fixed or a predetermined track) exclusively by means of thrusters” [7]. Traditionally, dynamic positioning systems have been associated with oil industry. They have been used extensively on offshore drilling vessels, mainly for the purposes of maintaining the position and orientation of the vessel. However, nowadays, DP is being employed on a variety of vessels including supply vessels [8], ice management vessels [9], under water vehicles and Remotely Operated Vehicles (ROVs), [10] and offshore floating wind turbines [11]. At present, DP is not only used to maintain the position and orientation, but also to follow a predetermined path.

Dynamic positioning systems use various sensors onboard the vessel in order to estimate and control the position, orientation and speed of the vessel in the presence of environmental forces. These sensors generally include, but are not limited to Global Positioning System (GPS), Hydro-acoustic Position Reference, Taut Wire, Gyro-compass and/or magnetic compass, Intertial Measurement Unit (IMU), and environmental sensors [12]. The detailed descriptions of these sensors are provided in the following chapter.

The essential part of dynamic positioning system, is the use of active thrusters. In contrast to sole dependence on the main propeller and a rudder, dynamic positioning systems utilize tunnel thrusters and azimuth thrusters in addition to the main propeller that is installed at the aft of the ship in order to achieve the desired states in surge, sway and yaw. The number of thrusters depends on various factors such as the type of the vessel, size, environmental conditions, etc. In some DP applications, vessels deploy as many as 8 thrusters to meet their needs [13]. It must be noted that in some DP configurations, rudders could also be employed [14].

A DP system is basically a control system that receives feedback from sensors, estimates the system states including speed, wind and wave forces, and controls the vessel's position and orientation. There is also one important module inside DP called thrust allocation, which is responsible for optimizing and distributing the commanded control signal to the active thrusters. Figure 1.1 shows the representation of DP as a feedback control system within the IPMS network.

1.1.3 Cyber Threats to Marine Vessels and Dynamic Positioning System

Owing to the increased reliance on digitization, integration and communication in Marine vessels, maritime operations are becoming more and more susceptible to cyber threats [15]. Adapting the paradigm of NCS and IPMS for marine vessels, in addition to producing

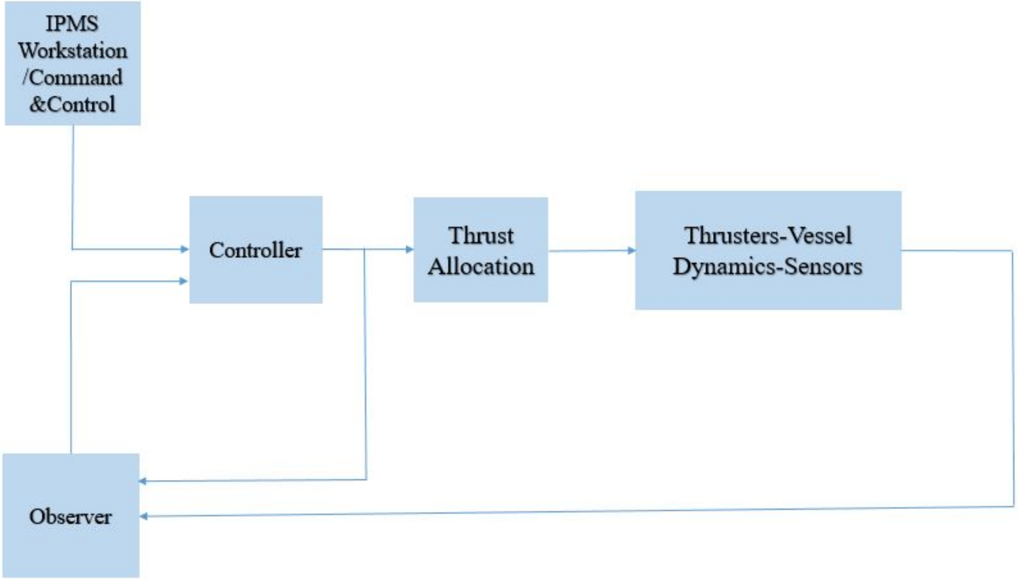

IPMS Workstation /Command &Control
Controller
Thrust Allocation
Thrusters-Vessel Dynamics-Sensors
Observer

Figure 1.1: Schematic of DP as a feedback control system.

the benefits that were discussed previously, brings about vulnerabilities in the cyber domain that need to be taken seriously and investigated.

The infamous cyber-attack on Maersk in 2017 led to increased attention and concern over cyber threats in maritime industry. Approximately 300 million dollars worth of losses were caused by the ransomware attack on Maersk's operations in 13 international ports [16]. Although previously there had been several recorded cyber-attacks targeting the maritime industry, the attack on Maersk was a turning point. This mentioned attack had targeted the IT domain of the ship. The systems onboard the vessels that could suffer from cyber-attacks are either information technology (IT), or operational technology (OT). Hackers could also target the OT systems as well.

An OT system is a cyber-physical system that interacts with the environment and controls the physical devices onboard the vessel [15]. Power management system, navigation system and dynamic positioning system are OT systems. Due to the fact that OT technologies are transforming into cyber-physical systems which are not standalone physical systems anymore that used to be controlled separately from other subsystems, and that these systems operate in the context of the IPMS network, they directly interact with the cyber domain and rely on communications to accomplish their objectives. And exactly as a consequence of this, OT systems could be compromised and pose a greater risk to human life compared to the situations where IT systems are compromised. There are multiple ways an attacker could penetrate the IPMS network of the ship and target the operational technologies onboard the vessel, which are provided in the following:

- **Connection to upper level networks**: IPMS of ship operate in similar manner as Supervisory Control and Data Acquisition (SCADA) in industrial settings. The field network which accommodates controller that control the physical plants, is connected to the process network where the supervisory control is situated. Sometimes This supervisory layer is connected to a higher level network such as a corporate network or data processing network [17]. A malicious entity could exploit this connections and find its way into the process network and eventually the field network and impose a directed cyber-attack on the OT systems.

- **Autonomous ships**: There is a growing interest for deploying autonomous ships, especially in oil and gas industry. Once ships are operational in an autonomous manner, human operators need to have access to and control the vessel remotely [18]. This remote access and control poses a great threat to the cyber-security of the autonomous ship. A hacker could intercept the communication signals between the remote controller and the ship and exert a cyber-attack.

- **Internet of Things (IoT)**: Implementing IoT solutions onboard vessels is considered highly profitable for the ships. Data analytics and the accessibility to a wealth of information from multiple sources are two of the significant benefits of IoT. A modern ship could host several sensors that could operate as as an IoT device. Some of these sensors are part of the IPMS network. A network is as secure as its most vulnerable device. Therefore, embracing IoT onboard vessels, will lead to huge risks associated with cyber threats [18].

- **Spreading malicious malware through USB** : The periodic updating of electronic navigational charts, or other software onboard the bridge or IPMS workstations, commonly performed by plugging a USB memory stick, may lead to infection and propagation of malicious code. This malicious code could target the OT systems of the vessel in the form of a cyber-attack. [19].

In 2013, a malware penetrated the communication links between the controller and thrusters in dynamic positioning system of a drilling vessel in the gulf of Mexico, leading to a halt in operation. The malware had found its way into the DP system when the crew onboard the ship accidentally connected malware infected PCs and USB devices to the ship's local network [20]. If a malicious entity breaches into the IPMS network, which houses the dynamic positioning system, through any of the means mentioned above, it could launch targeted and devastating attacks at the DP and it components. The most significant components of the DP that could be compromised, are the thrusters. Dynamic positioning systems of modern vessels that have adopted IPMS as a networked control system rely on communication channels to send the commanded signals to each of the thrusters. Once the attackers manage to compromise these communication channels, they could attack the thrusters and manipulate the vessel's position, heading and speed. The attackers might be motivated by financial gains or geopolitical interests. In any case, a compromised DP, and by extension, compromised thrusters, is serious issue and must be investigated in depth.

1.2 Problem Statement

The problem of interest in this thesis is to develop methodologies in order to maintain a secure control of thrusters in dynamic positioning systems in the scenarios where dynamic positioning system has been compromised and as a result, the communication links between the controller and the thrusters are infected with cyber-attacks.

In dynamic positioning system, each thruster receives their set-points or control commands from the thrust allocation module. The thrust allocation module receives the demanded force vector from the controller, or the operator. The thrust allocation module, controller and the operator reside at the bridge of the ship or the command and control center of the ship, where the IPMS work stations are situated. As a result, the set-point for each thruster is transmitted through communication lines. The vulnerabilities here lie at the communication links between the controller-Thrust allocation and each thruster. A malicious entity could intercept these links and inject a false data injection attack on the control signal.

In dynamic positioning systems, we are dealing with an over-actuated vessel, meaning the number actuators/thrusters is bigger than the degree of freedom (DoF). In dynamic positioning systems, the DoF is usually 3 - surge, sway, and yaw. The number of thrusters is usually around 7-8. Therefore, if there are n number of thrusters onboard the vessel, there would be n communication lines between the thrusters and command and control. In other words, the problem turns into an actuator attack, where n number of control commands from the controller to the actuator/thrusters are susceptible to false data injection attacks. The purpose of this thesis is threefold: 1- Developing a novel framework for thrust allocation in a way that the number of communication lines from the controller to the thrusters is minimized, hence reducing the attack surface. 2- Should a false data injection attack occurs in the new thrust allocation framework, the system remains resilient to those attacks, meaning that the attacks would not destabilize the system and would only result in a bounded deviation . 3- Securely estimating the true systems states along with estimating the attack vector in the presence of false data injection attack. Accomplishing these goals will result in a secure estimation and control of the thrusters in the dynamic positioning system in the presence of false data injection attacks.

1.3 Literature Review

1.3.1 Cyber-security of CPS

As mentioned previously, OT systems onboard the vessel which operate within the framework of IPMS, are considered Cyber-physical systems. The dynamic positioning system of a vessel is one of them. In the event of potential corruption by cyber-attacks, leveraging the literature on cyber-security of CPS from a control theoretic perspective, would be useful in order to diagnose the attack and continue controlling the system notwithstanding the presence of attacks [21].

Attack Models

As far as the cyber-security of the IT domain is concerned, three main properties are considered. Confidentiality, availability and integrity [1]. Confidentiality is defined as privacy and non-disclosure of the data to unauthorized entities. Integrity of the data refers to the their authenticity and availability is defined as timely access to the data. Unlike in IT systems where cyber-security is mainly concerned with the protection of the information and data, cyber-attacks on networked control systems and OTs, would affect the physical parts due to the communication links in the control and feedback loops. Therefore, the attack space would be wider in cyber-physical systems compared to traditional IT systems. Teixeira et al. [1] exhibits the attack space in cyber-physical systems in Figure 1.2.

Figure 1.2 shows three axes, each representing a dimension in the capabilities of the

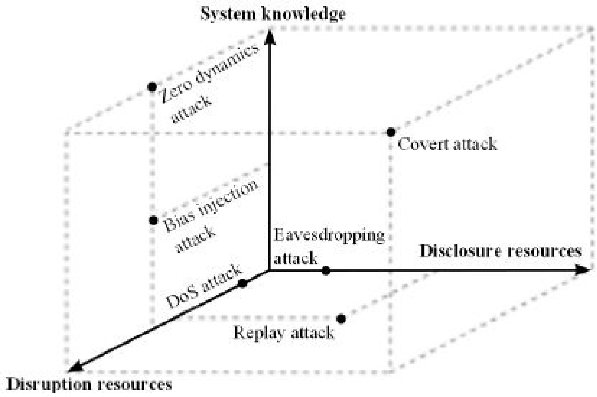


Figure 1.2: The attack space in cyber-physical systems [1].

attacker. These capabilities or resources are categorized into disclosure resources, disruption resources and system knowledge. Depending on which of these resources or their combinations the attacker possesses, different attacks could occur.

Disruption attacks are cyber-attacks that disrupt the flow of information in networked control systems through compromising their availability or integrity. If the attacker disrupts the availability of data, the attack is called Denial of Service attack (DoS), and if the integrity of the data is compromised, it is a False Data Injection (FDI) attack.

DoS attacks prevent the delivery of control and measurement packets. This could happen when a malicious node sends multiple requests to a server in the network. Another form of DoS attack is jamming the wireless transmissions. These attacks could be performed on various wireless technologies such as GPS, Wi-Fi, and mobile communications [22].

Long et al. [23] Proposed two models to approximate the behaviour of packet transmission under the influence of DoS attacks in NCSs and two methods were presented. In the first model, the attack occurs on an endpoint in the network. This endpoint could be a controller or a plant. In the second model, DoS attack is approximated in the case where the attacks is performed remotely on the service-provider-edge routers. Experiments show that these attacks tend to slow down the transmission of data between a controller and a remote plant.

FDI cyber-attack is a cyber-attack that happens when a malicious entity gains access to the communication links between the components of a NCS and injects erroneous data packets. The injection of inaccurate data packets on the original signal can cause the state estimation signals in the NCS to generate false values, which may bring about unpredictable and unstable responses, hence disrupting a system's desired operation [24]. False data injection attack is generally modeled as a bias signal being superimposed on the original signal. Therefore, this attack is also referred to as bias injection attack. FDI attacks are one of the most common and feasible cyber-attacks due to their low resource requirements, making them a significant threat to cyber-physical systems (CPS).

There are other forms of cyber-attacks in the attack space which are more sophisticated and demand more resources. These are replay attacks, covert attacks and zero dynamics attacks. Replay attacks are cyber-attacks in which the attacker replaces the real-time data with the former recorded data. The recorded data are usually sensor measurements. Two steps are carried out in this attack. Firstly, the attacker gathers sensor readings and stores the data, and then replays the gathered data to the controller [25]. In order to perform this attack, the attacker does not need to know the model of the system.

The most challenging type of attack to deal with is covert attack. Covert attacks rely on the complete system knowledge and complete access to the control input signals as well as to the sensor signals transmitted over the network. Here, the attacker injects corrupt signals on the control channel and cancels the influence of the additive attack signal by calculating the resulting output and subtracting it from the sensor readings. Hence, the diagnosis system located on the controller side of the network receives data which does not contain any information on the attack [25]. Another destructive and hard to detect attack is zero-dynamics attack. In this type of attack, the attacker knows the complete knowledge of system and does not need to read the input and output of the system. They just inject an attack as input on actuator channel in the same frequency of the right-half zero of the non-minimum phase system, which makes the system internally unstable [25].

Preventive Measures

The first layer of defense against malicious entities that intend to inflict cyber-attacks on the system is the preventive measures. These measures are employed to safeguard against the breach of attackers and securely transmit the data within the network. Two of the most notable preventive measures are authentication and encryption-decryption methods. Towards this end, the works of [26] and [27] have employed authentication based methodologies. More specifically, Jovanov and Pajic [26] have applied intermittent authentication to enforce the integrity of sensor readings that are transmitted from sensors to controller. Another notable preventive measure which has been used extensively in the IT domain but has found its way into the cyber-security measures for CPS, is encryption-decryption. In [28] the authors have employed a fully homomorphic encryption algorithm for the transmission of the data between the sensor and the controller, and also between the controller and the actuator. In [29] the authors have performed encryption on the sensor readings of the agents before sending them over to the neighboring agents. The data then is decrypted and processed for the consensus protocol.

The major problem with the use of encryption-decryption methods in real-time applications such as DP, is the computational and communication overhead that is brought about, which could hinder the desired operation of the system.

Attack Surface Shifting and Reduction

Sometimes adversaries manage to bypass the authentication and encryption-decryption protocols and breach into the system. Therefore, further measures are needed to safeguard the system. Another way of increasing the security of the system against cyber-attacks, is manipulating the attack surface. Attack surface is defined as the set of the resources of the system that can be exploited through cyber-attacks [30], or a set of system properties that can be used to launch cyber-attacks [31]. Therefore, it follows that in order to increase the security of the system, changing the attack surface would be a reasonable idea.

Changing the attack surface would increase uncertainty for the attacker and will make it harder for them to perform a successful attack. Towards this end, moving target approaches have been employed as a means to shift the attack surface [32], [31]. Moving target defense (MTD), is a proactive cyber-defense mechanism which aims at continually changing a system's attack surface over time through reconfiguration – which could be executed periodically and erratically– with the goal of increasing complexity and cost for attackers during reconnaissance activities, restricting the exposure of vulnerabilities, and increasing the overall system resilience [32].

Furthermore, attack surface could also be reduced, in order to reduce the vulnerabilities. CAI et al. [31] have considered applying MTD as a means to reduce the attack surface. In [33] the authors have developed a decentralized event-triggered mechanism in order to reduce the attack surface. The event-triggered mechanism was developed for network connection and communication, and agents were allowed to disconnect from the network in order to minimize the attacker's window of opportunity when exerting attacks on the agents. In [34], a security system for next generation industrial control systems was introduced in order to reduce the attack surface. The reduction in the attack surface was accomplished through limiting the communication paths to only the necessary paths, depending on the present configuration of the system.

Detection and Isolation of Cyber-Attacks

Once an attacker manages to evade the prevention protocols and overcome the attack surface manipulation schemes, they can compromise the network and launch their attack. As a result, detecting and isolating the attacks in the system becomes of vital importance. FDI cyber-attack could be modeled as an additive signal, hence leaving behind signatures similar to that of faults. Therefore, implementing the methods that have been developed in the fault detection and isolation literature could be beneficial. The detection and isolation procedure involves generating residuals using observers, and then developing a decision making process using the residual signals to detect and isolate the faults [35], [36].

There have been tremendous amount of work in the cyber-security of cyber-physical systems from a control theoretic point of view in order to adopt the methedologies used in the fault detection and isolation literature and use them to detect FDI cyber-attacks. A χ^2 detector based on Kalman filter has been developed in [37] to detect FDI cyber-attacks on sensors in smart grid. In [38] a Bayesian detection scheme considering binary hypothesis based on Kalman filter was developed to detect FDI cyber-attack on sensor measurements. A weighted least square method was used in [39] to detect sensor FDI cyber-attack in power systems.

The mentioned papers and the majority of observer-based detection schemes in the literature are designed for sensor attacks, which use Luenberger like observers. However, a decent amount of work has also been done with regards to detection and isolation of actuator attacks. In [40], a bank of unknown-input observers (UIO) have been employed to detect and isolate FDI cyber-attacks on both actuator and sensor channels. In [41], the notion of UIO observers was used along with graph-theoretic and system theoretic methods to detect FDI cyber-attacks on sensor and actuators in descriptor systems. In [42], the authors have applied UIOs to detect actuator attacks and estimate the attack magnitudes. A data driven approach was applied to detect and identify FDI cyber-attacks on only a subset of actuator channels in [43]. A hybrid data driven and model based approach was employed in [24] to detect FDI cyber-attack in NCS. For a more extensive review of the model-based methodologies foe detection of FDI cyber-attacks, the reader is referred to [44, 45, 46].

Secure Control and Estimation

Detection and isolation of the cyber-attacks are not the end goal when it comes to constructing security measures in cyber-physical systems. One reason for that is that while we have detected and isolated the attacks, the states of the system will not be properly estimated, or the controller will not lead to the desired operation of the system. It must be mentioned that most of the time, the observers used for detection and residual generation do not necessarily yield true estimates of the system [47].

In [48], estimation of the states were performed for CPSs subject to FDI cyber-attacks on sensors with measurement noise. Jeong and Eun [49] developed UIO based observers that estimated true system states in the presence of FDI cyber-attacks on sensors. The estimators were also robust to disturbances. In the work of Fawzi et al. [50], the exact system states were recovered in the presence of FDI cyber-attacks on a subset of sensors, and a state feedback controller was designed based on the estimated states. They investigated the maximum number of sensors that could be attacked while managing to recover the true system states.

Among the methods for secure control of the system under FDI cyber-attack, is the approach of attack compensation via attack estimation. In [51], using a novel sliding mode observer, true system states were estimated and the controller was compensated using reconstructed attack signals.

Gao et al. [52] tackled the problem of state estimation in the presence of sparse sensor and actuator attacks and secure control using both time driven and event triggered mechanisms. As far as passive resilient control of CPS against actuator attacks is concerned, there are various works that deals mitigating the effect of cyber-attacks only by focusing on designing the controller, such as the work by yang et al. [53]. Not every secure control problem needs true system states to maintain stability or desired behavior. In [54], an adaptive controller was designed to mitigate the effects of FDI cyber-attacks on sensors and actuators without estimating correct system states. One of the drawbacks of these works is the fact the passive resilient controller rely on feedback to achieve its objectives, whereas in some DP operation modes, the control command is set manually by the operator at the command and controller.

One could also resort to control reconfiguration via attack isolation. In this method, the corrupted actuators and sensors are removed once detected and isolated, and the controller is reconfigured accordingly. The work of yang et al. [40] is a good example of that.

1.3.2 Cyber-security of Dynamic Positioning (DP) Systems and Over-Actuated Systems

Few works have investigated cyber-security of DP from a control theoretic perspective. In [55], using fuzzy modeling, an event-triggered secure control mechanism was developed for an autonomous DP system that was controlled remotely and was subjected to DoS attacks on both actuator and sensor channels. Another major work relevant to secure control of DP is [56], where an H_∞ based hybrid triggered control mechanism was developed using observers to control an unmanned DP system that was under deception (FDI) cyber-attack on the actuator channel. Non of these works have considered over-actuated DP systems, which constitute the majority of DP systems. The challenge lies at the large attack surface that arises from several number of communication lines from controller to the individual actuators-thrusters. Furthermore, control-thrust allocation module is often neglected in fault diagnosis and fault tolerant of over-actuated systems.

Active isolation of actuators were performed considering the control allocation module, for over-actuated systems using a family of UIO observers in [57]. The limitations of isolating simultaneous faults were also studied. Cristofaro and Johansen [58] incorporated the thrust allocation module in their work and investigated isolating thruster faults using bank of UIOs. They took it on step forward and performed fault tolerant control by reconfiguring the allocation mechanism. In [59], the authors have assumed that the detection and isolation module has already identified the faulty actuators, and using a novel control allocation algorithm and a robust controller, they have managed to satisfy the closed loop stability and H_2 specification in the presence of uncertainties.

1.3.3 Attack Estimation and Compensation

The controllers that are used in the DP systems of marine vessels, generally use LQR controller and feed forward controller. The feed-forward control law is generated using the estimated wind force, which is acquired through a separate observer specifically tailored for measuring and estimating disturbance forces. However, the controller lacks the ability to counteract the effect of possible FDI attacks. As discussed in Section 1.3.2, the only methods developed to deal with anomalies are those that require isolation of the attacks. No paper has investigated attack estimation and compensation using estimated attack signals.

Zhu et al.[60] used unknown input observers to estimate system states and then reconstructed the unknown inputs, which were considered disturbances, to compensate for their effect. However, the observer matching condition that holds true in this work, does not apply to the DP system's state space configuration. In [51], states were estimated using adaptive sliding mode observer in the presence of actuator attacks and sensor attacks, and the attack signals were reconstructed and compensated for at the controller. The only issue is that the modeling the effect of the attacks in this work is not readily applicable to over-actuated systems subject to actuator attacks.

1.3.4 Control and Thrust Allocation

The primary objective of a control allocation module is to compute a control input $\boldsymbol{u} \in \mathbb{R}^n$, where n is the number of actuators or effectors, and $\boldsymbol{u}$ is the vector which consists of control inputs for individual actuators. This control input must ensure that the commanded control $\boldsymbol{u}_c$ is consistently generated by the effectors at all times according to the following relation: $\boldsymbol{u}_c = T\boldsymbol{u}$ [61], where T is the allocation matrix, representing a linear mapping between the actuator commands and the commanded control input. It can be either time-varying or time-invariant. In the case of DP systems in marine vessels or ROVs, T is the thruster configuration matrix that is dependent on the location and types of the thrusters onboard the vessel. In this thesis, this matrix is assumed to be time-invariant.

The thrust allocation problem in vessels usually turn into quadratic programming. The methods of solving this are divided into explicit and iterative approaches. Most of the explicit solutions use Lagrangian multipliers and least square to solve the allocation problem. There are also explicit methods based on linear piece-wise functions to avoid nonlinearities that could arise in the optimization problem in the case of rotating thrusters, i.e. azimuth thrusters. The other method is the iterative method, which has an advantage of having more flexibility in online reconfiguration, lower computational complicity by an appropriate initialization, and better suited to large scale control allocation problems [62]. Notable iterative methods used in quadratic programming are active set method and interior point method.

All of the aforementioned methodologies fall under centralized thrust allocation problem. The main complication arising from this type of allocation in terms of cyber-security, is the large attack surface. As the number of thrusters increase, the communication links between the controller and thrusters increase, leaving behind huge vulnerabilities to FDI cyber-attacks. The other complications have to with large memory needed for computation, and not being able to easily reconfigure it in case additional thrusters need to be deployed.

1.3.5 Decentralized Control Allocation

In light of the challenges arising from memory requirements, computational overhead and lack of scalability, some papers attempted at solving the control allocation problem in a distributed manner. In [63, 64] the problem of developing decentralized algorithms for control allocation for spacecraft was addressed. The authors used distributed estimation methods, each in a different way, to solve this issue. However, the optimization problem of interest in these works have an equality constraint which has one dimensional demand term. The dimension of the equality constraint is of significant importance in distributed optimization problems, and the solutions and methodologies developed for one dimensional and separable equality constraints are not applicable to optimization problems which have multi-dimensional and inseparable equality constraints. Moreover, their works only considered undirected topologies.

Lu et al. [65] approached the problem of developing distributed algorithm for thrust allocation for surface vessels. They utilized the concept of economic dispatch. The shortcoming of this work is the fact that the thrusters need to receive virtual expected force every time from the central controller. Moreover, they need to be initialized by another entity, presumably the controller. These necessitates communications between the thruster and the controller.

1.3.6 Distributed Optimization and Resource Allocation

The concept of distributed allocation is directly associated with distributed optimization. There have been tremendous amount of work dedicated to this area in the literature. A lot of these works are concerned with the economic dispatch problem, such as [66] and [67]. Lin et al. [68] and Bai et al. [69] used the concept of saddle point dynamics in constructing their distributed optimization protocols. The limitations of these works is related to the structure of their equality constraint, where each node already knows its own demand, and therefore the equations for the equality constraints are separable. This is not the case for the thrust allocation of the DP system for marine vessels.

To tackle this problem, recently, the concept of Alternating Direction Method of Multipliers (ADMM) have been popular and proven to be effective in distributed optimization problems with inseparable equality constraints [70]. However, in these methods, every agent computes an estimate of the whole problem and is not suited to the nature of the thrust allocation, where each thruster has to eventually figure out its own desired thrust value.

The work in [2] has used the concepts of average tracking consensus, multi time scale dynamics and saddle point dynamics in order to address distributed optimization with inseparable multi-dimensional equality constraints. This work could be applicable to the thrust allocation for the DP system. However, the major problem that could arise here is the vulnerability of this network to cyber-attacks, as the protocol is not resilient to FDI cyber-attacks.

1.4 Research Gaps

In the literature of cyber-security of over-actuated systems, several research gaps remain unaddressed. No existing work has successfully achieved secure estimation and isolation in the presence of FDI attacks without relying on the strict and often unrealistic observer matching condition. This is a significant limitation given the practical challenges in meeting this condition. Furthermore, Beyond the challenge of the observer matching condition, there is an absence of comprehensive solutions that simultaneously address secure state estimation, attack isolation, attack signal reconstruction, and compensation in over-actuated systems. This gap is particularly critical for over-actuated systems like dynamic positioning (DP), where achieving a holistic and integrated approach is essential for maintaining system performance and resilience in the face of cyber-attacks. Moreover, the current mitigation approaches often exclude compromised thrusters after isolation, resulting in sub-optimal performance and potential loss of control. This demonstrates a clear need for more robust and inclusive strategies to handle FDI attacks effectively.

Additionally, the problem of attack surface shifting/reduction and the development of proactive preventive measures to deter the adversary from using its disruptive capabilities and inject FDI attacks, have not been adequately addressed for over-actuated systems. Moreover, existing passive resilient control methods often fail to be effective during open-loop operations of the system. Transforming the centralized thrust allocation scheme into a decentralized thrust allocation scheme could be a potential solution that addresses both resiliency and attack surface shifting and reduction. Yet, no research has focused on developing such decentralized methods specifically for dynamic positioning to mitigate the security issues inherent in conventional centralized frameworks subjected to potential FDI cyber-attacks.

1.5 Thesis Contributions

In this thesis, a comprehensive solution for secure state estimation, attack reconstruction, isolation, and compensation in over-actuated systems is offered, all integrated within a centralized thrust allocation framework. Key contributions include achieving isolation and attack signal estimation on individual channels without the observer matching condition. Additionally, it introduces a novel compensation method using the estimated signals, addressing the limitations of exclusion-based methods.

Furthermore, a new thrust allocation algorithm has been introduced as a means of reducing and also shifting the attack surface to prevent potential adversaries from injecting FDI attacks between the controller and thrusters. This new decentralized allocation algorithm is also resilient to FDI cyber-attacks. Therefore, any FDI cyber-attacks that could have occurred on the communication lines between the controller and the thrusters in the conventional allocation schemes, if happened in this new allocation module, its effect would be attenuated. This resilient decentralized method is effective for both open-loop and closed-loop operations. Moreover, some of the methodologies developed in this thesis for the centralized scheme are extended and applied to the decentralized scheme to account for the considerable residual effects of attacks of significant amplitude that could not be sufficiently attenuated by the resilient consensus protocol, thereby enhancing the overall resilience of the system.

1.6 Thesis Outline

The rest of this thesis consist of four more chapters. In Chapter 2, the background information related to the development of the secure estimation in the presence of unknown inputs, attack reconstruction and compensation, and resilient decentralized allocation algorithm are provided. More specifically, the chapter starts off by explaining the DP system model and its parameters, which is the benchmark of this work. Next, the notions of unknown input observers and input observability and how they pertain to the problem of secure estimation and attack signal reconstruction, are discussed. Followed by that, the basics of thrust allocation are laid out. Thrust allocation is a key aspect of this thesis, especially for Chapter 4. Later, in order to develop the decentralized allocation scheme, the concept of saddle point dynamics and its relation to optimization are sufficiently explained in Chapter 2. Chapter 2 will be concluded by necessary explanations on the graph theory and dynamic average consensus, as development of the resilient decentralized allocation architecture heavily relies on them.

In Chapter 3, using sliding mode observers, the true system states are estimated in the presence of FDI attacks on the communication lines between the controller and the thrusters. Next, isolation of the FDI attacks, attack signal reconstruction and compensation methodologies are provided. Finally, the drawbacks of the mitigation methodologies in the literature that rely on excluding the affected thrusters and their channels from operation, are investigated.

In Chapter 4, the necessity for a new allocation strategy is examined in greater detail, along with its desired characteristics. Next, the resilient decentralized thrust allocation scheme is developed. Furthermore, the effects of the FDI cyber-attacks that could not be sufficiently attenuated by the resilient algorithm, are estimated and compensated for using the methodologies developed in Chapter 3.

Finally, Chapter 5 is dedicated to summarizing the conclusions of the thesis and proposing potential avenues for addressing unresolved issues in future research.

Chapter 2

Background Information

2.1 Mathematical Model of The ship

The equations of motion for a surface vessel with 6 degrees of freedom (DoF) is as follows [71]:

$$\boldsymbol{M}\dot{\boldsymbol{\nu}} + \boldsymbol{C}(\boldsymbol{\nu})\boldsymbol{\nu} + \boldsymbol{D}(\boldsymbol{\nu})\boldsymbol{\nu} + \boldsymbol{g}(\boldsymbol{\eta}) = \boldsymbol{g}_0 + \boldsymbol{\tau} + \boldsymbol{w} \tag{2.1}$$

where vector $\boldsymbol{\eta}$ represents the positions and the Euler angles: $\boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{\Theta} \end{bmatrix}$, with $\boldsymbol{P}$ denoting

Variables and Parameters	Descriptions
$\boldsymbol{M}$	system inertial matrix
$\boldsymbol{C}(\boldsymbol{\nu})$	Coriolis-centripetal matrix
$\boldsymbol{D}(\boldsymbol{\nu})$	damping matrix
$\boldsymbol{g}(\boldsymbol{\eta})$	vector of gravitational and buoyancy forces and moments
$\boldsymbol{\tau}$	vector of control inputs
$\boldsymbol{g}_0$	forces for ballast control
$\boldsymbol{w}$	vector of environmental forces

Table 2.1: Description of the variables and matrices involved in the dynamics of a 6 DoF surface vessel.

the position in earth fixed frame, i.e. motion in X, Y, Z directions. The Euler angles $\boldsymbol{\Theta}$, also defined in earth fixed frame, represent the orientation of ship in roll ϕ, i.e. rotation about the X axis, pitch θ, i.e. rotation about the Y axis, and yaw ψ, i.e. rotation about the Z axis.

The vector $\boldsymbol{\nu}$ represents the linear velocities, i.e. v_x, v_y, v_z and angular velocities, i.e. p, q, r. These linear and angular velocities are defined in the body fixed reference frame.

Therefore, $\boldsymbol{P} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$ and $\boldsymbol{\Theta} = \begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix}$. Furthermore, the velocity vector $\boldsymbol{\nu}$ becomes:

$$\boldsymbol{\nu} = \begin{bmatrix} v_x \\ v_y \\ v_z \\ p \\ q \\ r \end{bmatrix} \tag{2.2}$$

The relationship between the velocities, and the time derivative of the positions and Euler angles is described below:

$$\dot{\boldsymbol{\eta}} = \boldsymbol{J}(\boldsymbol{\eta})\boldsymbol{\nu} \tag{2.3}$$

Equation (2.3) represents the kinematics of a surface vessel in 6 DoF, and $\boldsymbol{J}(\boldsymbol{\eta})$ is the transformation matrix that is a function of the position and Euler angles.

In DP application, the motion is horizontal and limited to movement in X, Y, ψ, or surge, sway and yaw, respectively. This means $v_z = p = q = 0$. Furthermore, in DP operation, the vessel is supposed to operate at very low speeds. These assumptions turn Equation (2.4) into the following equation:

$$\boldsymbol{M}\dot{\boldsymbol{\nu}} + \boldsymbol{D}(\boldsymbol{\nu})\boldsymbol{\nu} = \boldsymbol{\tau} + \boldsymbol{w} \tag{2.4}$$

where the centripetal-Coriolis related forces along with the gravitational and buoyancy forces are neglected. Furthermore, $\boldsymbol{\eta}$ is reduced to $\boldsymbol{\eta} = [X \quad Y \quad \psi]^T$ and $\boldsymbol{\nu}$ changes to $\boldsymbol{\nu} = [v_x \quad v_y \quad r]^T$. In horizontal motion, the transformation matrix $\boldsymbol{J}$ becomes:

$$\boldsymbol{J}(\boldsymbol{\eta}) = \boldsymbol{R}(\boldsymbol{\Theta}) = \begin{bmatrix} cos(\psi) & -sin(\psi) & 0 \\ cos(\psi) & -sin(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.5}$$

Furthermore, the motion of a vessel in DP mode is divided into two motions: low frequency and wave frequency modes. Low frequency mode is due to the current, wind and second-order mean and slowly varying wave loads, whereas wave frequency mode or high frequency mode is driven by first order wave disturbances [7]. To avoid wear and tear of thrusters, the control action of the DP system will be designed to only account for the low frequency motion. Moreover, in most operations, wind force becomes the dominant disturbance term.

Hence, in this study only the low frequency motion is investigated. In low frequency mode, the matrices $\boldsymbol{M}$ and $\boldsymbol{D}$ become:

$$M = \begin{bmatrix} m - F_{X\dot{v_X}} & 0 & 0 \\ 0 & m - F_{Y\dot{v_y}} & mX_G - F_{Y\dot{r}} \\ 0 & mX_G - F_{Y\dot{r}} & I_z - M_{\psi_{\dot{r}}} \end{bmatrix} \text{and } D = \begin{bmatrix} -F_{X_{v_x}} & 0 & 0 \\ 0 & -F_{Y_{v_y}} & -F_{Y_r} \\ 0 & -M_{\psi_{v_y}} & -M_{\psi_r} \end{bmatrix}.$$

where $F_{X\dot{v_x}}, F_{Y\dot{v_y}}, F_{Y\dot{r}}, M_{\psi_{\dot{r}}}$ are added masses. For instance, $F_{X\dot{v_X}} = \frac{\partial F_X}{\partial \dot{v_X}}$ where F_X is the hydrodynamic added mass force in X direction, m is the mass of the ship, I_z is the moment of inertial about the Z axis, X_G is the distance of the center of gravity of the ship from the origin of the body fixed frame along the X axis. As far as the coefficients in the D matrix, $F_{X_{v_x}}, F_{Y_{v_y}}, F_{Y_r}, M_{\psi_{v_y}}, M_{\psi_r}$ are hydrodynamic linear damping coefficients [71]. In DP operations, it is common to choose the earth fixed frame as the reference frame.

Furthermore, it is common practice to linearize the model of the DP system around the operating point.

Given all the above considerations, $\boldsymbol{R}(\Theta) = I_3$ and the linearized DP system model is obtained as follows:

$$\begin{aligned} \dot{\boldsymbol{x}} &= A\boldsymbol{x} + B\boldsymbol{u}_c + B\boldsymbol{u}_w \\ \boldsymbol{y} &= C\boldsymbol{x} \end{aligned} \tag{2.6}$$

where $\boldsymbol{\tau}$ has been repalced by $\boldsymbol{u}_c$ as the commanded control signal or the commanded force for the thrusters, and $\boldsymbol{w}$ has been replaced by $\boldsymbol{u}_w$ which is the wind force as the disturbance term. The states vector is defined as : $\boldsymbol{x} = [X \quad Y \quad \psi \quad v_x \quad v_y \quad r]^T$. The matrices A and B are defined as $A = \begin{bmatrix} 0_{3\times3} & I_{3\times3} \\ 0_{3\times3} & -M^{-1}D \end{bmatrix}$ and $B = \begin{bmatrix} 0_{3\times3} \\ M^{-1} \end{bmatrix}$.

In this work, the parameters of the ship have been chosen for supply vessel according to [72]. For this vessel, $m = 4.5 \times 10^6, L = 1.225 \times 70$. Matrices M and D have been defined based on the bis model as following:

$$\begin{aligned} M &= mE^{-2}(EM''E^{-1}) \\ D &= m\sqrt{\frac{g}{L}}E^{-2}(ED''E^{-1}) \end{aligned} \tag{2.7}$$

where $M'' = \begin{bmatrix} 1.1274 & 0 & 0 \\ 0 & 1.8902 & -0.0744 \\ 0 & -0.0744 & 0.1278 \end{bmatrix}$ and $D'' = \begin{bmatrix} 0.0358 & 0 & 0 \\ 0 & 0.1183 & -0.0124 \\ 0 & -0.0041 & 0.0308 \end{bmatrix}$ and $E = diag(1, 1, L)$.

Regarding the outputs in (Equation 2.6), the GNSS and gyro-compass measure the position X, Y and orientation ψ respectively. Therefore, $C = [I_{3\times3} \quad 0_{3\times3}]$.

2.2 Control of DP

Linear Quadratic Regulator (LQR) is a widely used in DP controllers. It functions based on the principle of optimal control by determining control inputs that minimize the following cost function:

$$J = \int_0^\infty (\mathbf{x}^T Q \mathbf{x} + \mathbf{u}^T R \mathbf{u}) \, dt \tag{2.8}$$

Given the state-space equations in (2.6), the LQR control law is:

$$\boldsymbol{u}_c = -K\boldsymbol{x}$$

where K is the optimal feedback gain matrix obtained from the solution of the Riccati equation:

$$\begin{aligned} & K = R^{-1} B^T P \\ & P : A^T P + PA - PBR^{-1} B^T P + Q = 0 \end{aligned}$$

where P is the solution to the Riccati equation, and Q and R are positive definite weighting matrices usually chosen to represent the control objectives and constraints.

Depending on the sensors onboard the vessel, either output feedback or full state feedback can be employed for the LQR controller. If only GNSS and gyro-compass are involved, output feedback should be used alongside state observers, which can estimate the system states in finite time. In this scenario, the control input u_c is given by:

$$\boldsymbol{u}_c = -K\hat{\boldsymbol{x}} \tag{2.9}$$

where $\hat{x}$ is the estimated state vector provided by the state observer.

In more practical scenarios, only position sensors like GNSS and gyro-compass are available. Therefore, state estimation becomes essential to reconstruct the full state vector needed for effective control.

If the velocities are also measured, then full state feedback in the control law is justified. This means all relevant states, including positions and velocities, are directly measured, allowing the control input u_c to be computed as:

$$\boldsymbol{u}_c = -K\boldsymbol{x} \tag{2.10}$$

Here, $\boldsymbol{x}$ is the state vector measured directly by the sensors. The direct measurement of all states simplifies the control design and enhances system responsiveness.

The choice between output feedback and full state feedback depends on the available sensor suite. Full state feedback is preferable when velocity measurements are available, as it eliminates the need for state estimation and provides more precise control. On the other hand, output feedback is necessary when only position measurements are available, necessitating the use of state observers to reconstruct the full state vector.

2.3 UIO and Input Observability

Considering the following LTI system:

$$\begin{aligned} \dot{\boldsymbol{x}} &= A\boldsymbol{x} + B\boldsymbol{u} \\ \boldsymbol{y} &= C\boldsymbol{x} \end{aligned} \tag{2.11}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the vector of system states, $\boldsymbol{u} \in \mathbb{R}^m$ is the unknown input vector, and $\boldsymbol{y} \in \mathbb{R}^p$ is the output vector. It is considered that B has full column rank m.

