UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA DE POSTGRADO Y EDUCACIÓN CONTINUA
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

DETECTION AND MITIGATION OF FALSE DATA INJECTION ATTACKS IN AN AC MICROGRID UNDER A DIGITAL TWIN BASED APPROACH

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIA DE DATOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

WERNER IGNACIO GONZÁLEZ HOLTHEUER

PROFESOR GUÍA:
MARCOS ORCHARD CONCHA

PROFESOR CO-GUÍA:
CLAUDIO BURGOS MELLADO

COMISIÓN:
DORIS SÁEZ HUEICHAPÁN

Este trabajo ha sido parcialmente financiado por:
FONDECYT 1210031
FONDECYT 11220989
Advanced Center for Electrical and Electronic Engineering (AC3E)

SANTIAGO DE CHILE
2024

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE
MAGISTER EN CIENCIA DE DATOS Y MEMORIA PARA
OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO
POR: WERNER IGNACIO GONZÁLEZ HOLTHEUER
FECHA: 2024
PROF. GUÍA: MARCOS ORCHARD CONCHA
PROF. CO-GUÍA: CLAUDIO BURGOS MELLADO

DETECCIÓN Y MITIGACIÓN DE ATAQUES DE INYECCIÓN DE DATOS FALSOS EN UNA MICRORRED DE CORRIENTE ALTERNA BAJO UN ENFOQUE BASADO EN GEMELO DIGITAL

Esta tesis introduce un novedoso marco de gemelo digital diseñado para detectar y mitigar ataques de inyección de datos falsos dentro de un sistema de control distribuido basado en consenso de microrredes aisladas con inversores de corriente alterna. De esta manera, aborda eficazmente el problema urgente de las vulnerabilidades de ciberseguridad, que se ha convertido en una preocupación primordial con la creciente integración de microrredes en el panorama energético. Para abordar este problema, el marco propuesto consiste en índices residuales entre la microrred y su modelo virtual, lo que permite identificar ataques en curso. Además, estos índices residuales son procesados por un controlador PID que crea variables de mitigación que se inyectan de nuevo en el sistema de control para mitigar los ataques.

La metodología se centra en el impacto de los ataques de inyección de datos falsos en los controladores secundarios de los recursos energéticos distribuidos de la microrred. Un esquema de control distribuido requiere una red ciberfísica a través de la cual se comparten variables de consenso entre controladores locales, siendo estas variables el objetivo de estos ataques. Al secuestrar un enlace de comunicación, las variables de consenso pueden ser alteradas causando cortes de energía, pérdidas económicas o inestabilidad del sistema. Este trabajo expone que los patrones obtenidos de los índices residuales de frecuencia y amplitud de voltaje revelan indicadores claros de ataques de inyección de datos falsos y pueden ser utilizados para su mitigación. Esto permite la identificación precisa del recurso energético distribuido atacado y las variables de consenso específicas, incluso si diferentes puntos de la microrred están siendo atacados simultáneamente y de manera superpuesta. Los resultados de las simulaciones demuestran la efectividad del gemelo digital como una herramienta para detectar y mitigar amenazas cibernéticas en sistemas de potencia modernos como las microrredes.

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE
MAGISTER EN CIENCIA DE DATOS Y MEMORIA PARA
OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO
POR: WERNER IGNACIO GONZÁLEZ HOLTHEUER
FECHA: 2024
PROF. GUÍA: MARCOS ORCHARD CONCHA
PROF. CO-GUÍA: CLAUDIO BURGOS MELLADO

DETECTION AND MITIGATION OF FALSE DATA INJECTION ATTACKS IN AN AC MICROGRID UNDER A DIGITAL TWIN BASED APPROACH

This thesis introduces a novel digital twin framework designed to detect and mitigate False Data Injection Attacks within a consensus-based distributed control system of isolated inverter-based AC microgrids. By doing so, it effectively tackles the pressing issue of cybersecurity vulnerabilities, which have become a paramount concern with the growing integration of microgids into the energy landscape. Addressing this issue, the proposed framework consists of residual indices between the microgrid and its virtual model, allowing the identification of ongoing attacks. Furthermore, these residual indices are processed by a PID controller which create mitigation variables injected back to the control system to mitigate the attacks.

The methodology focuses on the impact of false data injection attacks on the distributed secondary controllers of the microgrid's distributed energy resources. A distributed control scheme requires a cyber-phsysical network through which consensus variables are shared between local controllers, these variables being the target of these attacks. By hijacking a communication link, consensus variables can be altered causing power outage, economic loss or system instability. This work exposes that the patterns obtained from the residual indices of frequency and voltage amplitude reveal clear indicators of false data injection attacks and be used for mitigation. The latter allows precise identification of the distributed energy resources being attacked and the specific consensus variables targeted, even if different points of the microgrid are being attacked at the same time and in an overlapping manner. The simulation results demonstrate the digital twin's effectiveness as a tool for detecting and mitigating cyber threats in modern energy systems as microgrids.

Mucha tesis,
pues no habíamos trabajado en una tesis tanto este año
como este año tanta tesis.

Agradecimientos

Primero, quiero agradecer a mis papás por siempre apoyarme, por guiarme en la vida y por darme una educación de primera; gracias a ustedes hoy estoy aquí logrando mis metas profesionales. A mi mamá por dedicarse 100 % a mi desarrollo desde muy chico, inventándome juegos, llevándome al parque, creando experimentos conmigo y despertando mi curiosidad por el mundo. Eres una mamá muy especial. A mi papá por dedicarse a mi y ser mi guía. Siempre te esforzaste para que estemos bien y no nos faltara nada. Me enseñaste, que ante las adversidades de la vida, siempre hay que resistir. Tú canción favorita del Dúo Dinámico, *Resistiré*, ejemplifica esta lección, como siempre me recuerdas cuando las escuchamos. Estaré agradecido de ustedes toda la vida. Este logro también es de ustedes.

En segundo lugar, agradecer al Nano. Tú nos acogiste a mis papás y a mi cuando llegamos a Chile. Fuiste mi segundo padre, enseñándome tu visión de la vida. Aunque siempre supe que iba a ser ingeniero, gracias a ti pude consolidar esa decisión, y aquí estoy, sacando la Ingeniería Eléctrica con magíster. Tú como beauchefiano, siempre pudiste enseñarme lo que era ser ingeniero. Nunca olvidaré tus clases de matemáticas, física, química, astronomía y hasta historia, sacando libros de tu estante, amarillos de lo antiguos que eran, aunque no los necesitabas porque tú eras la enciclopedia. Todos te recordaremos como el personaje que siempre fuiste, un hombre extravagante. Tu último deseo antes de irte fue verme titulado de ingeniero, y ahora lo estás viendo.

También quiero agradecer a la Javi por acompañarme todos estos años, desde el primer semestre. Ahora estoy saliendo de la u en mi 16° semestre y sigues aquí conmigo, cuántas cosas vivimos juntos. Creo que no puedo pedirte más como mi polola. Me encanta como siempre te preocupas de mi y has estado para apoyarme, en lo bueno y en lo malo, tanto en la tesis, como en la vida. Tú me has hecho ser mejor, aprendí a preocuparme más de mí, de mi familia y de mis amigos. Eres una persona muy especial, por eso la gente y tus amigos te quieren tanto. Gracias infinitas.

Gracias también a mis profes guías, Marcos y Claudio. Creo que es díficil encontrar profesores como ustedes, tuve suerte. El nivel de conocimientos, el feedback constante, la rapidez en las respuestas de sus mails y su habilidad para encontrar mejores formas de hacer las cosas y siempre guiarme hacia el camino correcto cuando yo estaba teniendo problemas en encontrarlo. Sin su mentoría tampoco podría haber logrado esto.

Gracias a mi primo Vitoco. Tuve suerte que la vida decidió juntarnos el año pasado. Hemos sido buenos partners este último tiempo, siempre aconsejándonos, pasando las penas, viajando y pasándolo la raja. Contigo conseguí algo que nunca tuve al ser hijo único: un hermano. Gracias por siempre escucharme con mis problemas de la tesis y darme tus consejos.

Gracias a mi madrina Andrea. Aunque nos veamos poco en el año, siempre he sentido tu preocupación. Gracias por regalarme la calculadora TI; fue vital cuando entré a Eléctrica en tercer año. Nunca voy a olvidar cuando me robaron la mochila con la calculadora. Tú, inmediatamente, me diste una nueva. Gracias por todo.

Gracias a mi cuñi Nacho. Con los años nos hemos hechos buenos amigos, con gran complicidad. Siempre me preguntabas cómo iba con la tesis y escuchaste mis altibajos. Esos pequeños gestos hacen una gran diferencia. Que buen cuñado que me tocó, gracias.

Gracias a Pichigang, mis amigos de la u. Que gran grupete formamos. Buenas sesiones de estudio hasta las 4am, las mejores juntas, excelentes viajes a la playa y carretes extraordinarios. Gracias a ustedes la universidad será una experiencia inolvidable.

Gracias especiales a los mateitos, Javi, Pascal, Cata, Coloma y Coni. Coloma, mi amigo de la vida, estoy muy feliz que pudimos retomar nuestra amistad entrando a la u. Pascal, que bueno que cruzamos nuestros caminos aventurando en Carén. Gracias Javi y Cata por salvarme en Intro a la Ingeniería, ahí empezó todo. Y gracias más especiales a Cata y Coni, porque gracias a ustedes estoy con la Javi.

También agradecer a los pibardos. Coloma, Pascal, Joaco, Maniega, Brook, Max, Nico y Rai. Que bueno que nos seguimos juntando. Ya casi todos salimos de la u y muchos están desarrollando su vida laboral, pero aún así nos vemos, carreteamos, jugamos poker o simplemente nos juntamos a conversar. Los quiero caleta cabros.

Por último agradecer a los K del D. Ustedes también serán una amistad para toda la vida. Tuve suerte de encontrarme con un curso como ustedes cuando llegué al Andrée en 4to básico; hasta el día de hoy nos juntamos con tanta frecuencia, a hacer asados, conversar, carretear o tomarnos unas chelas. Es muy lindo que sigamos viéndonos con la misma confianza de siempre. También los quiero caleta.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The increasing adoption of microgrids (MGs) in modern energy systems has brought numerous benefits, such as enhanced reliability, cost efficiency, and environmental sustainability. These localized grids, which can operate independently or in conjunction with the main power grid, are essential for integrating renewable energy sources and providing power to remote areas. However, the complex cyber-physical nature of MGs introduces significant cybersecurity challenges. This thesis explores the development and implementation of a Digital Twin (DT) framework to enhance the cybersecurity of MGs by detecting and mitigating False Data Injection Attacks (FDIAs). By using real-time data and detection and mitigation algorithms, the proposed DT framework aims to ensure the stable and secure operation of MGs.

1.1. Motivation

The study of cyberattacks in cyber-physical systems has become important in today's industry. In the past, hackers would attack companies, industries or governments to obtain critical information, but today they are capable of affecting critical physical infrastructure. For example, in 2010 a cyberattack was carried out on a nuclear plant in Natanz, Iran. A virus called Stuxnet was able to damage 1,000 centrifuges used to enrich uranium by manipulating their operation [1]. This case was historic, as it was the first time a cyberattack managed to damage physical infrastructure.

Given these risks, it is crucial to further investigate cybersecurity issues in modern electrical systems. This thesis addresses these concerns by exploring the effects of cyberattacks on microgrids (MGs). This electrical system was chosen because they are acquiring importance in today's energy landscape and are expected to have great impact in the future. Given MG's importance for today and future's energy terrain, protecting them from cyberattacks becomes imperative.

This thesis focuses on cyberattacks on the communication network of the distributed control system of a microgrid. By intruding a communication link a hacker can alter the control system. Some types of cyberattacks are 1) replay attacks [2], 2) FDIAs [3–8] and 3) denial of service (DoS) [9]. In this study the attacker alters control data in communication links using False Data Injection Attacks (FDIAs). This type of attacks are studied in different electrical systems such as MGs, smart grids and modular multilevel converters [3–7]. In MGs, FDIAs are the most studied cyberattacks. In a recent work a review of cyberattack on MGs is studied [10], showing various research on cyberattacks on DC and AC MGs. The review divides the studies in two main categories: 1) detection and mitigation approaches and 2) resilient control system designs. The detection methods are classified into i) signal-based methods, ii) model-based methods such as Kalman Filter schemes, observer-based schemes and sliding mode observers, and iii) data-based methods which use different kinds of neural networks.

This thesis contributes to the category of digital twin (DT) based detection and mitigation approaches, a relatively novel area unexplored in the previous review [10]. DTs have been increasingly being investigated during the last 9 years, and consist of a dynamic virtual model that replicates a physical entity using real-time data. They can be used for predictive analysis and diagnostics [11], or continuous control and monitoring [12]. In this work a DT is used as: 1) part of a detection method, where it compares the DT's output with the real system, creating residual indices as indicators of FDIAs, and 2) part of a mitigation method, where a PID controller processes the residual indices to create mitigation variables injected back to the control system to counter attacks.

While the review by Shafei et al. [10] does not specifically mention DTs for cyberattack detection, relevant literature exists. In [13] Saad et al. propose an IoT-based DT technology for detecting and mitigating FDIAs and DoS attacks across networked DC MGs with distributed consensus-based secondary and tertiary control. However, this approach does not consider intra-microgrid dynamics, internal control schemes (primary and secondary control levels), or AC MGs. To the best of the authors' knowledge, this thesis is the first to propose a DT-based FDIA detection framework targeting the secondary control level of internal dynamics in islanded AC MGs, aiming to restore voltage and frequency values and power sharing between DERs.

1.2. Hypothesis

The hypothesis of this thesis is that the digital twin (DT) framework can effectively detect and mitigate False Data Injection Attacks (FDIAs) in isolated inverter-based AC Microgrids by using residual indices of voltage amplitude and frequency, even under simultaneous and overlapping attack conditions.

1.3. General Objectives

The primary objective of this thesis is to develop and validate a DT-based detection and mitigation framework that enhances the cybersecurity of AC microgrids by accurately identifying and mitigating FDIAs. This framework aims to provide a reliable method for real-time monitoring and protection of microgrids, ensuring their stable and secure operation.

1.4. Specific Objectives

To achieve the general objective, the following specific objectives are outlined:

1. **Design a Digital Twin Framework**: Develop a virtual model that replicates the physical behavior of an isolated inverter-based AC microgrid, including its control systems and communication networks. The digital twin uses the real state of the MG and the virtual model's state to create residual indices used as attack indicator and for further mitigation.

2. **Devise a Distinction of the Virtual Model**: A very specific and clever distinction from the physical entity must be designed in the virtual model to generate residual indices that make sense.

3. **Simulate Attack Scenarios**: Conduct extensive simulations of various FDIA scenarios, including simultaneous and overlapping attacks, and step-like and ramp-like attacks, to test the effectiveness of the DT framework.

4. **Validate the Detection and Mitigation Method**: analyze the DT-based detection through the residual indices and validate the mitigation system. The latter must be confirmed through voltage and frequency levels going back to nominal values and recovering normal power sharing during an attack.

1.5. Thesis Structure

The structure of this thesis is as follows:

- **Chapter 2**: This chapter corresponds to the Theoretical Framework. This chapter introduces the main concepts to understand the novel DT framework proposed in this thesis. Concepts like micgrogrids, distributed control systems, cyber-physical systems, digital twin, among others are explained.

- **Chapter 3**: This chapter corresponds to the Methodology, where the MG topology, its distributed control system and the complete DT framework design are described in detail.

- **Chapter 4**: This chapter corresponds to the Experimental Results. Various experiments under different attack scenarios are carried out with the aim of effectively demonstrating that the digital twin framework fulfills its function perfectly.

- **Chapter 5**: This chapter corresponds to the Analysis and Discussion. The results from Chapter 4 are analyzed and discussed thoroughly to demonstrate that the results make sense and DT framework effectively detects and mitigates FDIAs.

- **Chapter 6**: This chapter corresponds to the Conclusions and Future Work. The main conclussions are presented and guidelines for future work are provided.

Chapter 2

Theoretical Framework

2.1. Microgrids

An MG is a localized electrical distribution network that operates at low and medium voltages and is composed of Distributed Energy Resources (DERs) which include Renewable Energy Sources and Energy Storage Systems as well as loads that operate locally [14]. MGs are particularly useful in isolated areas where the main power grid may be untrustworthy or non-existent, offering reliability, environmental benefits, reduced costs, and flexibility [15]. They are also closely linked with smart grids, which together will shape the future of smart cities, revolutionizing the manner energy is managed and distributed [16].

MGs have the capability to work in two modes:

1. **Grid-Tied Mode**: In this mode the MG is connected to the main grid via a Point of Common Coupling (PCC).
2. **Islanded Mode**: In this mode the MG functions independently from the main grid. This mode is particularly useful in remote or disaster-affected areas where the main grid is unreliable or unavailable.

In islanded mode, MGs suffer from stability issues due to the low inertia provided by DERs [14, 17–19]. In grid-tied mode, the main grid's substantial inertia helps stabilize the MG, maintaining voltage and frequency levels despite fluctuations in power demand or supply [14].

Microgrids can be classified based on the type of power they distribute. This classification is based on the nature of the electrical current they use. The three types of power are:

1. **Alternating Current (AC)**: this is the standard form of electricity supplied by traditional power grids, where current changes direction periodically, generally at frequencies of 50 or 60 Hz.
2. **Direct Current (DC)**: in this configuration current flows in only one direction. This approach has been taking more importance today due to its efficiency and simplicity [20].
3. **Hybrid**: hybrid MGs combine AC and DC distribution systems and exploit advantages of both types of power.

2.2. Control System of a Microgrid

An MG's control system is crucial for ensuring its stable and reliable operation. It manages the generation, distribution, and consumption of electrical energy within the MG, maintaining voltage and frequency levels and even power-sharing quality. Effective control systems enable MGs to integrate DERs, respond to dynamic changes in load and generation, and ensure system resilience.

2.2.1. Hierarchical Control Levels

Microgrid control systems are typically organized into three hierarchical levels: primary, secondary, and tertiary controls [21].

- **Primary Control**: The primary control level is responsible for the immediate, local control of DERs. It employs droop control schemes to regulate voltage and frequency. Primary control ensures that DERs can operate autonomously and share loads proportionally based on their capacities.

- **Secondary Control**: Secondary control aims to restore the voltage and frequency to their nominal values after deviations caused by load changes or disturbances. It operates on a slower timescale compared to primary control and uses feedback mechanisms to fine-tune the system's performance. For example, Distributed-Averaging Proportional-Integral (DAPI) controllers [21, 22] can be used in this level to enhance power-sharing among DERs.

- **Tertiary Control**: The tertiary control level focuses on optimizing the overall operation of the microgrid, including economic dispatch, energy management, and congestion management. It coordinates the interaction between the microgrid and the main power grid, ensuring efficient energy exchange and system stability.

2.2.2. Control Strategies

Microgrid control strategies can be classified into centralized and distributed approaches.

- **Centralized Control**: In centralized control systems, a single controller manages all the elements of the microgrid. This approach requires a high-bandwidth communication network to gather data from all DERs and send control commands. While centralized control can effectively manage the system, it has several drawbacks, including vulnerability to single points of failure and scalability issues [15, 21].

- **Distributed Control**: Distributed control involves multiple controllers working collaboratively across different points of the microgrid to achieve global objectives. Each controller is responsible for a local area and communicates with neighboring controllers to maintain system stability. Distributed control enhances scalability and reliability by distributing decision-making processes, is effective in managing the variability of DERs and reduces the risk of single points of failure [21, 23].

2.2.3. Consensus Based Distributed Control

When working with distributed control, the distributed controllers can be regarded as "agents" working in a collaborative manner. The agents would be associated with a specific DER and LC. If the MG consists of N DERs and their respective LCs, there would be N agents.

Based on [21], the communication network of the distributed control system can be regarded as an undirected cybergraph $\mathcal{G} = (U, E, A)$. $U = \{1, \ldots, N\}$ denotes the agents or nodes of the graph which correspond to the DERs of the MG; $E = \{(i, j)/(i, j) \in U\}$ represent the edges or links between the nodes; $A = [a_{ij}]_{N\text{x}N}$ is an adjacency matrix where each element $a_{ij} = a_{ji} \geq 0$ represents the connection weight between nodes i and j, where $a_{ij} > 0$ only if $(i, j) \in E$, otherwise, $a_{ij} = 0$. In this work the communication links are binary, so $a_{ij} = \mathbb{1}_E [(i, j)]$, meaning there is or there is not communication. On the other hand, the set of neigbours of the ith node is defined as $\mathcal{N}_i = \{j\, /j \in U \wedge (i, j) \in E\}$. For example, Fig. 2.1 depicts an example of a fully connected graph with three agents (X_1, X_2 and X_3).