The Rosenbrock matrix of this system is defined as follows:

$$P(s) = \begin{bmatrix} sI_n - A & -B \\ C & 0 \end{bmatrix} \tag{2.12}$$

In order to investigate the unknown input observability and input observability of the system, the invariant zeros of the system must be determined. The invariant zeros of the system 2.11, are values s in the complex domain which cause the Rosenbrock matrix to lose rank. In other words, its rank become less than normal:

$$rank(P(s_0)) < normrank(P) \tag{2.13}$$

where the normal rank of P is its maximum rank and has a relationship with the rank of the transfer function of the system $G(s)$ as following:

$$normrank(P(s)) = n + normrank(G(s)) \tag{2.14}$$

where $G(s) = C(sI_n - A)^{-1}B$. If the normal rank of the transfer function is m, and $p \geq m$, then the system is left invertible. Here are the definitions for strong observablity, strong detectability and $strong^*$ observability:

- **Strong observability:** Strongly observable, if $y(t) = 0$ for all $t \geq 0$ implies $x(t) = 0$ for all $t \geq 0$, all $u(t)$ and all $x(0) = x_0$.

- **Strong detectability:**Strongly detectable, if $y(t) = 0$ for all $t \geq 0$ implies $x(t) \to 0$ as $t \to \infty$, all $u(t)$ and all $x(0) = x_0$.

- $strong^*$ **detectability:** $strong^*$ detectable, if $y(t) \to 0$ as $t \to \infty$ implies $x(t) \to 0$ as $t \to \infty$, all $u(t)$ and all $x(0) = x_0$.

According to [73], the relationship between the three notions of observalities and system zeros are provided below:

- The system is strongly observable if and only if $rank(P(s)) = n + m \quad \forall s \in \mathbb{C}$. In other words, it must have no invariant zeros.

- The system is strongly detectable if and only if $rank(P(s)) = n + m \quad \forall s \in \mathbb{C}$ and $Re(s) > 0$. In other words, the system should not have any non-minimum phase zeros, i.e. zeros in the right half-plane of the complex domain.
- The system is $strong^*$ detectable if and only if it is strongly detectable and $rank(CB) = rank(B) = m$. This criterion is called the observer matching condition.

If the system is $strong^*$ detectable, then a linear UIO in the following form could be constructed to estimate the system states without the inputs [74]:

$$\begin{aligned} \dot{\boldsymbol{z}} &= N\boldsymbol{z} + G\boldsymbol{u} + L\boldsymbol{y} \\ \hat{\boldsymbol{x}} &= \boldsymbol{z} - H\boldsymbol{y} \end{aligned} \tag{2.15}$$

However, if the observer matching condition does not hold, but the system is still strongly observable, then it is still possible to estimate the states by including the higher order derivatives of the measurements, or augmenting the system outputs by additional virtual outputs and estimating them using higher order sliding mode observers [73].

The issue of estimating system states without the inputs, is different than estimating the unknown inputs themselves. The input $u(t)$ is considered observable when $y(t) \geq 0$ for $t \geq 0$ leads to $u(t) \geq 0$ for $t \geq 0$.

The input signal $u(t)$ is considered detectable if whenever $y(t) \geq 0$ for $t \geq 0$, it implies that $u(t) \rightarrow 0$ as $t \rightarrow \infty$ [75].

According to [75], the unknown input $\boldsymbol{u}(t)$ is observable if and only if $normrank(P(S)) = n + m$. Meaning the system should be left invertible. Moreover, the unknown input $\boldsymbol{u}(t)$ is detectable if and only if the system has only non-minimum phase zeros.

Finally, as stated by [75], input observability is a necessary and sufficient condition for the existence of an estimator which reconstructs the unknown inputs. In Chapter 4, the DP system and its observability and input observability will be further investigated and its required observer will be developed.

2.4 Thrust Allocation

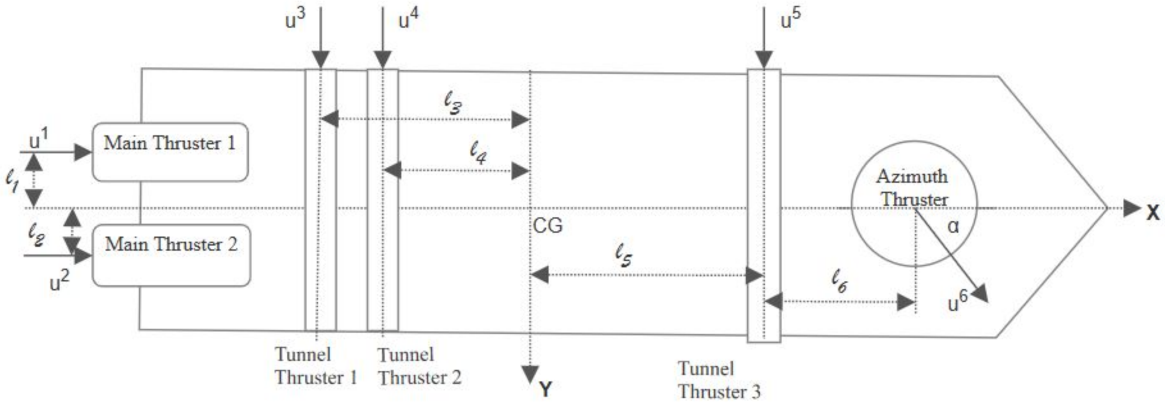


Figure 2.1: Configuration of the thrusters onboard the vessel. l_1 and l_2 are the vertical distances of thruster 1 and 2, respectively, from the center of gravity of the vessel, i.e CG. l_3 and l_4 are the horizontal distances of thruster 3 and 4, respectively, from CG. l_5 is the horizontal distance of thruster 5 from CG, and l_6 is the horizontal distance of thruster 6 from CG. α is the angle between the direction of the thrust force produced by thruster 6 and the horizontal axis. X is the horizontal axis, i.e. surge, and Y is the vertical axis, i.e sway, in the body-fixed frame.

The generated control command $\boldsymbol{u_c}$ must be distributed among the actuator of the system which are the thrusters. The configuration of the thrusters in vessel of interest in this work are depicted in Figure 2.1. There are 6 thrusters, 2 of which are the main thrusters, 3 are tunnel thrusters and 1 in azimuth thruster. Since the generated control command is three dimensional and there are six thrusters, the system is over-actuated and there are more unknowns than there are equations. To derive the equations, the generated forces in surge, sway and yaw by the thrusters are written as follows:

$$\sum \boldsymbol{F_x} = u^1 + u^2 + u^6 cos(\alpha) \tag{2.16a}$$

$$\sum \boldsymbol{F_Y} = u^3 + u^4 + u^5 + u^6 sin(\alpha) \tag{2.16b}$$

$$\sum \boldsymbol{M_\psi} = u^1\ell_1 - u^2\ell_2 - u^3\ell_3 - u^4\ell_4 + u^5\ell_5 + u^6\ell_6 sin(\alpha) \tag{2.16c}$$

where u^1 and u^2 are the forces produced by the main thrusters, which are mounted aft of the hull, usually in conjunction with rudders. u^3, u^4 and u^5 are produced by the tunnel thrusters that are transverse thrusters going through the hull of the vessel. These thrusters are mounted inside a transverse tube and produce forces alongside the tube they are in. Tunnel thrusters are only effective at low speeds, which makes them applicable to DP operations [72]. There is only one azimuth thruster in the configuration and it generates u^6. Azimuth thruster can be rotated an angle α about the z axis and generate two force components in the horizontal plane. Azimuth thrusters are frequently employed in DP systems since they are capable of producing thrusts in different directions.

The Equations (2.16) can we written in a matrix format as follows:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & cos(\alpha) \\ 0 & 0 & 1 & 1 & 1 & sin(\alpha) \\ \ell_1 & -\ell_2 & -\ell_3 & -\ell_4 & \ell_5 & \ell_6 sin(\alpha) \end{bmatrix} \begin{bmatrix} u^1 \\ u^2 \\ u^3 \\ u^4 \\ u^5 \\ u^6 \end{bmatrix} = \begin{bmatrix} \sum \boldsymbol{F_x} \\ \sum \boldsymbol{F_Y} \\ \sum \boldsymbol{M_\psi} \end{bmatrix} \tag{2.17}$$

where the matrix of coefficients is called the thruster configuration matrix, and is denoted by T. Each column of this matrix corresponds to one thruster only. The vector of unknowns which signify the produced thrusts by the thrusters is denoted $\boldsymbol{u} = [u^1 \quad u^2 \quad u^3 \quad u^4 \quad u^5 \quad u^6]^T$. Obviously in order for the thrusters to satisfy the control command $\boldsymbol{u_c}$, then it should follow that:

$$T\boldsymbol{u} = \boldsymbol{u_c} \tag{2.18}$$

Furthermore, the thrusters need to work efficiently and consume the minimum fuel possible. If a total cost function is to be minimized, which is the sum of individual cost functions of the thrusters associated with their fuel consumption, the thrust allocation problem turns into an optimization problem as follows:

$$\begin{aligned} \underset{\boldsymbol{u}}{\text{minimize}} \quad & f(\mathbf{u}) = \boldsymbol{u}^T W \boldsymbol{u} \\ \text{subject to} \quad & T\boldsymbol{u} = u_c \end{aligned} \tag{2.19}$$

It is a constrained optimization problem with an equality constraint, where W is a weight matrix reflecting the share of each thruster in fuel consumption.

The lengths in Figure 2.1 is chosen to be: $\ell_1 = \ell_2 = 4, \ell_3 = 35, \ell_4 = 33, \ell_5 = 40, \ell_6 = 28$ and the azimuth thruster is considered to be fixed with $\alpha = 90°$.

Consequently, the thruster configuration matrix becomes:

$T = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 4 & -4 & -35 & -33 & 40 & 28 \end{bmatrix}$. The thrust allocation problem for vessels, which is a quadratic programming problem, could be solved using either explicit or iterative solutions. Both of them are centralized allocation methods, since one module is in charge of solving the optimization problem and allocation each thruster their specific thrust value. The explicit solution to this allocation problem is obtained using pseudo-inverses and is as follows:

$$\boldsymbol{u} = W^{-1}T^T[Tw^{-1}T^T]^{-1}\boldsymbol{u_c} \tag{2.20}$$

However, there are serious issues associated with the explicit method, such as computational complexity, memory requirements, and vulnerability to ill-conditioned matrices.

Instead, iterative methods have an advantage of showing more flexibility for online deployment and reconfiguration. One of the notable iterative methods to solve the quadratic programming problem is interior point method.

2.5 Optimization and Saddle Point Dynamics

The optimization problem with equality constraints could be formulated as follows [76]:

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad f_0(\boldsymbol{x}) \\ &\text{subject to} \quad h_i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, p \end{aligned} \tag{2.21}$$

where $x \in \mathbb{R}^n$. The Lagrangian of this problem will be:

$$L(x, \boldsymbol{\mu}) = f_0(x) + \sum_{i=1}^{p} \mu_i h_i(\boldsymbol{x}) \tag{2.22}$$

where $\boldsymbol{x}$ is the vector of the optimization variables, $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_p]$ is the vector of Lagrange multipliers associated with the equality constraints $h_i(\boldsymbol{x}) = 0$, and $f_0(\boldsymbol{x})$ is the objective function to be minimized.

The Lagrangian dual function is then defined as:

$$g(\boldsymbol{\mu}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\mu}) = \inf_{x} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{p} \mu_i h_i(\boldsymbol{x}) \right) \tag{2.23}$$

and finally, the Lagrange dual problem associated with 2.21is formulated as follows:

$$\begin{aligned} &\text{maximize} \quad g(\mu) \\ &\text{subject to} \quad \mu \in \mathbb{R}^p \end{aligned} \tag{2.24}$$

The Lagrange dual problem aims at maximizing the Lagrange dual function $g(\boldsymbol{\mu})$ with respect to the Lagrange multipliers $\boldsymbol{\mu}$. The optimal solution $\boldsymbol{\mu}^*$ to the dual problem, which maximizes $g(\boldsymbol{\mu})$, provides the tightest lower bound for the optimal value of the primal problem 2.21. To put it differently, $g(\boldsymbol{\mu}^*)$ represents the maximum achievable value of the cost function subject to the given constraints, and it serves as a lower bound for the optimal objective value of the primal problem.

To ensure that the optimal value of the Lagrange dual function coincides with the primal problem's optimal value, the duality gap mus be zero. This condition shows that the primal and dual problems are optimally solved, ensuring that the Lagrange dual function precisely captures the primal problem's behavior under the given constraints.

Slater's condition states that:

$$\exists x \text{ such that } h_i(x) = 0, \text{ for } i = 1, \ldots, p \tag{2.25}$$

If this condition holds, the duality gap is zero. Therefore, if $\boldsymbol{x^*}$ and $\boldsymbol{\mu}^*$ are primal and dual optimal points for the optimization problem if interest, they constitute a saddle point for the Lagrangian. In other words:

$$L(\boldsymbol{x}^*, \boldsymbol{\mu}) \leq L(\boldsymbol{x}^*, \boldsymbol{\mu}^*) \leq L(\boldsymbol{x}, \boldsymbol{\mu}^*) \tag{2.26}$$

The KKT (Karush-Kuhn-Tucker) conditions are necessary conditions for a point to be an optimal solution in the aforementioned constrained optimization problem. These conditions ensure that an optimal solution satisfies the following:

- The gradient of the objective function aligns with adjustments required by the equality constraint.
- The equality constraints are satisfied.

The KKT conditions for the constrained optimization problem are:

$$\begin{aligned}
&1. \quad \nabla f_0(x^*) + \sum_{i=1}^{p} \mu_i^* \nabla h_i(x^*) = 0, \\
&2. \quad h_i(x^*) = 0, \quad \text{for } i = 1, \ldots, p, \\
&3. \quad \mu_i^* \geq 0, \\
&4. \quad \mu_i^* h_i(x^*) = 0, \quad \text{for } i = 1, \ldots, p.
\end{aligned} \tag{2.27}$$

where x^* is the optimal solution, μ_i^* are the corresponding Lagrange multipliers, and ∇ denotes the gradient operator. In order to find these optimal points $\boldsymbol{x^*}$ and $\boldsymbol{\mu}^*$, KKT conditions can be used. It must however be noted that in order to utilize KKT conditions, the equality constraint must be affine and the objective function must be convex. In the case of the thrust allocation defined in 2.5, the cost function is indeed convex and the equality constraint $T\boldsymbol{u} = \boldsymbol{u_c}$ is obviously affine. Consequently, the optimal points could be obtained using the following equations that are defined as KKT conditions:

$$\begin{aligned}
h_i(\boldsymbol{x}^*) &= 0 \\
\nabla f_0(\boldsymbol{x}^*) + \sum_{i=1}^{p} {\mu_i}^* \nabla h_i(\boldsymbol{x}^*) &= 0
\end{aligned} \tag{2.28}$$

Once the optimal points are obtained, according to [77]:

- if the Lagrangian L is globally convex-concave, and linear in $\boldsymbol{\mu}$,
- for each $(\boldsymbol{x^*}, \boldsymbol{\mu}^*) \in Saddle(L)$, if $L(\boldsymbol{x}, \boldsymbol{\mu}^*) = L(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$, then $(\boldsymbol{x}, \boldsymbol{\mu}^*) \in Saddle(L)$,

then the following dynamics ensure that starting from any arbitrary point $\boldsymbol{x^0}, \boldsymbol{\mu}^0$, the trajectories converge to a point, which is $(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$:

$$\dot{\mathrm{x}} = -\nabla_x L(\mathrm{x}, \boldsymbol{\mu}) \tag{2.29a}$$

$$\dot{\boldsymbol{\mu}} = \nabla_\mu L(\mathrm{x}, \boldsymbol{\mu}) \tag{2.29b}$$

This ideas was first used in [78]. For more details regrading the stability and convergence criteria of the saddle point dynamics, refer to [77].

2.6 Graph Theory and Dynamic Average Consensus

2.6.1 Graph Theory

In the context of consensus of multi-agent systems, graph theory is the core subject, and in this section its fundamental concepts pertinent to this work are reviewed. A directed graph, or digraph, denoted as $G = (V, E)$, consists of a set of nodes $V = \{1, \dots, N\}$ and a set of edges $E \subseteq V \times V$. In this representation, an edge (i, j) indicates that agent j can send information over to agent i. Each edge connects two vertices, with i termed as an in-neighbor of j, and j as an out-neighbor of i. An undirected graph, on the other hand, satisfies $(i, j) \in E$ whenever $(j, i) \in E$, meaning if i can send information to j, the reverse is also true.

Within this framework, a directed path is an ordered sequence of vertices where consecutive pairs constitute edges of the digraph. A digraph is called strongly connected if there exists a directed path between every pair of vertices in the digraph. Considering weights for the edges of the graph, a weighted digraph $G = (V, E, A)$, where (V, E) forms the digraph, and $A \in \mathbb{R}^{N\times N}$ represents the weighted adjacency matrix. For any edge $(i, j) \in E$, $a_{ij} > 0$, while $a_{ij} = 0$ otherwise.

The weighted out-degree and weighted in-degree of a node i are respectively defined as $d_{\text{in}}(i) = \sum_{j=1}^{N} a_{ji}$ and $d_{\text{out}}(i) = \sum_{j=1}^{N} a_{ij}$. Notably, $d_{\text{out}}^{\max} = \max_{i\in\{1,\dots,N\}} d_{\text{out}}(i)$ represents the maximum weighted out-degree across all nodes. In addition, a digraph is considered weight-balanced if, at every node i in V, the weighted out-degree equals the weighted in-degree.

The out-Laplacian matrix is defined as $L = D_{\text{out}} - A$, where D_{out} represents the weighted out-degree diagonal matrix. It is worth mentioning that $L\mathbf{1}_N = \mathbf{0}$, i.e. kernel of the Laplacian matrix is a vector of all ones. Moreover, a weighted digraph G is considered weight-balanced if and only if $\mathbf{1}_N^T L = \mathbf{0}$. Laplacian matrix has important characteristics. For instance, at least one of its eigenvalues is zero, while the other eigenvalues possess non-negative real components. Denoting the eigenvalues of L as λ_i, $i \in \{1, \ldots, N\}$, where $\lambda_1 = 0$, and Re$(\lambda_i) \leq$ Re(λ_j), for $i < j$. For a digraph that is strongly connected, the presence of zero as a simple eigenvalue of L is evident [79].

2.6.2 Dynamic Average Consensus

Given a directed graph G that consists of N agents, where G is both strongly connected and weight-balanced, in which it is assumed that each node i belonging to the set V has a reference input denoted as $r^i \in \mathbb{R}^p$, if for $\beta > 0$ each agent is embedded with the following protocol:

$$\begin{aligned}
\dot{\boldsymbol{v}}^i &= \beta \sum_{j=1}^{N} a_{ij}(\boldsymbol{y}^{\boldsymbol{i}} - \boldsymbol{y}^{\boldsymbol{j}}), \\
\dot{\boldsymbol{y}}^i &= (\boldsymbol{r}^i - \boldsymbol{y}^i) - \beta \sum_{j=1}^{N} a_{ij}(\boldsymbol{y}^{\boldsymbol{i}} - \boldsymbol{y}^{\boldsymbol{j}}) - \boldsymbol{v}^i + \dot{r}^i
\end{aligned} \tag{2.30}$$

then according to [79], starting at any $\boldsymbol{y}^i(0), \boldsymbol{v}^i(0) \in R^p$, the states $\boldsymbol{y}^{\boldsymbol{j}}$ will converge to $\frac{1}{N}\sum_{j=1}^{N} \boldsymbol{r}^{\boldsymbol{j}}$, i.e. the average sum of the reference inputs. It must be noted that one more condition must be satisfied, which is $\sum_{j=1}^{N} \boldsymbol{v}^i(0) = 0$. For further details on the stability and convergence of this protocol, the reader is referred to [79].

2.7 Conclusion

In this Chapter, the prerequisites of addressing the cyber-security of the over-actuated DP system from a control-theoretic point of view are laid out. Firstly, the standard mathematical model of the DP system which will server as a benchmark for the rest of the analysis in this thesis is provided. The main characteristics of this model is found in the majority of the over-actuated system, like ROVs, spacecrafts, etc.

Next, the notions of UIO and input observability, which will be used in Chapter 3, are presented. The necessary conditions for strong observability are studied, which will be used for estimating the true system states in the presence of FDI attacks. The concept of UIOs are mentioned, since the process of attack isolation directly relies on the design of these observers.

Thrust allocation module, as an important aspect of this work, has been given especial attention to, since it lies between the thrusters and the controller of the DP system, and the signals are sent from this module to the thrusters. Furthermore, one of the main contributions of this work, is introducing a new decentralized thrust allocation framework in Chapter 4, as a method to reduce and shift the attack surface and deter the adversary from launching an attack on the communication links between the controller and the thrusters.

Additionally, the optimization problem, which is directly associated with the thrust allocation problem is discussed. The notion of saddle point dynamics is explained, which will be utilized in developing the decentralized thrust allocation framework in Chapter 4. Finally, since the decentralized thrust allocation framework is developed within a multi-agent paradigm, the graph theory and the concept of dynamic average consensus is also touched upon in this chapter.

Chapter 3

Secure Estimation, Isolation and Cyber-Attack Compensation in Dynamic Positioning System

3.1 State Awareness and Operational Normalcy

In cyber-physical systems such as DP, state estimation is of significant importance, especially in scenarios where the integrity of the input signals have been compromised by FDI cyber-attacks on the actuator channels. Having access to genuine system states is necessary for performing state feedback control, i.e. LQR control in DP. Furthermore, the significance of maintaining access to authentic system states extends beyond its involvement in the control loop. Having access to real system states is crucial for achieving a thorough awareness of the DP system's operation. If the command and control center at the IPMS station loses access to real system states, the operators might make inappropriate decisions. In DP, conventionally, the sensors measure the position vectors and all the other states must be estimated. Traditional Luenberger observers will be ineffective in estimating the true system states in DP in the presence of FDI attacks. Unlike in healthy scenarios when there is no cyber-attacks and the observer receives accurate input signals reflecting the actual control commands applied to the system, in the face of actuator attacks, the observer does not have direct access to the real input signals. It has access to the control signal originating from the controller, before being corrupted by a potential FDI cyber-attack. Therefore, that information would not be reliable. The observer has solely access to the measurements of the system's output and some knowledge of the system dynamics. Therefore, the observers employed in the system in presence of actuator attacks must be able to reconstruct the real system states without the need for inputs. Moreover, in DP systems, the isolation of FDI cyber-attacks is also related to the broader concept of state awareness within the cyber-physical framework. Within this concept, state awareness essentially includes detecting and isolating malicious additive signals affecting its communication links from the controller-thrust allocation module, to the thrusters. By successfully isolating a corrupted channel, the DP system enhances its overall awareness and paves the way for further necessary security measures.

Furthermore, in addition to state awareness, operational normalcy is another factor that must be considered when safeguarding the system against cyber-attacks. Operational normalcy is ensuring that the system does not deviate substantially from its desired behaviour, in the face of cyber-attacks. This chapter focuses on addressing the challenge of mitigating the effects of FDI attacks on the communication links between the controller and the thrusters within a centralized thrust allocation framework. Specifically, it develops a method to estimate the attack signals and isolate the corrupted thrusters, securing continuous and reliable operation of the DP system. This approach is implemented despite the limitations regarding the observer matching condition and introduces a new compensation strategy that addresses the issues that the existing compensation methods in the literature have failed to solve.

3.2 Problem Formulation

According to Figures 3.1a, 3.1b, the problem of estimating the true system states in DP system having been equipped with a centralized thrust allocation module and corrupted by FDI cyber-attacks on the communication links between the controller and thrusters, is equivalent to estimating the true system states of a DP without an allocation module that is being compromised by an additive attack signal onto the control command.

Furthermore, to estimate and compensate for the FDI attacks in Figure 3.1a, one way to address it is to estimate it. Therefore, instead of estimating the FDI cyber-attacks occurring on the communication lines between the thrusters, their overall effect on the control command in the equivalent problem demonstrated in Figure 3.1b could be estimated and compensated for. Not only estimating the attack signal is needed for compensation, but also it enhances the state awareness of the DP system. To sum up, given the state space equations of the DP system:

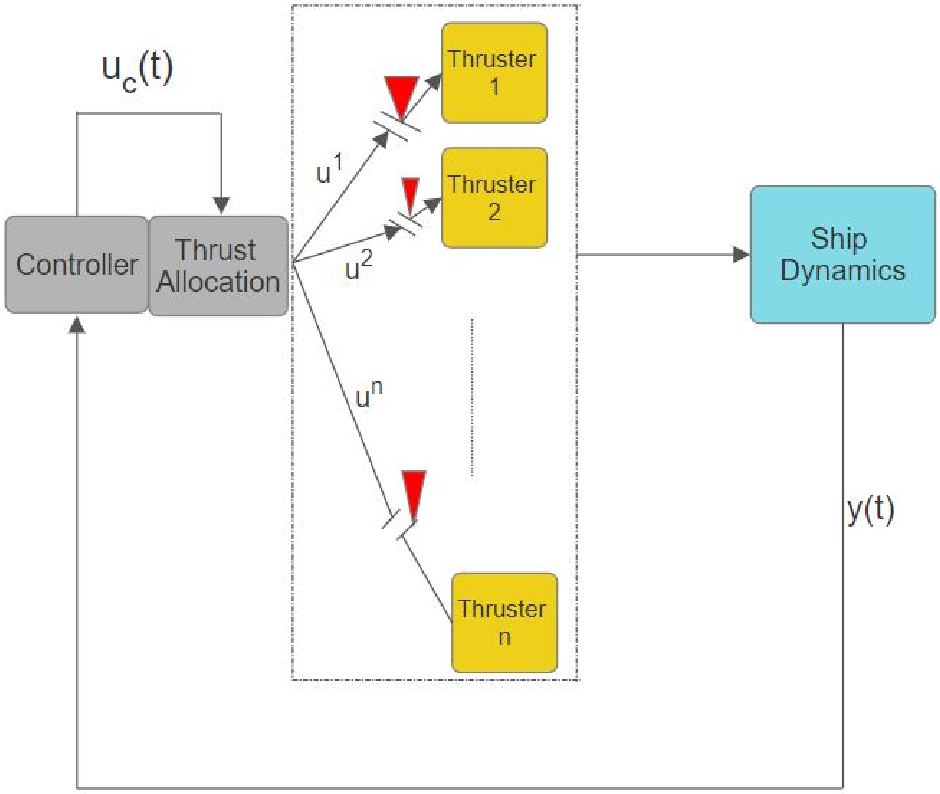
$$\dot{\boldsymbol{x}} = A\boldsymbol{x} + B\boldsymbol{u}_c + B\boldsymbol{u}_a$$

Problem 1: Designing an Unknown Input Observer

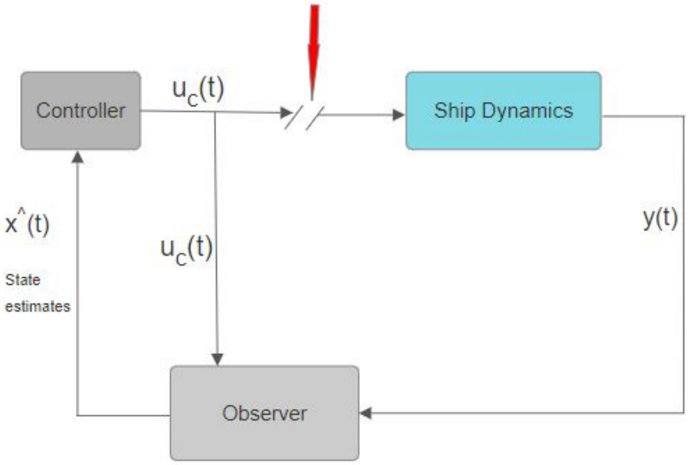
The first objective is to design an observer to estimate the true system states $\boldsymbol{x}$ using only the system outputs. Let $\hat{\boldsymbol{x}}$ represent the estimated state vector. The error between the true and estimated states is denoted as the estimation error: $\boldsymbol{e} = \boldsymbol{x} - \hat{\boldsymbol{x}}$.

To minimize the estimation error $\boldsymbol{e}$ without the need for the genuine input measurements, i.e. $\boldsymbol{u}_c + \boldsymbol{u}_a$, in a way that it converges to zero, the observer must ensure:

$$\lim_{t \to \infty} \boldsymbol{e}(t) = \boldsymbol{0}$$



(a) DP system equipped with a centralized thrust allocation architecture subject to FDI cyber-attacks on the communication channels from the controller-thrust allocation module to the thrusters. The red arrows represent FDI attacks on those channels.



(b) Equivalent problem where the FDI cyber-attacks on the communication channels from the controller-thrust allocation module to the thrusters are considered as a single bias injection vector $\mathbf{u_a}$ on the commanded control signal $\mathbf{u_c}$.

Figure 3.1: Equivalence of the problems

which can be achieved by appropriately tuning the observer gains.

Problem 2: Estimating/Reconstructing the Attack Signal $\boldsymbol{u}_a$

Having estimated the true system states $\hat{\boldsymbol{x}}$, the second objective is to reconstruct the attack signal $\boldsymbol{u}_a$. Let $\hat{\boldsymbol{u}}_a$ denote the reconstructed attack signal. The error between the actual and reconstructed attack signals is denoted as $\boldsymbol{e}_a = \boldsymbol{u}_a - \hat{\boldsymbol{u}}_a$.

Hence, the objective is to minimize the error $\boldsymbol{e}_a$ such that it approaches zero, i.e.,

$$\lim_{t \to \infty} \boldsymbol{e}_a(t) = \mathbf{0}$$

This should be accomplished by selecting appropriate parameters for reconstructing $\hat{\boldsymbol{u}}_a$ from the observed states.

Problem 3: Isolating the Corrupted Channels/Thrusters Once the true systems states have been estimated, another significant objective will be to isolate or identify the location of the FDI attacks, i.e. which channels or thrusters have been compromised by FDI attacks. This task should be accomplished without relying on the strict and limiting observer matching criterion.

Problem 4: Compensating for the Attack Signal Effect

Ultimately, once the attack signal $\hat{\boldsymbol{u}}_a$ is reconstructed, the final objective would be to compensate for its effect on the DP system. This compensation involves adjusting the commanded control input $\boldsymbol{u}_c$ to counteract the effect of the attack, making sure that the system outputs, which represent positions (denoted as X, Y, and ψ), remain close to the origin. In other words, the goal is to design the compensated control input $\boldsymbol{u}_c'$ such that the system outputs $X(t)$, $Y(t)$, and $\psi(t)$ remain close to zero despite the presence of FDI cyber-attacks. Therefore $\boldsymbol{u}_c'$ must ensure that:

$$\lim_{t \to \infty} |X(t)| \approx 0, \quad \lim_{t \to \infty} |Y(t)| \approx 0, \quad \text{and} \quad \lim_{t \to \infty} |\psi(t)| \approx 0$$

where $X(t)$, $Y(t)$, and $\psi(t)$ represent the outputs of the DP system, corresponding to the positrons and heading of the ship. They must approach zero since in DP applications the objective is to keep the vessel at the origin.

An important assumption here is that the attacker is not aware of this strategy or does not have disclosure resources to identify this additive compensation signal on the control signal.

3.3 FDI Attack Model

In this section, the effect of the FDI attack on the overall system model in the over-actuated DP system is outlined. The state-space equation of the system is outlined as follows:

$$\dot{\boldsymbol{x}} = A\boldsymbol{x} + B\boldsymbol{u}_c \tag{3.1}$$

where $\boldsymbol{x}$ represents the state vector, A is the system matrix, and B is the input matrix. The control command $\boldsymbol{u}_c$, generated by the LQR controller, is a 3-dimensional vector.

The thrust allocation problem is formulated as an optimization problem as follows:

$$\begin{cases} \text{minimize} & \boldsymbol{u}^T W \boldsymbol{u} \\ \text{subject to} & T\boldsymbol{u} = \boldsymbol{u}_c \end{cases}$$

where T is the thrust allocation matrix mapping the n-dimensional control signal vector $\boldsymbol{u}$ to the 3-dimensional control command $\boldsymbol{u}_c$, and W is the weighting matrix. The optimization is solved using the interior point method.

The vector $\boldsymbol{u}$ is defined as:

$$\boldsymbol{u} = \begin{bmatrix} u^1 \\ u^2 \\ \vdots \\ u^n \end{bmatrix} \tag{3.2}$$

In the presence of FDI attacks, which are additive signals injected into the communication lines between the controller-thrust allocation module and the thrusters, the system dynamics are affected as follows:

$$T(\boldsymbol{u} + \boldsymbol{a}) = T\boldsymbol{u} + T\boldsymbol{a} = \boldsymbol{u}_c + \boldsymbol{u}_a \tag{3.3}$$

Here, $\boldsymbol{a}$ is the n-dimensional vector of additive signals corresponding to the attack, defined as:

$$\boldsymbol{a} = \begin{bmatrix} a^1 \\ a^2 \\ \vdots \\ a^n \end{bmatrix} \tag{3.4}$$

and $\boldsymbol{u}_a$ represents the resultant additional input due to the attack.

With the presence of FDI attacks, the state equation can also be expressed as:

$$\dot{\boldsymbol{x}} = A\boldsymbol{x} + B\boldsymbol{u}_c + B\boldsymbol{u}_a = A\boldsymbol{x} + \overline{B}\boldsymbol{u} + \overline{B}\boldsymbol{a} \tag{3.5}$$

where $\overline{B} = B \times T$, modifying the influence of both $\boldsymbol{u}$ and $\boldsymbol{a}$ on the system state $\boldsymbol{x}$.

3.4 Secure State Estimation

First, the observer matching condition is examined for the following DP system with the parameters given in Section 2.1:

$$\begin{aligned}\dot{\boldsymbol{x}} &= A\boldsymbol{x} + B\boldsymbol{u}_c + B\boldsymbol{u}_a \\ \boldsymbol{y} &= C\boldsymbol{x};\end{aligned} \tag{3.6}$$

Since $rank(CB) \neq rank(B)$, the system is not $strong^*$ detectable, and a traditional linear unknown input observer cannot be designed for it.

Next, the invariant zeros of the triplet (A,B,C) is looked into by using the Rosenbrock matrix:

$$P(s) = \begin{bmatrix} sI_n - A & -B \\ C & 0 \end{bmatrix} \tag{3.7}$$

After close examination, it is observed that:

$$Rank(P(s)) = 9 \quad \forall s \in \mathbb{C} \tag{3.8}$$

In other words, the normal rank of the system is equal to the sum of the dimension of state space and the input space. Furthermore, the system does not have any invariant zeros, i.e. the matrix $P(s)$ does not lose rank for any value of s. Consequently, the system is strongly observable and according to [73], the true system states could be estimated in finite time using higher-order sliding mode observer approaches without having access to the unknown input signal.

Constructing the observer is based on the work of [80]. First step is to augment the output matrix in a way that the observer matching condition holds true. Towards that end, the relative degree of each of the original outputs with respect to the unknown input $\boldsymbol{u_a}$ is investigated. According to [81], for each output j, its relative degree μ_j with respect to the input $\boldsymbol{u_a}$ is the number of times it should be differentiated in order for the unknown input to explicitly appear. In technical terms, μ_j is determined as following:

$$C_j A^k B = 0 \quad \forall k < \mu_j - 1 \tag{3.9a}$$

$$C_j A^{\mu_j - 1} B \neq 0 \tag{3.9b}$$

Then the system 3.6, will have vector relative degree of $[\mu_1 \quad \mu_2 \quad \mu_3]$.

For y_1:

$$\begin{aligned} C_1 B &= 0 \\ C_1 A B &\neq 0 \end{aligned} \tag{3.10}$$

And for y_2:

$$\begin{aligned} C_2 B &= 0 \\ C_2 A B &\neq 0 \end{aligned} \tag{3.11}$$

And finally for y_3:

$$\begin{aligned} C_3 B &= 0 \\ C_3 A B &\neq 0 \end{aligned} \tag{3.12}$$

Therefore, $\mu_1 = \mu_2 = \mu_3 = 2$. A new output matrix C_a could be constructed in the following form:

$$C_a = \begin{bmatrix} C_1 \\ : \\ C_1A^{\lambda_1-1} \\ --- \\ C_2 \\ : \\ C_2A^{\lambda_2-1} \\ --- \\ C_3 \\ : \\ C_3A^{\lambda_3-1} \end{bmatrix} \tag{3.13}$$

such that $1 \leq \lambda_j \leq \mu_j$, and $Rank(C_aB) = Rank(B)$. Therefore, the following new output matrix:

$$C_a = \begin{bmatrix} C_1 \\ C_1A \\ C_2 \\ C_2A \\ C_3 \\ C_3A \end{bmatrix} \tag{3.14}$$

satisfies the observer matching condition, since $Rank(C_aB) = Rank(B) = 3$.

Now, a new system will be considered with the same state space dynamics and a new auxiliary output vector that contains the outputs of the original system:

$$\begin{aligned} \dot{\boldsymbol{x}} &= A\boldsymbol{x} + B\boldsymbol{u}_c + B\boldsymbol{u}_a \\ \boldsymbol{y_a} &= C_a\boldsymbol{x}; \end{aligned} \tag{3.15}$$

As stated by [81], the invariant zeros of the triplet (A, B, C_a) is the same as that of (A, B, C). Therefore, the construction of the sliding mode based unknown input observer will be based on this new system, which has additional auxiliary outputs that need to be estimated.

from 3.15:

$$\boldsymbol{y}_{ai} = \begin{bmatrix} y_{ai,1} \\ y_{ai,2} \end{bmatrix} = C_{ai}\boldsymbol{x} = \begin{bmatrix} C_i\boldsymbol{x} \\ C_iA\boldsymbol{x} \end{bmatrix} \quad i \in \{1, 2, 3\} \tag{3.16}$$

Differentiating 3.16 with respect to time, i.e. $\dot{\boldsymbol{y}}_{ai} = C_{ai}\dot{\boldsymbol{x}}$, yields:

$$\dot{\boldsymbol{y}}_{ai} = \begin{bmatrix} C_iA \\ C_iA^2 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} C_iB \\ C_iAB \end{bmatrix} \boldsymbol{u_c} + \begin{bmatrix} C_iB \\ C_iAB \end{bmatrix} \boldsymbol{u_a} \tag{3.17}$$

Given that $C_iB = 0$, it follows that:

$$\dot{\boldsymbol{y}}_{ai} = \begin{bmatrix} y_{ai,1} \\ C_iA^2\boldsymbol{x} \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 0 \\ C_iAB \end{bmatrix} \boldsymbol{u_c} + \begin{bmatrix} 0 \\ C_iAB \end{bmatrix} \boldsymbol{u_a} \tag{3.18}$$

These equations for the new output y_{ai} could themselves be written in a state-space format. Rewriting 3.18 in state-space format yields:

$$\begin{aligned} \dot{\boldsymbol{y}}_{ai} &= \Lambda_i \boldsymbol{y}_{ai} + e_i f_i(\boldsymbol{x}, \boldsymbol{u_a}) + B_i' \boldsymbol{u_c} \\ y_{i1} &= \bar{C} \boldsymbol{y}_{ai} \end{aligned} \tag{3.19}$$

where $\Lambda_i = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $e_i = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $f_i(\boldsymbol{x}, \boldsymbol{u_a}) = C_iA(A\boldsymbol{x} + B\boldsymbol{u_a})$, $B_i' = C_{ai}B$. Furthermore, y_{i1} is supposed to be the output of the state-space description of each y_{ai} and $\bar{C} = [1 \quad 0]$.

Clearly, this state-space description has two variables, i.e. $y_{ai,1}, y_{ai,2}$. Having constructed this state-space description, another variable is introduced, denoted $y_{ai,3}$. To accommodate for this variable, it is defined as $y_{ai,3} = f_i(\boldsymbol{x}, \boldsymbol{u_a})$. Consequently, the new state-space representation is constructed as follows:

$$\begin{aligned}
\dot{y}_{ai,1} &= y_{ai,2} \\
\dot{y}_{ai,2} &= y_{ai,3} + C_i AB\boldsymbol{u_c} \\
\dot{y}_{ai,3} &= \dot{f}_i \\
y_{i1} &= y_{ai,1}
\end{aligned} \tag{3.20}$$

For the above system, the following high-order sliding mode observer is constructed based on the work in [80]:

$$\begin{aligned}
\dot{\hat{y}}_{ai,1} &= \hat{y}_{ai,2} - w_{i,1} \\
\dot{\hat{y}}_{ai,2} &= \hat{y}_{ai,3} - w_{i,2} + C_i AB\boldsymbol{u_c} \\
\dot{\hat{y}}_{ai,3} &= -w_{i,3}
\end{aligned} \tag{3.21}$$

where

$$\begin{aligned}
w_{i,0} &= \hat{y}_{ai,1} - y_{i1} \\
w_{i,j} &= \lambda_{i,j} |w_{i,j-1}|^{\frac{\lambda_i - j+1}{\lambda_i - j+2}} sign(w_{i,j-1})
\end{aligned} \tag{3.22}$$

and $i \in \{1,2,3\}$ and $j \in \{1,2,3\}$. Moreover, the observer gains $\lambda_{i,j}$ are positive scalars that need to be tuned for this sliding mode observer.