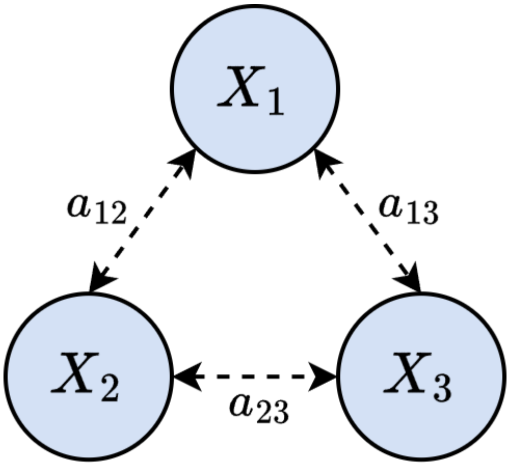

Figure 2.1: Example of a graph of the communication network of three agents

Consensus-based distributed control is a strategy where LCs or agents work together to achieve a common goal, such as maintaining voltage and frequency stability. Consensus algorithms ensure that all controllers agree on certain key variables, such as active and reactive power, by continuously exchanging information over the communication network.

In a consensus-based distributed control system, each LC adjusts its output based on local measurements and information received from neighboring controllers. This collaborative approach helps balance the power distribution and enhances the overall stability of the microgrid. The following equations describe the consensus-based control for frequency (ω) and voltage amplitude (V) that will be used in this work:

- Voltage control:

$$V_i^{ref} = V^* - n_i Q_i + \delta V_i \tag{2.1}$$

$$k_i^V \frac{d\,\delta V_i}{dt} = \beta_i \left(V^* - V_i\right) - \gamma_i \sum_{j \in \mathcal{N}_i} a_{ij} \left(Q_i - Q_j\right) \tag{2.2}$$

- Frequency control:

$$\omega_i = \omega^* - m_i P_i + \Omega_i \tag{2.3}$$

$$k_i^\omega \frac{d\Omega_i}{dt} = \alpha_i \left(\omega^* - \omega_i\right) - \eta_i \sum_{j \in \mathcal{N}_i} a_{ij} \left(P_i - P_j\right) \tag{2.4}$$

These equations are a modified version of the equations presented in [22], where active and reactive power, P and Q, are the consensus variables shared between agents to improve power-sharing between DERs. In particular, the first terms of equations (3.8) and (3.10) are in charge of the frequency and voltage restoration, respectively. The second terms are responsible of improving the active and reactive power sharing and corespond to the protocols to achieve consensus:

$$u_{q_i} = -\gamma_i \sum_{j \in \mathcal{N}_i} a_{ij} \left(Q_i - Q_j\right) \tag{2.5}$$

$$u_{p_i} = -\eta_i \sum_{j \in \mathcal{N}_i} a_{ij} \left(P_i - P_j\right) \tag{2.6}$$

Adapting what is said in [21, 22], variables Q_i achieve consensus if $Q_i(t) - Q_j(t) \to 0$ as $t \to \infty$. The same way variables P_i achieve consensus if $P_i(t) - P_j(t) \to 0$ as $t \to \infty$. Given the latter, what should happen is:

$$\lim_{t \to \infty} u_{q_i}(t) = 0 \tag{2.7}$$

$$\lim_{t \to \infty} u_{p_i}(t) = 0 \tag{2.8}$$

2.3. Cyber-Physical Systems (CPS)

Cyber-Phsyical Systems (CPS) are integrated systems which consist of computational (cyber) and physical elements interacting closely. NIST SP 1500-201 [24] defines CPSs as "*smart systems that include engineered interacting networks of physical and computational components*". These systems leverage the integration between physical processes and computer capabilities to improve the functionality and efficiency of industrial applications.

Cyberspace, as defined by NIST SP 800-30 Rev.1 standards [25], is "*a global domain within the information environment consisting of the interdependent network of information systems infrastructures including the Internet, telecommunications networks, computer systems, and embedded processors and controllers.*" It is immediate to link this definition to a CPS, as it is the medium through which a CPS can integrate the physical and cyber world.

A CPS's architecture comprises several key components:

1. **Physical Processes**: These include all physical components and processes that are being controlled and monitored. Some examples are machinery in a factory, biological processes in medical applications or electric grids in power systems.

2. **Sensors**: devices that collect data from physical processes. These measurements are transformed into digital signals that can be processed and analyzed by the computational elements.

3. **Actuators**: Devices that convert control signals from the computational elements into physical actions. They are key to reaching an operating point commanded by a control algorithm.

4. **Computation Elements**: These include all software and hardware components responsible for processing data and making decisions. This is the cyber component and it often encompasses embedded systems and microcontrollers running control algorithms.

5. **Communication Network**: manages data exchange between sensors, actuators, and computational elements. These networks can be wired or wireless and must be reliable and secure to ensure timely and accurate data transfer.

Any system including all these elements can be considered a CPS. For example an autonomous vehicles is a physical entity with physical processes which uses a combination of sensors, actuators, computers and communication networks to navigate and operate without human intervention. Another example are microgrids, which are physical power distribution systems, which have these same elements to manage power generation and consumption.

2.4. Cyber-Physical Networks (CPN)

A Cyber-Physical Network (CPN) is closely related to a CPS. It is the network infrastructure that manages the communication and coordination between the components of a CPS. A CPN ensures the data exchange and communication between the physical and cyber elements is reliable and secure.

The main components of a CPN are:

1. **Nodes**: Individual units within the network that perform specific tasks. For example, sensors, actuators and controllers can be nodes of the CPN.

2. **Communication Link**: paths for data transmission between nodes. These can be wired or wireless connections.

3. **Network Protocols**: Standards that govern data exchange in the network (TCP/IP, Bluetooth, etc.).

4. **Data Management**: Systems for storing, processing and retrieving data within the network.

2.5. False Data Injection Attacks (FDIAs)

Connecting latter concepts, it is clear that the control system of an MG, centralized or distributed, needs a CPN to function. This network is a vulnerability to the MG's operation as it can be a target of cyberattacks.

Some types of cyberattacks are 1) replay attacks [2], 2) FDIAs [3–8] and 3) denial of service (DoS) [9]. In particular, FDIAs are studied in different electrical systems such as MGs, smart grids and modular multilevel converters [3–7]. False Data Injection Attacks are a type of cyberattack that specifically target the data integrity of CPSs, such as power grids, smart grids, and microgrids. More specifically, they can target the CPN of the system, looking for intentional introduction of false or manipulated data into the network's data streams to cause the system to behave erroneously or inefficiently.

The principal characteristics of FDIAs are:

1. **Data Manipulation**: In FDIAs, attackers alter the data being fed into the control system. This can include measurements from sensors, control signals, or other data used in the decision-making processes of the system.

2. **Targeted Variables**: FDIAs typically target specific variables crucial to the system's operation. For example, in power systems, these variables can include voltage levels, frequency, active power, and reactive power.

3. **Stealthiness**: These attacks can be designed to be stealthy, making them difficult to detect. By carefully crafting the false data to resemble normal variations or anomalies, attackers can evade detection mechanisms that rely on threshold-based or anomaly detection.

4. **Impact on Control Systems**: The goal of an FDIA is to disrupt the normal operation of the control system. This disruption can lead to incorrect decisions by the control system, causing suboptimal performance, inefficiencies, or even physical damage if protective measures are not triggered.

2.6. Digital Twin (DT)

The Digital Twin (DT) was first presented by Micheal Grieves in collaboration with John Vikcers of NASA. In his paper [26] he characterizes the concept of a Digital Twin as an entity that consists of a virtual representation of a physical system, receiving data from the physical representation and sending data back, in a bi-directional manner. They are used in industrial systems for continuous monitoring, diagnostics and optimization of the processes of the physical system.

In a more recent study from Jones et al. [11] the main components of a DT are characterized as:

1. **Physical Entity (PE)**: real-world system being modeled.

2. **Virtual Model (VM)**: dynamic simulation mirroring the physical entity's behavior.

3. **Data Interface**: mechanism for continuous data exchange between the virtual model and the physical entity.

4. **Analytics and Control**: Tools to analyze data, giving useful information about the systems state and control the physical system based on the virtual model's outputs.

The analytics and control component of the digital twin can be any tool to analyze the data of the real system and the VM. In particular, it can use AI/Machine Learning, and be considered an Intelligent Digital Twin [27].

The data transfer within a DT framework is important to detail. It consists of multiple steps:

1. **Data Collection**: Sensors and IoT devices collect real-time data from the physical entity.

2. **Data Transmission**: The collected data is then transmitted to the VM through communication networks.

3. **Data Integration**: The VM integrates the measured data to update its state trying to reflect the PE's actual state.

4. **Feedback Loop**: Information given by the analytics and control of the VM are sent back to the PE to optimize performance, predict failures and mitigate risks.

This last step is key for completing the DT framework, where it takes real actions in the physical system. Without the feedback loop the digital twin can be used exclusively for monitoring of the state and health of the system. Fig. 2.2 illustrates how the data transfer works between the principal components of the DT.

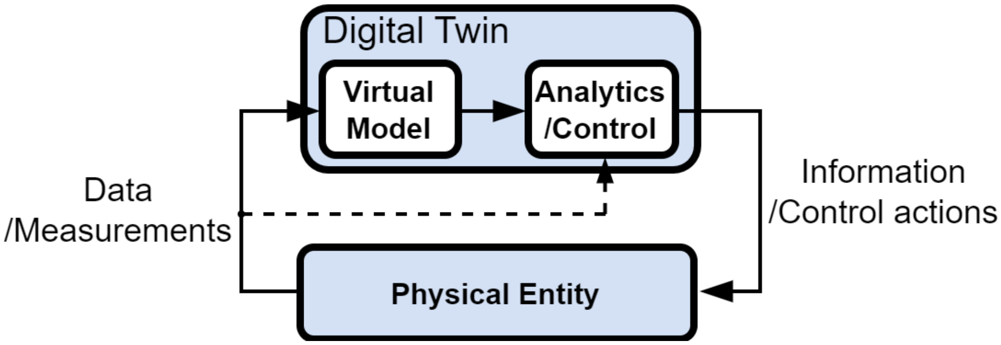

Figure 2.2: General illustration of the components and data transfer of a digital twin framework

Chapter 3

Methodology

3.1. Microgrid Topology and its Consensus-based Distributed Control Strategy

As illustrated in Fig. 3.1 this thesis uses an MG topology comprising N DERs. Each DER is a three-phase power generating unit connected parallely to the PCC which operates at nominal values of voltage V^* and frequency f^*. They consist of a DC power source which is converted to three-phase AC power through an inverter. The inverter is coupled with an LC filter characterized by L_f and C_f. As the MG operates in islanded mode, the energy sources provide low stability due to low inertia, making the primary and secondary control of vital importance. In addition, the MG works with a consensus-based distributed secondary control scheme based on [22], explained in Section 2.2.3. Tertiary control will not be mentioned nor implemented in the simulation and experiments. The parameter values of the MG used in the experiments are shown in Table 3.1.