From Equations (3.21) and (3.22), the error dynamics of the observer is obtained:

$$\begin{aligned}
\dot{e}_{ai,1} &= e_{ai,2} - w_{i,1} \\
\dot{e}_{ai,2} &= e_{ai,3} - w_{i,2} \\
\dot{e}_{ai,3} &= -\dot{f}_i - w_{i,3}
\end{aligned} \tag{3.23}$$

According to [80], by carefully tuning the observer gains, a sliding mode emerges on the manifold $e_{ai,1} = e_{ai,2} = e_{ai,3} = 0$ in finite time. In other words, by appropriately choosing the observer gains, the observer in 3.21, estimates the exact values of $y_{ai,j}$ and we have: $\hat{y}_{ai,j} - y_{ai,j} = 0 \quad \exists T \in \mathbb{R}^+ : T > 0$.

In the case of the DP system, the system states are $[X \quad Y \quad \psi \quad \nu_x \quad \nu_y \quad r]^T$, and the outputs are the first three states to be measured, i.e. X, Y, ψ. The C matrix would be $[I_3 \quad 0_{3\times3}]$. Hence, from 3.14 and 3.16, the following deductions are established:

$$\begin{cases} y_{a1,1} = X \\ y_{a1,2} = \dot{X} = \nu_x \\ y_{a2,1} = Y \\ y_{a2,2} = \dot{Y} = \nu_y \\ y_{a3,1} = \psi \\ y_{a3,2} = r \end{cases}$$

Therefore, the correlation between the estimates of y_{ai} and estimates of $\boldsymbol{x}$ is:

$$\begin{cases} \hat{y}_{a1,1} = \hat{x}_1 \\ \hat{y}_{a1,2} = \hat{x}_4 \\ \hat{y}_{a2,1} = \hat{x}_2 \\ \hat{y}_{a2,2} = \hat{x}_5 \\ \hat{y}_{a3,1} = \hat{x}_3 \\ \hat{y}_{a3,2} = \hat{x}_6 \end{cases}$$

Consequently, the estimates of y_{ai} could readily be turned into the estimates of $\boldsymbol{x}$.

Having estimated the true system states, the next step is estimating the unknown/attack signal $\boldsymbol{u}_a$.

3.5 Attack Signal Estimation

Given Equations (3.20), the following expression for $y_{ai,3}$ can be written:

$$y_{ai,3} = \dot{y}_{ai,2} - C_i A B \boldsymbol{u}_{\boldsymbol{c}} \tag{3.24}$$

and since $y_{ai,3} = f_i = C_i A(A\boldsymbol{x} + B\boldsymbol{u}_{\boldsymbol{a}})$, therefore the following equation is obtained:

$$C_i A(A\boldsymbol{x} + B\boldsymbol{u}_{\boldsymbol{a}}) = \dot{y}_{ai,2} - C_i A B \boldsymbol{u}_{\boldsymbol{c}} \tag{3.25}$$

Expanding 3.25, will yield:

$$C_i A^2 \boldsymbol{x} + C_i A B \boldsymbol{u}_a = \dot{y}_{ai,2} - C_i A B \boldsymbol{u}_{\boldsymbol{c}} \tag{3.26}$$

Rearranging the above equation, will result in the following equation:

$$C_i A B \boldsymbol{u}_{\boldsymbol{a}} = \dot{y}_{ai,2} - C_i A(A\boldsymbol{x} + B\boldsymbol{u}_{\boldsymbol{c}}) \tag{3.27}$$

Consequently, writing 3.27 for all $i \in \{1, 2, 3\}$ results in the following:

$$\begin{aligned} C_1 A B \boldsymbol{u}_{\boldsymbol{a}} &= \dot{y}_{a1,2} - C_1 A(A\boldsymbol{x} + B\boldsymbol{u}_{\boldsymbol{c}}) \\ C_2 A B \boldsymbol{u}_{\boldsymbol{a}} &= \dot{y}_{a2,2} - C_2 A(A\boldsymbol{x} + B\boldsymbol{u}_{\boldsymbol{c}}) \\ C_3 A B \boldsymbol{u}_{\boldsymbol{a}} &= \dot{y}_{a3,2} - C_3 A(A\boldsymbol{x} + B\boldsymbol{u}_{\boldsymbol{c}}) \end{aligned} \tag{3.28}$$

The above equations could be represented in a matrix format as follows:

$$C' B \boldsymbol{u}_a = \bar{\boldsymbol{y}}_{\boldsymbol{a2}} - C'(A\boldsymbol{x} + B\boldsymbol{u}_{\boldsymbol{c}}) \tag{3.29}$$

where $C' = \begin{bmatrix} C_1A \\ C_2A \\ C_3A \end{bmatrix}$, and $\bar{\boldsymbol{y}}_{\boldsymbol{a2}} = \begin{bmatrix} \dot{y}_{a1,2} \\ \dot{y}_{a2,2} \\ \dot{y}_{a3,2} \end{bmatrix}$. However, it is not a good practice to use the derivative of outputs for input reconstruction. Therefore, in 3.29, instead of $\dot{y}_{ai,2}$, its estimate will be used, and needless to say that $\boldsymbol{x}$ will be replaced by $\hat{\boldsymbol{x}}$.

Once the estimates of y_{ai} are obtained by using 3.21 and 3.22, the estimates of their derivative could also be achieved as follows:

$$\begin{aligned} \hat{\dot{y}}_{ai,1} &= \hat{y}_{ai,2} \\ \hat{\dot{y}}_{ai,2} &= \hat{y}_{ai,3} + C_iAB\boldsymbol{u_c} \end{aligned} \tag{3.30}$$

It follows that $\hat{\bar{\boldsymbol{y}}}_{a2} = \begin{bmatrix} \hat{y}_{a1,3} + C_1AB\boldsymbol{u_c} \\ \hat{y}_{a2,3} + C_2AB\boldsymbol{u_c} \\ \hat{y}_{a3,3} + C_3AB\boldsymbol{u_c} \end{bmatrix}$. By defining a new matrix $G = C'B$, and considering 3.29, it follows that:

$$G\hat{\boldsymbol{u}}_a = \hat{\bar{\boldsymbol{y}}}_{a2} - C'(A\hat{\boldsymbol{x}} + B\boldsymbol{u_c}) \tag{3.31}$$

Paying close attention to the matrix $G = \begin{bmatrix} C_1AB \\ C_2AB \\ C_3AB \end{bmatrix} \in \mathbb{R}^{3\times3}$ reveals that $rank(G) = rank(C_aB) = rank(\begin{bmatrix} C_1B \\ C_1AB \\ C_2B \\ C_2AB \\ C_3B \\ C_3AB \end{bmatrix}$)=3, since $C_iB = 0$. Therefore, matrix G is invertible.

Consequently, the estimate of the attack signal $\hat{\boldsymbol{u}}_a$ is obtained as follows:

$$\hat{\boldsymbol{u}}_a = (G^T G)^{-1} G^T (\hat{\bar{\boldsymbol{y}}}_{a2} - C'(A\hat{\boldsymbol{x}} + B\boldsymbol{u}_c)) \tag{3.32}$$

To sum up, by appropriately choosing the values for $\lambda_{i,j}$, not only are the true system states estimated, but also the attack signal $\boldsymbol{u}_a$ is reconstructed.

3.6 Attack Isolation

In this section, the aim is to identify which links from the controller to the thrusters have been compromised by FDI attacks. To isolate FDI attacks, a bank of Unknown Input Observers (UIOs) is constructed. Each UIO is designed to be sensitive to a broad set of attack scenarios, while being specifically insensitive to FDI attacks affecting its associated thruster. In a system where only a single communication channel may be compromised at any one time by an adversary, n UIOs are configured at the command and control side, matching the number of thrusters in the DP system. Each observer is tailored in a way in which a UIO linked to a particular thruster is designed to be robust to FDI attacks on that thruster's channel while being affected to FDI attacks on other channels for the remaining thrusters. This approach enables precise localization of FDI attacks affecting the system.

3.6.1 Isolation of Single Attacks

The design procedure for the family of UIOs first addresses scenarios associated with single attacks, where only one communication channel may be corrupted at any one time interval by an adversary, outlining the criteria for sensitivity and insensitivity. Next, the explanation of the isolation logic that underlies the isolation process is provided. Followed by that, the case of isolation of simultaneous FDI attacks is addressed.

As discussed in chapter 2, in order to design UIOs in the form of 2.15, the observer matching condition needs to hold. meaning, $rank(CB) = rank(B)$. As seen earlier, this condition does not hold with the current C matrix, as is the case with the majority of over-actuated systems. However, in section 3.4, it was observed that since the system is strongly observable, the output matrix could be augmented using a sliding mode observer, and the additional output variables could be constructed in finite time and this could be achieved fast by choosing appropriate gains for the sliding mode observer. Hence, the state and output equations could be rewritten as following:

$$\begin{aligned} \dot{\boldsymbol{x}} &= A\boldsymbol{x} + B\boldsymbol{u}_c + B\boldsymbol{u}_a \\ \boldsymbol{y_a} &= C_a\boldsymbol{x}; \end{aligned} \tag{3.33}$$

Having n thrusters and n communication channels, the design procedure for the ith UIO associated with the ith thruster goes as follows: First, the attack vector $\mathbf{a}$ is considered as the general form of the FDI attack and its elements contain the additive signals affecting each communication line. It is of the same dimension as the $\mathbf{u}$ vector, meaning $n = 6$. The state equation is written in the following form:

$$\begin{aligned} \dot{\boldsymbol{x}} &= A\boldsymbol{x} + \bar{B}^i\boldsymbol{u}^i + \bar{B}^i\boldsymbol{a}^i + \bar{b}_i(u_i + a_i) \\ \boldsymbol{y_a} &= C_a\boldsymbol{x}; \end{aligned} \tag{3.34}$$

where $\bar{B}^i$ is obtained by deleting the ith column of the $\bar{B}$ matrix, and $\bar{b}_i$ is the ith column of the $\bar{B}$ matrix. $\boldsymbol{u}^i$ is the vector consisting of healthy thruster commands excluding the ith element, and $\boldsymbol{a}^i$ is the vector consisting of additive attack signals on the individual communication channels, excluding the i the attack signal on the ith channel. u_i and a_i are the command signal and the additive attack signal associated with the ith thruster. Associated with this thruster, the following UIO is designed:

$$\begin{aligned} \dot{\boldsymbol{z}} &= N^i\boldsymbol{z} + G^i\boldsymbol{u}^i + L^i\boldsymbol{y_a} \\ \hat{\boldsymbol{x}} &= \boldsymbol{z} - H^i\boldsymbol{y_a}; \end{aligned} \tag{3.35}$$

The error signal for the ith UIO, which is defined as $\boldsymbol{e}^{(i)} = \boldsymbol{x} - \hat{\boldsymbol{x}}$, can be rewritten as: $\boldsymbol{e}^{(i)} = \boldsymbol{x} - \boldsymbol{z} + H^i\boldsymbol{y}_a$. Therefore, the error dynamics becomes: $\dot{\boldsymbol{e}}^{(i)} = \dot{\boldsymbol{x}} - \dot{\boldsymbol{z}} + H\dot{\boldsymbol{y}}_a$. Hence, using Equations (3.35) and (3.34), the error dynamics of the ith UIO associated with ith channel becomes as follows:

$$\dot{\boldsymbol{e}}^i = N^i\boldsymbol{e^i} + (P^iA - K^iC_a - N^i)\boldsymbol{x} + (P^i\bar{B}^i - G^i)\boldsymbol{u}^i + P^i\bar{B}^i\boldsymbol{a}^i + P^i\bar{b}_i(u_i + a_i) \tag{3.36}$$

where $P^i = I + H^iC_a$ and $K^i = L^i + N^iH^i$. Also, the following relationship holds between K^i and N^i, where $N^i = P^iA - K^iC_a$. From Equation (3.36) it is implied that in order for this UIO to be sensitive to FDI attacks on the other channels besides the ith channel, and be insensitive to FDI attacks on its own associated channel, the following conditions must hold true:

$$\begin{cases} N^i \text{ be Hurwitz,} & (1) \\ P^iA - K^iC_a - N^i = 0, & (2) \\ P^i\bar{B}^i = G^i, & (3) \\ P^i\bar{b}_i = 0, & (4) \\ P^i\bar{B}^i \neq 0, & (5) \end{cases}$$

Therefore, the design procedure for the ith UIO is as follows:

- First, H^i is obtained by solving the 4th condition: $P^i\bar{b}_i = 0 \rightarrow (I + H^i)\bar{b}_i = 0 \rightarrow H^i = -\bar{b}_i(C_a\bar{b}_i)^+$, where $(W)^+$ is the pseudo-inverse of matrix W and W is an arbitrary matrix.

- Next, K^i is obtained in a way to make matrix N^i Hurwitz. This could be achieved using pole placement.

- N^i is solved for according to the 2nd condition: $N^i = P^iA - K^iC_a$.
- Once K^i and N^i are calculated, L^i is solved for based on the definition $K^i = L^i + N^iH^i$ which was mentioned earlier. Therefore: $L^i = K^i - N^iH^i$.

Subsequently, the residual of the ith UIO will be $\boldsymbol{r}^i = \boldsymbol{y}_a - C_a\hat{\boldsymbol{x}} = C_a\boldsymbol{e}^i$. Therefor, the isolation module at the command and control side checks the residuals of all n UIOs. If the residual of the UIO associated with a specific thruster does not surpass its associated threshold value (TV) while the residuals of the UIOs associated with the other thrusters do surpass their threshold values, it can be deduced that there is an attack on that specific thruster and no attack on the others. This process is about isolating the affected thruster rather than merely detecting the presence of an attack. In other words:

$$\text{Isolation Rule:}\quad \begin{cases} \text{Attack on the } i^{th} \text{ channel} & \text{if } \|r^i\| < \text{TV and } \|r^j\| > \text{TV } \forall j \neq i, \text{ where } i \in \{1,2,\ldots,n\}, \\ \text{No attack} & \text{if } \|r^i\| < \text{TV for all } i \text{ and } i \in \{1,2,\ldots,n\}, \\ \text{Attack, location undetermined} & \text{if } \|r^i\| > \text{TV for all } i \text{ and } i \in \{1,2,\ldots,n\} \end{cases}$$

where the threshold TV is obtained by conducting multiple simulation scenarios considering noises for each UIO. The threshold is taken to be the minimum value which the residuals will not surpass when there are no attacks with different noise levels.

It must be mentioned that in the isolation rule, the first two scenarios are the most informative, identifying exactly the corrupted thrusters, while the last scenario only signals of an attack, not determining which channel or thruster has been compromised. Despite tuning the UIOs to minimize sensitivity to additive attack signals on their respective channels and maximize sensitivity to attacks on other channels, each UIO remains vulnerable to attacks on its designated channel. This vulnerability arises because Equation (3.36) presumes direct access to the signal $\boldsymbol{y}_a$ which in practice it is estimated under the influence of the very unknown inputs—the FDI attacks—that the system aims to isolate. Consequently, during the initial transient response following an attack, residuals across all channels may exceed their predefined thresholds, compromising the observer's ability to maintain accurate detection.

3.6.2 Isolation of Simultaneous Attacks

To effectively isolate two simultaneous FDI attacks affecting two communication channels at the same time, within a DP system, another bank of UIOs is needed. Within this bank of UIOs and each UIO each tuned is to a pair of thrusters. The number of such pairs, given n thrusters, is equal to the number of pairs that could be chosen from n channels or thrusters, i.e. $\binom{n}{2} = \frac{n!}{2!(n-2)!}$. Each UIO must satisfy the following conditions:

- Insensitive to FDI attacks on its associated two thrusters.

- Sensitive to FDI attacks on all the other two thrusters.

The design procedure for these UIOs for isolating simultaneous FDI attacks, is quite similar to that of UIOs for single attacks, except for the fact that here the state space is written as:

$$\dot{\boldsymbol{x}} = A\boldsymbol{x} + \bar{B}^{ij}\boldsymbol{u}^{ij} + \bar{B}^{ij}\boldsymbol{a}^{ij} + \bar{b}_{ij}(\boldsymbol{u}_{ij} + \boldsymbol{a}_{ij}) \tag{3.37}$$

where $\bar{B}^{ij}$ is obtained by deleting the ith and jth column of the $\bar{B}$ matrix, and $\bar{b}_{ij}$ is the matrix consisting of the ith and jth column of the $\bar{B}$ matrix. $\boldsymbol{u}^{ij}$ is the vector consisting of healthy thruster commands excluding the ith and jth element, and $\boldsymbol{a}^{ij}$ is the vector consisting of additive attack signals excluding those on the ith and jth communication channels. $\boldsymbol{u}_{ij}$ and $\boldsymbol{a}_{ij}$ are the command signal and the additive attack signal associated with the ith and jth thrusters. For this pair, the following UIO is designed:

$$\begin{aligned} \dot{\boldsymbol{z}} &= N^{ij}\boldsymbol{z} + G^{ij}\boldsymbol{u}^{ij} + L^{ij}\boldsymbol{y_a} \\ \hat{\boldsymbol{x}} &= \boldsymbol{z} - H^{ij}\boldsymbol{y_a}; \end{aligned} \tag{3.38}$$

The error signal for the ijth Unknown Input Observer (UIO), defined as $\boldsymbol{e}^{ij} = \boldsymbol{x} - \hat{\boldsymbol{x}}$, can be expressed as $\boldsymbol{e}^{ij} = \boldsymbol{x} - \boldsymbol{z} + H^{ij}\boldsymbol{y}_a$. Consequently, the error dynamics become $\dot{\boldsymbol{e}}^{ij} = \dot{\boldsymbol{x}} - \dot{\boldsymbol{z}} + H^{ij}\dot{\boldsymbol{y}}_a$. Using Equations (3.37) and (3.38), the dynamics of the ijth UIO are given by:

$$\dot{\boldsymbol{e}}^{ij} = N^{ij}\boldsymbol{e}^{ij} + (P^{ij}A - K^{ij}C_a - N^{ij})\boldsymbol{x} + (P^{ij}\bar{B}^{ij} - G^{ij})\boldsymbol{u}^{ij} + P^{ij}\bar{B}^{ij}\boldsymbol{a}^{ij} + P^{ij}\bar{b}_{ij}(u_{ij} + a_{ij}) \tag{3.39}$$

where P^{ij} is defined as $I + H^{ij}C_a$ and K^{ij} is given by $L^{ij} + N^{ij}H^{ij}$. The matrix relationship $N^{ij} = P^{ij}A - K^{ij}C_a$ is also established. To ensure that the ijth UIO is sensitive to FDI attacks on channels other than channels i and j while remaining insensitive to attacks on its own associated channels, the following conditions must be satisfied:

$$\begin{cases} N^{ij} \text{ must be Hurwitz,} & (1) \\ P^{ij}A - K^{ij}C_a - N^{ij} = 0, & (2) \\ P^{ij}\bar{B}^{ij} = G^{ij}, & (3) \\ P^{ij}\bar{b}_{ij} = 0, & (4) \\ P^{ij}\bar{B}^{ij} \neq 0, & (5) \end{cases} \tag{3.40}$$

The design procedure for the ijth UIO is as follows:

- First, H^{ij} is determined by solving the 4th condition: $P^{ij}\bar{b}_{ij} = 0 \rightarrow (I + H^{ij})\bar{b}_{ij} = 0 \rightarrow H^{ij} = -\bar{b}_{ij}(C_a\bar{b}_{ij})^{+}$, where $(W)^{+}$ represents the pseudo-inverse of matrix W.

- Next, K^{ij} is derived to ensure that matrix N^{ij} is Hurwitz.

- N^{ij} is solved for based on the 2nd condition: $N^{ij} = P^{ij}A - K^{ij}C_a$.

- With K^{ij} and N^{ij} computed, L^{ij} is obtained using the definition $K^{ij} = L^{ij} + N^{ij}H^{ij}$. Consequently, $L^{ij} = K^{ij} - N^{ij}H^{ij}$.

The residual of the ijth UIO is $\boldsymbol{r}^{ij} = \boldsymbol{y}_a - C_a\hat{\boldsymbol{x}} = C_a\boldsymbol{e}^{ij}$. The isolation module at the command and control side evaluates the residuals from all $\binom{n}{2}$ UIOs. If the residual of the UIO associated with channels i and j does not exceed its threshold value while the residuals of the UIOs for other pair of channels exceed their thresholds, it can be concluded that channels i and j are compromised by FDI attacks. The isolation logic for 2 simultaneous FDI attacks corrupting two channels is as follows:

$$\text{Isolation Rule:}\quad \begin{cases} \text{Attack on channels } i \text{ and } j & \text{if } \|r^{ij}\| < \text{TV and } \|r^{kl}\| > \text{TV } \forall (kl) \neq ij,\ i, j, l, k \in \{1, 2, \ldots, n\}, \\ \text{No attack} & \text{if } \|r^{ij}\| < \text{TV for all } ij \text{ and } ij \in \{1, 2, \ldots, n\}, \\ \text{Attack, location undetermined} & \text{if } \|r^{ij}\| > \text{TV for all } ij \text{ and } ij \in \{1, 2, \ldots, n\} \end{cases}$$

The threshold for each UIO is determined in a way similar to calculating thresholds for isolating single attacks.

Similar to the single attacks case, for UIOs associated with paired thrusters, despite tuning to reduce sensitivity to attacks on their respective channels, each UIO can still be affected by attacks on its respected channels. Since Equation (3.39) assumes $\boldsymbol{y}_a$ is known but is actually estimated under FDI attacks, transient behavior following an attack may cause all residuals to exceed their thresholds.

In the case of the aforementioned DP system, matrix B has full column rank, which is the same number as the degrees of freedom of the vessel, i.e 3. The matrix $\bar{B}$ which is the multiplication of the input matrix B and the thruster configuration matrix, will have the same column rank. According to [58], the limit to the isolation capacity of the family of UIOs is less than the column rank of matrix $\bar{B}$. Therefore, the bank of UIOs designed here, can isolate up to two simultaneous FDI attacks on the communication channels.

3.7 Cyber-Attack Estimation on Individual Channels

Previously, the overall attack signal $\mathbf{u}_a$ was estimated. Once the affected communication channels are isolated, the exact additive signals on those channels can be determined by solving the following system of equations:

$$T_{ij}\mathbf{u}_a^{ij} = \mathbf{u}_a \tag{3.41}$$

Here, T_{ij} is obtained by concatenating the columns of T associated with the indexes of the affected thrusters, and $\mathbf{u}_a^{ij}$ is the column vector consisting of the additive signals to each affected channel. Given that $\mathbf{u}_a$ is generally 3-dimensional and $\mathbf{u}_a^{ij}$ is 2-dimensional, this problem is over-determined and should be solved using the least squares method.

The least squares problem can be written as:

$$\min_{\mathbf{u}_a^{ij}} \left\| T_{ij}\mathbf{u}_a^{ij} - \mathbf{u}_a \right\|_2^2 \tag{3.42}$$

The solution to this problem is given by:

$$\mathbf{u}_a^{ij} = \left(T_{ij}^T T_{ij}\right)^{-1} T_{ij}^T \mathbf{u}_a \tag{3.43}$$

Estimating these attack signals in addition to the overall attack signal $\mathbf{u}_a$ is necessary for overall awareness of the DP system.

Since the isolation capacity of the system is two simultaneous attacks maximum, $\mathbf{u}_a^{ij}$ is 2-dimensional. Let's first describe the case for two-dimensional $\mathbf{u}_a^{ij}$:

The least squares problem for two-dimensional $\mathbf{u}_a^{ij}$ is:

$$\min_{\mathbf{u}_a^{ij}} \left\| T_{ij} \mathbf{u}_a^{ij} - \mathbf{u}_a \right\|_2^2 \tag{3.44}$$

Given $T_{ij} \in \mathbb{R}^{3\times 2}$ and $\mathbf{u}_a \in \mathbb{R}^3$, we solve for $\mathbf{u}_a^{ij} \in \mathbb{R}^2$ using the least squares method:

$$\mathbf{u}_a^{ij} = \left(T_{ij}^T T_{ij}\right)^{-1} T_{ij}^T \mathbf{u}_a \tag{3.45}$$

For one-dimensional $\mathbf{u}_a^i$, where only one thruster is affected, the least squares problem simplifies to:

$$\min_{\mathbf{u}_a^{i}} \left\| T_{i} \mathbf{u}_a^{i} - \mathbf{u}_a \right\|_2^2 \tag{3.46}$$

Given $T_i \in \mathbb{R}^{3\times 1}$ and $\mathbf{u}_a \in \mathbb{R}^3$, we solve for $\mathbf{u}_a^i \in \mathbb{R}$ using the least squares method:

$$\mathbf{u}_a^{i} = \left(T_{i}^T T_{i}\right)^{-1} T_{i}^T \mathbf{u}_a \tag{3.47}$$

3.8 Cyber-Attack Compensation

As stated in Chapter 1, existing compensation solutions for FDI cyber-attacks often involve excluding the isolated thrusters and their communication channels [40, 58]. This approach has serious drawbacks, especially in scenarios where consecutive cyber-attacks occur. For instance, if two thrusters are compromised and subsequently excluded from operation, the system may still have enough redundancy to satisfy the control command. However, if other attack occurs on different thrusters shortly after, they need to be excluded as well, and then the system may not have sufficient active thrusters to realize the commanded control signal due to the excluded thrusters and reduced redundancy. Additionally, once a thruster and its communication channel are excluded, bringing them back into operation is not straightforward. It requires extensive troubleshooting of both the thruster and the communication channel, which involves diagnosing and resolving potential issues in both hardware and network communication. This process can be time-consuming. Therefore, bringing the affected thrusters back into operation immediately after the attack has stopped, is not practical.

In contrast, in this thesis a more effective approach is considered, in which the compensation is carried out by estimating the overall attack signal $\boldsymbol{u_a}$, filtering out high-frequency noise or chattering using a low-pass filter, and then feeding the negative of this filtered signal back to the controller, provided that the attacker is not aware of this strategy and cannot read the control command. This approach maintains system redundancy by keeping all thrusters operational, hence enhancing the system's resilience and ensuring the control objectives can still be satisfied despite being subjected to consecutive FDI cyber-attacks.

In this section, by using the estimated states and the reconstructed attack signals, the new control law is designed to account for the effect of the FDI cyber-attacks in the network of thrusters. Given the state-space description of the system in below:

$$\dot{\boldsymbol{x}} = A\boldsymbol{x} + B\boldsymbol{u}_c + B\boldsymbol{u}_a$$

and having estimated the attack signal $\mathbf{u}_a$ as $\hat{\mathbf{u}}_a$, the objective is to devise a compensator to attenuate its effects. Initially, the control signal $\mathbf{u}_c$ was generated using a Linear Quadratic Regulator (LQR) controller. The revised control law will incorporate the following modifications:

(1) **Observer-based Feedback Controller**: The feedback controller used the estimated state vector $\hat{\boldsymbol{x}}$ obtained through the sliding mode observer. This substitution ensures that the controller operates based on the most accurate state information, compensating for any discrepancies introduced by the attack signal.

(2) **Design of Low Pass Filter**:A low pass filter characterized by a transfer function

$F(s)$ with a cut-off frequency ω_c to attenuate high-frequency modes of the estimated attack signal $\hat{\mathbf{u}}_a$ caused by time-varying FDI cyber-attacks in the network of thrusters and the fluctuations in the estimation process. The output of the filter, denoted $\overline{\hat{\mathbf{U}}}_a(t)$, is given by:

$$\overline{\hat{\boldsymbol{U}}}_a(s) = \begin{bmatrix} F_1(s) & 0 & 0 \\ 0 & F_2(s) & 0 \\ 0 & 0 & F_3(s) \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{U}}_{a1}(s) \\ \hat{\boldsymbol{U}}_{a2}(s) \\ \hat{\boldsymbol{U}}_{a3}(s) \end{bmatrix} \tag{3.48}$$

where $F_i(s) = \frac{1}{1+\frac{s}{\omega_{ci}}}$ represents the transfer function of the low-pass filter corresponding to the i-th component of $\boldsymbol{u}_a$. This low pass filter helps prevent the high frequency modes entering the compensator to avoid wear and tear of the actuators. The resulting signal will be $\overline{\hat{\boldsymbol{u}}}_a = \mathcal{L}^{-1}\{\overline{\hat{\boldsymbol{U}}}_a(s)\}$.

(3) **Compensation Strategy**: The filtered estimate of the attack signal is then combined with the LQR controller output. The negative of the filtered estimate is added to the control signal to counteract the impact of the attack signal. This results in the final compensated control signal, denoted by $\boldsymbol{u}_c$:

$$\boldsymbol{u}_c = \boldsymbol{u}_{\text{LQR}}(\hat{\boldsymbol{x}}) - \overline{\hat{\boldsymbol{u}}}_a \tag{3.49}$$

where $\boldsymbol{u}_{\text{LQR}}(\hat{\mathbf{x}})$ represents the output of the LQR controller.

By implementing these modifications in the control law, the controller nullifies the effect of the FDI attacks and becomes more resilient against cyber-attacks or disturbances, ensuring the stability and performance of the DP system in the face of adversaries. In the following section, simulations are provided to verify the effectiveness of the mentioned estimation and compensation schemes.

3.9 Results and Simulations

3.9.1 Estimation, Isolation and Compensation

In this section the effectiveness of the developed methodologies for securely estimating the true system states in the presence of FDI cyber-attacks on the communication channels is examined through simulations in Matlab and Simulink. The linear model of a ship is used for simulations. The model is obtained based on the mathematical model of Section 2.1. LQR controller has been used as the controller, with $Q = 100000I_6, R = 0.00002I_3$. Only the positions and yaw are measured (X, Y, ψ), i.e $C = [I_3 \quad 0_{3\times3}]$. The thruster configuration matrix is:

$$T = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 4 & -4 & -35 & -33 & 40 & 28 \end{bmatrix} \tag{3.50}$$

Three FDI attacks compromise the communication channels, in the following way:

- From t=15s to 55s, the channels associated with thruster 1 and 3 are compromised simultaneously.
- From t=55 to 160s the channel associated with thruster 2 is compromised.

Mathematically, the attacks are represented based on the model given in Section 3.3.

$$a^1(t) = \begin{cases} 0 & \text{if } 0 \leq t < 15, \\ 2 \times 10^6 & \text{if } 15 \leq t < 55, \\ 0 & \text{if } t \geq 55. \end{cases}$$

$$a^2(t) = \begin{cases} 0 & \text{if } 0 \leq t < 55, \\ 2 \times 10^6 & \text{if } t \geq 55. \end{cases}$$

$$a^3(t) = \begin{cases} 0 & \text{if } 0 \leq t < 15, \\ 2 \times 10^6 & \text{if } 15 \leq t < 55, \\ 0 & \text{if } t \geq 55. \end{cases}$$

The parameters chosen for the sliding mode observer are as follows: $\lambda_{1,1} = 40$, $\lambda_{1,2} = 30$, $\lambda_{1,3} = 20$, $\lambda_{2,1} = 50$, $\lambda_{2,2} = 50$, $\lambda_{2,3} = 180$, $\lambda_{3,1} = 150$, $\lambda_{3,2} = 1.8 \times 10^6$, and $\lambda_{3,3} = 200$. As shown in Figure 3.2, the FDI cyber-attacks indicated above cause the vessel to diverge significantly from their operating point, which is the origin for DP applications. These attacks also led to considerable deviation of the yaw angle of the vessel.

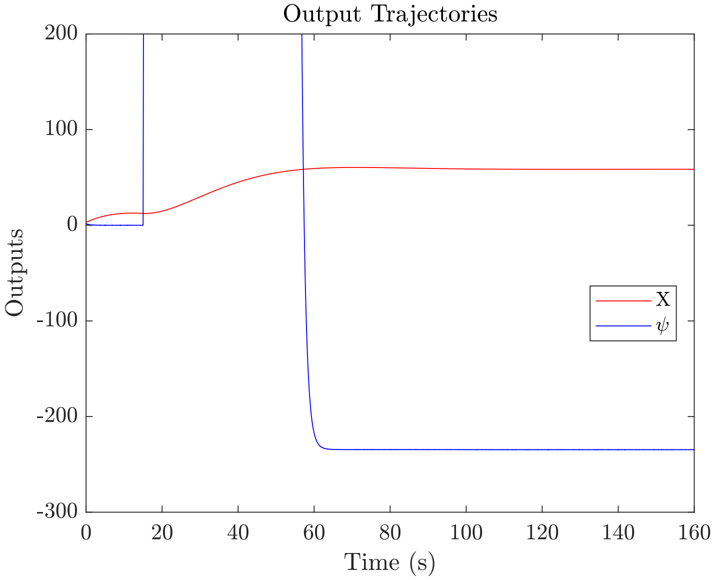


Figure 3.2: Effect of the FDI attacks on the surge movement and on the yaw.

Next, the effectiveness of the sliding mode observer is verified by showing the estimation error.

As shown in figures 3.3, 3.4, 3.5, 3.6, 3.7, and 3.8, the sliding mode observer manages

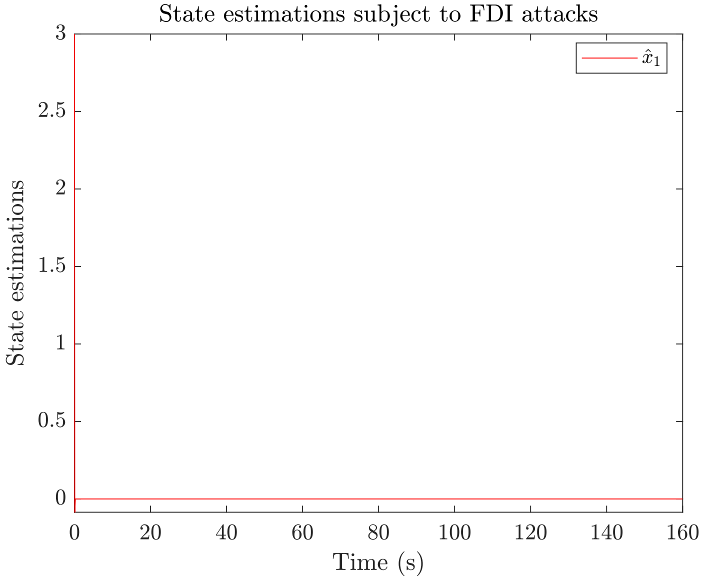


Figure 3.3: Estimation error of $\hat{x}_1$.

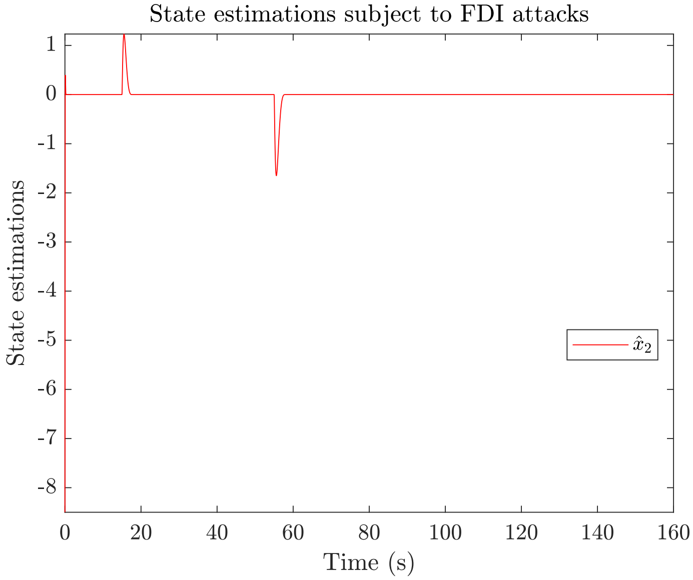


Figure 3.4: Estimation error of $\hat{x}_2$.

to successfully estimate the true system states in the presence of FDI cyber-attacks on the communication lines between the controller and the thrusters. As the estimation error remains close to origin, it means that the observer estimate the real states of the DP system by having access to only the healthy control input and system outputs.

Next, the performance of the observer is investigated with regards to the estimation of

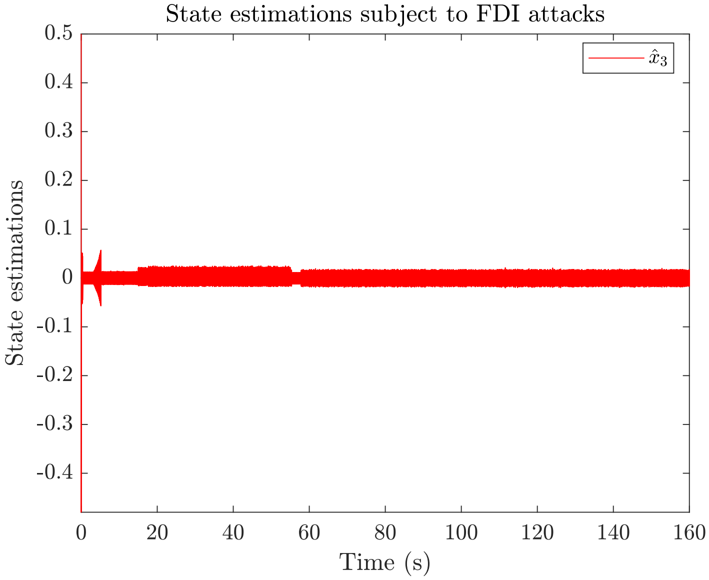


Figure 3.5: Estimation error of $\hat{x}_3$.

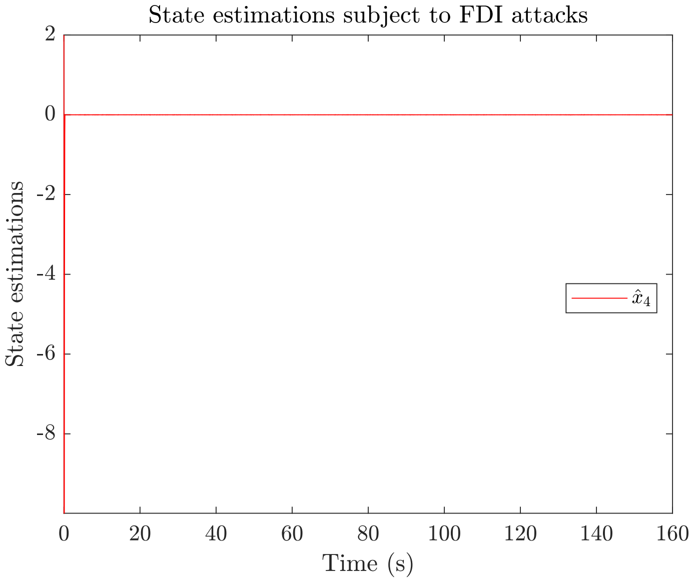


Figure 3.6: Estimation error of $\hat{x}_4$.

the unknown input $\boldsymbol{u}_a$. As stated before, the signal $\boldsymbol{u}_a$ is the net effect of any number of FDI cyber-attacks occurring on the communication channels between the controller and the thrusters, on the control command. In other words, how accurate the observer is able to estimate the additional signal that was injected onto the control signal $\boldsymbol{u_c}$.

As shown in figures 3.9a, 3.9b and 3.9c, the attack reconstruction scheme presented

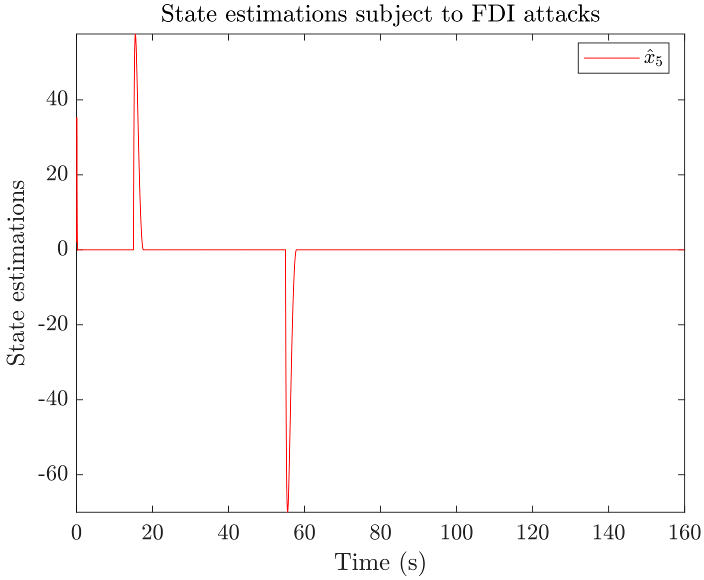


Figure 3.7: Estimation error of $\hat{x}_5$.