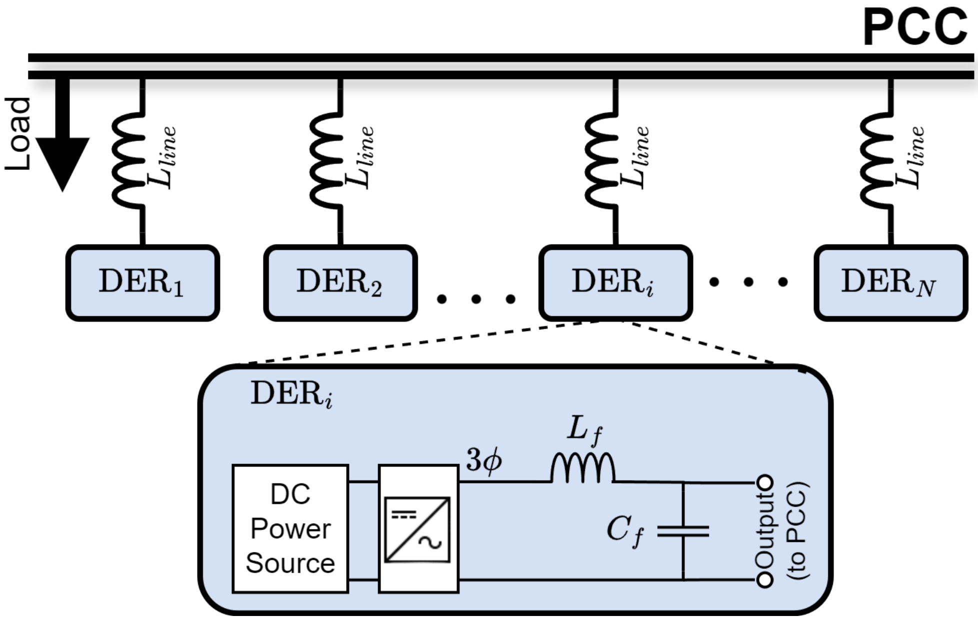

Figure 3.1: Circuit topology of the islanded AC MicroGrid with N DERs and a constant load.

Table 3.1: MG parameters used for Matlab/Simulink Simulation

Description	Value
Nominal voltage amplitude (V^*)	110[V]
Nominal frequency (f^*)	50[Hz]
Load resistance (R_{load})	10 [Ω]
Load inductance (L_{load})	3 [mH]
Line inductance (L_{line})	1.25 [mH]
DER filter inductance (L_f)	1.8 [mH]
DER filter capacitance (C_f)	20 [μF]
Number of DERs (N)	3

The DERs shown in Fig. 3.1 are controlled by the distributed control strategy to restore their output voltage amplitude and frequency and balance power-sharing, control which is described in equations (3.7)-(3.10). For it to function there must be a distributed communication network, as depicted in Fig. 3.2, so each LC can receive the other LC's consensus variables, in this case, active and reactive power (P_i and Q_i). The control loop must seek the balance of voltage and frequency at some point of the MG to nominal values (frequently the voltage of the PCC) and improve active and reactive power sharing among the DERs. In particular, the frequency is global parameter of the MG, while the voltage is not. In this

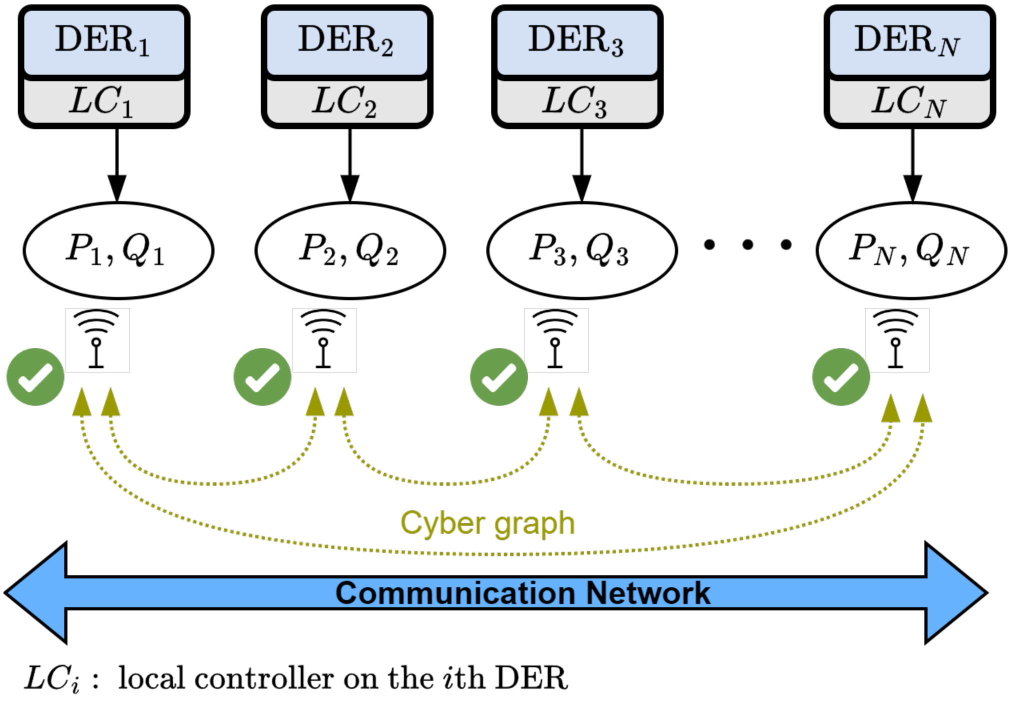

Figure 3.2: Illustration of local controllers sharing active and reactive power as consensus variables through the communication network creating a cyber graph.

sense, in steady-state, the value of voltages can vary at different points of the MG [21]. Additionally, the PCC is connected to a local load that consists of a resistive load R_{load} and inductive load L_{load} connected in series.

Fig. 3.3 shows how the N DERs and LCs interact with the communication network. At the bottom the ith DER and LC is expanded showing in detail its local control scheme. Focusing on the role of the communication network, Fig. 3.3 illustrates that each LC transmits its active and reactive power measurements P_i and Q_i into the network. Simultaneaously, the LCs receive active and reactive power measurements $P_{j \in \mathcal{N}_i}$ and $Q_{j \in \mathcal{N}_i}$ from neighbor controllers. It can be noticed that local active and reactive power measurements are calculated using the sensor measurements of output voltage V_i and current I_i.

The ith LC's scheme in Fig. 3.3 shows how the DAPI control from equations (3.7)-(3.8) and (3.9)-(3.10) are implemented together with the DER circuit and the zero control level.

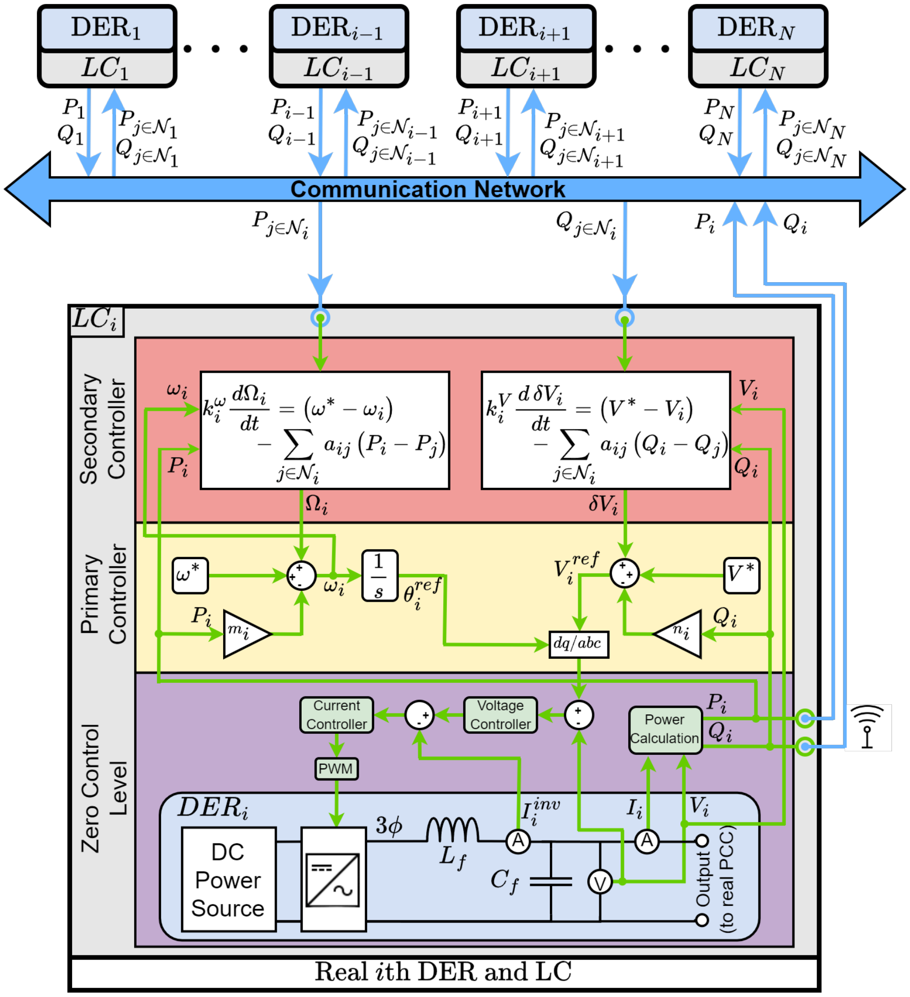

Figure 3.3: Ilustration of N DERs with their LCs interacting through the communication network. The real ith DER with its LC is expanded at the bottom.