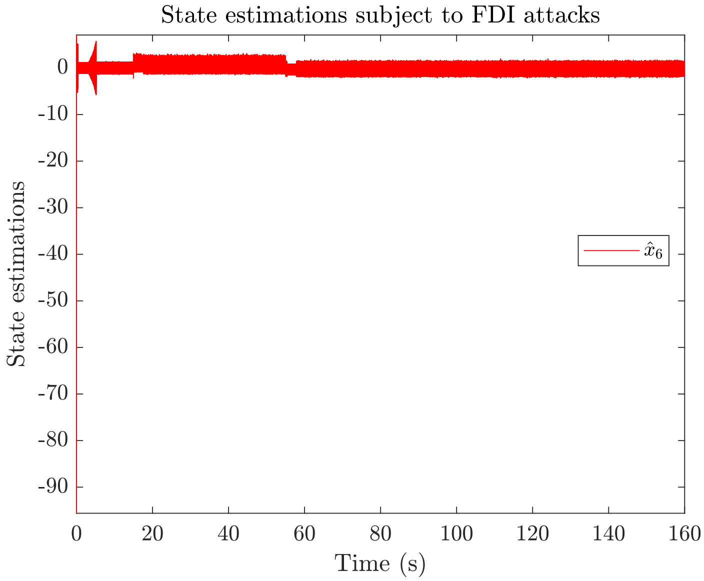
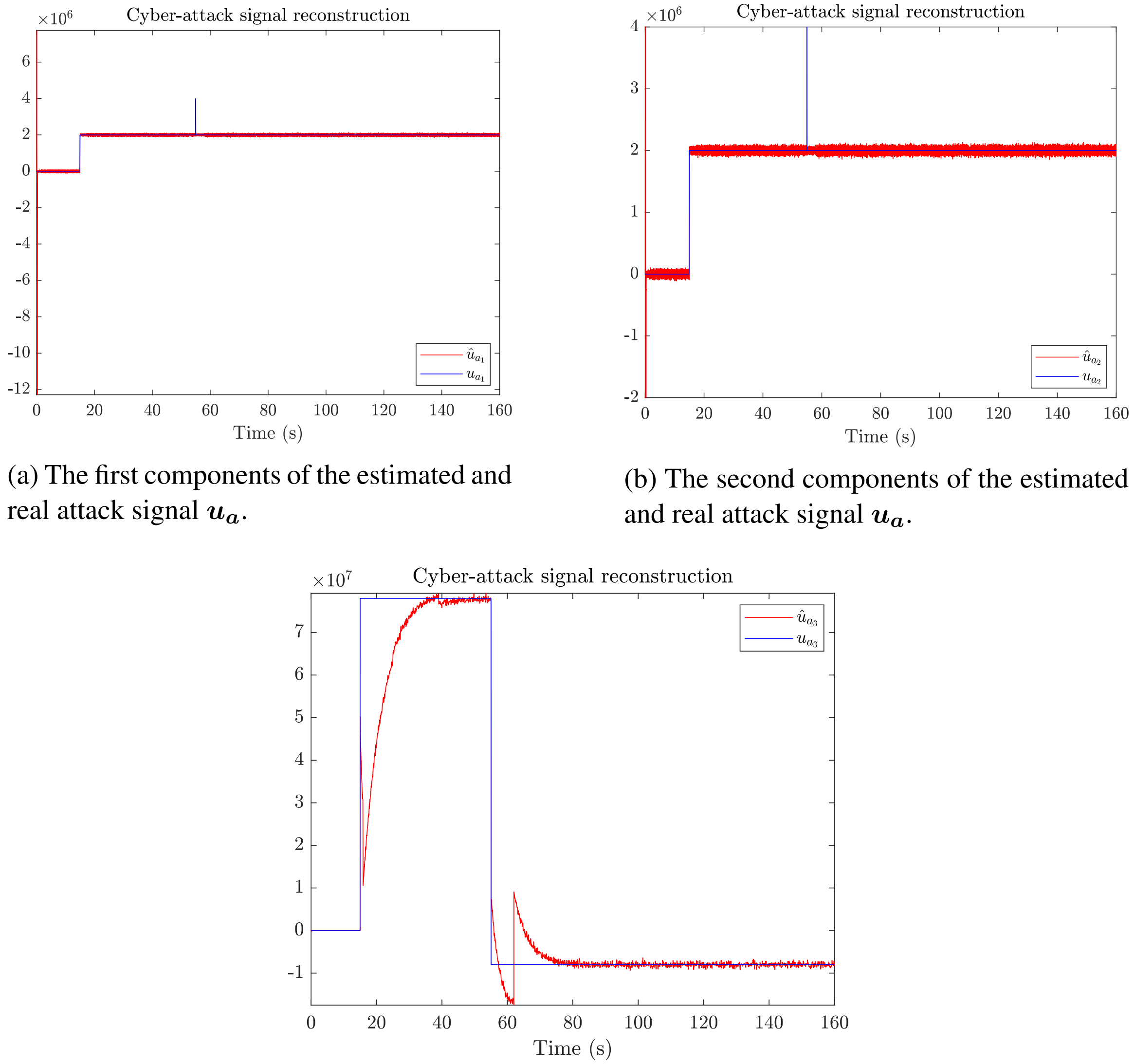


Figure 3.8: Estimation error of $\hat{x}_6$.

in Section 3.5, has managed to estimate the unknown signal $\boldsymbol{u_a}$ to an acceptable degree. The scheme, which was based on estimating the derivative of the outputs, rather than using them directly, successfully captures the sudden variations in the attack signal and remains sufficiently close to its trajectory.

Afterwards, the performance of the isolation scheme provided in Section 3.6 is tested.



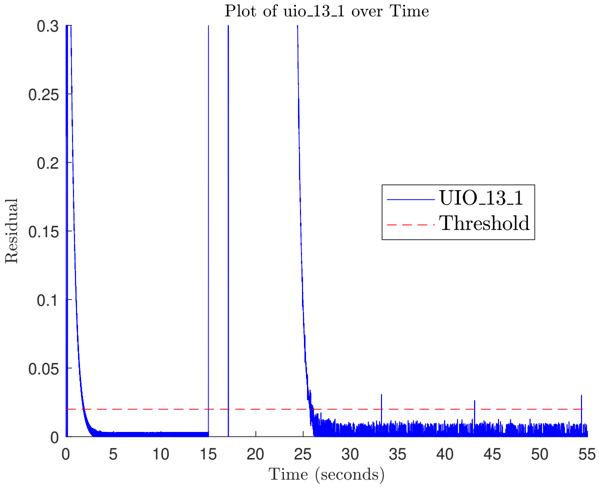
(a) The first components of the estimated and real attack signal $\boldsymbol{u_a}$.

(b) The second components of the estimated and real attack signal $\boldsymbol{u_a}$.

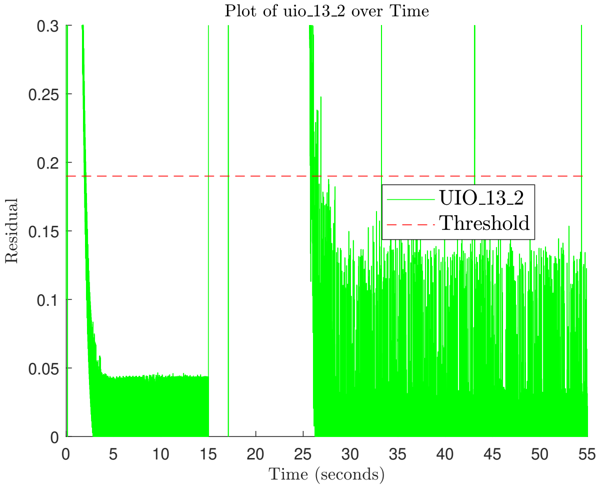
(c) The third components of the estimated and real attack signal $\boldsymbol{u_a}$.

Figure 3.9: Reconstruction of the unknown cyber-attack signal $\boldsymbol{u_a}$ using the sliding mode observer.

The system is equipped with two set of bank of UIOs. The first set, consists of UIOs each associated with a pair of thrusters. Since from t=15s to t=55s, thrusters 1 and 3 are corrupted by a^1 and a^3 simultaneously, in order to isolate their associated channels as being corrupted, the residuals in other UIOs for other pairs must surpass their thresholds, whereas the residual in the UIO associated with thruster 1 and 3 must not exceed the threshold.



(a) First component of the residual.



(b) Second component of the residual.

Figure 3.10: residual of the UIO for the pairs 1 and 3

Before presenting the figures related to the isolation, it is important to clarify the notation used in the legends. For instance, `UIO_i_k` refers to the kth component of the residual derived from the UIO associated with the ith thruster or channel. Similarly, `UIO_ij_k` denotes the kth component of the residual derived from the UIO associated with the ith and jth thrusters or channels.

As shown in figures 3.10a and 3.10b, the residuals exceed the threshold at t=15s, and then return below the threshold, and occasionally exceed the threshold afterwards. When performing the isolation logic for the possible two simultaneous attacks in Section 3.6, since the residuals do not stay above the threshold after exceeding it, it is interpreted as not having gone beyond the threshold. Next, the other UIOs must be looked at in order to reach a conclusion in terms of isolation of the pair of corrupted channels.

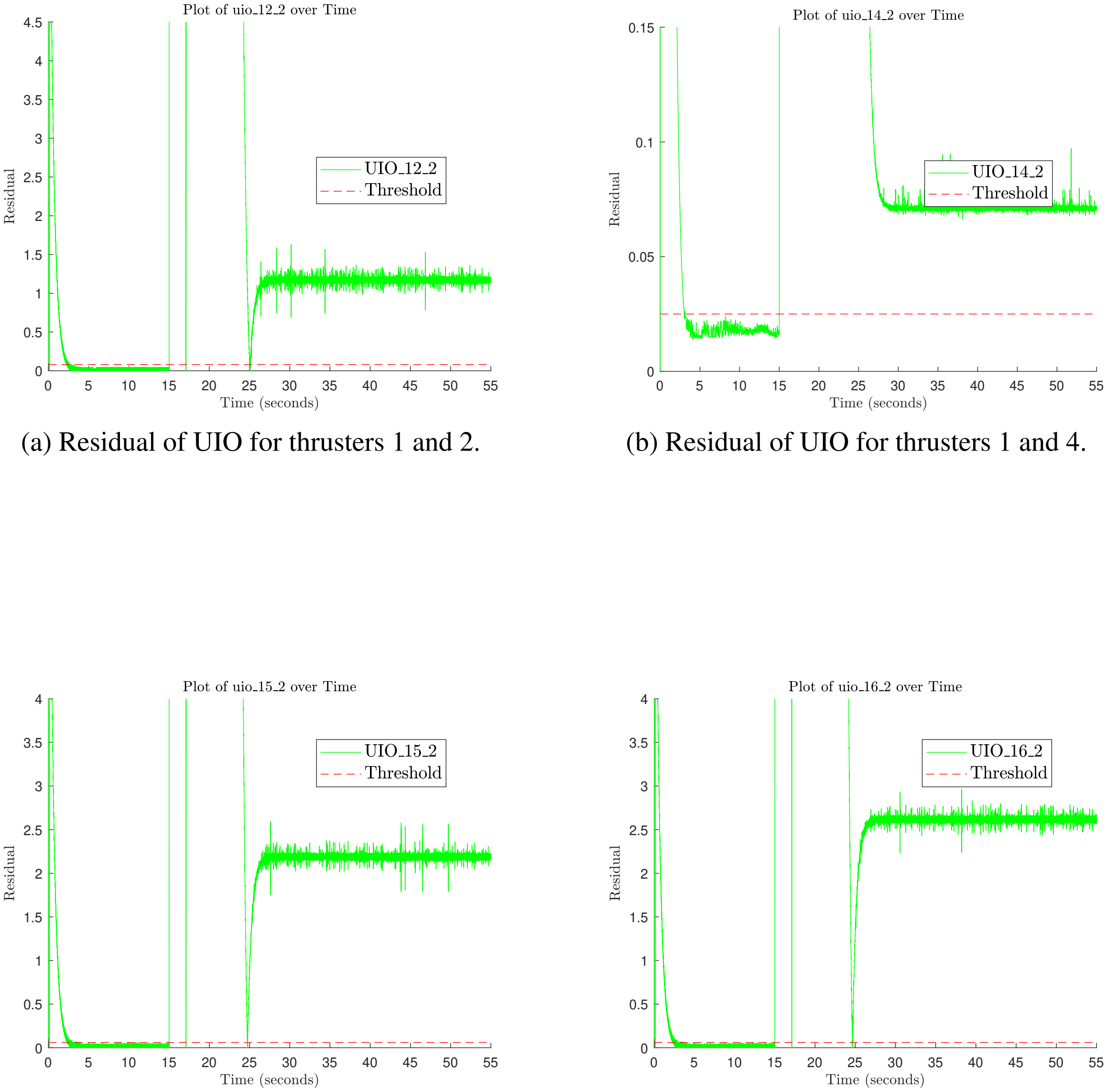
As shown in figures 3.11a, 3.11b, 3.11c, and 3.11d, the residuals surpass the threshold at t=15. The residuals plummet quickly below threshold at t=25 and rise back above the threshold instantly, which could be attributed to the transient behaviour of the sliding mode observer after the injection of the attack, and since it happens for a very small period, it could be ignored. In other words, in terms of interpreting these residuals, it can be concluded that the residuals surpass the threshold in these UIOs.

The same pattern is also apparent in figures 3.12a, 3.12b, 3.12c, and 3.12d which represent the residuals associated with UIOs for the following pairs: thruster 2 and 3, thruster 2 and 4, thruster 2 and 5, and thruster 2 and 6. Therefore, the same conclusion can be made about this set of thrusters, i.e. the residuals surpass the threshold, albeit later than the occurrence of the attacks.

In figures 3.13a, 3.13b, and 3.13c the residual exceeds the threshold at the time of the occurrence of the FDI attacks, and remain above the threshold. Therefore, without any hesitation, it can be stated that in these UIOs–which are associated with thrusters 3 and 4, thrusters 3 and 5, and thrusters 3 and 6–the residuals go beyond the threshold.

The same pattern is also observed for the UIOs associated with pairs 4 and 5, 4 and 6, and thrusters 5 and 6, i.e. the residuals exceed the threshold right after t=15s.

Therefore, of all the possible pairs of thrusters for which a UIO and a residual signal can be developed for, i.e. 15 UIOs, only the one associated with thrusters 1 and 3 did not



(a) Residual of UIO for thrusters 1 and 2.

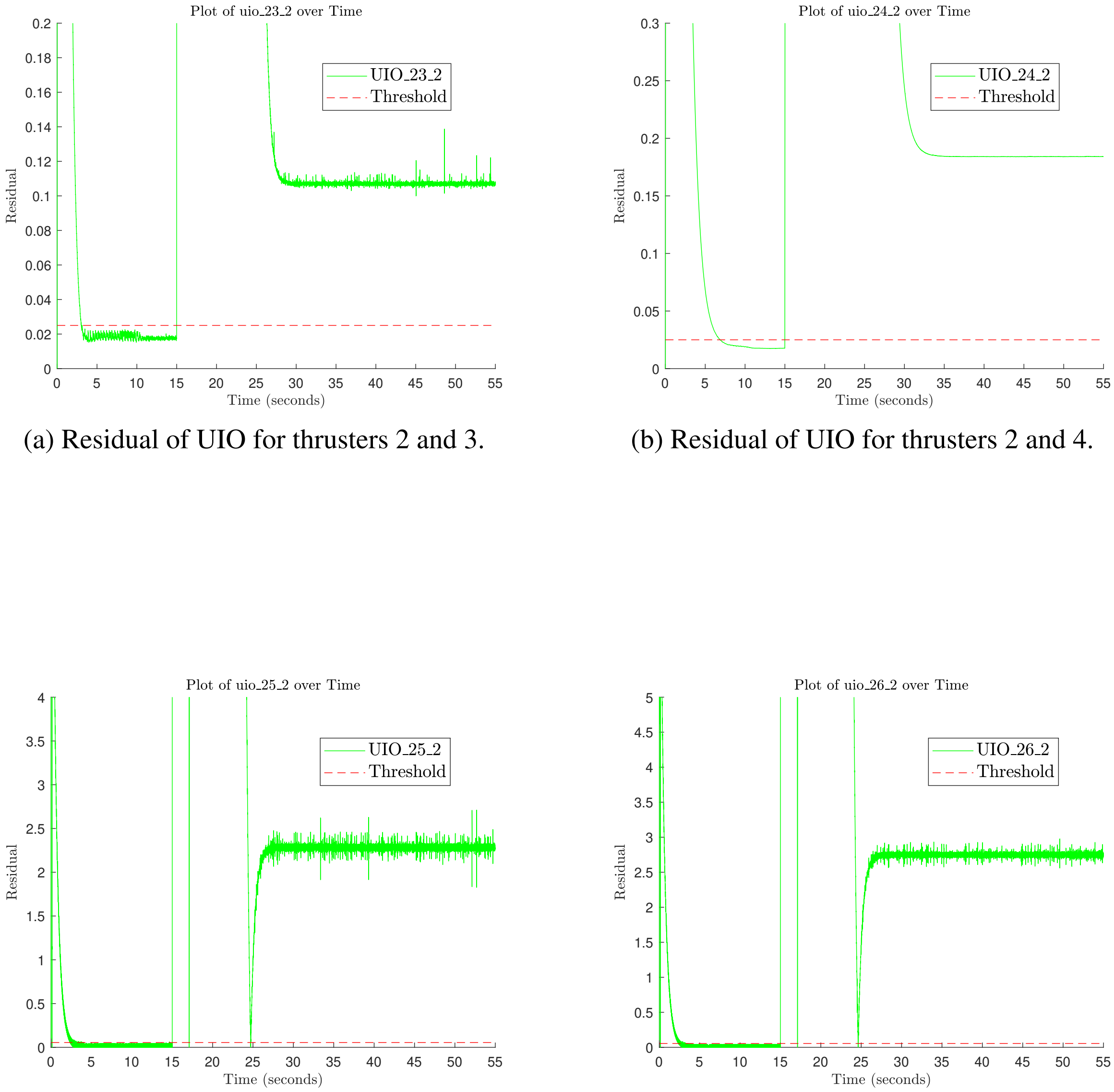
(b) Residual of UIO for thrusters 1 and 4.

(c) Residual of UIO for thrusters 1 and 5.

(d) Residual of UIO for thrusters 1 and 6.

Figure 3.11: Residual of UIOs associated with the thrusters paired with thruster 1 except for UIO 13, i.e. UIO 12, UIO 14, UIO 15, UIO 16.

produce residuals that would exceed their corresponding thresholds. Due to the transient behaviour of sliding mode observer that is constantly estimating $\boldsymbol{y}_a$, following the injection of attacks on channels 1 and 3, the residual becomes sensitive to attacks on the aforementioned channels from $15 < t < 25$. According to the isolation logic, table 3.1 demonstrates the outcome of the isolation : ,



(a) Residual of UIO for thrusters 2 and 3.

(b) Residual of UIO for thrusters 2 and 4.

(c) Residual of UIO for thrusters 2 and 5.

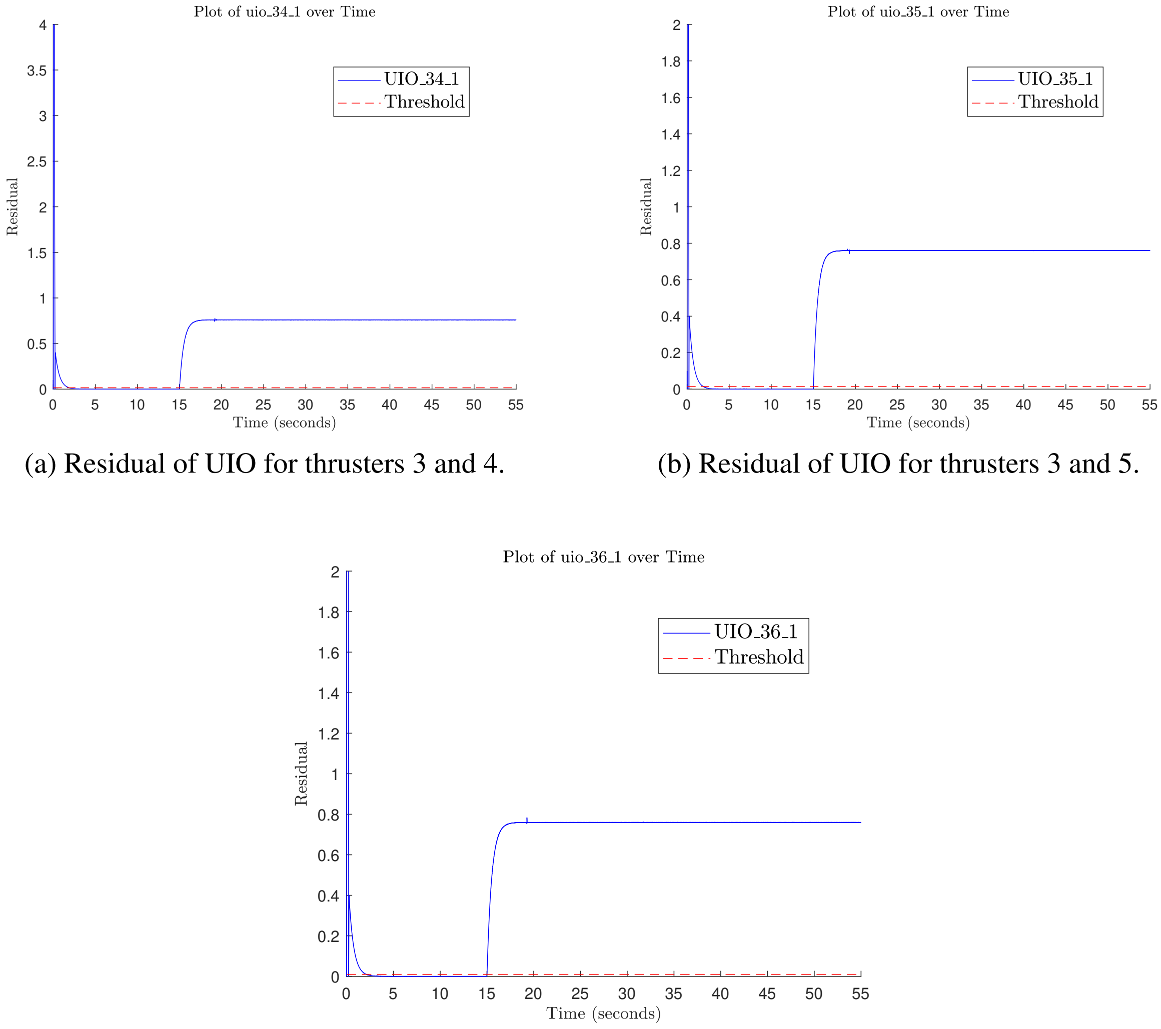
(d) Residual of UIO for thrusters 2 and 6.

Figure 3.12: Residual of UIOs associated with the thrusters paired with thruster 2, i.e. UIO23, UIO 24, UIO 25, UIO 26.

$0 < t < 15$	$15 < t < 25$	$25 < t$
No attack	Attack (location not determined)	Attack on thruster 1 and 3

Table 3.1: Outcome of the isolation for simultaneous attacks.

This isolation and the conclusion can be sent to the command and control center for further investigation and diagnosis. It can also be used to estimate the individual attack signals being injected onto the affected channels.



(a) Residual of UIO for thrusters 3 and 4.

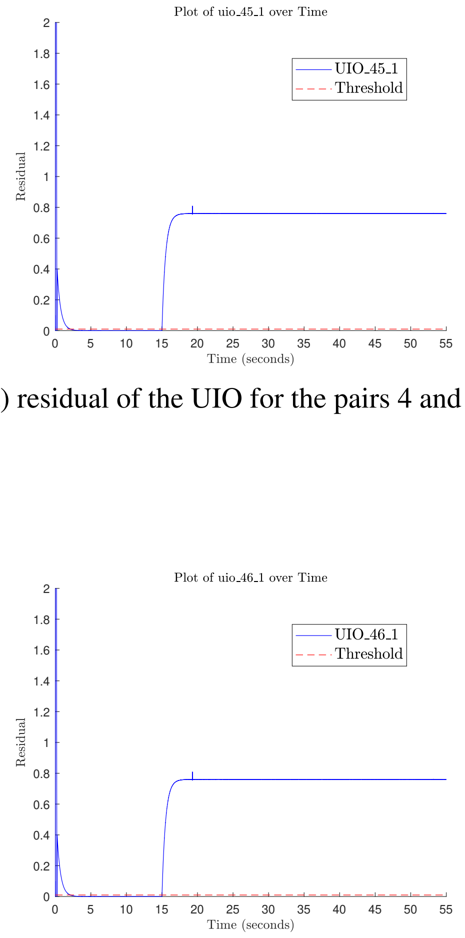
(b) Residual of UIO for thrusters 3 and 5.

(c) Residual of UIO for thrusters 3 and 6.

Figure 3.13: Residual of UIOs associated with the thrusters paired with thruster 3, i.e. UIO 34, UIO 35, UIO 36.

As far as the noisy behaviour of the residuals is concerned, the fluctuations in the residual signals could be attributed to the chattering that is a result of the sliding mode observer and also the solver that was chosen in the simulations in Matlab/Simulink.

The next step would to be to check for a possibility of single FDI attacks, meaning only one channel being compromised. Towards that end, we resort to the next set of family of UIOs, which are devised for isolating single FDI attacks corrupting the channels.



(a) residual of the UIO for the pairs 4 and 5.

(b) residual of the UIO for the pairs 4 and 6.

Figure 3.14: Residuals of UIOs associated with pair 4,5 and 4,6.

As shown in Figure 3.16, the residual for the UIO associated with the thruster 2 starts from above the threshold because of the initial condition, and exponentially fast goes below

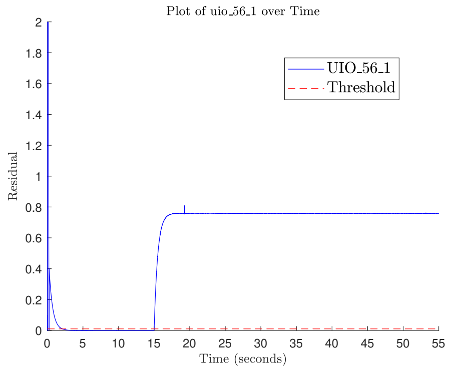


Figure 3.15: Residuals of UIO associated with pair 5,6.

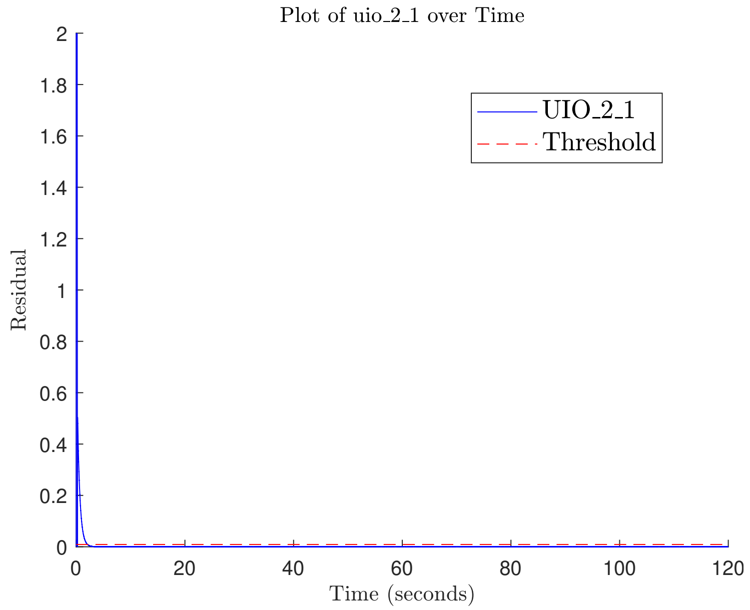


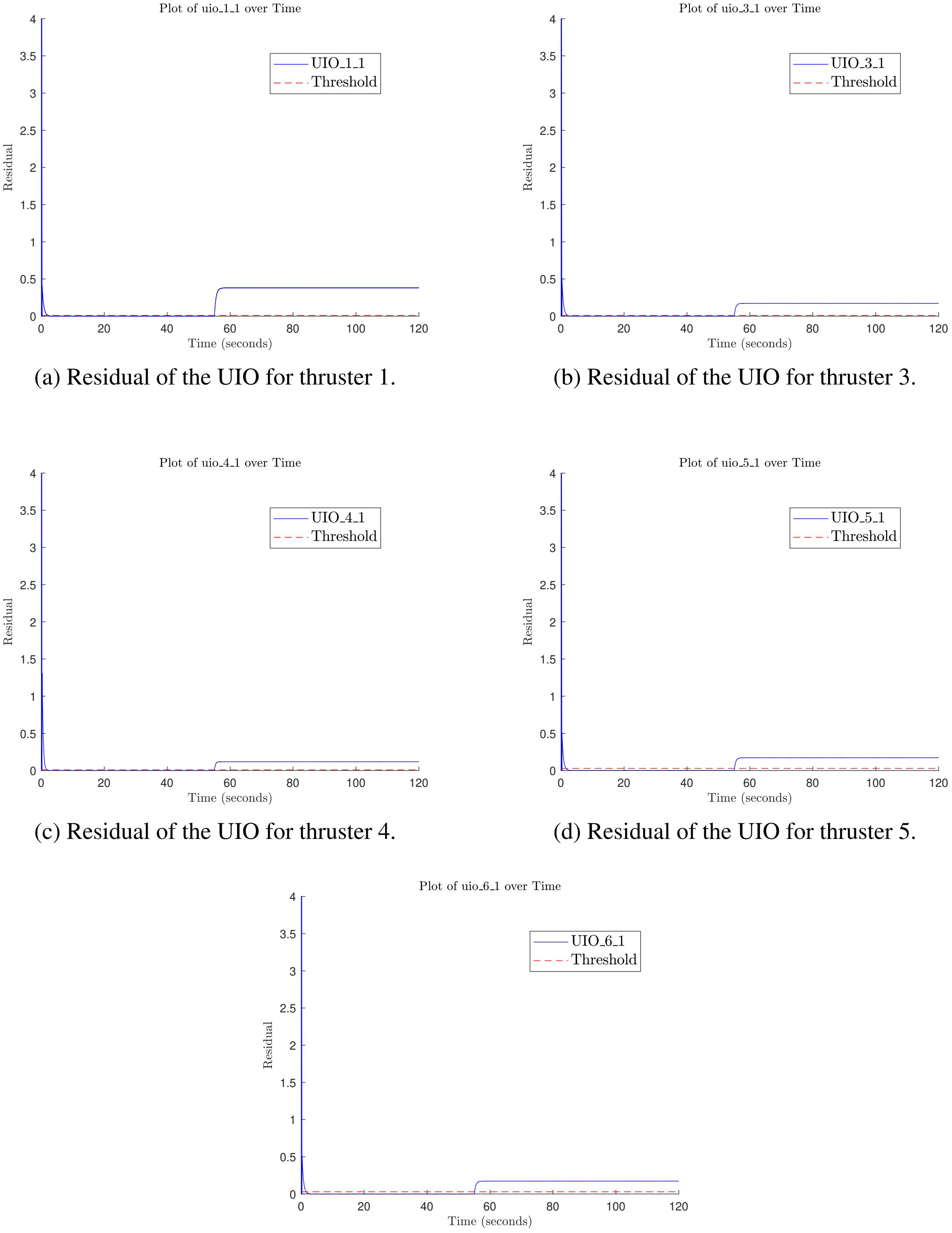
Figure 3.16: Residual of the UIO for thruster 2.

the threshold and remains there until the end. This thruster, i.e its associated communication channel could be a candidate for a possible FDI cyber-attack. However, to make that conclusion and to find out approximately when does this attack place, the residuals of all the other UIOs for single thrusters must also be examined. Paying attention to figures 3.17a, 3.17b, 3.17c, 3.17d and 3.17e, which represent the UIOs associated with thrusters 1, 3, 4, 5, and 6 respectively, it is revealed that all of the residuals start to surpass their associated thresholds at t=55s.

Consequently, by applying the isolation logic developed for isolating single FDI cyber-attacks, it is inferred that the communication channel from the controller to the thruster number is corrupted by an FDI cyber-attack at t=55s.

to sum up, two simultaneous cyber-attacks were isolated and is inferred to have corrupted channels 1 and 3 from at least t=25s and one single attack compromising channel 2 at t=55s.

Figures 3.18a, 3.18b, and 3.18c represent the estimation of the original FDI attack signals on the compromised channels 1, 3, and 2 respectively. The attacks on channel 1



(a) Residual of the UIO for thruster 1.

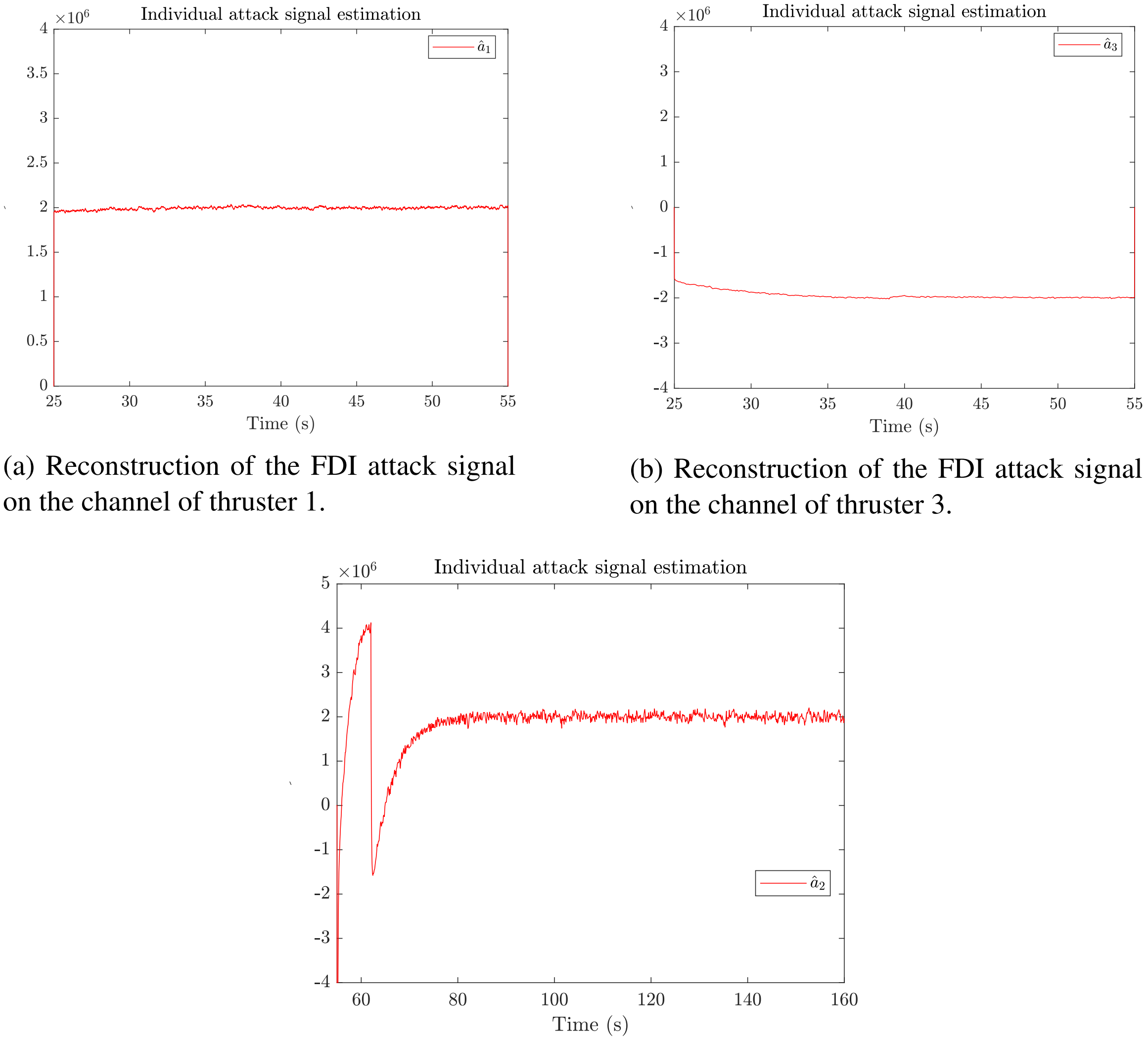
(b) Residual of the UIO for thruster 3.

(c) Residual of the UIO for thruster 4.

(d) Residual of the UIO for thruster 5.

(e) Residual of the UIO for thruster 6.

Figure 3.17: Residual of the UIOs for all thrusters except for thruster 2



(a) Reconstruction of the FDI attack signal on the channel of thruster 1.

(b) Reconstruction of the FDI attack signal on the channel of thruster 3.

(c) Reconstruction of the FDI attack signal on the channel of thruster 2.

Figure 3.18: Estimated and real states in the presence of FDI cyber-attacks in the network of thrusters using a sliding mode observer.

and 3 occur from t=15s to t=55s, with $a^1 = 2 \times 10^6$ and $a^3 = -2 \times 10^6$. As stated in the previous results, the isolation scheme was able to isolate the presence of FDI attacks on channels 1 and 3 after t=25s. Therefore, according to figures 3.18a and 3.18b, the original attack signals on the channels 1 and 3 were estimated by an acceptable accuracy from t=25s to t=55s. Furthermore, as mentioned previously, the isolation of the attack on channel 2 is concluded right after t=55s. Therefore, according to 3.18c, the original injected attack

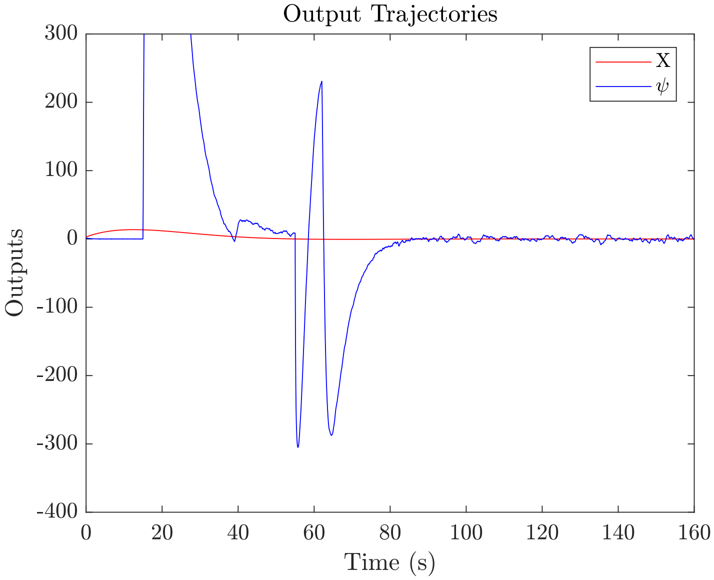


Figure 3.19: Effect of the FDI attacks on the surge movement and on the yaw.

signal on channel 2 is successfully reconstructed from t=55s.

Next, using the overall estimated attack signal $\hat{\boldsymbol{u}}_a$, the effect of the FDI cyber-attacks are compensated based on the methodology described in Section 3.8. The cut-off frequencies for the low-pass filter have chosen to be: $w_{c1} = 2, w_{c2} = 10$, and $w_{c3} = 10$.

As shown previously in Figure 3.2, the FDI attack had caused significant deviation of the surge and yaw movements from their operating point. By applying the compensation scheme, the X position and the yaw angle of the vessel is maintained around the origin in steady state, as shown in Figure 3.19. Therefore, the compensation scheme has managed to successfully attenuate the effect of the FDI attacks whiteout using redundancy and controllability.

3.9.2 Drawbacks of Reconfiguration Based Compensation Methods in the Literature

To illustrate the shortcoming of reconfiguration based compensation methods in the Literature that exclude the isolated thrusters from operation, the following scenario is considered:

From $t = 15s$ to $t = 55s$, the first and third communication channels are compromised. Using the isolation scheme developed in this work, these channels are isolated at $t > 25s$. If these compromised channels are excluded from operation, the thrust distribution matrix, originally given by

$$T = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 4 & -4 & -35 & -33 & 40 & 28 \end{bmatrix},$$

is modified to the following:

$$T' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ -4 & -33 & 40 & 28 \end{bmatrix}.$$

At $t = 55s$, the second channel is compromised and isolated immediately after $t = 55s$. Excluding the second channel results in another new thrust distribution matrix:

$$T'' = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ -33 & 40 & 28 \end{bmatrix}.$$

According to Section 3.3 , the effective $\bar{B}$ matrix after excluding these channels becomes

$$\bar{B} = B \times T'' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -0.0002 & 0.0002 & 0.0002 \\ -0.4323 & 0.5240 & 0.3668 \end{bmatrix}.$$

To check the controllability of the system, the rank of the controllability matrix $Co = [\bar{B}\ A\bar{B}\ A^2\bar{B}\ A^3\bar{B}\ A^4\bar{B}\ A^5\bar{B}]$ is computed. The rank of the controllability matrix is found to be 4, which is less than the number of states (6). Therefore, the system becomes uncontrollable. In the event of a disturbance that affects the surge movement:

According to Figure 3.20, using this reallocation methodology in the literature, particularly the work of [40], the X position of the ship diverges from the origin and the controller is not able to bring it back. This clearly demonstrates the fact that excluding compromised

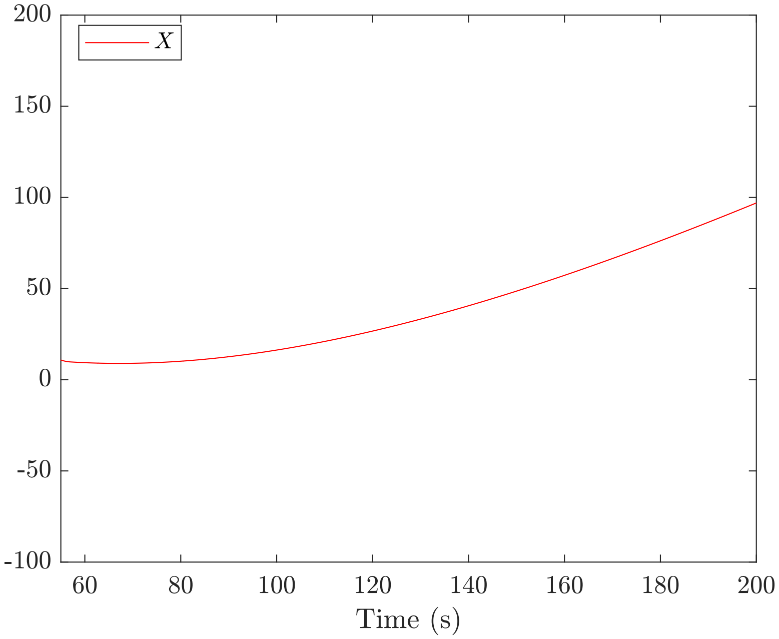


Figure 3.20: Divergence of the X position from the origin.

channels can lead to a situation where the remaining thrusters fail to fulfil the control objectives, particularly under consecutive FDI attack scenarios.

3.10 Conclusion

In this chapter, using high-order sliding mode observer, the true system states and the overall attack signals were estimated in a centralized thrust allocation framework. Due to the fact that observer matching condition does not hold true for the DP system, conventional unknown input observers could not be implemented. First, the system outputs were augmented, and then estimated. Followed by that, banks of UIOs were used the isolate the FDI attacks on the communication channels. Furthermore, the overall attack signal and the original attacks signals on the affected communication channels were reconstructed. As far as the mitigation scheme, the estimated attack signals along with low pass filters were used to modify the controller and to compensate for the effect of the FDI attacks. The observer managed to estimate the true system states despite the cyber-attacks and the attack signal with a proper precision. This chapter demonstrated that a comprehensive estimation, isolation and compensation methodology could be achieved without the strict requirements of the output matrix.

Chapter 4

A Decentralized Thrust Allocation Framework as a Security Measure

4.1 The Decentralized Architecture

The conventional centralized thrust allocation module in over-actuated systems including DP systems, that operate within a CPS framework, raises serious vulnerabilities to cyber-attacks. In order to send the commanded thrust to each thruster, a communication link is needed. We summarize below the security challenges of utilizing conventional centralized thrust allocation mechanism.

(1) Communication lines from the controller and the thrust allocation module, both of which reside at the command and controller center at the bridge, to the thrusters introduce a great attack surface that could be exploited by adversaries who have gained access to the integrated platform management system network and can inject false data injection attacks on the communication lines.

(2) Once the communication lines have been compromised, the best course of action is to detect, isolate the compromised thrusters and reconfigure the allocation module.

As shown in [58], this approach involves developing several observers in addition to a reconfiguration strategy. Furthermore, in this method there is a limitation as to isolation of simultaneous thruster faults. By extension, applying this method of detection, isolation and reconfiguration in an over-actuated system subjected to FDI attacks, not only are there a lot of steps to be carried, but also there are limitations to this approach.

(3) If the attacker compromises a few communication channels from the controller to the thrusters, and knows which channels belongs to which thrusters, they could launch a more targeted, sophisticated and damaging attack. This weakness is very much likely to be exploited in conventional centralized allocation frameworks.

The FDI attacks compromising the communication links from the thrust allocation module to the thrusters in the centralized allocation scheme and their effect on the overall control signal are modeled as follows:

The command signals from the thrust allocation module to the thrusters are denoted as u^i. The vector $\boldsymbol{u}$ comprises these signals:

$$\boldsymbol{u} = \begin{bmatrix} u^1 \\ u^2 \\ \vdots \\ u^n \end{bmatrix}$$

The false data injections on these communication channels are represented by the vector $\boldsymbol{a}$, consisting of individual injection signals. These injections can be variable or constant but are considered to be bounded:

$$\boldsymbol{a} = \begin{bmatrix} a^1 \\ a^2 \\ \vdots \\ a^n \end{bmatrix}$$

The final command signals sent to the thrusters, denoted as $\boldsymbol{u'}$, are obtained by adding the false data injections to the original command signals:

$$\boldsymbol{u'} = \boldsymbol{u} + \boldsymbol{a} = \begin{bmatrix} u^1 \\ u^2 \\ \vdots \\ u^n \end{bmatrix} + \begin{bmatrix} a^1 \\ a^2 \\ \vdots \\ a^n \end{bmatrix}$$

The thrust allocation module solves the optimization problem, where the objective function is $\boldsymbol{u}^T\boldsymbol{W}\boldsymbol{u}$ and the constraint is $T\boldsymbol{u} = \boldsymbol{u_c}$, using the interior point method. In the presence of an FDI attack, the relationship becomes:

$$T(\boldsymbol{u} + \boldsymbol{a}) = T\boldsymbol{u'} = \boldsymbol{u_c} + \boldsymbol{u_a}$$

This means that the original command signal has been corrupted, and it is this corrupted signal that affects the plant, not the original $\boldsymbol{u_c}$.

In Figure 4.1, the aforementioned vulnerabilities of the conventional thrust allocation module to FDI attacks on the communication lines from the controller-thrust allocation module to the actuator-thrusters is presented. The red arrows represent potential FDI attacks on each line. Considering that the sensor measurements from the IMUs, gyrocompass and GPS are transmitted safely to the controller, we are basically dealing with an actuator attack.

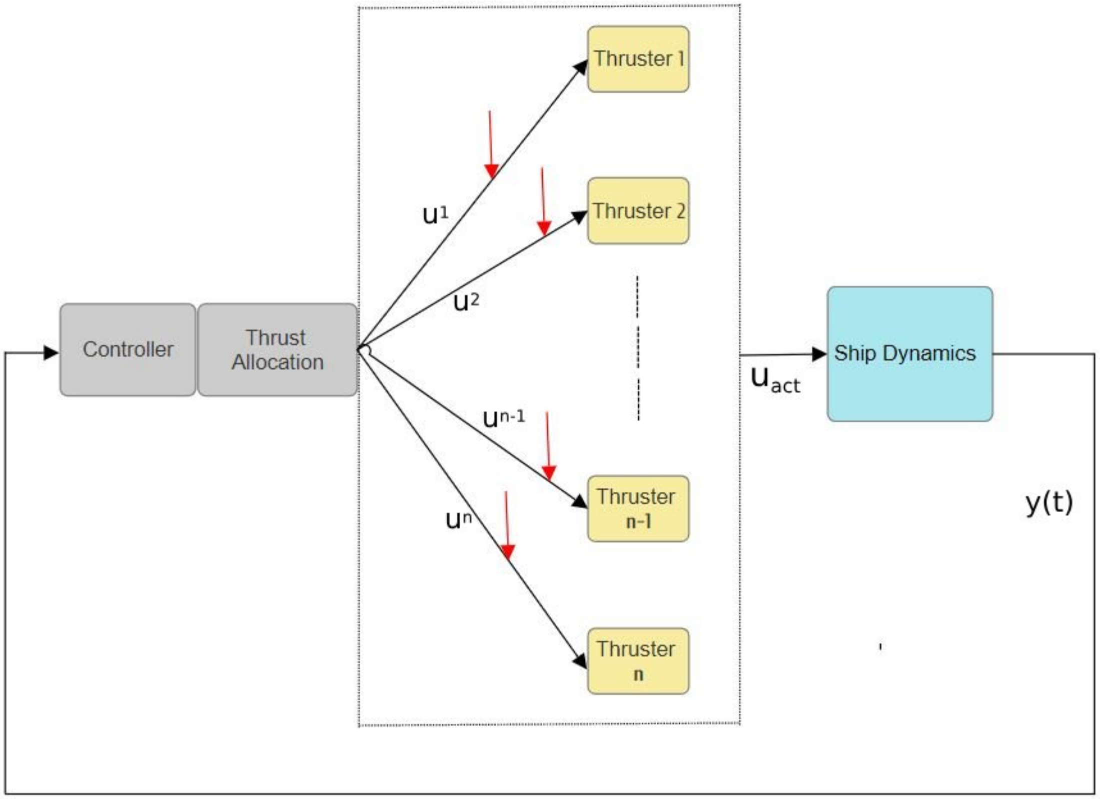


Figure 4.1: Vulnerabilities to FDI attacks in a centralized thrust allocation framework. In a DP system with n thrusters, there are n communication links that could be exploited and attacked. The red arrows in the figure represent individual FDI attacks on the communication links, which affect the individual desired thrust values of thrusters.

Therefore, in order to address the preceding challenges, the new thrust allocation module must reduce the attack surface. In order to achieve this, a new decentralized thrust allocation framework is presented in this work. In this framework, the thrusters achieve the desired commanded thrust collectively. There is an internal separate network between the thrusters and they aim at realizing the demanded thrust vector by communicating with the neighboring thrusters and a consensus protocol that is embedded at each thruster. This new framework should satisfy the following requirements:

- The communication channels between the controller-thrust allocation and the thrusters should be minimized.

- The thrusters should achieve the desired commanded thrust which are generally the desired thrusts in surge, sway and yaw.

- The thrusters should minimize the global objective function associated with fuel consumption, just like they do when configured in a centralized fashion.

- The thrusters must achieve all the above in a decentralized manner, meaning they would only have access to their own information and that of their neighbours.

- In order to establish the consensus protocol, each thruster must be equipped with an auxiliary system that has a computing capability along with a communication interface.

- The consensus protocol would not ensure consensus on the individual thrusts produced by each thruster, but rather consensus on a variable which would ensure the minimization of the fuel consumption and generation of the desired thrust vector.

Towards that end, this decentralized framework is represented in Figure 4.2. According to this figure, there is an internal network amongst the thrusters. There is no longer a thrust allocation module next to the controller. In this setup, the controller sends the commanded thrust to only one thruster. However, this configuration raises a few issues as following:

- The fact that there is only one communication line from the controller to a thruster, e.g. thruster i, introduces a vulnerability, as is shown in Figure 4.2. The only communication line from the controller to a thruster, is susceptible to FDI attack and this is not desirable.

- Although this new decentralized framework has managed to reduce the communication lines from n to only 1, the internal network between the thrusters could have as many communication lines as in centralized allocation, or even more.

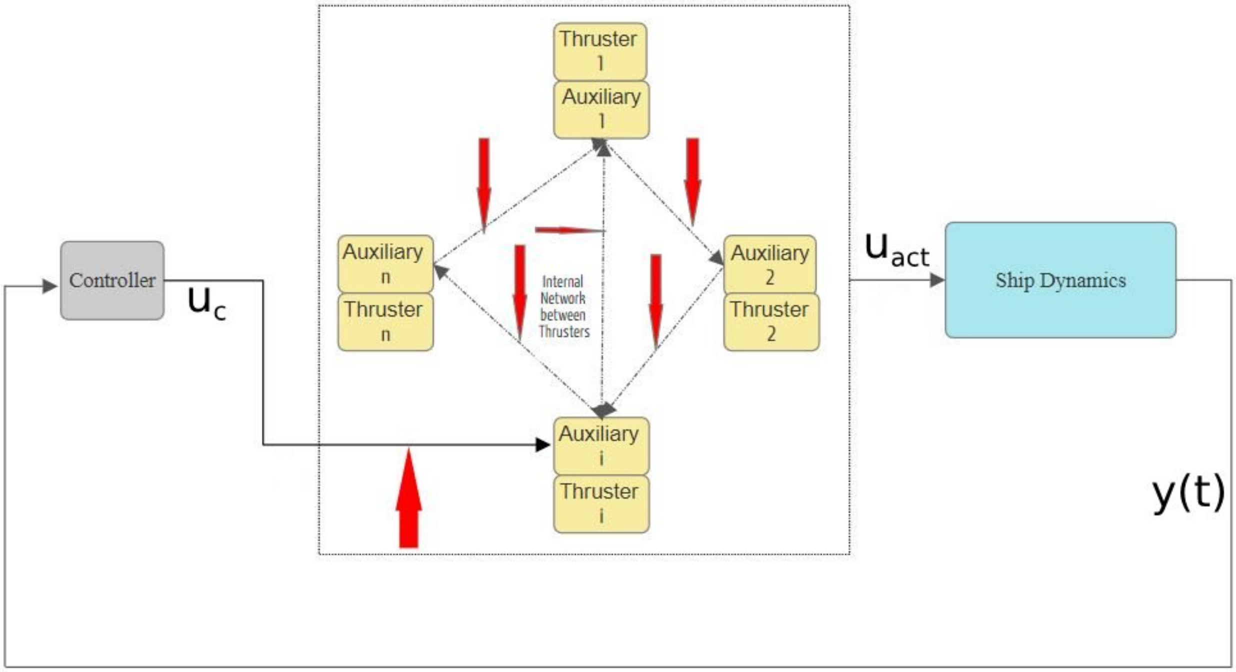


Figure 4.2: A new decentralized thrust allocation module with the aim of reducing the attack surface. The red arrows characterize the FDI attacks that could compromise the communication links between the thrusters.

- The communication lines within the internal network of thrusters, regardless of their number, are susceptible to FDI attacks.

Therefore, this decentralized scheme must be refined in order to resolve the issues that were discussed. The following solutions are proposed:

- **Randomized connectivity from the controller to the thrusters**: If the control command, i.e., commanded thrust, is sent from the controller to only one thruster, an FDI attack could manipulate this signal and disrupt the system. On that account, in the modified framework the control command could be sent to a subset of thrusters randomly. Implying that at a given period of time, the controller would send the commanded thrust to one thruster and that one thruster could be any thruster within a predefined subset of thrusters. The reason for this mechanism is to shift the attack surface and create uncertainty for the attacker, should they intend on compromising this signal.

- **Minimal communication lines between the thrusters**: As it has been stated several times in previous sections, one of the objectives of this thesis has been reducing the attack surface. Although the number of communication lines from the controller to the thrusters has been reduced, setting up a separate interval network between the thrusters would create another vulnerability. Thus, in order to fulfil the original objective, the topology of the internal network of the thrusters should have minimum number of communication lines possible. By virtue of that, the graph representing the network of the thrusters should have a small number of edges and be a directed graph.

- **Resilient network architecture**: Through reducing the communication lines from the controller to the thrusters the attack surface would diminish. By establishing a mechanism through which the controller could send the commanded thrust to any thruster randomly, the attack surface would be shifted. In a manner, the attacker will be pushed away from attacking the communication lines between the controller and the thruster(s), to compromising the communication lines between the thrusters. Therefore, the only course of action left to do as a defender is to prop up the communication lines between the thrusters against FDI attacks. Thus, the need to develop a resilient consensus protocol becomes evident. This resilient protocol must ensure that the final produced thrust vectors should not deviate substantially from the commanded thrust vector should FDI attacks occur on the communication lines within the network.

An instance of the modified configuration of a new decentralized scheme is presented in Figure 4.3. This representation is just an instance of the modified framework which satisfies the requirements mentioned above..

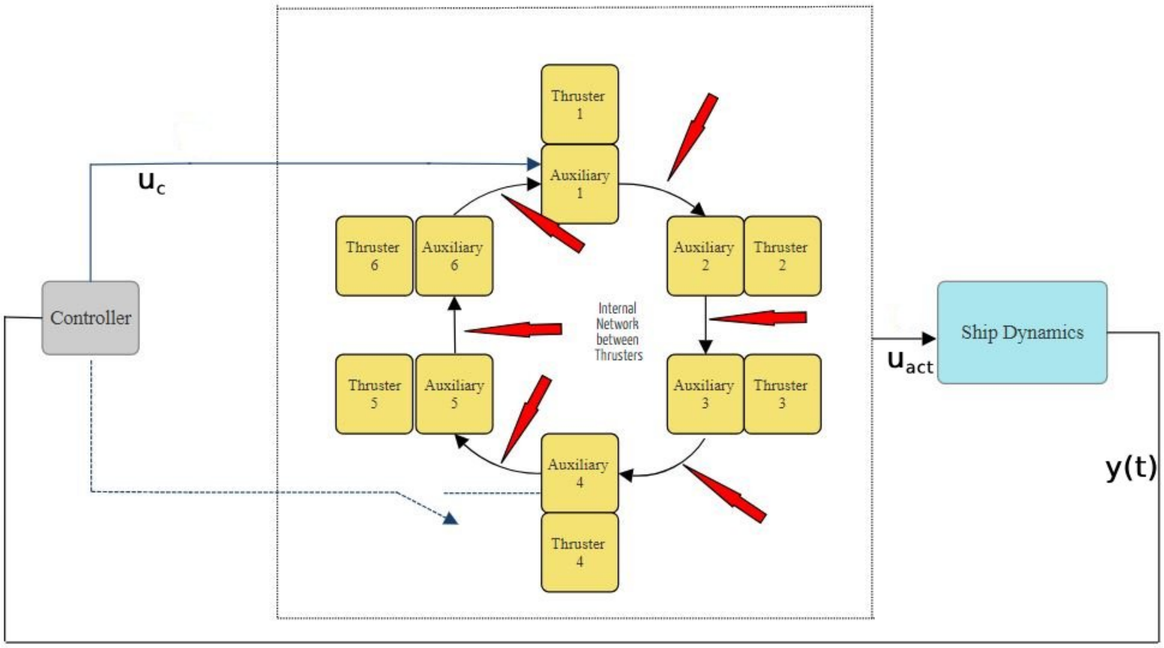


Figure 4.3: The final modified decentralized scheme. The red arrows show potential FDI attacks

As shown in this Figure 4.3, the controller can send the thrust command to thruster 1 (its auxiliary system). The command could also be sent to the thruster 4, but at this particular moment, it is being sent to the auxiliary system of thruster 1. The auxiliary systems of the thrusters are communicating their local states – which will be defined in the following sections– with their neighboring auxiliary systems based on a consensus algorithm to achieve the commanded thrust and satisfy the optimization requirement. The topology is a ring topology, which is strongly connected and balanced. The consensus algorithm should be resilient to FDI attacks.

Before presenting the consensus algorithm and the mathematical framework, further justifications are put forward for considering this new framework.

Primary Benefits

- Although we have reduced the communication lines from the controller to the thrusters down to 1, one might object that the internal network between the thrusters adds communication lines, therefore the attack surface has not really changed. However, it should be pointed out that in a conventional centralized scheme having n thrusters, there would be n communication lines. In this new decentralized framework, the number of communication lines between the thrusters is n, using a ring topology which is strongly connected and balanced. Hence, the total number of communication lines would be $n + 1$, which has only one more communication line than the centralized scheme.

- In this new framework, in order to achieve the commanded thrust vector and optimize the fuel consumption, the flow of information between the controller and the thrusters occurs in two segments. In the first segment, the commanded thrust is sent to only one thruster. The choice of which thruster is the commanded thrust being sent to is determined randomly by the controller. This randomness makes it unpredictable for potential adversaries, as they cannot reliably target a specific thruster. It adds an element of uncertainty, making it more challenging for attackers to formulate and carry out targeted FDI attacks. In the second segment, via a consensus protocol and communication topology, the thrusters communicate with each other only their states.

- The consensus protocol is supposed to be resilient against any number of bounded FDI attacks on the communication lines between the thrusters in the network. In the conventional centralized scheme, an attacker could inflict a bounded but a devastating FDI attack on the communication lines between the controller and the thrusters, causing disruption on the system. However, in this new framework, the adversaries could indeed impose large FDI attack signals on the internal communication lines between the thrusters, however the resilient algorithm will attenuate the effects of the attack signals significantly in a passive manner.

Secondary Benefits

- **Scalability**: Unlike the centralized framework, in this framework, adding or removing thrusters does not require substantial changes to the overall network configuration. Moreover, the consensus protocol would remain the same, along with the control signal being sent to a thruster. This flexibility facilitates system expansion or modification without disrupting existing functionalities.
- **Computational Load**: As the number of thrusters increases, the computational complexity of the centralized thrust allocation module grows, demanding significant computational resources for optimization processes. Using this decentralized framework, the computational load will be distributed among the auxiliary systems of the thrusters.

4.2 Problem Formulation

In thrust allocation, the problem of concern is to find the thrust values of individual thrusters that collectively result in the desired commanded thrust vector coming from the controller, while minimizing a cost function. In most cases, the cost function is chosen as the fuel consumption. Fuel consumption usually is considered to have a quadratic relationship with the generated thrust for each thruster. Hence, the optimization problem is as follows:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = \mathbf{x}^T W \mathbf{x} \\ \text{subject to} \quad & T\mathbf{x} = \mathbf{u_c} \end{aligned} \tag{4.1}$$

here $\mathbf{x}$ is the vector of desired thrusts of all the thrusters: $\mathbf{x} = [x^1...x^i...x^n]^T$, W is the weight matrix representing the contribution of each thruster to the overall fuel consumption, u_c is the commanded thrust vector originating in the controller and T is the thruster configuration matrix.

Evident from 4.1, the real produced thrust vector u, has been replaced by x, knowing that if 4.1 is satisfied, then 2.19 will be satisfied. The overall structure of the consensus protocol acting on N thruster will take the following form:

$$\dot{\boldsymbol{\xi}}^i(t) = \mathbf{g}(\boldsymbol{\xi}^i(t), \boldsymbol{\xi}^j(t), x^i(t), \forall j \in \mathcal{N}_i) \tag{4.2}$$

$$\dot{x}^i(t) = h(x^i(t), \boldsymbol{\xi}^i(t)) \tag{4.3}$$

where $\boldsymbol{\xi}_i$ is the virtual state vector of the auxiliary system associated with thruster i, x_i is the desired thrust of the thruster i, and $\mathcal{N}_i$ is the set of in-neighbours of agent i. The idea is to develop a protocol in which the dynamics of the states of the auxiliary systems associated with their thrusters evolve based on the exchange of these states with their neighbouring auxiliary systems' states and a consensus like law. The states should eventually reach a common value in a way that in steady state, the resulting vector $\mathbf{x}$ will have satisfied 4.1.

The dimension of the state vector for each auxiliary system should match that of $\mathbf{u_c}$ in 4.1, the commanded thrust vector. for the sake of convenience from here on, instead of auxiliary system of each thruster, the term "agent" will be used. It is important to note that the dynamics of x_i and $\boldsymbol{\xi}_i$ affect each other. It also needs to be clarified that the state vectors involved in the consensus algorithm on each agent doesn't correspond to real variables.

As the consensus algorithm is running, at each time step, the x_i of each agent will be the desired thrust value for that agent and hence, will be sent to the engine controller as the set-point. In 4.1, $\mathbf{x}$ is the desired thrust value, not the actual generated thrust. Each thruster has its own dynamics which could be approximated as follows [82]:

$$
\begin{aligned}
u_d^i &= x^i \\
\dot{u^i} &= \frac{(u_d^i - u^i)}{\tau_i}
\end{aligned}
\tag{4.4}
$$

where u_d^i is the desired thrust of thruster i, u^i is its generated thrust, and τ_i is its time constant.

If FDI attacks occur on the communication links between agent i and its neighbouring agents, the dynamics of the states of agent i will be:

$$\dot{\boldsymbol{\xi}}^i(t) = \mathbf{g}(\boldsymbol{\xi}^i(t), \boldsymbol{\xi}^j(t) + a^j, x^i(t), \forall j \in \mathcal{N}_i) \tag{4.5}$$

where a^j is the attack vector on the agent j's state vector that is being sent to agent i. It is well-known that in a regular consensus algorithm, if an FDI attack occurs on the communication lines, it could cause instability. Therefore, the objective is to develop the consensus protocol in a way that in the presence of bounded FDI attacks on the communication links between the agents, the difference between the resulting control force (i.e., thrust vector) and the commanded control force, is small.

In other words:

$$\|T\mathbf{u}(\mathbf{t}) - \mathbf{u_c}(\mathbf{t})\| < k \quad \forall t > T > 0 \tag{4.6}$$

where $\mathbf{u}(\mathbf{t}) = [u^1 \quad u^2 \quad ... \quad u^n]^T$, and K is a sufficiently small positive value. Therefore, if the consensus algorithm succeeds in developing a decentralized scheme that satisfies 4.1, with the functionality that the commanded control signal could be sent to any agent at random at each period of time, then it will have succeeded in reducing the attack surface from the controller to the thrusters and prevents a possible attack on the communication line from the controller to that one agent.

Once the attacker has only the communication lines between the thrusters to compromise, if the consensus algorithm manages to satisfy 4.6, then the algorithm will have succeeded in being resilient to FDI attacks. The development of the consensus algorithm will take place in the subsequent sections of this chapter.

4.3 Resilient Dynamic Average Consensus

In this work, the concept of dynamic average consensus is utilized, and developing the consensus algorithm is based on the work of [83] – which itself is inspired by competitive interaction [84] – with a few alterations and adjustments. The consensus protocol is stated as follows:

$$\begin{aligned} \dot{\mathbf{y}}^i &= \alpha(\mathbf{r}^i - \mathbf{y}^i) - \alpha\beta \sum_{j \in N_i} (\mathbf{y}^i - \mathbf{y}^j) + \beta \sum_{j \in N_i} (\mathbf{z}^i - \mathbf{z}^j) + \mathbf{v}_r^i \\ \dot{\mathbf{z}}^i &= \alpha(\mathbf{r}^i - \mathbf{z}^i) - \beta \sum_{j \in N_i} (\mathbf{y}^i - \mathbf{y}^j) - \alpha\beta \sum_{j \in N_i} (\mathbf{z}^i - \mathbf{z}^j) + \mathbf{v}_r^i \end{aligned} \tag{4.7}$$

where $\mathbf{y}^j \in \mathbb{R}^3$ and $\mathbf{z}^j \in \mathbb{R}^3$ are the states of each agent, $\mathbf{r}^i \in \mathbb{R}^3$ is the reference signal of each agent and $\mathbf{v}_r^i$ being its time derivative, α and β are constant parameters that need to be designed.

The protocol is laid out over a strongly connected and balanced digraph *G*, due to the justifications mentioned in 4.1. The reason for choosing the dimensions of $\mathbf{y}^j$ and $\mathbf{z}^j$ as 3, is due to the commanded control signal having 3 dimensions (surge, sway, and yaw). The objective is for the agents to track the average of these signals. In other words:

$$\|\mathbf{y}^i - \frac{1}{n}\sum_{j=1}^{n} \mathbf{r}^j\| \quad and \quad \|\mathbf{z}^i - \frac{1}{n}\sum_{j=1}^{n} \mathbf{r}^j\| < M \quad \forall t > T > 0 \tag{4.8}$$

Where M could be made small. Equations (4.7) could be represented in a compact form by using the Laplacian matrix of the connected and weight-balanced digraph. The equations below illustrate that form:

$$\begin{aligned}\dot{\mathbf{y}} &= \alpha(\mathbf{r}-\mathbf{y}) - \alpha\beta\mathbf{L}\mathbf{y} + \beta\mathbf{L}\mathbf{z} + \mathbf{v}_r^i \\ \dot{\mathbf{z}} &= \alpha(\mathbf{r}-\mathbf{y}) - \beta\mathbf{L}\mathbf{y} - \alpha\beta\mathbf{L}\mathbf{z} + \mathbf{v}_r^i\end{aligned} \tag{4.9}$$

where $\mathbf{L} = L \otimes I_3$, and L is the Laplacian matrix associated with the digraph *G*. The error signals $\mathbf{e_y}$ and $\mathbf{e_z}$ are defined as:

$$\begin{aligned}\mathbf{e_y} &= \mathbf{y} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T \otimes I_3\mathbf{r} \\ \mathbf{e_z} &= \mathbf{z} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T \otimes I_3\mathbf{r}\end{aligned} \tag{4.10}$$

where $\mathbf{y} \in \mathbb{R}^{3n}$ and $\mathbf{z} \in \mathbb{R}^{3n}$ combine every $\mathbf{y}^i$ and $\mathbf{z}^i$, respectively into one vector. The error signals represent the difference between the states of all agents, and average of the references. The dynamics of the error signals are outlined below:

$$\begin{aligned}\dot{\mathbf{e}}_{\mathbf{y}} &= \alpha(\mathbf{r}-\mathbf{y}) - \alpha\beta\mathbf{L}\mathbf{e}_y + \beta\mathbf{L}\mathbf{e}_z + \mathbf{v}_r^i - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T \otimes I_3\mathbf{v}_r \\ \dot{\mathbf{e}}_{\mathbf{z}} &= \alpha(\mathbf{r}-\mathbf{z}) - \beta\mathbf{L}\mathbf{e}_y - \alpha\beta\mathbf{L}\mathbf{e}_z + \mathbf{v}_r^i - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T \otimes I_3\mathbf{v}_r\end{aligned} \tag{4.11}$$

It should be noted that $\mathbf{L}\mathbf{e}_y = \mathbf{L}\mathbf{y}$ and $\mathbf{L}\mathbf{e}_z = \mathbf{L}\mathbf{z}$, since:

$$\mathbf{L} \times \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T \otimes I_3 = (L \otimes I_3) \times (\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T \otimes I_3) = \frac{1}{n}(L \times \mathbf{1}_n\mathbf{1}_n^T) \otimes (I_3) = 0 \tag{4.12}$$

Owing to the fact that $L \times \mathbf{1}_n = 0$, as $\mathbf{1}_n$ is the left eigenvector of the Laplacian matrix L associated with the zero eigenvalue of the Laplacian matrix. After further simplifications and manipulations, the dynamics of the error signals could be represented in the following form:

$$\begin{aligned}\dot{\mathbf{e}}_{\mathbf{y}} &= -(\alpha I_{3n} + \alpha\beta\mathbf{L})\mathbf{e}_y + \beta\mathbf{L}\mathbf{e}_z + P_{3n}(\alpha\mathbf{r} + \mathbf{v}_r) \\ \dot{\mathbf{e}}_{\mathbf{z}} &= -(\alpha I_{3n} + \alpha\beta\mathbf{L})\mathbf{e}_z - \beta\mathbf{L}\mathbf{e}_z + P_{3n}(\alpha\mathbf{r} + \mathbf{v}_r)\end{aligned} \tag{4.13}$$

where $P_{3n} = I_{3n} - \frac{1}{n}(\mathbf{1}_n\mathbf{1}_n^T) \otimes I_3$.

A new state vector is defined as $\zeta \in \mathbb{R}^{6n}$ which consists of $\mathbf{e}_y$ and $\mathbf{e}_z$: $\zeta = [\mathbf{e}_y^T \quad , \quad \mathbf{e}_z^T]^T$. Therefore, the error dynamics could be shown in the following way:

$$\dot{\zeta} = \Xi\zeta + [1, \quad 1]^T \otimes (P_{3n}(\alpha\mathbf{r} + \mathbf{v}_r)) \tag{4.14}$$

and $\Xi \in \mathbb{R}^{6n\times 6n}$ being:

$$\Xi = \begin{bmatrix} -(\alpha I_n + \alpha\beta L) & -(\alpha I_n + \alpha\beta L) \\ -(\alpha I_n + \alpha\beta L) & -(\alpha I_n + \alpha\beta L) \end{bmatrix} \tag{4.15}$$

In order to analyze the stability of this protocol, the zero-input stability of 4.14 is studied first. Specifically,

$$\dot{\zeta} = \Xi\zeta \tag{4.16}$$

In order to analyze the stability of 4.16, diagonalization of matrix Ξ and examining the eigenvalues of the Laplacian matrix is employed and two similarity transformations of T and S are used. Given that $\zeta = \mathbf{TS}\bar{\zeta}$, then it follows that 4.16 is similar to: $\dot{\bar{\zeta}} = S^{-1} \quad T^{-1} \quad \Xi \quad T \quad S \quad \bar{\zeta}$. The transformation matrices of T and S are defined as follows:

$$T = \begin{bmatrix} iI_{3n} & -iI_{3n} \\ I_{3n} & I_{3n} \end{bmatrix} \tag{4.17}$$

$$S = \begin{bmatrix} V & 0 \\ 0 & V \end{bmatrix} \tag{4.18}$$

where $V = v \otimes I_3$ and v is the matrix consisting of left eigenvectors of the Laplacian matrix. Applying the first transformation yields:

$$\Xi' = T^{-1} \quad \Xi \quad T = \begin{bmatrix} -\alpha I_{3n} - \alpha\beta\mathbf{L} - i\beta\mathbf{L} & 0 \\ 0 & -\alpha I_{3n} - \alpha\beta\mathbf{L} + i\beta\mathbf{L} \end{bmatrix} \tag{4.19}$$

So far, the matrix Ξ has been diagonalized. However, without the second transformation, relating the eigenvalues of the matrix L to this newly transformed matrix is not straightforward. Hence, the second transformation is applied as follows:

$$S^{-1}\Xi' S = \begin{bmatrix} -\alpha V^{-1}V - \alpha\beta V^{-1}\mathbf{L}V - i\beta V^{-1}\mathbf{L}V & 0 \\ 0 & -\alpha V^{-1}V - \alpha\beta V^{-1}\mathbf{L}V + i\beta V^{-1}\mathbf{L}V \end{bmatrix} \tag{4.20}$$

And given that:

$$\begin{aligned} V^{-1}V &= (v^{-1}v) \otimes I_3 = I_{3n}, \\ V^{-1}\mathbf{L}V &= (v^{-1}Lv) \otimes I_3 = j \otimes I_3 \end{aligned} \tag{4.21}$$

j being the Jordan normal form of the Laplacian matrix, the following equation is obtained:

$$\dot{\bar{\zeta}} = \begin{bmatrix} -\alpha I_{3n} - \alpha\beta J - i\beta J & 0 \\ 0 & -\alpha I_{3n} - \alpha\beta J + i\beta J \end{bmatrix} \bar{\zeta} = \Xi'' \bar{\zeta} \tag{4.22}$$

where $J = j \otimes I_3$. Therefore, in order for 4.16 to be stable, the diagonalized matrix Ξ'' needs to be Hurwitz. If $\lambda_i = \delta_i + i\omega_i$ is the ith eigenvalue of the Laplacian matrix, then in order for Ξ'' to be Hurwitz, the following condition must hold:

$$\begin{aligned} -(\alpha + \alpha\beta\delta_i - \beta\omega_i) < 0 \\ -(\alpha + \alpha\beta\delta_i + \beta\omega_i) < 0 \end{aligned} \tag{4.23}$$

It comes down to the following conditions:

$$\alpha + \alpha\beta\delta_i > \beta|\omega_i| \tag{4.24}$$

After showing the zero-input stability of 4.14, the input-to-state stability (ISS) of 4.14 is investigated. Applying the same transformation that was done on Ξ, on the whole of 4.14, the following equation is obtained:

$$\dot{\bar{\zeta}} = \Xi''\bar{\zeta} + \Theta(\alpha\mathbf{r} + \mathbf{v}_r) \tag{4.25}$$

where $\Theta = S^{-1}T^{-1}[1, \quad 1]^T \otimes P_{3n}$. The matrix Θ will have the following structure:

$$\Theta = \begin{bmatrix} 0[3,3n] \\ \Theta_{4:3n} \\ --- \\ 0[3,3n] \\ \Theta_{3n+4:6n} \end{bmatrix} \bar{\zeta} = \Xi''\bar{\zeta} \tag{4.26}$$

As shown in the above matrix, the first 3 rows and the rows from 3n+1 to 3n+3 become all zero. Furthermore, knowing that the digraph is supposed to be strongly connected, it has a 0 eigenvalue. Therefore:

$$j = \begin{bmatrix} 0 & \\ & j' \end{bmatrix} \tag{4.27}$$

And since $J = j \otimes I_3$, $\bar{\zeta}$ could be decomposed into two state vectors $\tilde{\zeta} \in \mathbb{R}^6 \quad and \quad \check{\zeta} \in$

$\mathbb{R}^{6n-6}$ in the following form: $\bar{\zeta} = [\tilde{\zeta}_1^T \quad \check{\zeta}_1^T \quad \check{\zeta}_2^T \quad ... \quad \check{\zeta}_{n-1}^T \quad \tilde{\zeta}_2^T \quad \check{\zeta}_n^T \quad ... \quad \check{\zeta}_{2n-2}^T]^T$. Therefore, 4.14 could be decomposed into two dynamics as described below:

$$\dot{\tilde{\zeta}} = (blockdiag(-\alpha \otimes I_3))\tilde{\zeta} \tag{4.28a}$$

$$\dot{\check{\zeta}} = \Xi'''\check{\zeta} + \Theta'(\mathbf{r}, \mathbf{v}_r) \tag{4.28b}$$

where:

$$\Xi''' = \begin{bmatrix} -\alpha I_{3(n-1)} - \alpha\beta(j' \otimes I_3) - i\beta(j' \otimes I_3) & 0 \\ 0 & -\alpha I_{3(n-1)} - \alpha\beta(j' \otimes I_3) + i\beta(j' \otimes I_3) \end{bmatrix} \tag{4.29}$$

and,

$$\Theta'(\mathbf{r}, \mathbf{v}_r) = \begin{bmatrix} \Theta_{4:3n} \\ \Theta_{3n+4:6n} \end{bmatrix} (\alpha\mathbf{r} + \mathbf{v_r}) \tag{4.30}$$

From Equation (4.28a), it is evident that the dynamics of $\tilde{\zeta}$ is stable as long as $\alpha > 0$. As far as the stability of $\check{\zeta}$, the solution to Equation (4.28b) becomes:

$$\check{\zeta}(t) = exp(t\Xi''') + \int_0^t \exp((t-\tau)\Xi''')\Theta'(\mathbf{r}, \mathbf{v}_r)\, d\tau \tag{4.31}$$

According to [85, 84], it follows that:

$$\|\check{\zeta}(t)\| \leq exp(-\underline{\lambda}t)\|\check{\zeta}(0)\| + \frac{1}{\underline{\lambda}} \sup_{0 \leq \tau \leq t} \|\Theta'(\mathbf{r}, \mathbf{v}_r)\| \tag{4.32}$$

where $\underline{\lambda}$ is a function of the measure of the matrix Ξ''' and its eigenvalues. If $\mathbf{r} and \mathbf{v_r}$ are bounded, then by choosing α and β appropriately in such a way that $\underline{\lambda}$ becomes large enough, the bound on the error dynamics will converge to zero, and hence in steady state:

$$\mathbf{y}^i = \mathbf{z}^i = \frac{1}{n}\sum_{j=1}^{n} \mathbf{r}^j \tag{4.33}$$

In the presence of cyber-attacks on the communication channels, the information coming from agent j to agent i, gets corrupted in the following way:

$$\begin{aligned} \mathbf{y}^{j'} &= \mathbf{y}^{j} + \mathbf{a_y}^{j} \\ \mathbf{z}^{j'} &= \mathbf{z}^{j} + \mathbf{a_z}^{j} \end{aligned} \tag{4.34}$$

Therefore, the consensus algorithm will take the following form:

$$\begin{aligned} \dot{\mathbf{y}}^i &= \alpha(\mathbf{r}^i - \mathbf{y}^i) - \alpha\beta \sum_{j \in N_i} (\mathbf{y}^i - \mathbf{y}^j) + \beta \sum_{j \in N_i} (\mathbf{z}^i - \mathbf{z}^j) + \mathbf{v}_r^i + \alpha\beta \sum_{j \in N_i} \mathbf{a_y}^j - \beta \sum_{j \in N_i} \mathbf{a_z}^j \\ \dot{\mathbf{z}}^i &= \alpha(\mathbf{r}^i - \mathbf{z}^i) - \beta \sum_{j \in N_i} (\mathbf{y}^i - \mathbf{y}^j) - \alpha\beta \sum_{j \in N_i} (\mathbf{z}^i - \mathbf{z}^j) + \mathbf{v}_r^i + \beta \sum_{j \in N_i} \mathbf{a_y}^j + \alpha\beta \sum_{j \in N_i} \mathbf{a_z}^j \end{aligned} \tag{4.35}$$

Without loss of generality, Equations (4.35) could be represented in a compact form in the following way:

$$\begin{aligned} \dot{\mathbf{y}} &= \alpha(\mathbf{r} - \mathbf{y}) - \alpha\beta \mathbf{L}\mathbf{y} + \beta \mathbf{L}\mathbf{z} + \mathbf{v}_r^i + \mathbf{d_y} \\ \dot{\mathbf{z}} &= \alpha(\mathbf{r} - \mathbf{y}) - \beta \mathbf{L}\mathbf{y} - \alpha\beta \mathbf{L}\mathbf{z} + \mathbf{v}_r^i + \mathbf{d_z} \end{aligned} \tag{4.36}$$

where $\mathbf{d_y}$ is the net effect of the communication attacks on the dynamics of $\mathbf{y}$, and $\mathbf{d_z}$ is the net effect of the communication attacks on the dynamics of $\mathbf{z}$.

It could be easily shown that the error dynamics in 4.14, could be represented in the following way in the presence of FDI attacks:

$$\dot{\zeta} = \Xi\zeta + [1, \quad 1]^T \otimes (P_{3n}(\alpha\mathbf{r} + \mathbf{v}_r)) + \begin{bmatrix} \mathbf{d_y} \\ \mathbf{d_z} \end{bmatrix} \tag{4.37}$$

Similar to before, the same transformation S and T are applied to 4.37 and then, the dynamics of $\bar{\zeta}$ is decomposed into two as follows:

$$\dot{\tilde{\zeta}} = (blockdiag(-\alpha \otimes I_3))\tilde{\zeta} + \begin{bmatrix} \tilde{\mathbf{d_y}} \\ \tilde{\mathbf{d_z}} \end{bmatrix} \tag{4.38a}$$

$$\dot{\check{\zeta}} = \Xi'''\check{\zeta} + \Theta'(\mathbf{r}, \mathbf{v}_r) + \begin{bmatrix} \check{\mathbf{d_y}} \\ \check{\mathbf{d_z}} \end{bmatrix} \tag{4.38b}$$

where $\tilde{\mathbf{d_y}} \in \mathbb{R}^3$ and $\tilde{\mathbf{d_z}} \in \mathbb{R}^3$ are obtained after applying the transformation. In the same manner, $\check{\mathbf{d_y}} \in \mathbb{R}^{3n-3}$ and $\check{\mathbf{d_z}} \in \mathbb{R}^{3n-3}$ are derived.

Examining the input-to-state stability of 4.38a, yields:

$$\|\check{\zeta}(t)\| \leq exp(-\alpha t)\|\check{\zeta}(0)\| + \frac{1}{\alpha}\sup_{0\leq\tau\leq t} \| \begin{bmatrix} \tilde{\mathbf{d_y}} \\ \tilde{\mathbf{d_z}} \end{bmatrix} \| \tag{4.39}$$

Therefore, if the FDI attacks are bounded, by choosing α large enough, the bound on $\check{\zeta}$ will converge to zero.