The control parameters of the DAPI expressions are shown in Table 3.2.

Table 3.2: Consensus-based controller parameters

Description	Value
Frequency secondary controller velocity coefficient (k_i^ω)	0.5
Frequency deviation coefficient (α_i)	10
Active power-sharing consensus coefficient (η_i)	1e-3
Voltage secondary controller velocity coefficient (k_i^V)	5
Voltage deviation coefficient (β_i)	20
Reactive power-sharing consensus coefficient (γ_i)	-0.1

In Fig. 3.3 the zero control level uses a standard voltage and current controller to control the inverter through PWM signals. As can be reviewed in [21], these control loops are usually implemented using resonant controllers [28, 29], controller within a synchronous rotating d-q frame [30, 31], predictive controller [32, 33], etc. For simulation simplicity, the inverter is modelled as a controlled voltage source instead of real switches. Given the latter, inner voltage and current control loops are implemented, but omitting the use of PWM signals.

Another point to mention is that V_i, I_i and I_i^{inv} measurements are discrete, sampling 256 data points per cycle assuming a nominal frequency $f^* = 50Hz$, so the sensor's sampling time is $T_s = 1/(256 \cdot f^*) = 78.125[\mu s]$. Furthermore, the MG is simulated in Matlab®/Simulink® integrated with PLECS Blockset®, so the sensors do not have associated noise. To imitate this behaviour, white noise was added to voltage and current measurements with standard deviation $\sigma = 0.01$.

3.2. Definition and Implementation of FDIAs in the MG

The previous consensus-based control scheme is used to restore voltage amplitude and frequency and improve power sharing of the DERs, but this works under normal conditions and can be disturbed. Even more, in this work the MG operates in islanded mode, which is traduced into low inertia and stability making it easier for the voltage and frequency to be thrown out of balance. For this reason a robust control system is required. However, even with such a system in place, a vulnerability remains: cyber-attacks. If an attacker is capable of hijacking one or more local communication links of the CPN, the control system can be compromised no matter how robust it is.

This work will focus on FDIA detection and mitigation since they are common and well known attacks in the literature. Fig. 3.4 shows an example of an FDIA being carried out on DER_3. This example tries to exhibit that variables P_3 and Q_3, vital for the consensus of the DERs, are being attacked, spreading destabilization through the distributed communication network, compromising the control system and the entire MG.

In this work, the attacker targets the communication network and not the control system

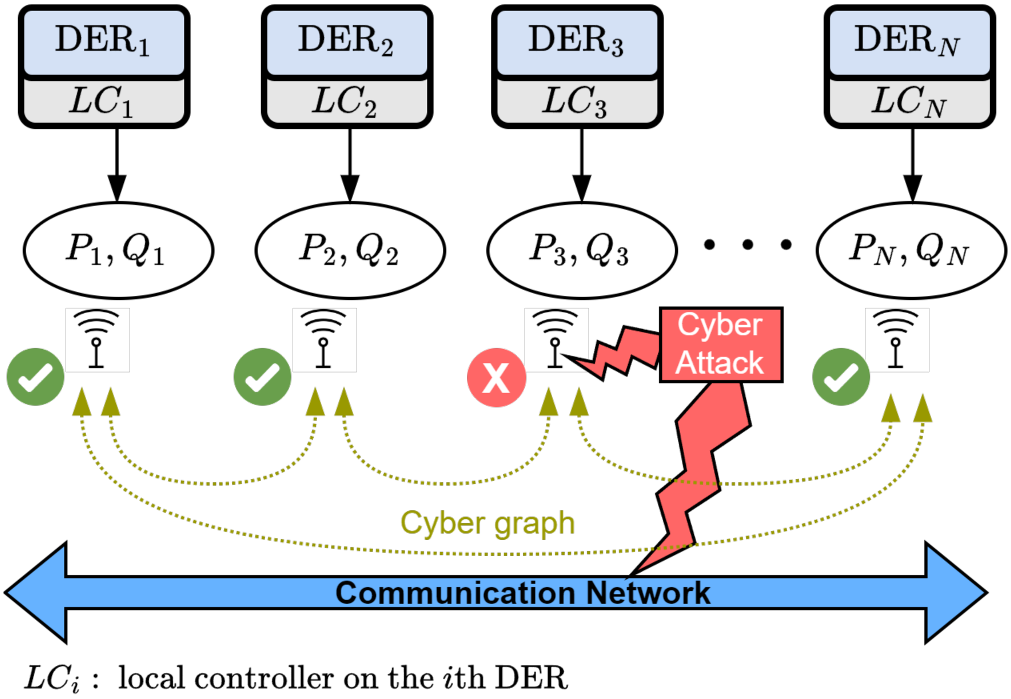

Figure 3.4: Illustration of local controllers sharing active and reactive power as consensus variables through the communication network creating a cyber graph.

directly. Given the latter, following the example from Fig. 3.4, LC$_3$ should not be directly affected by the attack since it calculates P_3 and Q_3 internally. This way the original variables are used by the local controller, however when sent to the communication network an FDIA can alter their values, either before sending them or directly by intercepting the data stream of the communication network. The neighbor LCs would receive false data affecting the consensus.

The DER's inverters voltage amplitude and frequency are regulated through the distributed consensus-based control scheme described in Section 3.1 and Fig. 3.3. This control scheme is based on the DER's active and reactive power as consensus variables Q_i and P_i, and FDIAs target either or both of these variables. In this study, the FDIA in LC$_i$ is modeled as shown in (3.1)-(3.2). An attack on active power variable will be referred to as "P-attack" and on reactive power variable as "Q-attack".

$$\text{Q-attack: } Q_i^f(t) = Q_i(t) + \kappa_i^q Q_i^a(t) \tag{3.1}$$
$$\text{P-attack: } P_i^f(t) = P_i(t) + \kappa_i^p P_i^a(t) \tag{3.2}$$

In the latter definitions, $\kappa_i^q, \kappa_i^p \in \{0, 1\}$, $\forall i \in U = \{1, \ldots, N\}$ are Boolean variables, representing the presence of an attack sequence $Q_i^a(t)$ or $P_i^a(t)$ on consensus variables. When there is a P-attack and/or Q-attack, following the example from Fig. 3.4, LC$_3$ uses Q_3 and P_3, while the faulty versions $Q_3^f(t)$ and $P_3^f(t)$ are sent to neighbor local controllers. This point will be explained with more detail later.

The study in this work focuses on testing the FDIAs defined in (3.1)-(3.2) on the communication network. In the experiments of this study, the attack sequences that reproduce the FDIAs are represented in (3.3)-(3.4). $P_i^a(t)$ is activated between $t_{p_i}^{\text{init}}$ and $t_{p_i}^{\text{final}}$ with an attack element $p_i(t)$; $P_i^a(t) = 0$ outside that temporal interval. The same occurs for $Q_i^a(t)$, where the attack element $q_i(t)$ appears between $t_{q_i}^{\text{init}}$ and $t_{q_i}^{\text{final}}$.

$$Q_i^a(t) = \begin{cases} 0 & \text{if } 0 \leq t < t_{q_i}^{\text{init}} \\ q_i(t) & \text{if } t_{q_i}^{\text{init}} \leq t < t_{q_i}^{\text{final}} \\ 0 & \text{if } t \geq t_{q_i}^{\text{final}} \end{cases} \tag{3.3}$$

$$P_i^a(t) = \begin{cases} 0 & \text{if } 0 \leq t < t_{p_i}^{\text{init}} \\ p_i(t) & \text{if } t_{p_i}^{\text{init}} \leq t < t_{p_i}^{\text{final}} \\ 0 & \text{if } t \geq t_{p_i}^{\text{final}} \end{cases} \tag{3.4}$$

In Fig. 3.5 the interception of the consensus variables Q_i and P_i is shown. It can be noticed that the primary and secondary controller receive the original variables calculated

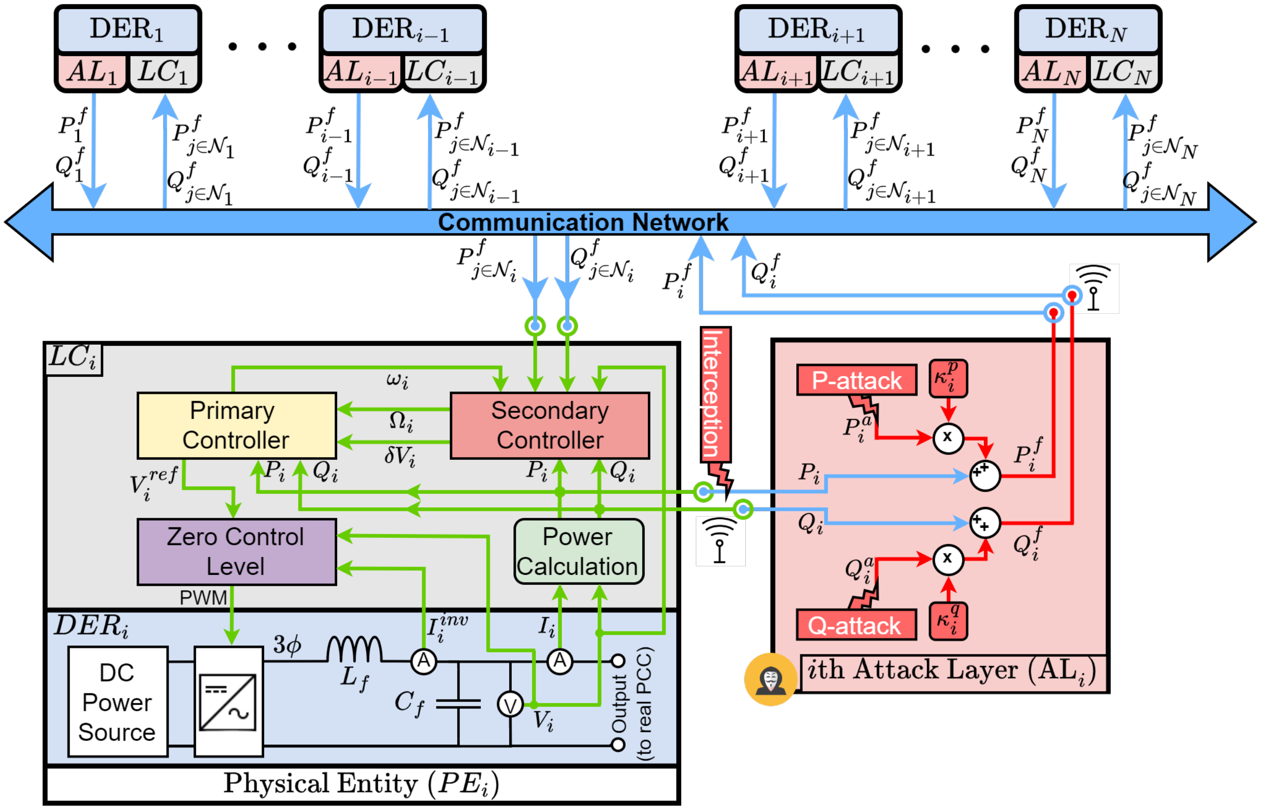

Figure 3.5: Illustration of how the consensus variables Q_i and P_i are intercepted to send the attacked variables Q_i^f and P_i^f to the communication network.