By the same token, examining the input-to-state stability of 4.38b, yields:

$$\|\check{\zeta}(t)\| \leq exp(-\underline{\lambda} t)\|\check{\zeta}(0)\| + \frac{1}{\underline{\lambda}}\sup_{0\leq\tau\leq t} (\|\Theta'(\mathbf{r}, \mathbf{v}_r)\| + \| \begin{bmatrix} \check{\mathbf{d_y}} \\ \check{\mathbf{d_z}} \end{bmatrix} \|) \tag{4.40}$$

As discussed previously, according to [84], $\underline{\lambda}$ is dependent on the measure of the matrix Ξ''' and its eigenvalues. If $\mathbf{r}, \mathbf{v_r}$ and the FDI attacks are bounded, by appropriately choosing α and β, the ultimate bound on the error dynamics decreases. Consequently, the error between the average of the reference signals and state vectors of $\mathbf{y^i}, \mathbf{z^i}$ of individual agents becomes bounded, and depending on the value of α and β, this bound could be made small, that is

$$\|\mathbf{y}^i - \frac{1}{n}\sum_{j=1}^{n}\mathbf{r}^j\| \quad and \quad \|\mathbf{z}^i - \frac{1}{n}\sum_{j=1}^{n}\mathbf{r}^j\| < M \quad \forall t > T > 0 \tag{4.41}$$

Therefore, no matter the magnitude of and the number of the FDI attacks, their effects could be attenuated.

4.4 Decentralizing the Optimization Problem

In order to develop the desired consensus algorithm that will ensure optimization and achieve the desired thrust vector, the optimization problem along with its equality constraint in 4.1 should be reframed in a decentralized manner first, that is

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^{\mathbf{n}}}{\text{minimize}} \quad & \sum_{i=1}^{n} f^i(x^i) \\ \text{subject to} \quad & T_j^1 x^1 + T_j^2 x^2 + ... + T_j^n x^n - u_{cj}, \quad j \in \{1,2,3\} \end{aligned} \tag{4.42}$$

where T_j^i is the jth row and ith column of the thruster configuration matrix, u_{cj} is the jth element of the commanded thrust vector. Since the commanded thrust vector has three directions (surge, sway and yaw), there are three rows only. The columns $1, 2, ...i, ..., n$ could be equal to or greater than three, which represents the number of thrusters. The elements inside the matrix is related to their location onboard the vessel, that is

$$T = \begin{bmatrix} T_1^1 & ... & T_1^i & ... & T_1^n \\ T_2^1 & ... & T_2^i & ... & T_2^n \\ T_3^1 & ... & T_2^i & ... & T_3^n \end{bmatrix} \tag{4.43}$$

Taking $\mathbf{f^i}(\mathbf{x^i})$ as in 4.1, the Lagrangian of this constrained optimization problem is as follows:

$$L(\mathbf{x}, \boldsymbol{\mu}) = \sum_{i=1}^{n} x^i W^i x^i + \boldsymbol{\mu}^T (Tx - \mathbf{u_c}) \tag{4.44}$$

Checking Slater's condition:

- The cost function (sum of local cost functions) is convex

- The equality constraint is affine
- Matrix T is of full row rank. Therefore, there exists at least one strictly feasible $x \in \mathbb{R}^n$ such that $T\mathbf{x} - \mathbf{u_c} = 0$

Therefore, the Slater's conditions hold. Therefore, the duality gap is zero and KKT conditions could be used to find the primal and dual optimal points of $\mathbf{x}^*$ and $\boldsymbol{\mu}^*$ [76].
Using the KKT conditions, the primal and dual optimal points $\mathbf{x}^* \in \mathbb{R}^n$ and $\boldsymbol{\mu}^* \in \mathbb{R}^3$ are:

$$\begin{aligned} 2W\mathbf{x}^* + T^T\boldsymbol{\mu} = 0 \\ T^1 x^{*^1} + ... + T^n x^{*^n} - \mathbf{u_c} = 0 \end{aligned} \tag{4.45}$$

Since the duality gap is zero, the optimal points $\mathbf{x}^* \in \mathbb{R}^n$ and $\boldsymbol{\mu}^* \in \mathbb{R}^3$ constitute the saddle points of the Lagrangian. In other words:

$$L(\mathbf{x}^*, \boldsymbol{\mu}) \leq L(\mathbf{x}^*, \boldsymbol{\mu}^*) \leq L(\mathbf{x}, \boldsymbol{\mu}^*) \tag{4.46}$$

Therefore, dynamics could be devised, according to [78], in order to ensure the asymptotic stability of the saddle points. In other words, starting from an arbitrary $\mathbf{x} \in \mathbb{R}^n$ and $\boldsymbol{\mu} \in \mathbb{R}^3$, the trajectory of $\mathbf{x}$ and $\boldsymbol{\mu}$ will converge to $\mathbf{x}^*$ and $\boldsymbol{\mu}^*$, respectively. The saddle point dynamics is as follows:

$$\dot{\mathbf{x}} = -\nabla_x L(\mathbf{x}, \boldsymbol{\mu}) = -2W^i x^{*^i} - T^{i^T}\boldsymbol{\mu} \quad i \in \{1, 2, .., n\} \tag{4.47a}$$

$$\dot{\boldsymbol{\mu}} = \nabla_\mu L(\mathbf{x}, \boldsymbol{\mu}) = T^1 x^1 + ... + T^n x^n - \mathbf{u_c} \tag{4.47b}$$

As shown in 4.51d, the dynamics of $\mathbf{x}$ could easily be decomposed into n dynamics, one for each x^i. However, it is not that straightforward to decompose the dynamics of $\boldsymbol{\mu}$ into n dynamics for each μ^i. Therefore, the strategy is to come up with an $\mathbf{r}^i$ for each agent, in a way that the virtual state vectors discussed in Section 4.3, i.e. $\mathbf{y}^i$ and $\mathbf{z}^i$, reach the average sum of those reference signals in steady state. In developing that candidate, the following factors must be taken into account:

- In order for the algorithm to be fully decentralized, each agent needs to know only its own location onboard the vessel. To put it technically, each thruster only knows its associated column in the configuration matrix. The ith agent has access to the ith column of the T matrix.

- To each agent a vector μ^i will be assigned, and they all have to reach the same value in steady state.

- At each phase only one agent has access to the u_c vector, and choosing which agent could have access, must be random. The agent is called rth agent.

Considering the above points, the candidate for the reference signal of each agent is $T^ix^i + \boldsymbol{\mu^i} - \tilde{\mathbf{u_c}}^{\mathbf{i}}$, Where $\tilde{\mathbf{u}}_c^i = \begin{cases} \mathbf{u_c} & \text{i = r} \\ 0 & i \neq r \end{cases}$

As stated before, the rth agent is the one having access to the commanded control vector. In order to examine if this choice is a viable candidate, the average sum of these reference signals in steady state will be:

$$\frac{1}{n}\sum_{i=1}^{n}(T^ix^{*^i} + \boldsymbol{\mu^*} - \tilde{\boldsymbol{u_c}}^{\boldsymbol{i}}) = \frac{1}{n}(T\mathbf{x}^* - \mathbf{u_c}) + \boldsymbol{\mu^*} = \boldsymbol{\mu^*} \tag{4.48}$$

It is shown that the average sum of the reference signal is equal to the optimal dual variable, i.e. the Lagrangian multiplier. Therefore, the goal is for the virtual variables of $\mathbf{y^i}$ and $\mathbf{z^i}$ in Section 4.3, to reach this value in steady state. Therefore, based on 4.51d, the dynamics of x^i could be devised as follows:

$$\dot{x^i} = -2W^ix^i - T^{i^T}\mathbf{z^i} \tag{4.49}$$

Here, $\boldsymbol{\mu}$ in 4.51d, has been replaced by z^i. Since both $\mathbf{y^i}$ and $\mathbf{z^i}$ are supposed to track the average of $\mathbf{r^i}$s, either of them could have been used here. But z^i will be the choice in this protocol.

Moreover, as stated before, each $\boldsymbol{\mu}^i$ will have to converge to $\boldsymbol{\mu}^*$. Moreover, as discussed earlier, each $\mathbf{y}^i$ and $\mathbf{z}^i$ will also eventually converge to $\boldsymbol{\mu}^*$. Therefore, the following dynamics could be proposed for the evolution of $\boldsymbol{\mu}^i$:

$$\dot{\boldsymbol{\mu}}^i = -\boldsymbol{\mu}^i + \mathbf{z}^i \tag{4.50}$$

4.5 Resilient Decentralized Thrust Allocation Algorithm

After presenting the resilient average consensus and decentralizing the optimization problem, the final resilient decentralized thrust allocation protocol, which is inspired by [2], alongside the dynamics of the thrusters is presented as follows:

$$\dot{\mathbf{y}}^i = \alpha(T^i x^i + \boldsymbol{\mu}^{\boldsymbol{i}} - \tilde{\mathbf{u}}^{\mathbf{i}}_{\mathbf{c}} - \mathbf{y}^i) - \alpha\beta \sum_{j\in N_i} (\mathbf{y}^i - \mathbf{y}^j) + \beta \sum_{j\in N_i} (\mathbf{z}^i - \mathbf{z}^j) \tag{4.51a}$$

$$\dot{\mathbf{z}}^i = \alpha(T^i x^i + \boldsymbol{\mu}^{\boldsymbol{i}} - \tilde{\mathbf{u}}^{\mathbf{i}}_{\mathbf{c}} - \mathbf{z}^i) - \beta \sum_{j\in N_i} (\mathbf{y}^i - \mathbf{y}^j) - \alpha\beta \sum_{j\in N_i} (\mathbf{z}^i - \mathbf{z}^j) \tag{4.51b}$$

$$\dot{\boldsymbol{\mu}}^i = -\boldsymbol{\mu}^i + \mathbf{z}^i \tag{4.51c}$$

$$\dot{x}^i = -2W^i x^i - {T^i}^T \mathbf{z^i} \tag{4.51d}$$

$$\tilde{\mathbf{u}}^i_c = \begin{cases} \mathbf{u_c} & \text{i = r} \\ 0 & i \neq r \end{cases} \tag{4.51e}$$

$$u^i_d = x^i \tag{4.51f}$$

$$\dot{u}^i = \frac{(u^i_d - u^i)}{\tau_i} \tag{4.51g}$$

In this thesis, the focus lies on examining the steady-state behavior of the controller and the thrusters. Consequently, the derivative of the reference signal has been omitted in the final protocol. Based on [2] that used the concept of multi time scale dynamics, if 4.51a and 4.51b converge faster than 4.51c and 4.51d –which could be achieved by choosing α large enough– based on the preceding discussions, every $\mathbf{y}_i$ and $\mathbf{z}_i$ will converge to $\frac{1}{n}$ $\sum_{i=1}^{n}(T^i x^i + \boldsymbol{\mu}^i - \tilde{\mathbf{u}}_c^i)$. Therefore, in fast dynamics in 4.51c, $\mathbf{z}^i$ could be replaced by $\frac{1}{n}$ $\sum_{i=1}^{n}(T^i x^i + \boldsymbol{\mu}^i - \tilde{\mathbf{u}}_c^i)$. Hence, it could be written that $\dot{\boldsymbol{\mu}}^i = \frac{1}{n}\sum_{i=1}^{n}(T^1 x^1 + ... + T^n x^n - \mathbf{u}_c)$. In a way, in this decentralized dynamics, each agent receives a copy of 4.47b. Furthermore, $\boldsymbol{\mu}^i \to \mathbf{z}^i$ could be used to arrive at $\dot{x}^i = -2W^i x^i - {T^i}^T \boldsymbol{\mu}^i$, in slow dynamics.

It is worth noting that the sum of $\tilde{\mathbf{u}}_c^i$'s should be equal to $\mathbf{u}_c$. Since one of the main objectives of the design was reducing the attack surface, and hence sending the commanded controller to only one agent, in this algorithm only one agent has a nonzero $\tilde{\mathbf{u}}_\mathbf{c}^\mathbf{i}$. The process of random selection of agents is as follows:

- Let $M = \{T_1, T_2, \ldots, T_n\}$ denote the set of all thrusters.
- Let $S \subseteq M$, where $S = \{T_{i_1}, T_{i_2}, \ldots, T_{i_k}\}$, $k \leq n$, represent the subset of thrusters that are designated to receive the control command, with k members.
- Let Δt denote the period of each time interval. After Δt, the controller sends the commanded control signal to another agent.
- Let $m \in \mathbb{N}$ denote the switching counter. By switching it is meant every time that the controller selects another agent to send the control command to.
- At the start of each switching, $r = S[\text{PRNG}(m\Delta t) \mod (k+1)]$ will be the selected thruster. Here, PRNG is the pseudo-random number generator function. $m\Delta t$ is the seed at each phase. The modular arithmatic is used to map the generated random number to the thrusters inside the set S. The result is r, which will be used in 4.51e.

The investigation now shifts to the equilibrium points of Equations (4.51) when there is no FDI cyber-attacks. The Equations (4.51a), (4.51b), (4.51c), (4.51d) could be represented in a compact form as follows:

$$\dot{\mathbf{y}} = \alpha(\Omega\mathbf{x} + \boldsymbol{\mu} - \tilde{\mathbf{U}}_{\mathbf{c}} - \mathbf{y}) - \alpha\beta\mathbf{L}\mathbf{y} + \beta\mathbf{L}\mathbf{z} \tag{4.52a}$$

$$\dot{\mathbf{z}} = \alpha(\Omega\mathbf{x} + \boldsymbol{\mu} - \tilde{\mathbf{U}}_c - \mathbf{z}) - \beta\mathbf{L}\mathbf{y} - \alpha\beta\mathbf{L}\mathbf{z} \tag{4.52b}$$

$$\dot{\boldsymbol{\mu}} = -\boldsymbol{\mu} + \mathbf{z} \tag{4.52c}$$

$$\dot{\mathbf{x}} = -2\mathbf{x}^T W - \Omega^T \mathbf{z} \tag{4.52d}$$

$$\dot{\mathbf{u}} = diag(\frac{1}{\tau_1}, \frac{1}{\tau_2}, ..., \frac{1}{\tau_n})(\mathbf{x} - \mathbf{u}) \tag{4.52e}$$

where $\Omega = diag(T^1, T^2, ...T^n)^T$. and $\tilde{\mathbf{U}}_c = [\tilde{\mathbf{u}^1}_c^T \tilde{\mathbf{u}^2}_c^T ... \tilde{\mathbf{u}^n}_c^T]^T$ Taking $\mathbf{z}^*$, $\mathbf{y}^*$, $\boldsymbol{\mu}^*$, $\mathbf{x}^*$ to be the equilibrium points of the above equations. In other words:

$$0 = \alpha(\Omega\mathbf{x}^* + \boldsymbol{\mu}^* - \tilde{\mathbf{U}}_{\mathbf{c}} - \mathbf{y}^*) - \alpha\beta\mathbf{L}\mathbf{y}^* + \beta\mathbf{L}\mathbf{z}^* \tag{4.53a}$$

$$0 = \alpha(\Omega\mathbf{x}^* + \boldsymbol{\mu}^* - \tilde{\mathbf{U}}_c - \mathbf{z}^*) - \beta\mathbf{L}\mathbf{y}^* - \alpha\beta\mathbf{L}\mathbf{z}^* \tag{4.53b}$$

$$0 = -\boldsymbol{\mu}^* + \mathbf{z}^* \tag{4.53c}$$

$$0 = -2\mathbf{x}^{*T}W - \Omega^T\mathbf{z}^* \tag{4.53d}$$

$$0 = \mathbf{x}^* - \mathbf{u}^* \tag{4.53e}$$

As previously discussed, the average consensus protocol of 4.7 that is embedded in 4.52a and 4.52b, will cause both $\mathbf{y}$ and $\mathbf{z} \to \Psi$, where Ψ is the consensus value. Therefore, $\mathbf{y}^* = \mathbf{z}^* = \Psi$. Consequently, from 4.53a and 4.53b, it follows:

$$\alpha\beta\mathbf{L}\Psi = 0 \tag{4.54}$$

Given that $rank(L) = n - 1$, and its null-space is $\mathbf{1}_n$, since $\mathbf{L} = L \otimes I_3$, then $rank(\mathbf{L}) = 3(n-1)$ and Ψ could be written as: $\Psi = \mathbf{1}_n \otimes \psi$, implying y^i and z^i of each agent will reach $\psi = [\psi^{1^T}\psi^{2^T}\psi^{3^T}]$. From 4.53c, it follows that $\boldsymbol{\mu}^* = \Psi$. Thus, every μ^i will also reach $\psi = [\psi^{1^T}\psi^{2^T}\psi^{3^T}]$. Finally, $2W^i x^{*i} T^{i^T}\psi = 0$. It is evident that $\mathbf{x}^*$ and ψ are the saddle points of the Lagrangian 4.44, and are the solution to the primal dual optimization problem. From 4.53e, $\mathbf{u}^* = \mathbf{x}^*$, implying the vector of actual thrusts of the thrusters also satisfies the primal problem 4.1: $T\mathbf{x}^* = T\mathbf{u}^* = \mathbf{u}_c$.
For more explanation on the stability of 4.51, the reader is referred to [2].

In the following, the behaviour of the consensus algorithm in presence of FDI cyber-attacks on the communication links is examined.

Since in this thesis, bounded FDI cyber-attacks are of concern, without loss of generality, the FDI attacks will be considered constant. This constant could be the upper bound or the lower bound of the cyber-attack. In the presence of FDI cyber-attacks on the communication channels, the information coming from agent j to agent i, gets corrupted in the following way:

$$\begin{aligned} \mathbf{y}^{j'} &= \mathbf{y}^j + \mathbf{a_y}^j \\ \mathbf{z}^{j'} &= \mathbf{z}^j + \mathbf{a_z}^j \end{aligned} \tag{4.55}$$

The resilient decentralized thrust allocation algorithm will take the following form:

$$\begin{aligned} \dot{\mathbf{y}}^i &= \alpha(T^i x^i + \boldsymbol{\mu}^i - \tilde{\boldsymbol{u}}_{\boldsymbol{c}}^i - \boldsymbol{y}^i) - \alpha\beta \sum_{j \in N_i} (\boldsymbol{y}^i - \boldsymbol{y}^j) + \beta \sum_{j \in N_i} (\boldsymbol{z}^i - \boldsymbol{z}^j) + \alpha\beta \sum_{j \in N_i} a_y^j - \beta \sum_{j \in N_i} a_z^j \\ \dot{\mathbf{z}}^i &= \alpha(T^i x^i + \boldsymbol{\mu}^i - \tilde{\boldsymbol{u}}_{\boldsymbol{c}}^i - \boldsymbol{z}^i) - \beta \sum_{j \in N_i} (\boldsymbol{y}^i - \boldsymbol{y}^j) - \alpha\beta \sum_{j \in N_i} (\boldsymbol{z}^i - \boldsymbol{z}^j) + \beta \sum_{j \in N_i} a_y^j + \alpha\beta \sum_{j \in N_i} a_z^j \end{aligned} \tag{4.56}$$

The final decentralized thrust allocation algorithm could be represented in the following compact form:

$$\begin{aligned} \dot{\mathbf{y}} &= \alpha(\Omega\mathbf{x} + \boldsymbol{\mu} - \tilde{\mathbf{U}}_c - \mathbf{y}) - \alpha\beta\mathbf{L}\mathbf{y} + \beta\mathbf{L}\mathbf{z} + \mathbf{d_y} \\ \dot{\mathbf{z}} &= \alpha(\Omega\mathbf{x} + \boldsymbol{\mu} - \tilde{\mathbf{U}}_c - \mathbf{y}) - \beta\mathbf{L}\mathbf{y} - \alpha\beta\mathbf{L}\mathbf{z} + \mathbf{d_z} \end{aligned} \tag{4.57}$$

The equations above are similar to the resilient dynamic average consensus protocol subject to FDI attacks. As shown in Equation (4.40), where the stability of the resilient dynamic average consensus was investigated, in the presence of FDI cyber-attacks on the communication links, with no limitation on the number of these cyber-attacks, the error between $\mathbf{y}^i$, $\mathbf{z}^i$ and the average sum of the reference signals of the agents becomes bounded by appropriately chosing the values for α and β.

Strictly speaking, based on 4.41 it follows:

$$\|\mathbf{y}^i - \frac{1}{n} \sum_{j=1}^{n} T^j x^j + \boldsymbol{\mu}^j - \tilde{\mathbf{u}}_c^j\| < M_y \quad \forall t > T > 0 \tag{4.58a}$$

$$\|\mathbf{z}^i - \frac{1}{n} \sum_{j=1}^{n} T^j x^j + \boldsymbol{\mu}^j - \tilde{\mathbf{u}}_c^j\| < M_z \quad \forall t > T > 0 \tag{4.58b}$$

From 4.51c and 4.58b, it can be inferred that:

$$\|\sum_{i=1}^{n}\dot{\boldsymbol{\mu}}^i - (T^1x^1 + ... + T^nx^n - \mathbf{u}_c)\| < M_\mu \quad \forall t > T > 0 \tag{4.59}$$

Analyzing the behaviour of the agents in steady state, from 4.59 it readily follows that:

$$\|T^1x^{*1} + ... + T^nx^{*n} - \mathbf{u}_c\| < M_\infty \tag{4.60}$$

Therefore, $\|T\mathbf{u} - \mathbf{u}_c\| < M_\infty$, and M_∞ could be made small, based on a careful selection of α and β. However, two factors must be taken into consideration:

- The ultimate bound between the commanded control, i.e. $\mathbf{u_c}$, and the actual generated force, i.e. $T\mathbf{u} = \mathbf{u}_{act}$.
- The time that it takes for the consensus algorithm to attenuate the effects of the FDI cyber-attacks, i.e. the interval between the onset of the cyber-attacks and their final effect on the generated force.

In the upcoming section, simulations will illustrate the algorithm's effectiveness and how different combinations of α and β handle the cyber-attacks.

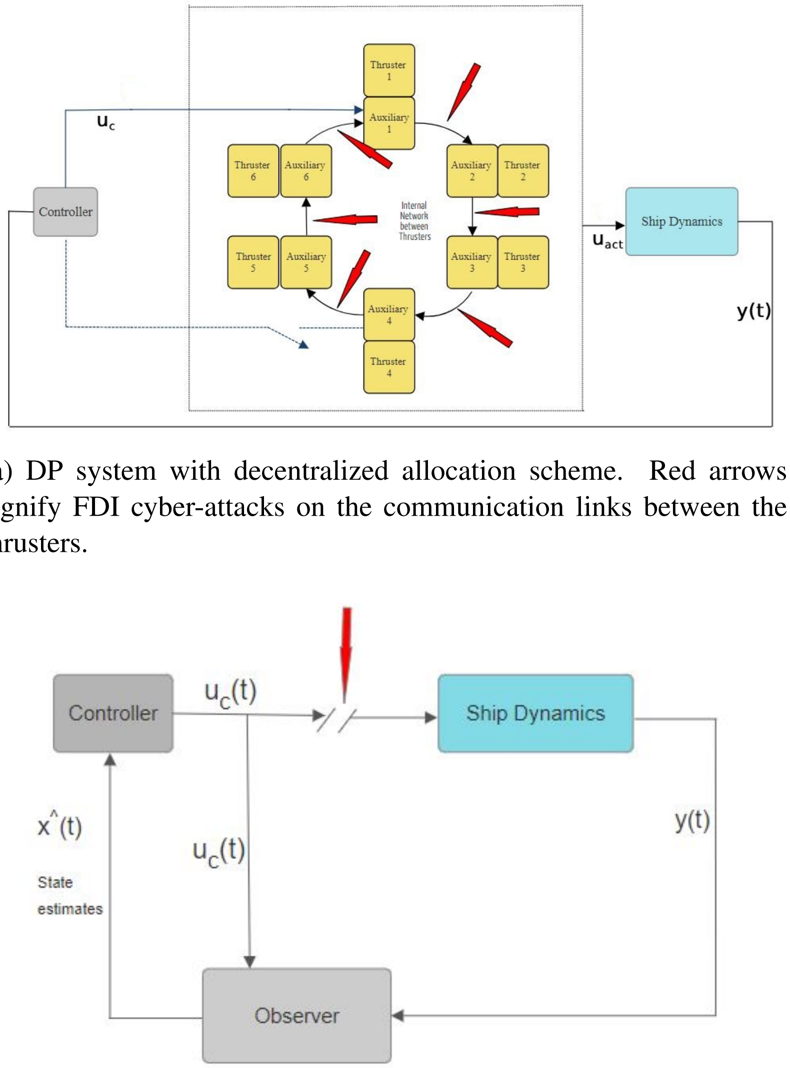
4.6 Extension of the Methodologies to Decentralized Scheme

In this section, the secure estimation, attack reconstruction, and compensation methods developed in Chapter 3 for the traditional centralized allocation scheme are extended to the decentralized thrust allocation framework. In the centralized scheme, the signature of the attacks on individual communication links from the controller to the thrusters are considered as a lumped three-dimensional attack vector $\boldsymbol{u}_a$. This vector represents the overall effect of the FDI cyber-attacks. For the decentralized scheme, the effects of attacks on the communication links between the thrusters can similarly be represented by $\boldsymbol{u}_a$. Therefore, except for the isolation procedure, the other developed methods are applicable to the decentralized scheme, specifically the compensation scheme.

As pointed out in Section 4.5, the resilient consensus protocol developed for the decentralized allocation scheme can attenuate the effects of FDI attacks on the links, regardless of their number or amplitude. However, for FDI attacks of significant amplitude, there is still some residual effect on the output, implying that the overall attack signal $\boldsymbol{u}_a$, which is correlated with the bound M_∞ mentioned in Section 4.5, is substantial. This suggests that the resilient protocol alone may not sufficiently attenuate the effect of significant amplitude FDI cyber-attacks on the communication links in the network of the thrusters.

Therefore, the methods developed in Chapter 3 can estimate the true system states, and also estimate $\boldsymbol{u}_a$ and compensate for it. By integrating the compensation scheme from the centralized framework, the decentralized system can effectively resolve the noticeable residual cyber-attack effects. This approach leverages the capabilities of both the secure estimation-compensation schemes and the resilient decentralized methods to enhance the overall resilience of the system.

Figures 4.4a and 4.4 illustrate the application of secure state estimation and Reconstruction of the cyber-attack signal $\boldsymbol{u_a}$. As shown, the red arrows in Figure 4.4a represent FDI attacks on individual communication links between the thrusters. In Figure 4.4, the overall effect of those attacks on the control command are replaced by a attack vector shown by a single red arrow representing $\boldsymbol{u_a}$.



(a) DP system with decentralized allocation scheme. Red arrows signify FDI cyber-attacks on the communication links between the thrusters.

(b) The FDI cyber-attacks in the decentralized scheme are represented by an overall attack signal $\boldsymbol{u_a}$ which is represented by the single red arrow. This attack signal is superimposed on the healthy control command $\boldsymbol{u_c}$.

Figure 4.4: Extension of the secure estimation methodologies to the decentralized scheme. In order to apply these methodologies, the FDI cyber-attacks on the communication channels between the thrusters have been omitted, and instead, their net effect on the control command, has been considered.

4.7 Simulation Results and Performance Analysis

This section is divided into four parts. First, the performance of the developed decentralized allocation algorithm is examined for a simple optimization problem with equality constraint. The simulations represent both attack free and under attack scenarios. Second, the performance of a notable decentralized allocation method existing in the literature is evaluated with the same topology and with and without the same FDI attack that was imposed on the developed resilient decentralized allocation scheme. Next, the effectiveness of the decentralized protocol operating within the DP system of a ship subject to FDI cyber-attack will be verified through simulations. The switching functionality of the algorithm will also be evaluated. Finally, some of the methodologies developed in Chapter 3 is applied to the new decentralized architecture. All simulations have been performed in Matlab-Simulink environment.

4.7.1 Decentralized Optimization

The effectiveness of the decentralized algorithm is shown for the constrained optimization problem, that is

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = \sum_{i=1}^{6} x^{i^2} \\ \text{subject to} \quad & T\mathbf{x} = \mathbf{u} \end{aligned} \tag{4.61}$$

$$T = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 & 1 & 1 \\ 3 & -4 & -35 & -33 & 40 & 28 \end{bmatrix} \tag{4.62}$$

$$\mathbf{u_c} = \begin{bmatrix} 600 \\ 400 \\ 120 \end{bmatrix} \tag{4.63}$$

The performance of the decentralized algorithm is tested against the *fmincon* solver of Matlab, which is considered to be a centralized algorithm. The *fmincon* function in MATLAB is a built-in optimization tool used for solving nonlinear constrained optimization problems. It utilizes a variety of optimization algorithms to find the minimum of a given objective function while satisfying the equality constraint. The decentralized algorithm is implemented over a digraph with a ring topology consisting of 6 agents. The configuration of this digraph is shown in the figure below. Each agent i will produce its own x^i.

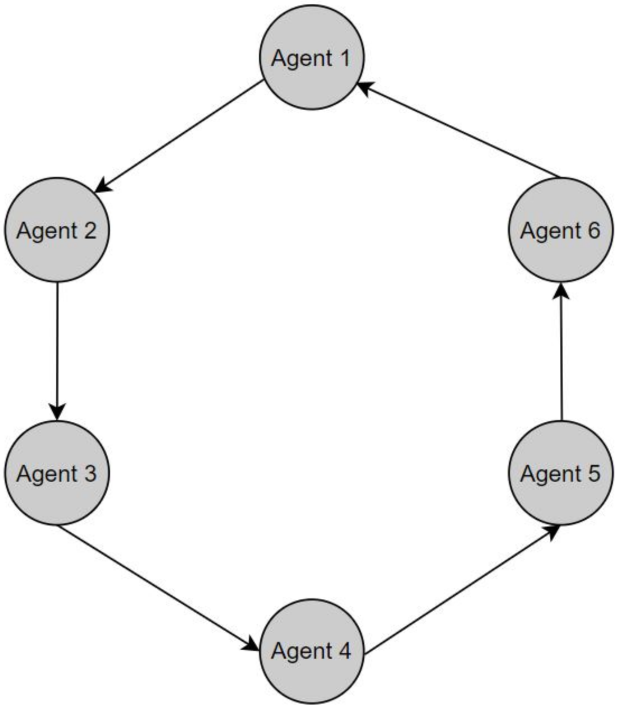
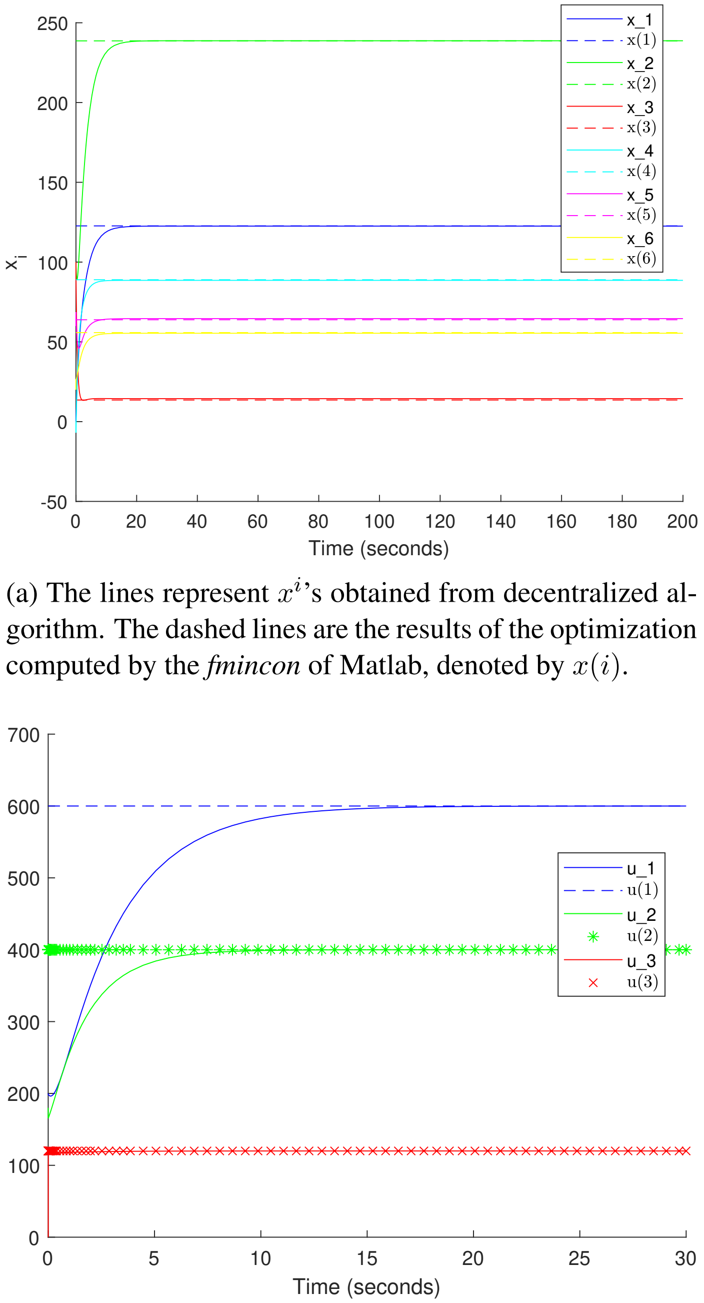


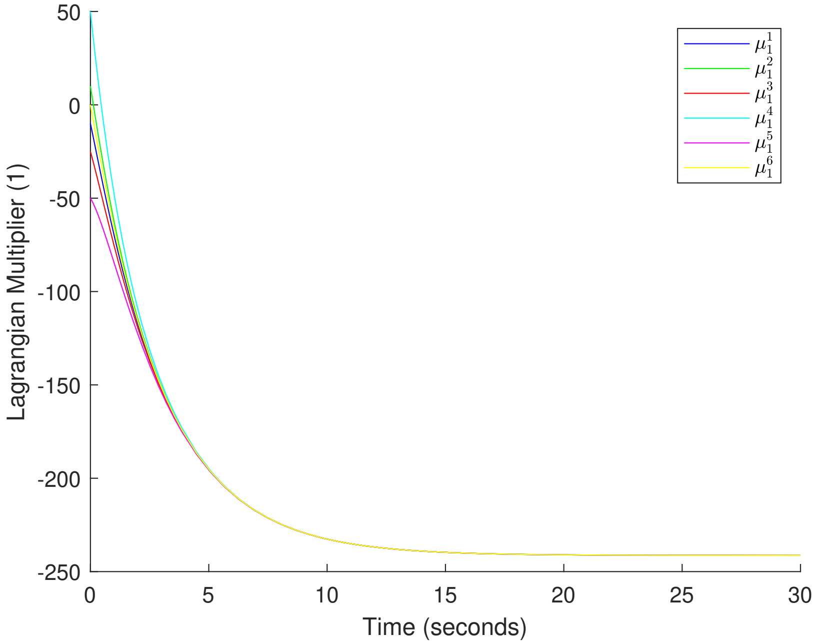
Figure 4.5: The digraph with a ring topology used in the simulations



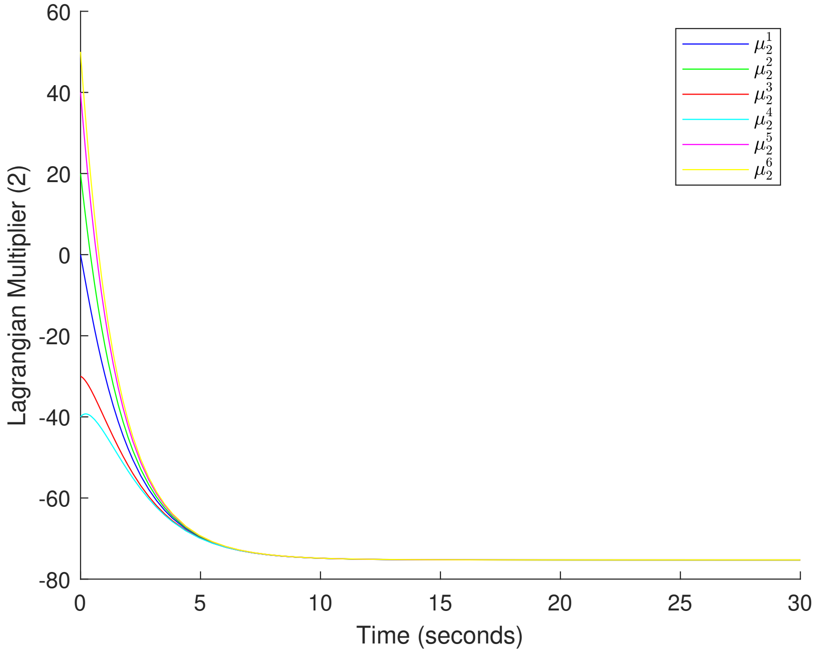
(a) The lines represent x^i's obtained from decentralized algorithm. The dashed lines are the results of the optimization computed by the *fmincon* of Matlab, denoted by $x(i)$.

(b) Satisfaction of the equality constraint. The lines represent the components of Tx over time, and the dashed lines represent the constant components of 4.61.

Figure 4.6: Results of the simulation of the decentralized algorithm 4.51a-4.51d, with $\alpha = 1100$ and $\beta = 300$. The algorithm successfully optimizes the cost function and meets the equality constraint, when there are no FDI cyber-attacks.

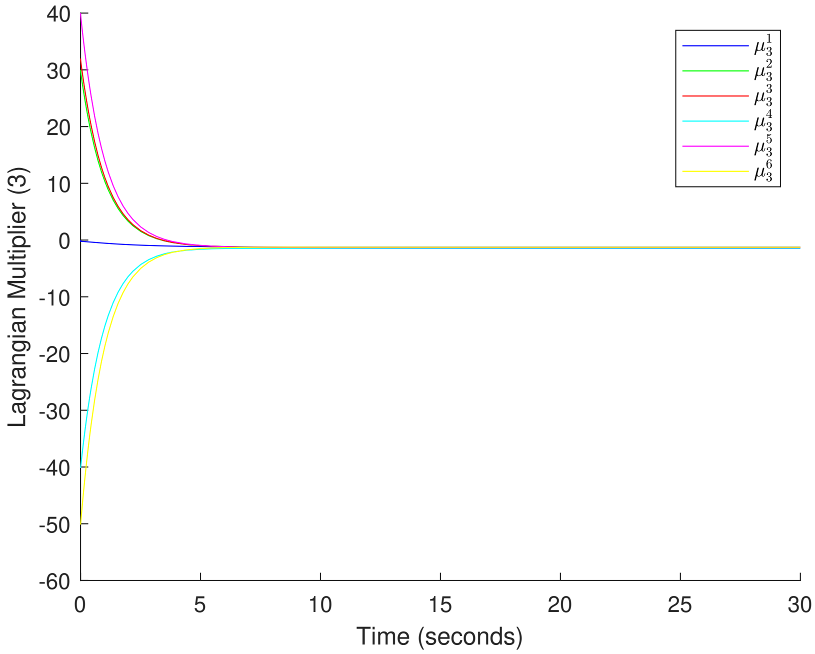


(a) First components of the Lagrangian multiplier of the agents.



(b) Second components of the Lagrangian multiplier of the agents.

Figure 4.7: Trajectories of the first two components of the Lagrangian multipliers of the agents, which converge to $\boldsymbol{\mu}^*$.



(a) Third components of the Lagrangian multiplier of the agents.

Figure 4.8: Trajectories of the third component of the Lagrangian multipliers of the agents, which converge to $\boldsymbol{\mu}^*$.

As stated in Section 4.5, in the absence of FDI cyber-attacks, every $\mathbf{y^i}$, $\mathbf{z^i}$, and $\boldsymbol{\mu}^i$ reaches consensus on a common value. That common value is $\boldsymbol{\mu}^*$, which is the answer to the dual problem of the optimization problem.