from the measurements. Before sending these variables to the communication network, they are intercepted by the ith attack layer and modified into faulty variables Q_i^f and P_i^f, following the definition from (3.1)-(3.2). For all N DERs, Q_i^f and P_i^f are sent to the communication network by the attack layers, and all LC's receive the attacked neighbor variables $Q_{j\in\mathcal{N}_i}^f$ and $P_{j\in\mathcal{N}_i}^f$. It must be noted that if $\kappa_i^q = 0$ and $\kappa_i^p = 0$, then $Q_i^f = Q_i$ and $P_i^f = P_i$, returning to the normal operation of the MG. Given these changes, the secondary control equations (3.8) and (3.10) turn into (3.5) and (3.6) shown below, while primary control remains the same.

$$k_i^V \frac{d\,\delta V_i}{dt} = \beta_i \left(V^* - V_i\right) - \gamma_i \sum_{j\in\mathcal{N}_i} a_{ij} \left(Q_i - Q_j^f\right) \tag{3.5}$$

$$k_i^\omega \frac{d\Omega_i}{dt} = \alpha_i \left(\omega^* - \omega_i\right) - \eta_i \sum_{j\in\mathcal{N}_i} a_{ij} \left(P_i - P_j^f\right) \tag{3.6}$$

The concepts in this section are fundamental for understanding how FDIAs disturb the system. When a P-attack and/or Q-attack occurs, voltage amplitude and frequency will show transients, affecting there values. The attacks can deviate them from nominal values activating the protection systems, or being stealthier, where voltage amplitude and/or frequency can be deviated a bit from its nominal values, not activating protection systems, but achieving sub-optimal performance by affecting the power-sharing between DERs.

In the next section the DT framework will be presented and how it works by analyzing discrepancies of voltage amplitude and frequency between the DT and the real system, referred as residual indices, and how these are used for mitigation.

3.3. Proposed Digital Twin Framework

A digital twin (DT) [11] is a digital representation of a real system which is dynamic and evolves and adapts over time. DTs provide a bridge between the physical and digital domains through a continuous exchange of data, explained with more detail in Section 2.6. As discussed in Section 3.1, all power consensus variables are flowing through the communication network, being the medium of data exchange between the real DERs and their corresponding DTs. In particular, as explained in Subsection 3.2, these variables are potentially attacked.

In the following sub-sections the whole framework will be explained. In the first subsection, the topology of the DT will be detailed, showing similarities and differences with the physical entity. The second sub-section exposes how the "analytics" block of the DT functions, creating residual indices between the real system and its virtual counterpart. The last subsection shows how the feedback loop of the digital twin is implemented for mitigation.

3.3.1. Digital Twin's Virtual Model

In this thesis the whole MG topology and its control system is emulated by a virtual model (VM). The MG topology from Fig. 3.1 is implemented in PLECS: each DER, their electrical connections and the PCC. For simplicity, the real loads are assumed known and their exact values are given to the DT to simulate their connection to the virtual PCC.

On the other hand, the control system is also emulated by the VM. The control system relies on two elements: DER output voltage and current measurements, and power consensus variables. As the VM's DER connections to the PCC are simulated, the virtual voltage and current measurements are used as inputs to the virtual control system. On the other hand, the virtual model does not have a communication network to transfer consensus variables. Instead, it integrates the real consensus variables from the real communication network. As the virtual model receives the consensus variables, it updates its output state, with the goal to represent (at least approximately) the real MG's state.

Although the VM to this point should be capable of imitating the MG's behavior, it still is not useful to detect or mitigate cyberattacks. If the virtual replica is assumed perfect, it

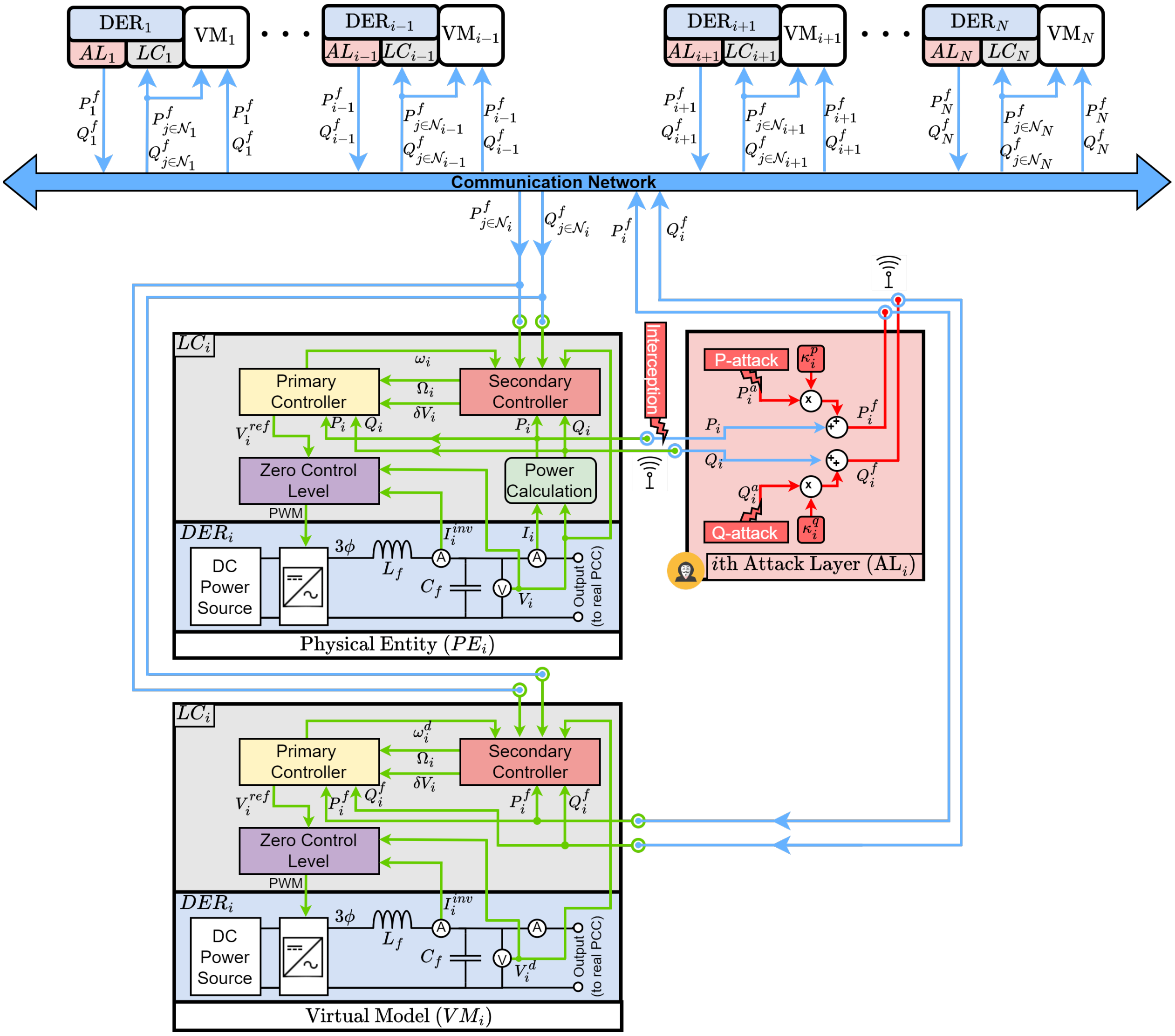

Figure 3.6: Illustration of VM$_i$, showing its interaction with the real communication network and its fundamental difference with its physical counterpart.

would output the exact same state as the real MG for the same inputted consensus variables. So, even if they are attacked, the virtual model would react the same way as the real system, showing the same output state dynamics. This way, no difference could be detected against a cyberattack. To overcome this problem, a clever distinction from the VM's control system is made therefore generating a different output state in the presence of an attack. This distinction is shown in Fig. 3.6, where the ith DER and LC of the virtual model is illustrated at the bottom. For explanatory simplicity, the VM will be separated into its DERs and LCs, where VM$_i$ corresponds to its ith DER and LC. When comparing it with the real DER and LC from the physical entity, the similarities can be noted, being almost identical, yet showing this one crucial distinction: unlike the real system, the VM's DER power variables (P_i and Q_i) are not internally calculated from voltage and current measurements. Instead, the ith LC's digital replica receives P_i^f and Q_i^f directly from the communication network. So forth, if P_i or Q_i are being attacked, the real system's ith LC will calculate P_i and Q_i internally and receive potentially attack versions $Q^f_{j\in\mathcal{N}_i}$ and $P^f_{j\in\mathcal{N}_i}$ from neighbors, while in contrast the VM's ith LC would receive and use directly all potentially attacked versions from the communication network $Q^f_{i\in U}$ and $P^f_{i\in U}$. This way, the control equations for the VM are:

- Voltage control:

$$V_i^{ref} = V^* - n_i Q_i^f + \delta V_i \tag{3.7}$$

$$k_i^V \frac{d\,\delta V_i}{dt} = \beta_i \left(V^* - V_i^d\right) - \gamma_i \sum_{j\in\mathcal{N}_i} a_{ij} \left(Q_i^f - Q_j^f\right) \tag{3.8}$$

- Frequency control:

$$\omega_i = \omega^* - m_i P_i^f + \Omega_i \tag{3.9}$$

$$k_i^\omega \frac{d\Omega_i}{dt} = \alpha_i \left(\omega^* - \omega_i^d\right) - \eta_i \sum_{j\in\mathcal{N}_i} a_{ij} \left(P_i^f - P_j^f\right) \tag{3.10}$$

The VM's zero control still needs V_i and I_i^{inv} to function. These virtual measurements, as in the physical entity, are discrete. In this manner, they have the same sampling rate $T_s = 78.125[\mu s]$ as the real MG measurements.

As noted in last subsection, $Q_i^f = Q_i$ and $P_i^f = P_i$ when $\kappa_i^q = 0$ and $\kappa_i^p = 0$. In this scenario, the VM should behave similar to the physical entity, resembling (approximately) the same output state. On the contrary, when $\kappa_i^q = 1$ and/or $\kappa_i^p = 1$, the states should be different. This point gives insight into the next section, where the difference between the real MG and its virtual model output state will be used as residual indices for attack detection.

3.3.2. Analytics: Residual Indices for Cyber-Attack Detection

Fig. 3.7 illustrates how the residual indices are obtained in the Analytics block in this DT framework. For simplicity, this figure only shows the ith DER and LC of the physical MG and its virtual counterpart VM$_i$, but this replicates for all N DERs. The ith LC outputs Q_i and P_i, which are intercepted by AL$_i$. They are transformed to Q_i^f and P_i^f which are delivered to the communication network. In return, the communication network contains the other LCs (potentially attacked) consensus variables $Q^f_{j\in\mathcal{N}_i}$ and $P^f_{j\in\mathcal{N}_i}$. The latter variables

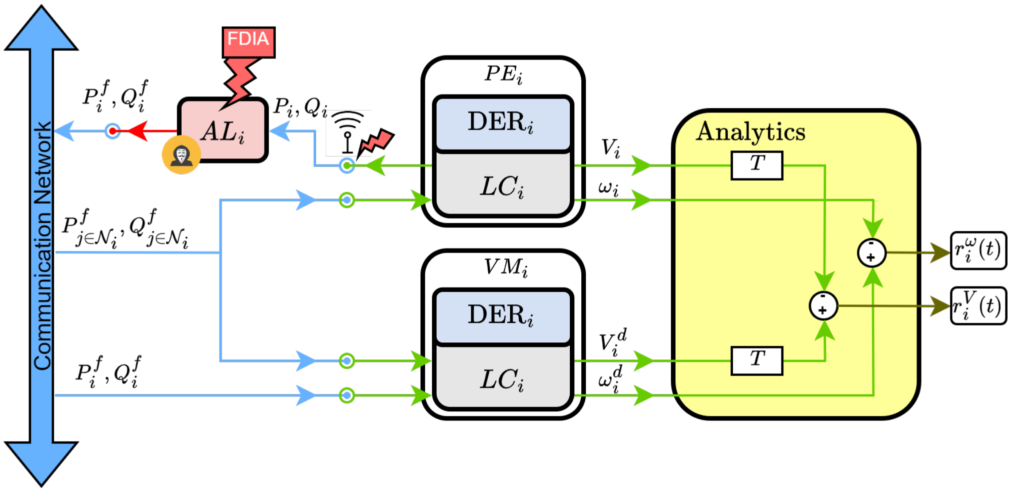

Figure 3.7: Voltage and frequency residual indices calculation from the ith DER and its VM.

are simultaneously inputted to the real LC and it virtual version. As discussed thoroughly before, VM$_i$ receives Q_i^f and P_i^f to be used in its control system.

To calculate the frequency residual indices, the frequency variable generated by the primary controller is outputted, ω_i for the real DER and ω_i^d for its virtual version. The frequency residual index is their difference as shown in equation (3.11). The voltage residual indices are obtained analogously as shown in equation (3.12).

$$r_i^\omega(t) = \omega_i^d - \omega_i \tag{3.11}$$
$$r_i^V(t) = T(V_i^d) - T(V_i) \tag{3.12}$$

It must be noted that the voltage output is a three-phase sinusoidal signal. To obtain the overall amplitude a T transformation is applied, where each phase amplitude is estimated calculating their RMS values with a sliding window, passing RMS to peak value, and then averaging:

$$T(V_i) = T(V_i^a, V_i^b, V_i^c) = \sqrt{2}\cdot\frac{RMS(V_i^a) + RMS(V_i^b) + RMS(V_i^c)}{3} \tag{3.13}$$

The indices from (3.11)-(3.12) are considered activated when $\omega_i^d \neq \omega_i$ and/or $V_i^d \neq V_i$, otherwise, they should just show white noise associated with the sensors. When these indices are activated, they will give precise information on which DER is being attacked and which consensus variable was targeted.

Finally it is important to emphasize the distinction made in the virtual model's implementation. If this distinction was not made, then $\omega_i^d \approx \omega_i$ and $V_i^d \approx V_i$ in any scenario, independent of the presence of an FDIA.

3.3.3. DT Control: Cyber-Attack Mitigation

In this section the DT framework is completed. To this point the virtual model and the analytics of the DT has been characterized, and now the control implementation is detailed. As before, the DT will be separated into N block: DT$_1$, . . . ,DT$_N$. In this sense, DT$_i$ contains VM$_i$, its own analaytics and control block. Fig. 3.8 and 3.9 show DT$_i$ and the whole DT framework implemented in the system, respectively. These will be explained later on.

The DT's mission is to mitigate the cyberattacks. This is accomplished by healing variables $Q_i^h(t)$ and $P_i^h(t)$. The communication network is potentially attacked, and as shown in Fig. 3.6, the real LC receives the faulty consensus variables $Q^f_{j\in\mathcal{N}_i}$ and $P^f_{j\in\mathcal{N}_i}$. The DT control should intercept these variables before sending them to the DER's control system and instead send the healed versions $Q_i^h(t)$ and $P_i^h(t)$. Equations (3.14)-(3.15) show how these variables are defined.

$$\text{Q-mitigation: } Q_i^h(t) = Q_i^f(t) + \lambda_i^q Q_i^m(t) \tag{3.14}$$

$$\text{P-mitigation: } P_i^h(t) = P_i^f(t) + \lambda_i^p P_i^m(t) \tag{3.15}$$

In the latter definitions, $\lambda_i^q, \lambda_i^p \in \{0,1\}$, $\forall i \in U = \{1,\dots,N\}$ are Boolean variables, representing the presence of a mitigation sequence $Q_i^m(t)$ or $P_i^m(t)$ on faulty consensus variables. By adding these mitigation variables to the faulty versions, there goal is:

$$\lim_{t\to\infty} Q_i^h(t) = \lim_{t\to\infty} Q_i(t) \tag{3.16}$$

$$\lim_{t\to\infty} P_i^h(t) = \lim_{t\to\infty} P_i(t) \tag{3.17}$$

Developing Equation (3.16) and assuming $\lambda_i^q = 1$:

$$\begin{aligned}
\lim_{t\to\infty} Q_i^h(t) &= \lim_{t\to\infty} Q_i(t) \\
\lim_{t\to\infty} (Q_i^f(t) + Q_i^m(t)) &= \lim_{t\to\infty} Q_i(t) \\
\lim_{t\to\infty} (\cancel{Q_i(t)} + \kappa_i^q Q_i^a(t) + Q_i^m(t)) &= \lim_{t\to\infty} \cancel{Q_i(t)} \\
\lim_{t\to\infty} -Q_i^m(t) &= \kappa_i^q \lim_{t\to\infty} Q_i^a(t)
\end{aligned}$$

Doing this analogously for Equation (3.17), the values the mitigation variables should reach are expressed in Equations (3.18)-(3.19).

$$\lim_{t\to\infty} -Q_i^m(t) = \begin{cases} 0 & \text{if } \kappa_i^q = 0 \\ \lim_{t\to\infty} Q_i^a(t) & \text{if } \kappa_i^q = 1 \end{cases} \tag{3.18}$$

$$\lim_{t\to\infty} -P_i^m(t) = \begin{cases} 0 & \text{if } \kappa_i^p = 0 \\ \lim_{t\to\infty} P_i^a(t) & \text{if } \kappa_i^p = 1 \end{cases} \tag{3.19}$$

Fig. 3.8 shows DT$_i$'s diagram, and specifically the mitigation control scheme. DT$_i$ has th-

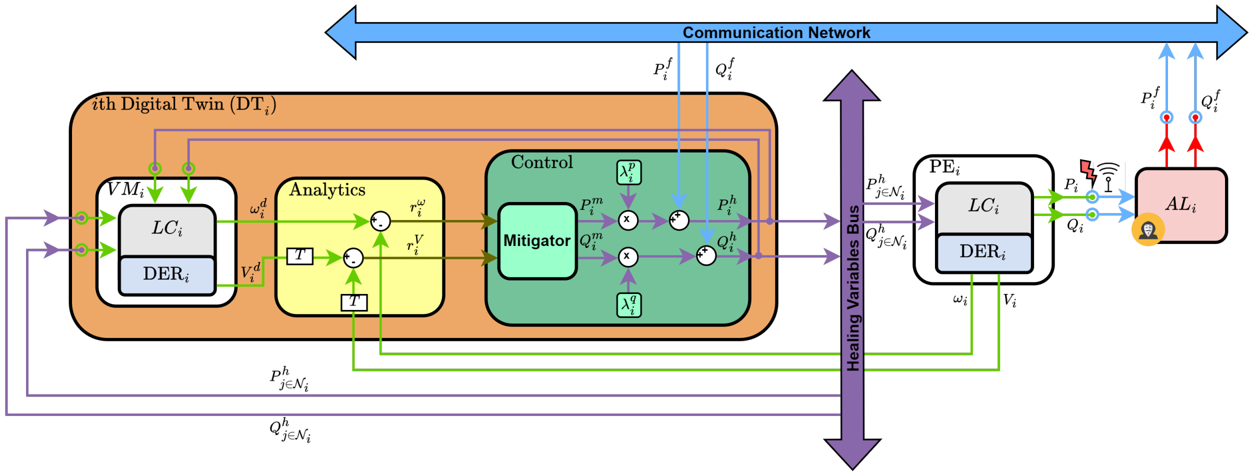

Figure 3.8: DT$_i$'s diagram and interaction with other elements of the study.

ree principal elements: 1) VM$_i$, 2) Analytics and 3) Control. The Control block's function is to return the MG back to it normal operation against a cyberattack. This could be traduced into making $r_i^V(t) \to 0$ and $r_i^\omega(t) \to 0$ when $t \to \infty$.