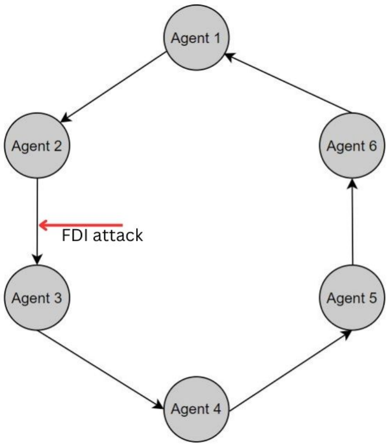


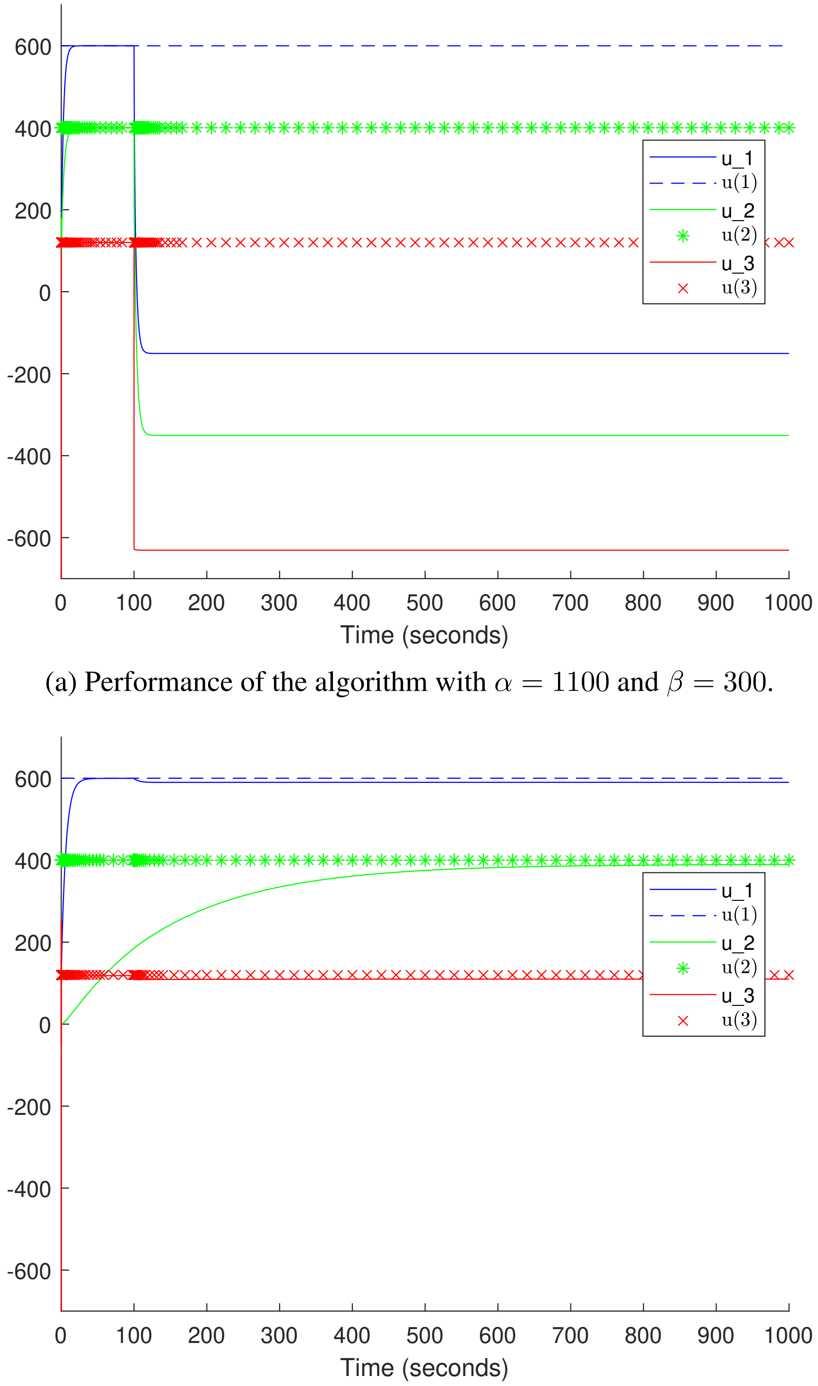
Figure 4.9: Injection of FDI attack between thruster 2 and thruster 3 in the ring topology.

In the following, the performance of the algorithm will be evaluated in the presence FDI cyber-attacks on the communication links. The communication link between agent 2 and agent 3 is compromised at time $100s$. FDI attack is modeled according to FDI model in 4.55, as a constant vector $\boldsymbol{a}_y^2 = \boldsymbol{a}_z^2 = [25 \quad 25 \quad 25]^T$ added to $\mathbf{y^2}$ and $\mathbf{z^2}$ being transmitted to agent 3, which is depicted in Figure 4.9.

In Figure 4.10a, the same set of values for α and β that were used for the attack free scenario is used for the scenario where FDI cuber-attacks have been injected at time $100s$ between agent 2 and 3. As it is shown in the figure, there is a considerable error between the desired constraint values and the resulting ones obtained from the decentralized algorithm. However, it is worth mentioning that the consensus algorithm has managed to keep the trajectories stable. In the conventional consensus algorithms, an FDI attack on the communication lines would destabilize the network and cause the variables to diverge. However, in 4.10b, by reducing the value of β, the decentralized consensus algorithm has managed to minimize the gap between the desired constraint values and the actual ones. As discussed in 4.3, the ultimate bound on $\|Tx - u\|$ depends on the choice of α and β. Therefore, the bound could be diminished by a careful selection of values for α and β. In this case, by keeping the value of α large enough, and reducing the value of β, the ultimate bound is decreased. Decreasing the value of beta leads to a reduction in error; however, this reduction is accompanied by a noticeable increase in the time required for convergence. It also must be noted that β could be decreased to a certain limit, beyond which the reduction will cause instability in the system as explained in 4.3.

Applicability to open-loop DP scenarios:

An important implication of the success of the proposed method so far is the fact that this method can be applied to open-loop DP operations, where the control signal is generated by the operator. Here, this signal is $\boldsymbol{u_c}$ and is relayed to the decentralized architecture. In the face of FDI cyber-attacks, the same control signal, with a small deviation manages



(a) Performance of the algorithm with $\alpha = 1100$ and $\beta = 300$.

(b) Performance of the algorithm with $\alpha = 1100$ and $\beta = 5$.

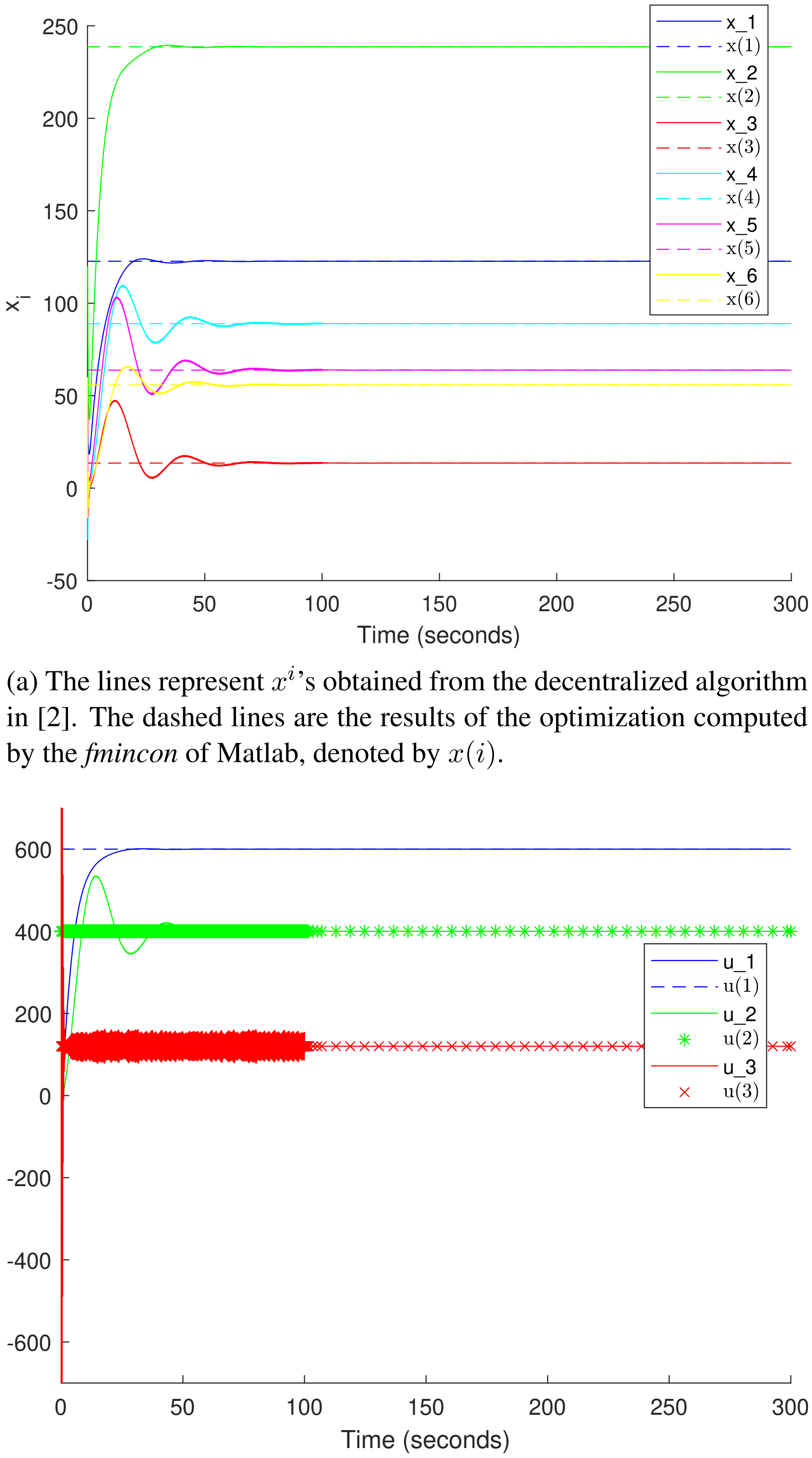
Figure 4.10: Results of the simulations of the decentralized algorithm 4.51a-4.51d, subjected to FDI cyber-attacks, for two sets of values for α and β. The lines represent the components of the resulting Tx obtained from the decentralized algorithm.

to be applied to the ship, without needing any feedback from the sensors of the ship. In the following, application of the resilient decentralized optimization method is examined in closed-loop DP system.

4.7.2 Evaluation Against Existing Decentralized Allocation Method

In this section, the performance of the decentralized optimization scheme in the work of [2], which among the decentralized allocation schemes in the literature, is the most suited for decentralized allocation schemes in over-actuated systems, is examined first without any FDI attacks and then in the presence of FDI attacks on the communication channels. The algorithm is set to solve the same constrained optimization problem as before, i.e Equations (4.61), and 4.62. The network topology is the same as the previous one in Figure 4.9, and the same FDI attack is imposed on the communication channel from agent 2 to agent 3, at time 100 s. This algorithm, as mentioned in Section 4.4, uses the same methodology as the one in this thesis, to achieve the allocation in a decentralized manner, namely saddle point dynamics and distributing the demand vector. However, as will be shown in the following simulations, it is not resilient to FDI attacks on the communication links between the agents.

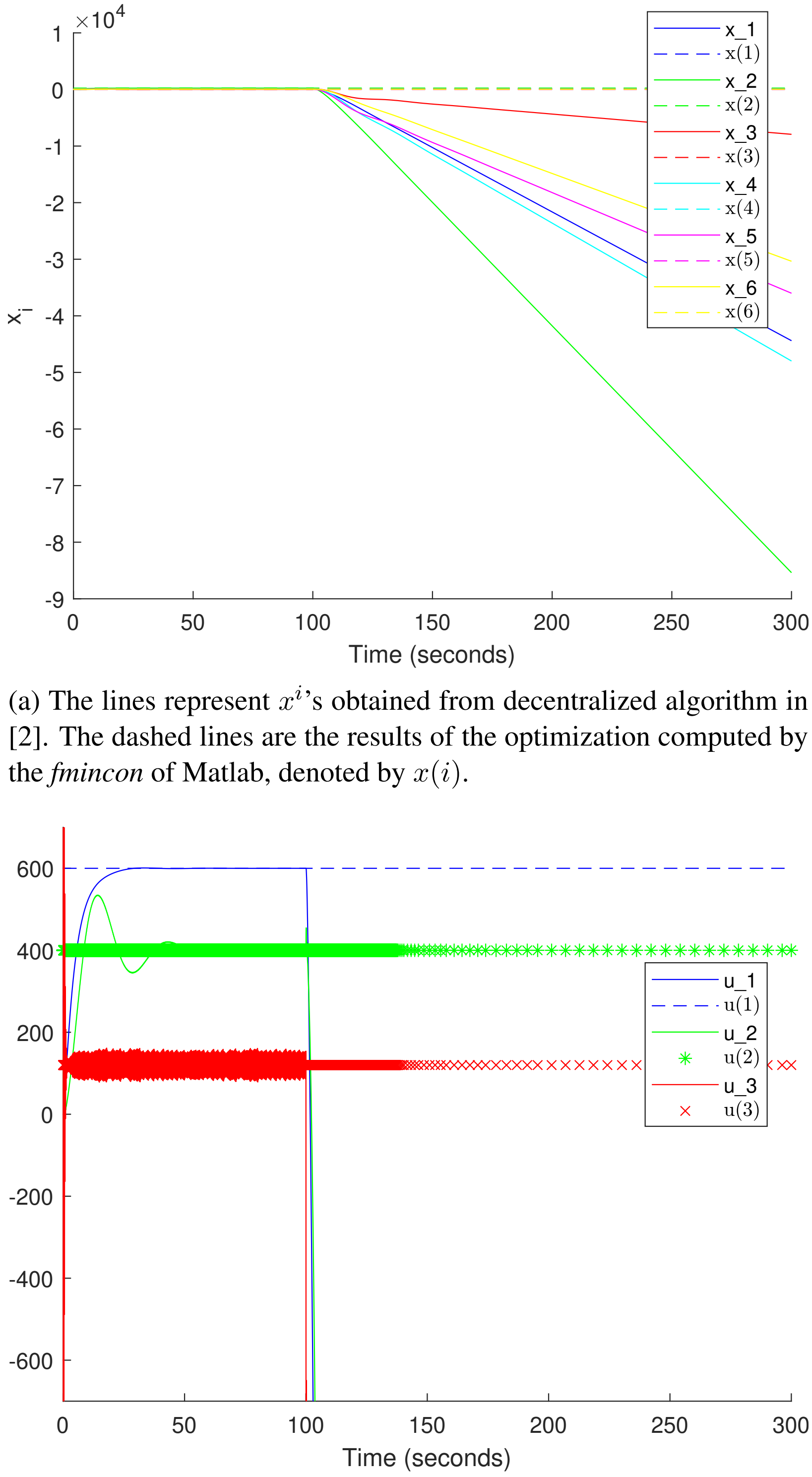
As seen in Figure 4.12, the consensus based allocation scheme provided in [2] not only does fail to meet the allocation or the equality constraint, but also becomes unstable in the presence of FDI attacks on the communication link from agent 2 to agent 3. Therefore, it is not suitable in terms of cyber-security to be implemented in an over-actuated system like DP, as the decentralized allocation protocol. However, the resilient decentralized allocation scheme managed to remain stable and meet the allocation demands when subjected to the same FDI attack, as shown in Section 4.7.1.



(a) The lines represent x^i's obtained from the decentralized algorithm in [2]. The dashed lines are the results of the optimization computed by the *fmincon* of Matlab, denoted by $x(i)$.

(b) Satisfaction of the equality constraint. The lines represent the components of Tx over time, and the dashed lines represent the constant components of 4.61.

Figure 4.11: Results of the simulation of the decentralized algorithm in [2]. The algorithm successfully optimizes the cost function and meets the equality constraint, when there are no FDI cyber-attacks.



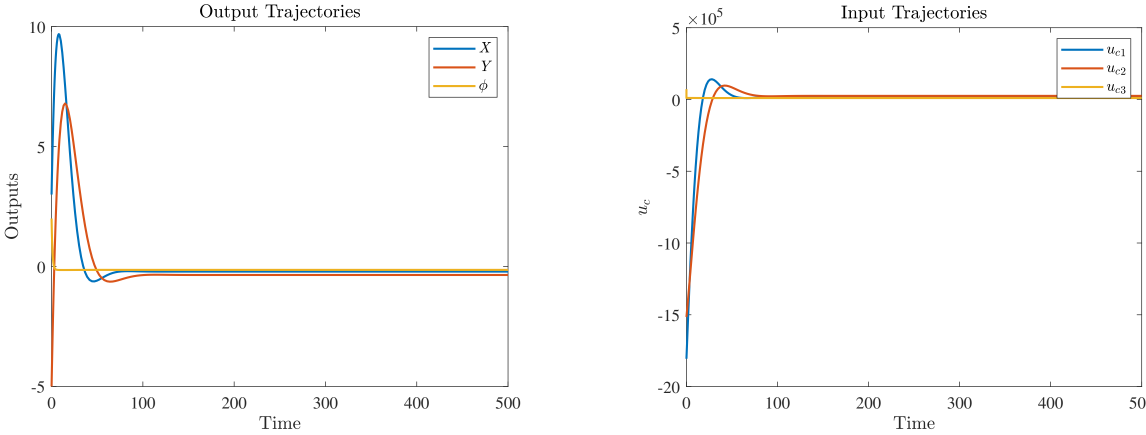
(a) The lines represent x^i's obtained from decentralized algorithm in [2]. The dashed lines are the results of the optimization computed by the *fmincon* of Matlab, denoted by $x(i)$.

(b) Failure in satisfaction of the equality constraint. The lines represent the components of Tx over time, and the dashed lines represent the constant components of 4.61.

Figure 4.12: Results of the simulation of the decentralized algorithm in [2]. The algorithm fails to satisfy the allocation, i.e., equality constraint, in the presence of an FDI attack, and the variables x^i's diverge.

4.7.3 Decentralized Thrust Allocation in DP

In this section, the decentralized algorithm is implemented in the thrust allocation of the DP of a ship. The model of the ship, the parameters of the controller and the thrust allocation module are the same as those in the previous chapter in Section 3.9, except that no observer is needed since full state feedback is considered here. First as a reference, the performance of the controller without the thrust allocation module is considered. implying that the control input $\mathbf{u_c}$ is instantly applied to the ship dynamics without going through the allocation process. As is shown in Figure 4.13, the output trajectories converge to around zero, and the input components converge to their steady state values. It should be noted that they do not converge to a same value, as it might appear from Figure 4.13 due to scaling factor.



(a) Output trajectories of the ship (X, Y, ϕ).

(b) Components of the input signal $\mathbf{u_c}$ to the ship.

Figure 4.13: Input and output trajectories of the ship dynamics in the case where there is no allocation module considered.

Next, the decentralized allocation scheme will be applied, by taking into account its switching functionality. To iterate, the purpose of the switching scheme using a pseudo-random number generator function, is to reduce and shift the attack surface. in this scenario, no attack is present. Furthermore, in this work, thruster dynamics have been neglected owing to their significantly shorter time constants compared to both the consensus network and the ship dynamics. The topology of the thrusters is the same as shown in Figure 4.5.

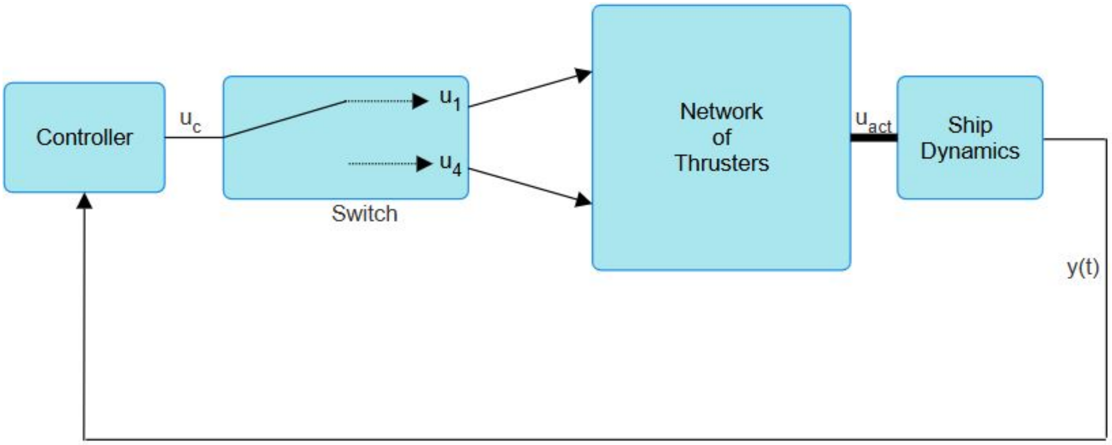
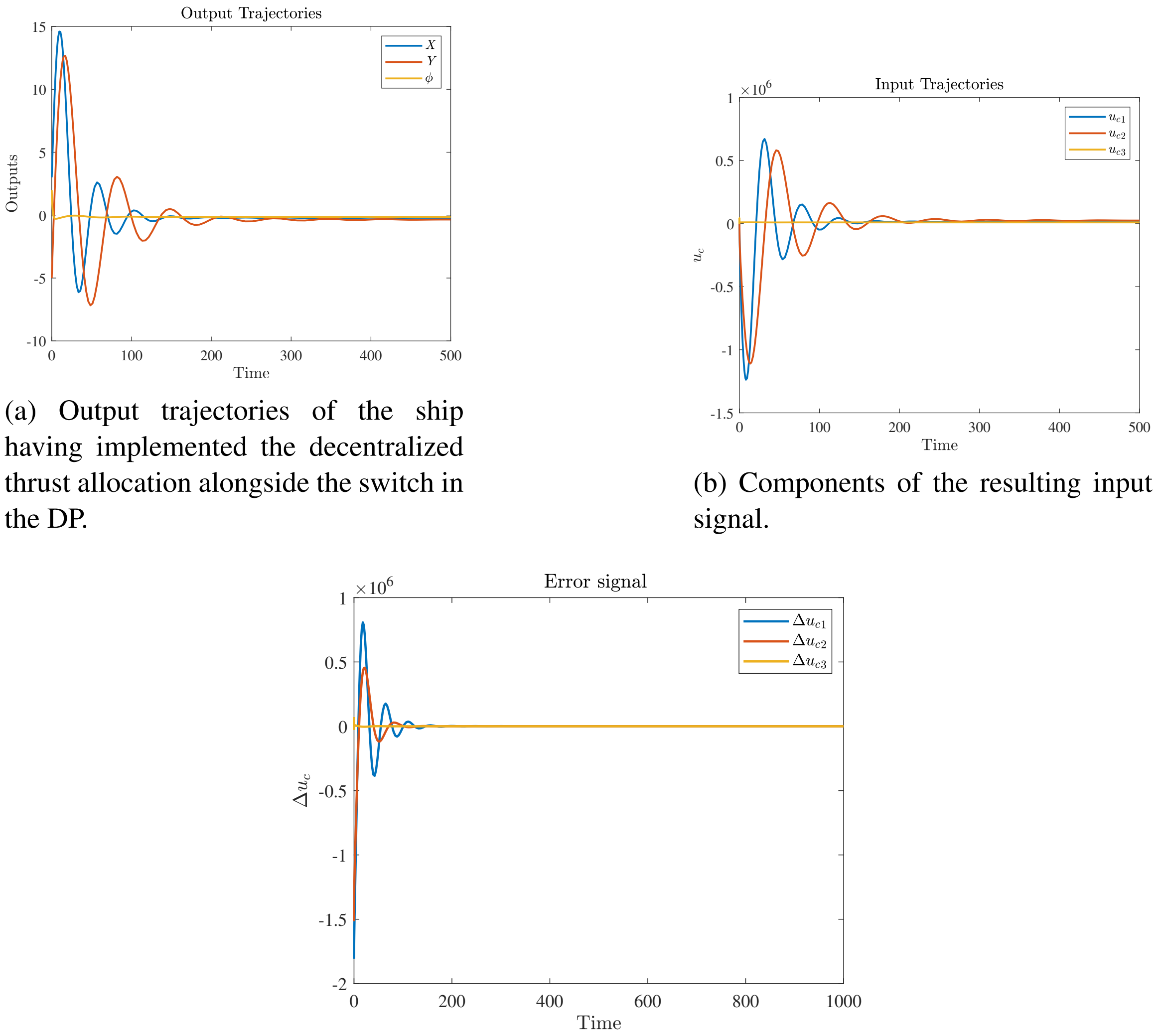


Figure 4.14: Block diagram representation of Simulink implementation of the model- The overall control system architecture.

In Figure 4.14, the configuration of the implementation of the ship dynamics and the decentralized thrust allocation module in the Simulink environment is provided. As shown in Figure 4.14, the task of the block ”switch” is to randomly switch between the thrusters inside the set S, i.e. the subset of thrusters that are designated to receive the control command. If the decentralized thrust allocation algorithm is included in the DP alongside its switching functionality, and the output trajectories converge to a small neighbourhood of zero and $\|T\mathbf{u} - \mathbf{u}_c = \mathbf{u}_{act} - \mathbf{u}_c\| \rightarrow 0$, then it can be concluded that the algorithm has managed to shift the attack surface, i.e. introduce uncertainty to the attacker and prevent it from injecting FDI cyber-attacks, while meeting the control demands. In this illustration, the set S consists of thruster 1 and thruster 4, and the switch block randomly switches between these two thrusters. The time interval of switching is taken to be $35s$, which is a little over the time it takes for the decentralized algorithm to reach steady state, according to Section4.6.



(a) Output trajectories of the ship having implemented the decentralized thrust allocation alongside the switch in the DP.

(b) Components of the resulting input signal.

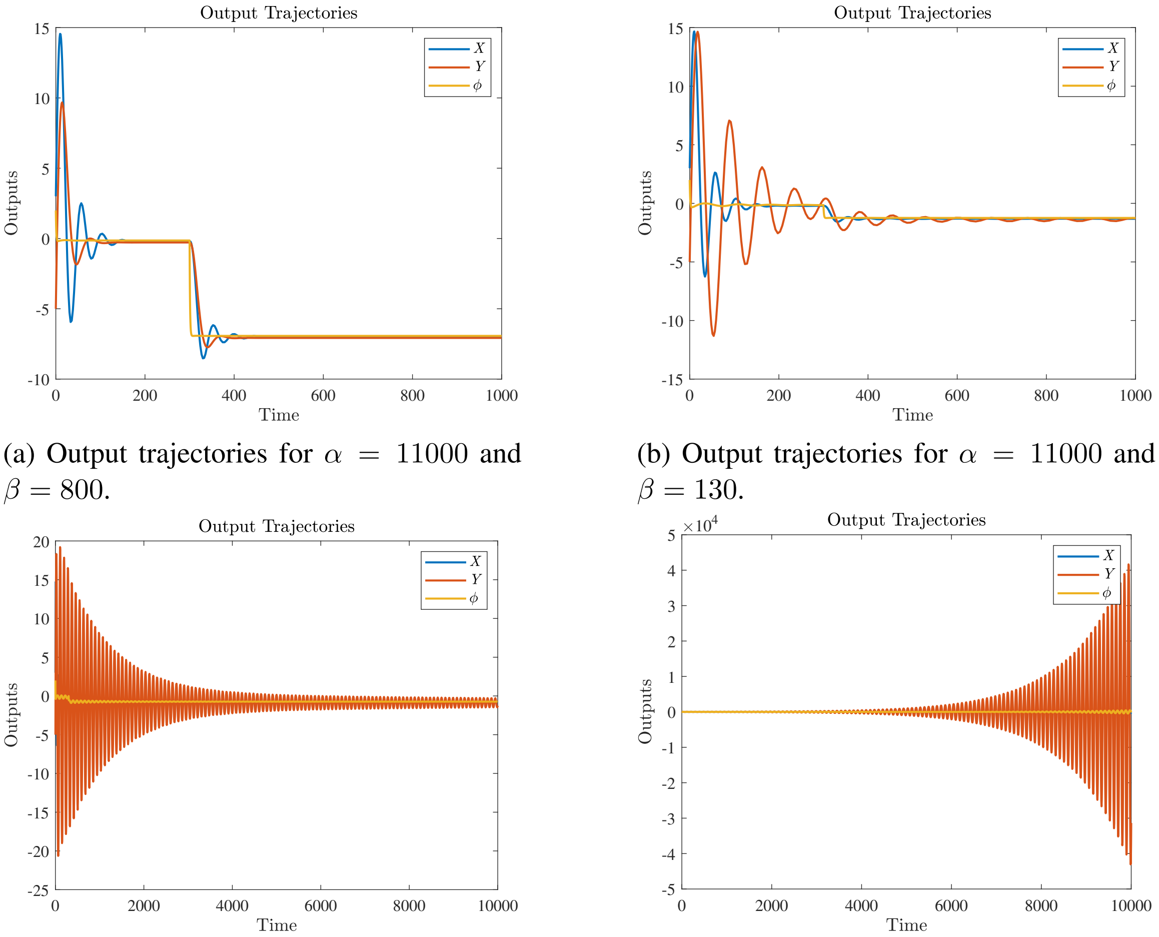
(c) Components of $\mathbf{u}_{act} - \mathbf{u_c}$.

Figure 4.15: Trajectories of the output signals, input signals and the error signals, with $\alpha = 11000, \beta = 800$.

As it is shown in Figures 4.15a, the outputs converge to around zero. Furthermore, according to Figure 4.15c, the error between the actual control input $\mathbf{u}_{act}$ that is obtained from the allocation algorithm, and the commanded control signal $\mathbf{u}_c$, converges to zero. Therefore, it could be deduced that the decentralized algorithm meets the desired operation of the system, while managing to shift the attack surface and deterring the adversaries from launching FDI cyber-attacks between the controller and the thrusters.

In the following scenario, an adversary injects FDI attacks at time 300 s, between thrusters 2 and 3 and between thrusters 4 and 5.

The additive attack signal is $\mathbf{a} = [30000 \quad 30000 \quad 30000]^T$.



(a) Output trajectories for $\alpha = 11000$ and $\beta = 800$.

(b) Output trajectories for $\alpha = 11000$ and $\beta = 130$.

(c) Output trajectories for $\alpha = 11000$ and $\beta = 70$.

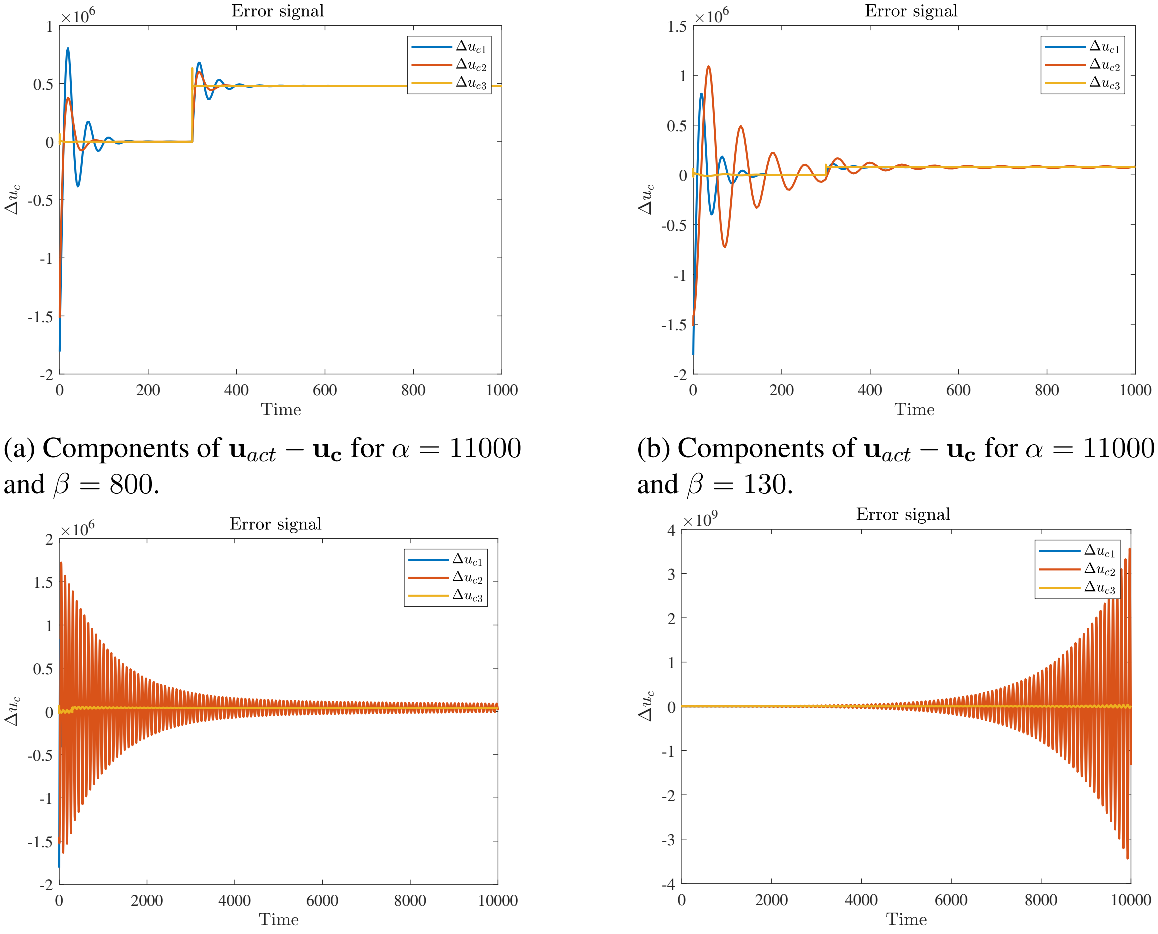
(d) Output trajectories for $\alpha = 11000$ and $\beta = 60$.

Figure 4.16: Output trajectories of the ship in the presence of FDI cyber-attacks between thruster 2 and 3, and between thruster 4 and 5 for four different pairs of α and β. The magnitude of the FDI attack signal is $[30000 \quad 30000 \quad 30000]^T$

.

As stated in Section 4.7.1, different pairs of α and β will have different attenuating effects, and there is a trade-off between how much the effects of the attacks have been attenuated, and instability and duration of convergence. As shown in Figures 4.16a-4.16d, when $\beta = 800$, although the system maintains stability, the output trajectories converge to around 7, which is far off from the operating point. Keeping α constant, when $\beta =$

130, the effects of the attacks have been attenuated, the system maintains stability and the output trajectories converge to a vicinity closer to the origin. Changing β to 70, the output trajectories converge closer to zero compared to the previous case, however the system behaves in oscillatory mode in the long term, which is not desirable. Finally, by setting β to 60, the system becomes unstable.



(a) Components of $\mathbf{u}_{act} - \mathbf{u_c}$ for $\alpha = 11000$ and $\beta = 800$.

(b) Components of $\mathbf{u}_{act} - \mathbf{u_c}$ for $\alpha = 11000$ and $\beta = 130$.

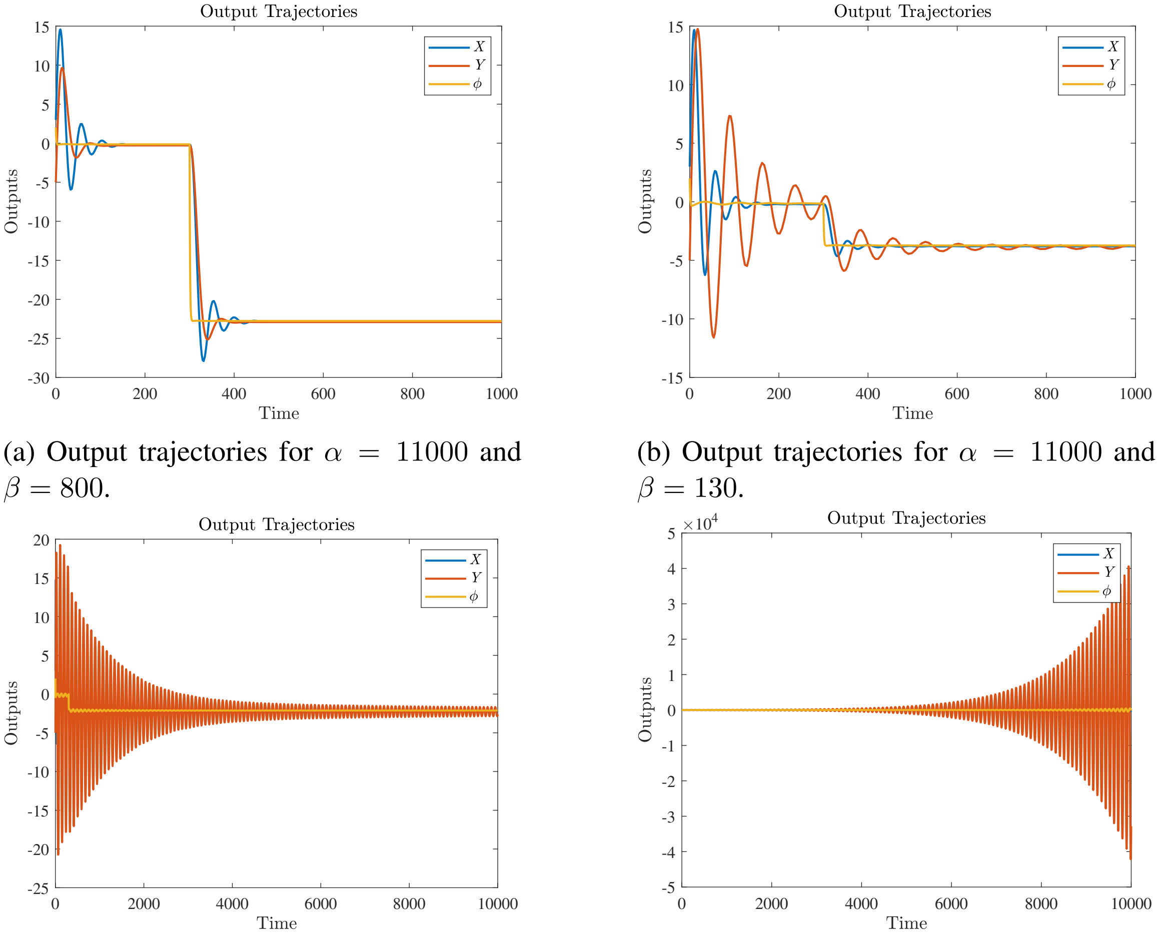
(c) Components of $\mathbf{u}_{act} - \mathbf{u_c}$ for $\alpha = 11000$ and $\beta = 70$.

(d) Components of $\mathbf{u}_{act} - \mathbf{u_c}$ for $\alpha = 11000$ and $\beta = 60$.

Figure 4.17: The difference between the resulting input signal $\mathbf{u}_{act}$ and the commanded control signal $\mathbf{u_c}$ in the presence of FDI cyber-attacks between thruster 2 and 3, and between thruster 4 and 5 for four different pairs of α and β. The magnitude of the FDI attack signal is $[30000 \quad 30000 \quad 30000]^T$.

According to Figures 4.17a-4.17d, the difference between the resulting input signal $\mathbf{u}_{act}$ and the commanded control signal $\mathbf{u_c}$ remains bounded for the first three cases, i.e. $\beta =$

$800, 130, 70$. Among these three, the best choice would be $\beta = 130$, since the difference between the resulting control signal and the commanded control signal has been minimized, while avoiding oscillatory behavior and instability. Obviously when $\beta = 60$, the system becomes unstable, therefore the value of β could only be reduced to a certain extent before the system becomes unstable. Finally, the magnitude of the FDI cyber-attack is increased to $[100000 \quad 100000 \quad 100000]^T$.



(a) Output trajectories for $\alpha = 11000$ and $\beta = 800$.

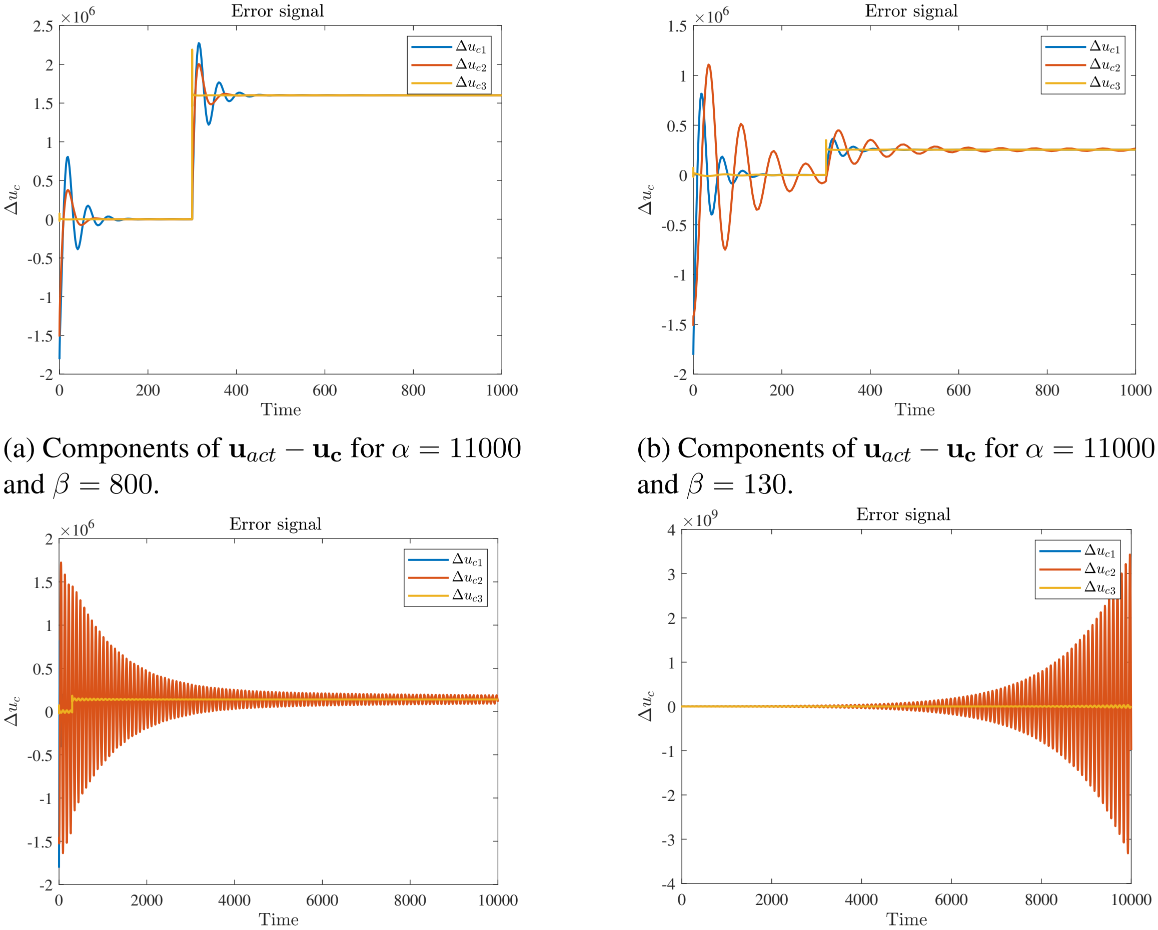
(b) Output trajectories for $\alpha = 11000$ and $\beta = 130$.

(c) Output trajectories for $\alpha = 11000$ and $\beta = 70$.

(d) Output trajectories for $\alpha = 11000$ and $\beta = 60$.

Figure 4.18: Output trajectories of the ship in the presence of FDI cyber-attacks between thruster 2 and 3, and between thruster 4 and 5 for four different pairs of α and β. The magnitude of the FDI attack signal is $[100000 \quad 100000 \quad 100000]^T$.

As shown in Figure 4.18a-4.17d, the same pattern of behaviour is observed for the case where the magnitude of the FDI cyber-attacks has been increased as the the previous case. By reducing the value of β, the magnitudes of the error signals decrease. However, as stated earlier, there is a compromise between the reduction in the error signals, and oscillatory and instability of the system.



(a) Components of $\mathbf{u}_{act} - \mathbf{u_c}$ for $\alpha = 11000$ and $\beta = 800$.

(b) Components of $\mathbf{u}_{act} - \mathbf{u_c}$ for $\alpha = 11000$ and $\beta = 130$.

(c) Components of $\mathbf{u}_{act} - \mathbf{u_c}$ for $\alpha = 11000$ and $\beta = 70$.

(d) Components of $\mathbf{u}_{act} - \mathbf{u_c}$ for $\alpha = 11000$ and $\beta = 60$.

Figure 4.19: The difference between the resulting input signal $\mathbf{u}_{act}$ and the commanded control signal $\mathbf{u_c}$ in the presence of FDI cyber-attacks between thruster 2 and 3, and between thruster 4 and 5 for four different pairs of α and β. The magnitude of the FDI attack signal is $[100000 \quad 100000 \quad 100000]^T$.

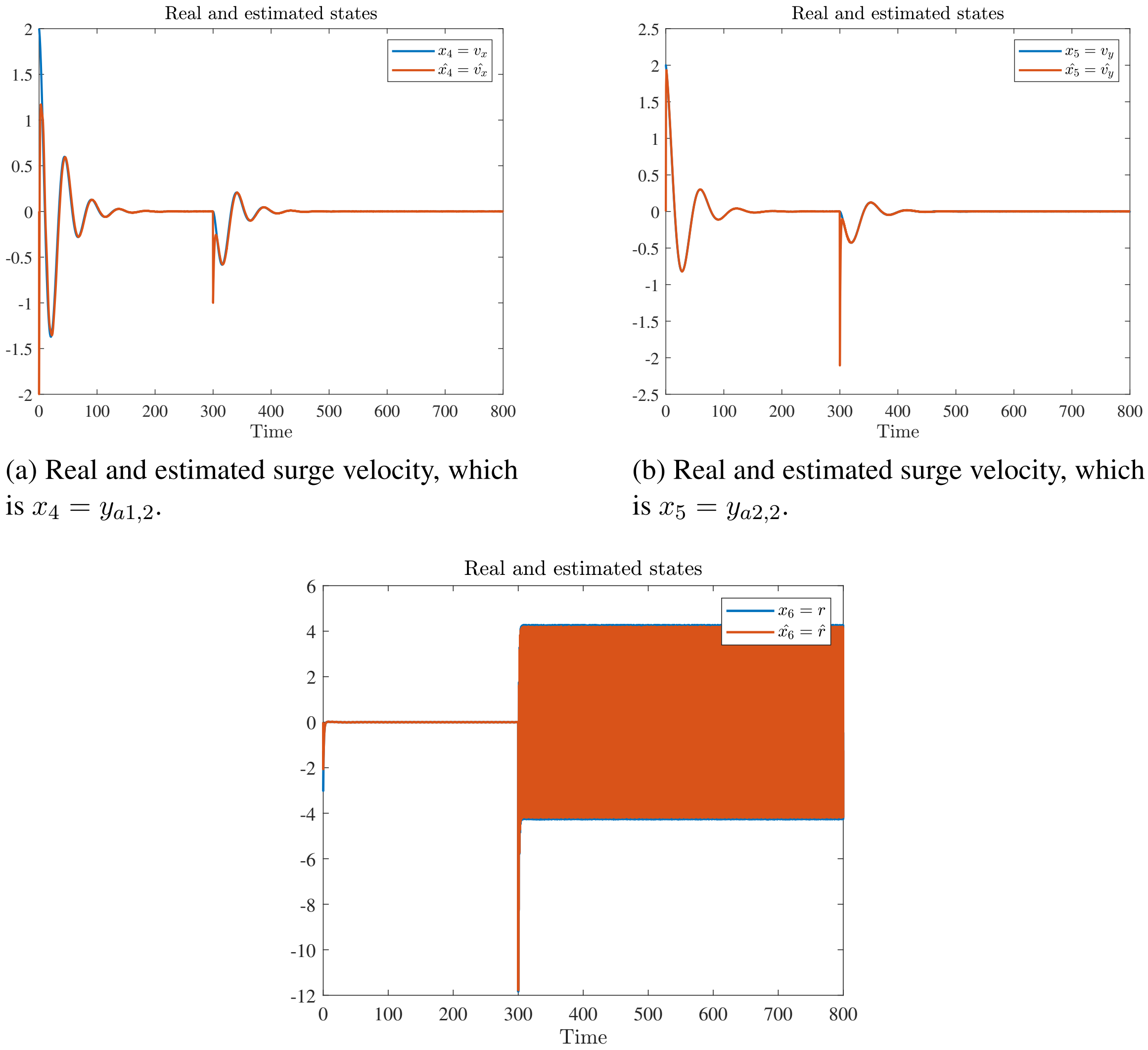
One important observation from the Figures 4.19a and 4.19b, is that effect of the FDI cyber-attacks on the communication links between the thrusters is a an additive signal to the commanded control signal $\mathbf{u}_c$. Although the effect of these attacks could be attenuated by appropriately tuning the values of α and β, as the magnitude of the attacks increases the magnitude of this additive signal also increases.

4.7.4 Application of the Secure estimation and Compensation Methodologies in the Decentralized Scheme

In this section, the performance of the secure estimation and compensation is applied to a DP system with decentralized scheme with the same specifications for the controller and the observer as in Section 3.9 and the same configuration as in Section 4.7.3, and in the presence of FDI cyber-attacks occurring in the network of thrusters. It should be mentioned that unlike in section 4.7.3, only the positions and yaw are measured, i.e. $C = [I_6 \quad 0_{6\times6}]$. The cyber-attacks compromise the communication lines between thrusters 2 and 3, and thrusters 4 and 5 at $t > 300s$. The additive attack signals are the same for every FDI cyber-attack and are modeled according to Section 4.5 and is specified by $\boldsymbol{a_y} = \boldsymbol{a_z} = [10000 + 5000sin(20t) \quad 10000 + 5000sin(20t) \quad 10000 + 5000sin(20t)]^T$.

As seen in Figures 4.20a-4.20c, the observer successfully has managed ti estimate the true system states despite the presence of unknown inputs caused by FDI cyber-attacks. It is worth mentioning that the first three states corresponding to X, Y, ψ are measured. According to Figures 4.21a-4.21c, in addition to successfully estimate the true system states while having no access to the unknown inputs, the observer also reconstructs the attack signal with a proper precision. The oscillatory cyber-attack signals cause oscillatory behavior on the third component of the input signal, and the observer rightfully reflects that. Next, the performance of the compensation scheme is examined.

In this part, the same FDI cyber-attacks are applied to the system on the communication lines between the thrusters at $t > 300s$. The cut-off frequencies for the low-pass filter have chosen to be: $w_{c1} = 2, w_{c2} = 10$, and $w_{c3} = 10$.



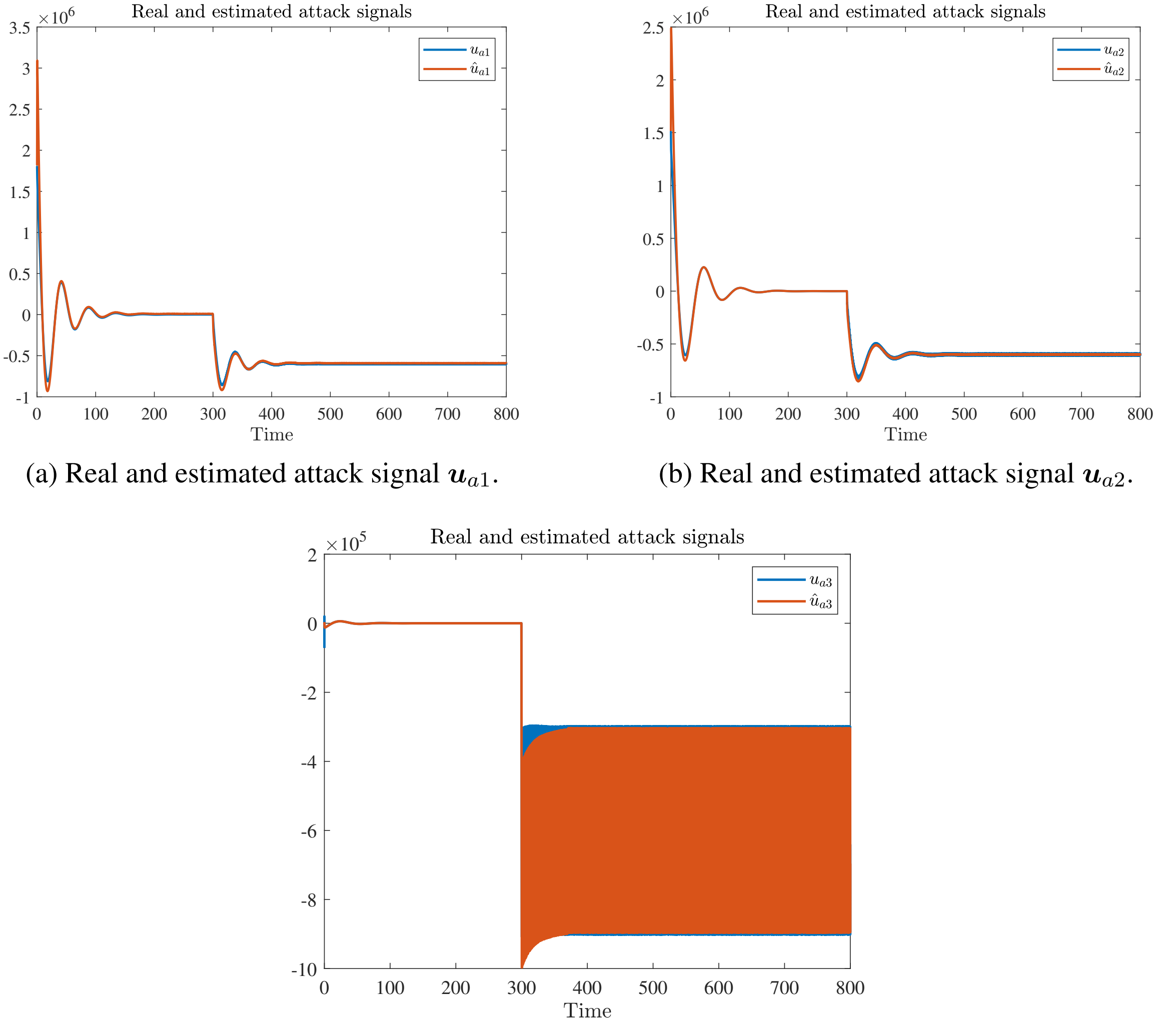
(a) Real and estimated surge velocity, which is $x_4 = y_{a1,2}$.

(b) Real and estimated surge velocity, which is $x_5 = y_{a2,2}$.

(c) Real and estimated surge velocity, which is $x_5 = y_{a3,2}$.

Figure 4.20: Estimated and real states in the presence of FDI cyber-attacks in the network of thrusters using a sliding mode observer.

The cut-off frequencies have been chosen in a such a way to handle the high-frequency behaviour of the second and third outputs. The network of thrusters that is subject to FDI cyber-attacks is running consensus algorithm as explained in Chapter 4. Here, α is chose to be 160 and β is set to 11000. Since the magnitude of the attacks are large, the resilient consensus algorithm causes the outputs to substantially deviate from their operating points, as shown in Figure4.22a.



(a) Real and estimated attack signal $\boldsymbol{u}_{a1}$.

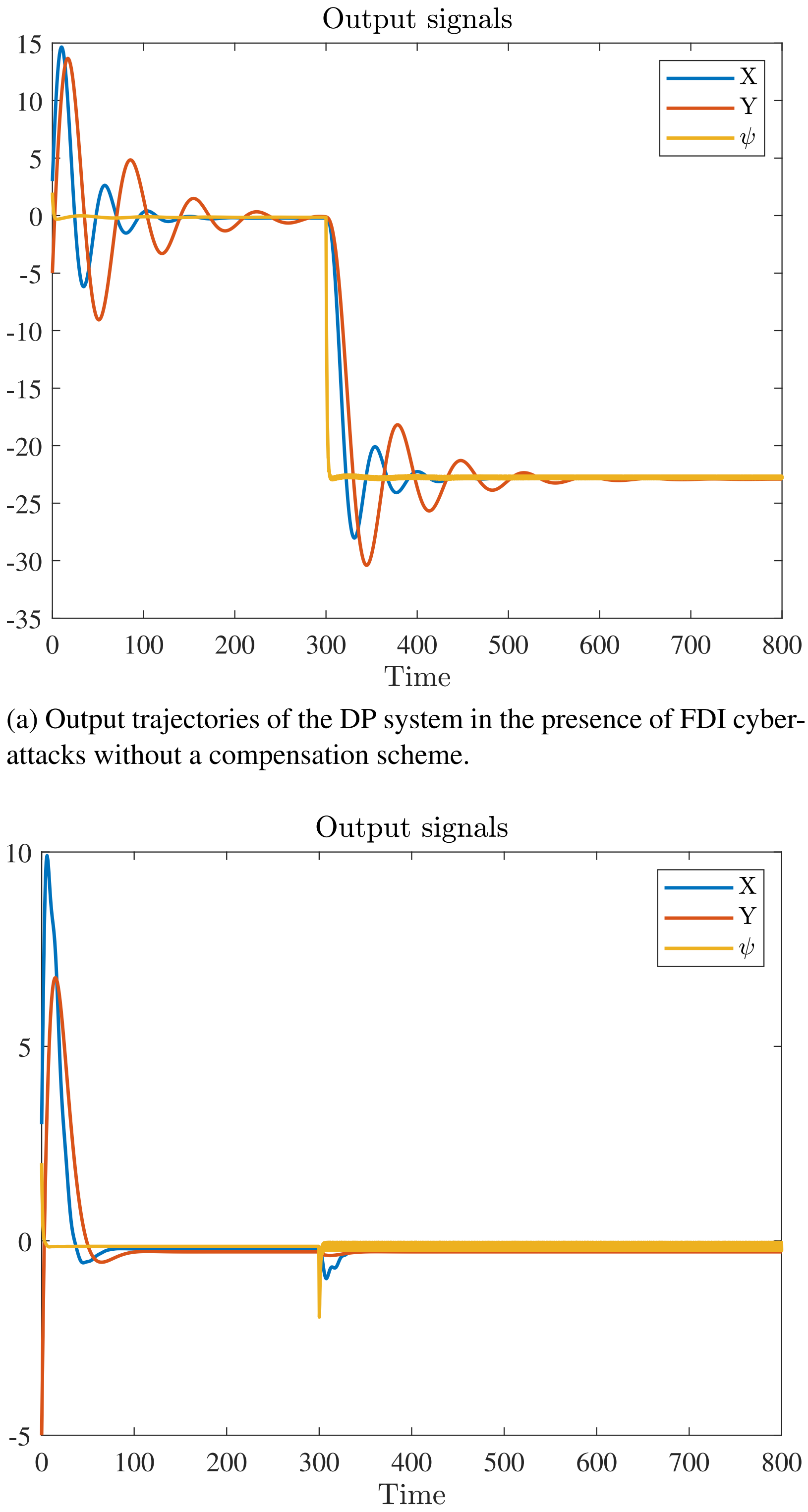
(b) Real and estimated attack signal $\boldsymbol{u}_{a2}$.

(c) Real and estimated attack signal $\boldsymbol{u}_{a3}$.

Figure 4.21: Estimated and real attack signal which signifies the effect of the FDI cyber-attacks between thrusters 2,3 and thrusters 4,5 at $t > 300s$.

However, after implementing the compensation scheme as provided in Section 3.8, the output trajectories remain close to the operating point. As shown in Figure 4.22b, the compensation method successfully attenuates the effect of multiple FDI cyber-attacks that the resilient consensus algorithm between the thrusters was unable to attenuate sufficiently.

Moreover, a few other points are worthy of discussion. First, the compensation scheme is always operating, even when there are no cyber-attack present in the system, since there are no detection algorithms employed. Therefore, this compensation algorithm is always



(a) Output trajectories of the DP system in the presence of FDI cyber-attacks without a compensation scheme.

(b) Output trajectories of the DP system in the presence of FDI cyber-attacks equipped with a compensation scheme.

Figure 4.22: Compensation of the FDI attack signals on the communication links between the thrusters.

accounting for discrepancies between the generated control command at the command and control center caused by disturbances and also by the decentralized allocation , i.e. the consensus algorithm. As demonstrated in Figure 4.22b, even before the launching of the FDI cyber-attacks, the compensation algorithm is causing a difference in the output trajectories compared to the case where there is co compensate. The reason is that the compensator is removing any discrepancies caused by the decentralized allocation algorithm, since this allocation algorithm takes some time to reach steady state and exactly follow the commanded control signal.

Second, even if the compensation scheme is always under operation, its major contribution is to account for the large FDI cyber-attacks, since the LQR controller does a good job attenuating the disturbances. Furthermore, in many DP applications, thanks to the wind sensors and observers, feed-forward control laws could also be incorporated in order to account for the effect of disturbances. Hence, the main benefit of the compensation scheme is to attenuate the effect of FDI cyber-attacks. The compensator does a good job at both reducing the oscillatory behaviour of the system and also removing any deviation from the desired operating points.

4.8 Conclusion

In this chapter, a decentralized thrust allocation module was introduced as a security measure for over-actuated systems, specifically DP systems of ships, against FDI cyber-attacks. This decentralized scheme aims at reducing the attack surface from the controller to the thrusters, and shifting the attack surface periodically in order to deter the adversary from executing FDI attacks between the controller and the thrusters. Furthermore, since this decentralized allocation scheme relies on a consensus network of thrusters, it would be susceptible to FDI attacks. Therefore, the consensus algorithm must have the minimum number of communication lines and also be resilient to FDI cyber-attacks on the communication links. This objective was achieved by taking advantage of the concepts of distributed optimization and resilient average tracking consensus. Simulations were performed on a linearized model of a DP system of a ship for various scenarios. Five major conclusions have been drawn.

- The switching functionality of the decentralized algorithm, which sends the control command randomly to one thrusters in each time interval among a subset of thrusters, proved to operate successfully. This functionality introduced uncertainty for the attackers and could prevent them from successfully launching an FDI cyber-attack on the communication links between the controller and the thrusters.

- By appropriately tuning the values of α and β according to stability conditions achieved in this section, the effects of any number of FDI cyber-attacks on the communication lines could be attenuated, and the output variables of the DP system do not deviate from the operating points substantially. Therefore, the decentralized thrust allocation has achieved resiliency.

- By keeping the value of α large enough, the deciding factor in attenuating the effect of the FDI cyber-attacks is the value of β. The simulations have shown that reducing the value of β will lead to a better attenuation. However, reducing the value of β beyond a certain limit will cause oscillatory behavior for the overall system, which is not desirable. Moreover, further reducing the value of β will lead to instability of the DP system.

- No matter the number of the FDI cyber-attacks on the communication links between the thrusters, their effect will eventually be an additive signal onto the command control signal $\mathbf{u_c}$. Therefore, if the magnitude of these attacks becomes too large, this additive signal becomes considerable and cannot be sufficiently attenuated by solely tuning α and β, necessitating the development of a mechanism to eliminate this additive signal.

- Using the existing consensus protocols that are suitable for decentralized optimization, as the decentralized thrust allocation scheme of the DP system, would not be a desirable choice as FDI cyber-attacks on the communication links between the thrusters make the whole DP system unstable.
- The resilient Decentralized architecture could also be used in open-loop scenarios and be effective in the presence of FDI cyber-attacks.
- The methodologies developed in Chapter 3, namely secure estimation, overall attack signal reconstruction and compensation, were applied to the DP system in conjunction with the resilient decentralized allocation framework developed in this chapter, to securely estimate the true system states, reconstruct the effect of the FDI attacks on the communication channels between thrusters and compensate for their effect.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

This thesis focused on the problem of the cyber-security of over-actuated systems, while considering the DP system of surface vessels as the benchmark, as it is the foremost example of over-actuated cyber-physical systems. The motivations for this work is addressing some of the unresolved problems in cyber-security of over-actuated cyber-physical system, i.e. attack surface reduction/shifting, resiliency, secure estimation, isolation and compensation.

The motivation for Chapter 3 is twofold. First, the estimation and isolation schemes in centralized allocation frameworks in over-actuated systems rely on the strict and somehow unrealistic condition in DP systems that is associated with the observer matching condition, which itself is related to the input and output matrices of the system. Moreover, most of the existing mitigation approaches rely on excluding the corrupted channels and reallocating the control signal after the isolation. The problem with this approach is the possible loss of controllability of the overall system. Therefore, efforts were made to construct an observer to estimate the system's true states in the face of FDI cyber-attacks in the communication links. To tackle the observer matching condition issue, in this work, a different approach should have been made. Using the concept of vector relative degree, the outputs of the system were augmented with virtual variables. The new output vector was then estimated using higher order sliding mode observer. Subsequently, the system states were extracted from the estimated augmented output vector. Then, the affected communication channels were isolated, the overall attack signal system was estimated and the real attack signals on the affected channels were reconstructed based on the isolation scheme and the estimated overall attack signal. Furthermore, using the estimated overall attack signal and low-pass filters, the effect of the FDI attacks were compensated. This compensation scheme does not suffer from the limitations of the exclusion based methods described earlier.

Chapter 4 addresses critical gaps in the literature on the cyber-security of over-actuated systems from a control-theoretic perspective, specifically focusing on reducing and shifting the attack surface. This work responds to the severe vulnerabilities inherent in these systems, where the communication links between the controller and thrusters, which are abundant, are exposed to potentially compromising FDI cyber-attacks. If an adversary could compromise every single communication lines, they can wield too much control over the DP system and could impose catastrophic consequences. Therefore, the need for altering this conventional allocation module was imperative. A new allocation framework was developed to reduce and shift the attack surface. This framework is a decentralized allocation scheme, in which the controller only sends the commanded control vector to one thruster, which it chooses randomly and periodically. The thrusters, through a consensus network, collectively reach that commanded control force by sharing only their own information with their neighbouring thrusters. The topology of the thrusters have the minimum number of communication links, so as to not add additional vulnerabilities to the system.

This new decentralized thrust allocation module is also resilient to FDI cyber-attacks. Meaning that if an adversary was prevented from launching attacks on the communication line between the thruster and the controller, they could still compromise the internal network of the thrusters, through injecting bias signals onto the communication channels. The adversaries could inject limitless attacks with large magnitudes, however bounded, and the resilient algorithm managed to diminish its effects. Not to mention that this architecture proved to be successful even in open-loop operation modes of DP. However, as observed in Chapter 4, the resilient protocol could only attenuate the large attacks to some degree. Although it managed to maintain the system's stability at all times, attacks with large amplitudes could not be sufficiently attenuated.

Finally, the effect of the FDI cyber-attacks that had not been sufficiently attenuated by the resilient protocol, were estimated and compensated for by the compensation strategy developed in Chapter 3.

One noteworthy conclusion of Chapter 4 is the fact that the resilient decentralized architecture could be developed independently from the controller design in a modular way and can be applied to a wide variety of control system that have multiple actuators, just by changing the allocation scheme.

The methods in Chapter 3 and 4, collectively managed to maintain the operational normalcy and state awareness of the system in the face of FDI cyber-attack and equip over-actuated CPSs with security measures that address the gaps in the literature.

5.2 Future Directions

- The resilient algorithm developed in this work could not account for all FDI attacks. Some of them has residue effects on the overall system behaviour. Therefore, employing other resilient methodologies to the decentralized allocation algorithm will be beneficial.

- The parameters of the resilient algorithm of the decentralized allocation scheme were not designed in conjunction with the controller of the DP. This is why unnecessary oscillatory behavior and instabilities were observed for some values of α and β. It is recommended that in future works, these two modules be designed simultaneously.

- The decentralized allocation algorithm in this thesis did not consider thruster limitations and saturation. Therefore, future works must focus on developing decentralized allocation algorithm by taking into account the inequality constraints as well.
- The isolation scheme could be improved to isolate more simultaneous FDI attacks on the communication channels between the controller and the thrusters, without meeting the observer matching condition.
- In this work, it was assumed that the measurements from the GPS and gyro-compass were safe and not susceptible to FDI cyber-attacks. For a more comprehensive security measure for over-actuated systems, sensor attacks should also be accounted for.
- This decentralized allocation scheme in conjunction with the compensation strategy could be used for other over-actuated systems as well, including ROVs, spacecrafts, etc.

Bibliography

[1] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *Proceedings of the 1st international conference on High Confidence Networked Systems*, pp. 55–64, 2012.

[2] S. S. Kia, "Distributed optimal in-network resource allocation algorithm design via a control theoretic approach," *Systems & Control Letters*, vol. 107, pp. 49–57, 2017.

[3] T. C. Yang, "Networked control system: a brief survey," *IEE Proceedings-Control Theory and Applications*, vol. 153, no. 4, pp. 403–412, 2006.

[4] Y. Tipsuwan and M.-Y. Chow, "Control methodologies in networked control systems," *Control engineering practice*, vol. 11, no. 10, pp. 1099–1111, 2003.

[5] M. Taheri, K. Khorasani, I. Shames, and N. Meskin, "Cyberattack and machine-induced fault detection and isolation methodologies for cyber-physical systems," *IEEE Transactions on Control Systems Technology*, 2023.

[6] Y. Yin, L. Xia, L. Song, and Z. Ren, "The ship ipms networked control system modelling and design," *International Journal of Modelling, Identification and Control*, vol. 20, no. 3, pp. 234–241, 2013.

[7] T. I. Fossen, "Guidance and control of ocean vehicles," *University of Trondheim, Norway, Printed by John Wiley & Sons, Chichester, England, ISBN: 0 471 94113 1, Doctors Thesis*, 1999.

[8] F. Benetazzo, G. Ippoliti, S. Longhi, and P. Raspa, "Advanced control for fault-tolerant dynamic positioning of an offshore supply vessel," *Ocean Engineering*, vol. 106, pp. 472–484, 2015.

[9] S. Islam, D. Watson, J. Brown, and M. Sayed, "Modeling of a full scale dp in ice scenario using an advanced ice dynamics model," in *Proceedings of the International Conference on Port and Ocean Engineering Under Arctic Conditions*, 2019.

[10] E. C. De Souza and N. Maruyama, "Intelligent uuvs: Some issues on rov dynamic positioning," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 1, pp. 214–226, 2007.

[11] S. Xu, M. Murai, X. Wang, and K. Takahashi, "A novel conceptual design of a dynamically positioned floating wind turbine," *Ocean Engineering*, vol. 221, p. 108528, 2021.

[12] A. J. Sørensen, "Marine control systems," *Propulsion and Motion Control of Ships and Ocean Structures*, vol. 3, 2013.

[13] F. Mauro and R. Nabergoj, "Advantages and disadvantages of thruster allocation procedures in preliminary dynamic positioning predictions," *Ocean Engineering*, vol. 123, pp. 96–102, 2016.

[14] L. Zhang, X. Peng, N. Wei, Z. Liu, C. Liu, and F. Wang, "A thrust allocation method for dp vessels equipped with rudders," *Ocean Engineering*, vol. 285, p. 115342, 2023.

[15] M. H. Larsen and M. S. Lund, "Cyber risk perception in the maritime domain: a systematic literature review," *IEEE Access*, vol. 9, pp. 144895–144905, 2021.

[16] C. Baraniuk, "How hackers are targeting the shipping industry. bbc news (aug 18, 2017)."

[17] L. Fillatre, I. Nikiforov, P. Willett, *et al.*, "Security of scada systems against cyber–physical attacks," *IEEE Aerospace and Electronic Systems Magazine*, vol. 32, no. 5, pp. 28–45, 2017.

[18] K. Tam and K. Jones, "Situational awareness: Examining factors that affect cyber-risks in the maritime sector," 2019.

[19] B. Svilicic, M. Kristić, S. Žuškin, and D. Brčić, "Paperless ship navigation: cyber security weaknesses," *Journal of Transportation Security*, vol. 13, pp. 203–214, 2020.

[20] P. H. Meland, K. Bernsmed, E. Wille, Ø. J. Rødseth, and D. A. Nesheim, "A retrospective analysis of maritime cyber security incidents," 2021.

[21] E. Miehling, M. Rasouli, and D. Teneketzis, "Control-theoretic approaches to cyber-security," *Adversarial and Uncertain Reasoning for Adaptive Cyber Defense: Control-and Game-Theoretic Approaches to Cyber Security*, pp. 12–28, 2019.

[22] A. Cetinkaya, H. Ishii, and T. Hayakawa, "An overview on denial-of-service attacks in control systems: Attack models and security analyses," *Entropy*, vol. 21, no. 2, p. 210, 2019.

[23] M. Long, C.-H. Wu, and J. Y. Hung, "Denial of service attacks on network-based control systems: impact and mitigation," *IEEE Transactions on Industrial Informatics*, vol. 1, no. 2, pp. 85–96, 2005.

[24] A. Sargolzaei, K. Yazdani, A. Abbaspour, C. D. Crane III, and W. E. Dixon, "Detection and mitigation of false data injection attacks in networked control systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4281–4292, 2019.

[25] A. Hoehn and P. Zhang, "Detection of replay attacks in cyber-physical systems," in *2016 American Control Conference (ACC)*, pp. 290–295, IEEE, 2016.

[26] I. Jovanov and M. Pajic, "Relaxing integrity requirements for attack-resilient cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 12, pp. 4843–4858, 2019.

[27] A. Eslami, F. Abdollahi, and K. Khorasani, "Stochastic fault and cyber-attack detection and consensus control in multi-agent systems," *International Journal of Control*, vol. 95, no. 9, pp. 2379–2397, 2022.

[28] J. Kim, C. Lee, H. Shim, J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Encrypting controller using fully homomorphic encryption for security of cyber-physical systems," *IFAC-PapersOnLine*, vol. 49, no. 22, pp. 175–180, 2016.

[29] C. Gao, Z. Wang, X. He, and H. Dong, "Fault-tolerant consensus control for multi-agent systems: An encryption-decryption scheme," *IEEE Transactions on Automatic Control*, vol. 67, no. 5, pp. 2560–2567, 2021.

[30] A. Khaled, S. Ouchani, Z. Tari, and K. Drira, "Assessing the severity of smart attacks in industrial cyber-physical systems," *ACM Transactions on Cyber-Physical Systems*, vol. 5, no. 1, pp. 1–28, 2020.

[31] G.-l. Cai, B.-s. Wang, W. Hu, and T.-z. Wang, "Moving target defense: state of the art and characteristics," *Frontiers of Information Technology & Electronic Engineering*, vol. 17, no. 11, pp. 1122–1153, 2016.

[32] V. Casola, A. De Benedictis, C. Mazzocca, and R. Montanari, "Designing secure and resilient cyber-physical systems: a model-based moving target defense approach," *IEEE Transactions on Emerging Topics in Computing*, 2022.

[33] P. Griffioen, R. Romagnoli, B. H. Krogh, and B. Sinopoli, "Reducing attack opportunities through decentralized event-triggered control," *IEEE Transactions on Control of Network Systems*, 2023.

[34] S. Obermeier, M. Wahler, T. Sivanthi, R. Schlegel, and A. Monot, “Automatic attack surface reduction in next-generation industrial control systems,” in *2014 IEEE Symposium on Computational Intelligence in Cyber Security (CICS)*, pp. 1–8, IEEE, 2014.

[35] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, “A survey of fault detection, isolation, and reconfiguration methods,” *IEEE transactions on control systems technology*, vol. 18, no. 3, pp. 636–653, 2009.

[36] G. Zhiwei, C. Cecati, S. X. Ding, *et al.*, “A survey of fault diagnosis and fault-tolerant techniques—part ii: Fault diagnosis with knowledge-based and hybrid/active approaches,” *IEEE transactions on industrial electronics*, 2015.

[37] K. Manandhar, X. Cao, F. Hu, and Y. Liu, “Detection of faults and attacks including false data injection attack in smart grid using kalman filter,” *IEEE transactions on control of network systems*, vol. 1, no. 4, pp. 370–379, 2014.

[38] A. Gupta, A. Sikdar, and A. Chattopadhyay, “Quickest detection of false data injection attack in remote state estimation,” in *2021 IEEE International Symposium on Information Theory (ISIT)*, pp. 3068–3073, IEEE, 2021.

[39] Z. Qu, J. Zhang, Y. Wang, P. M. Georgievitch, and K. Guo, “False data injection attack detection and improved wls power system state estimation based on node trust,” *Journal of Electrical Engineering & Technology*, pp. 1–15, 2021.

[40] T. Yang, C. Murguia, M. Kuijper, and D. Nešić, “An unknown input multiobserver approach for estimation and control under adversarial attacks,” *IEEE Transactions on Control of Network Systems*, vol. 8, no. 1, pp. 475–486, 2020.

[41] F. Pasqualetti, F. Dörfler, and F. Bullo, “Attack detection and identification in cyber-physical systems,” *IEEE transactions on automatic control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[42] D. Muniraj and M. Farhood, “Detection and mitigation of actuator attacks on small unmanned aircraft systems,” *Control Engineering Practice*, vol. 83, pp. 188–202, 2019.

[43] Z. Zhao, Y. Xu, Y. Li, Y. Zhao, B. Wang, and G. Wen, “Sparse actuator attack detection and identification: A data-driven approach,” *IEEE Transactions on Cybernetics*, 2023.

[44] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths, N. O. Tippenhauer, H. Sandberg, and R. Candell, “A survey of physics-based attack detection in cyber-physical systems,” *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, pp. 1–36, 2018.

[45] D. Zhang, Q.-G. Wang, G. Feng, Y. Shi, and A. V. Vasilakos, “A survey on attack detection, estimation and control of industrial cyber–physical systems,” *ISA transactions*, vol. 116, pp. 1–16, 2021.

[46] D. Ding, Q.-L. Han, Y. Xiang, X. Ge, and X.-M. Zhang, “A survey on security control and attack detection for industrial cyber-physical systems,” *Neurocomputing*, vol. 275, pp. 1674–1683, 2018.

[47] N. Mtukushe, A. K. Onaolapo, A. Aluko, and D. G. Dorrell, “Review of cyberattack implementation, detection, and mitigation methods in cyber-physical systems,” *Energies*, vol. 16, no. 13, p. 5206, 2023.

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

[49] Y. Jeong and Y. Eun, "A robust and resilient state estimation for linear systems," *IEEE Transactions on Automatic Control*, vol. 67, no. 5, pp. 2626–2632, 2021.

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

[51] S. Nateghi, Y. Shtessel, and C. Edwards, "Resilient control of cyber-physical systems under sensor and actuator attacks driven by adaptive sliding mode observer," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 15, pp. 7425–7443, 2021.

[52] Y. Gao, G. Sun, J. Liu, Y. Shi, and L. Wu, "State estimation and self-triggered control of cpss against joint sensor and actuator attacks," *Automatica*, vol. 113, p. 108687, 2020.

[53] H. Yang, H. Han, S. Yin, H. Han, and P. Wang, "Sliding mode-based adaptive resilient control for markovian jump cyber–physical systems in face of simultaneous actuator and sensor attacks," *Automatica*, vol. 142, p. 110345, 2022.

[54] X. Jin, W. M. Haddad, and T. Yucelen, "An adaptive control architecture for mitigating sensor and actuator attacks in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 6058–6064, 2017.

[55] D. Zhang, Z. Ye, G. Feng, and H. Li, "Intelligent event-based fuzzy dynamic positioning control of nonlinear unmanned marine vehicles under dos attack," *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 13486–13499, 2021.

[56] W. Song, Y. Li, and S. Tong, "Fuzzy finite-time hybrid-triggered dynamic positioning control of nonlinear unmanned marine vehicles under cyber-attacks," *IEEE Transactions on Intelligent Vehicles*, 2023.

[57] F. Cao, Z. Zhang, and X. He, “Active fault isolation of over-actuated systems based on a control allocation approach,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.

[58] A. Cristofaro and T. A. Johansen, “Fault tolerant control allocation using unknown input observers,” *Automatica*, vol. 50, no. 7, pp. 1891–1897, 2014.

[59] A. Argha, S. W. Su, and B. G. Celler, “Control allocation-based fault tolerant control,” *Automatica*, vol. 103, pp. 408–417, 2019.

[60] F. Zhu, W. Zhang, J. Zhang, and S. Guo, “Unknown input reconstruction via interval observer and state and unknown input compensation feedback controller designs,” *International Journal of Control, Automation and Systems*, vol. 19, no. 1, pp. 145–157, 2021.

[61] T. A. Johansen and T. I. Fossen, “Control allocation—a survey,” *Automatica*, vol. 49, no. 5, pp. 1087–1103, 2013.

[62] T. I. Fossen and T. A. Johansen, “A survey of control allocation methods for ships and underwater vehicles,” in *2006 14th Mediterranean Conference on Control and Automation*, pp. 1–6, IEEE, 2006.

[63] H. Chang, P. Huang, Y. Zhang, Z. Meng, and Z. Liu, “Distributed control allocation for spacecraft attitude takeover control via cellular space robot,” *Journal of Guidance, Control, and Dynamics*, vol. 41, no. 11, pp. 2499–2506, 2018.

[64] X. Lang and A. de Ruiter, “A control allocation scheme for spacecraft attitude stabilization based on distributed average consensus,” *Aerospace Science and Technology*, vol. 106, p. 106173, 2020.

[65] X. Lu, Y. Chen, X. Wang, and J. Zhao, “A multi-agent distributed optimization

method for thrust allocation," in *2021 China Automation Congress (CAC)*, pp. 4725–4730, IEEE, 2021.

[66] S. Kar and G. Hug, "Distributed robust economic dispatch in power systems: A consensus+ innovations approach," in *2012 IEEE Power and Energy Society General Meeting*, pp. 1–8, IEEE, 2012.

[67] G. Binetti, M. Abouheaf, F. Lewis, D. Naso, A. Davoudi, and B. Turchiano, "Distributed solution for the economic dispatch problem," in *21st Mediterranean Conference on Control and Automation*, pp. 243–250, IEEE, 2013.

[68] W.-T. Lin, Y.-W. Wang, C. Li, and X. Yu, "Distributed resource allocation via accelerated saddle point dynamics," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 9, pp. 1588–1599, 2021.

[69] L. Bai, M. Ye, C. Sun, and G. Hu, "Distributed economic dispatch control via saddle point dynamics and consensus algorithms," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 2, pp. 898–905, 2017.

[70] R. Carli and M. Dotoli, "Distributed alternating direction method of multipliers for linearly constrained optimization over a network," *IEEE Control Systems Letters*, vol. 4, no. 1, pp. 247–252, 2019.

[71] T. I. Fossen, "Marine control systems–guidance. navigation, and control of ships, rigs and underwater vehicles," *Marine Cybernetics, Trondheim, Norway, Org. Number NO 985 195 005 MVA, www. marinecybernetics. com, ISBN: 82 92356 00 2*, 2002.

[72] T. I. Fossen, *Handbook of marine craft hydrodynamics and motion control.* John Wiley & Sons, 2011.

[73] M. Tranninger, H. Niederwieser, R. Seeber, and M. Horn, "Unknown input observer

design for linear time-invariant systems—a unifying framework," *International Journal of Robust and Nonlinear Control*, vol. 33, no. 15, pp. 8911–8934, 2023.

[74] J. Chen and R. J. Patton, *Robust model-based fault diagnosis for dynamic systems*, vol. 3. Springer Science & Business Media, 2012.

[75] M. Hou and R. J. Patton, "Input observability and input reconstruction," *Automatica*, vol. 34, no. 6, pp. 789–794, 1998.

[76] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[77] A. Cherukuri and J. Cortés, "Asymptotic stability of saddle points under the saddle-point dynamics," in *2015 American Control Conference (ACC)*, pp. 2020–2025, IEEE, 2015.

[78] K. J. Arrow, L. Hurwicz, and H. B. Chenery, "Studies in linear and non-linear programming," *(No Title)*, 1958.

[79] S. S. Kia, J. Cortés, and S. Martinez, "Dynamic average consensus under limited control authority and privacy requirements," *International Journal of Robust and Nonlinear Control*, vol. 25, no. 13, pp. 1941–1966, 2015.

[80] F. Zhu, "State estimation and unknown input reconstruction via both reduced-order and high-order sliding mode observers," *Journal of Process Control*, vol. 22, no. 1, pp. 296–302, 2012.

[81] T. Floquet, C. Edwards, and S. K. Spurgeon, "On sliding mode observers for systems with unknown inputs," *International Journal of Adaptive Control and Signal Processing*, vol. 21, no. 8-9, pp. 638–656, 2007.

[82] A. J. Sørensen, S. I. Sagatun, and T. I. Fossen, “Design of a dynamic positioning system using model-based control,” *Control Engineering Practice*, vol. 4, no. 3, pp. 359–368, 1996.

[83] M. Iqbal, Z. Qu, and A. Gusrialdi, “Resilient dynamic average-consensus of multiagent systems,” *IEEE Control Systems Letters*, vol. 6, pp. 3487–3492, 2022.

[84] B. Gharesifard and T. Başar, “Resilience in consensus dynamics via competitive interconnections,” *IFAC Proceedings Volumes*, vol. 45, no. 26, pp. 234–239, 2012.

[85] C. A. Desoer and M. Vidyasagar, *Feedback systems: input-output properties*. SIAM, 2009.