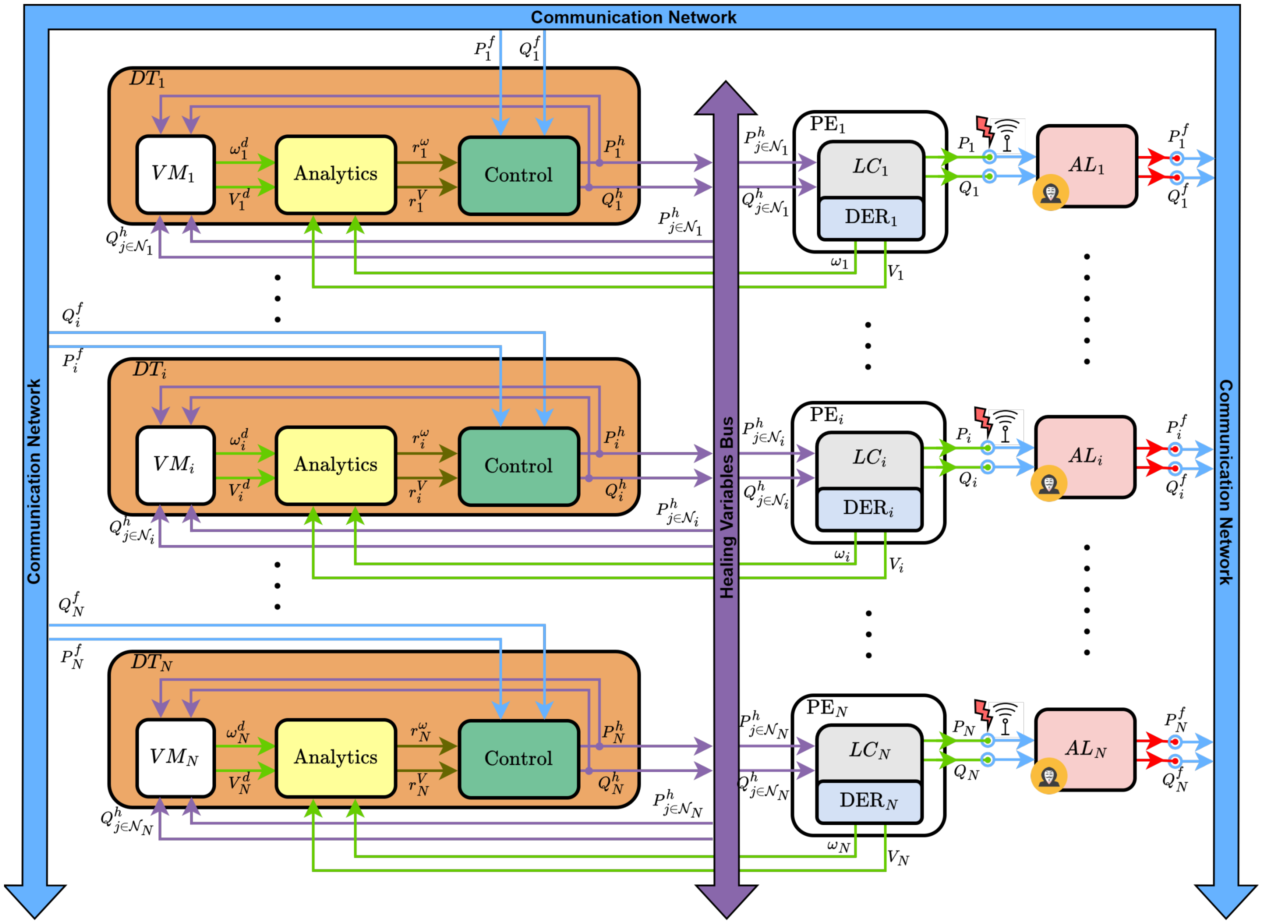

Figure 3.9: Complete DT scheme implemented in conjunction with the real MG.

To achieve the latter, it is important that VM$_i$ and PE$_i$ now receive $Q_i^h(t)$ and $P_i^h(t)$ instead of $Q_i^f(t)$ and $P_i^f(t)$. This is achieved by intercepting $Q_i^f(t)$ and $P_i^f(t)$ from the communication network before sending them to VM$_i$'s and PE$_i$'s control system, as depicted in Fig. 3.8.

The control block has a Mitigator, which essentially is a function M_i that maps the residual indices into the mitigation variables: $[Q_i^m(t), P_i^m(t)] = M_i(r_i^V(t), r_i^\omega(t))$. On the other hand, VM$_i$ and PE$_i$ receive the neighbor consensus variables, but as explained before, these should be the healed versions $Q_{j\in\mathcal{N}_i}^h$ and $P_{j\in\mathcal{N}_i}^h$. Each DT$_i$ returns $Q_i^h(t)$ and $P_i^h(t)$, which are delivered to a DT internal communication bus. This way $Q_{j\in\mathcal{N}_i}^h$ and $P_{j\in\mathcal{N}_i}^h$ are received by VM$_i$ and PE$_i$ through this communication bus. This complete scheme is depicted in Fig. 3.9.

The Mitigator of DT$_i$ must comply with the restrictions imposed in equations (3.18)-(3.19). This can be reached if $r_i^\omega(t) \to 0$ and $r_i^V(t) \to 0$ when $t \to \infty$. As the residual indices are the error between the real MG and the VM, it is natural that the mitigator can be a PID controller, as shown in Fig. 3.10. The gain values of each component of the PID controllers are detailed in Table 3.3.

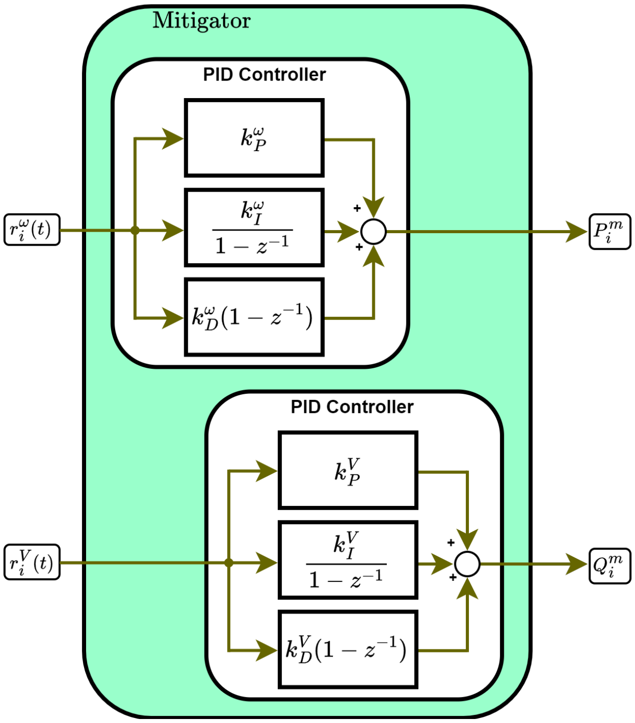

Figure 3.10: Mitigator of DT$_i$ with two PID controller: one for P-mitigation and another for Q-mitigation.

Table 3.3: PID mitigation control parameters.

Description	Value
P-mitigation proportional gain (k_P^ω)	2716
P-mitigation integral gain (k_I^ω)	61288
P-mitigation derivative gain (k_D^ω)	50.36
Q-mitigation proportional gain (k_P^V)	140.8
Q-mitigation integral gain (k_I^V)	819.27
Q-mitigation derivative gain (k_D^V)	33.43

3.4. Detection Under Different DT Operating Configurations

As detailed in all previous sections, the DT framework operates according to the value of κ_i^q, κ_i^p, λ_i^q and λ_i^p. In this section each configuration is recapitulated making emphasis how to detect cyberattacks under different configurations of the boolean variables that indicate attack and mitigation sequences activation.

In the next two subsections, the configuration of the DT framework is characterized based on the presence of mitigation sequences. In each configuration the detection method changes.

3.4.1. DT Configuration 1: Mitigation Sequences Deactivated

This configuration is set with $\lambda_i^q = 0$ and $\lambda_i^p = 0$, in other words, in absence of mitigation sequences and no intervention of the DT control block on the control system. Still, attack sequence can be injected, testing this configuration in their presence or not.

3.4.2. In Absence of Attack Sequences

From Q-attack and P-attack definitions from Equations (3.1)-(3.2), if $\kappa_i^q, \kappa_i^p = 0$ then $Q_i^f(t) = Q_i(t)$ and $P_i^f(t) = P_i(t)$. On the other hand, from Q-mitigation and P-mitigation definitions from Equations (3.14)-(3.15), if $\lambda_i^q, \lambda_i^p = 0$, then $Q_i^h(t) = Q_i^f(t) = Q_i(t)$ and $P_i^h(t) = P_i^f(t) = P_i(t)$. This way, the virtual model should behave the same as the physical entity, and the DT's analytics should output $r_i^V(t) = 0$ and $r_i^\omega(t) = 0$, $\forall i \in U$.

In this case then, the control system can be tested to validate its functioning in normal conditions, while observing how the VM replicates its operation and validate preliminarily that the residual indices from the analytics block behave correctly against no attack sequences.

3.4.3. In Presence of Attack Sequences

If $\kappa_i^q = 1$ and/or $\kappa_i^p = 1$, from Equations (3.1)-(3.2) and (3.14)-(3.15), the DT would not be mitigating the attacks, but attack sequences would be injected as $Q_i^h(t) = Q_i^f(t)$ and $P_i^h(t) = P_i^f(t)$.

Reviewing Fig. 3.8 and 3.9, it can be noticed that in this configuration VM$_i$ would receive $Q^f_{j\in\mathcal{N}_i}$, $P^f_{j\in\mathcal{N}_i}$, Q^f_i and Q^f_i while PE$_i$ would receive $Q^f_{j\in\mathcal{N}_i}$, $P^f_{j\in\mathcal{N}_i}$. With this configuration and the analytics block working, the residual indices would be the detection indicators when $r^V_i(t) \neq 0$ and/or $r^\omega_i(t) \neq 0$.

So in this case, the crucial distinction made in the VM is tested (explained in detail in section 3.3.1). The validation of the residual indices as detection indicators would imply the validation of the distinction element of the VM.

3.4.4. DT Configuration 2: Mitigation Sequences Activated

In this configuration $\lambda^q_i = 1$ and $\lambda^p_i = 1$. As the mitigation sequences are activated in this configuration, in the presence of a Q-attack and/or P-attack, the PID controller mitigator would make $r^V_i(t) \to 0$ and $r^\omega_i(t) \to 0$ when $t \to \infty$. The residual indices tend rapidly to zero, as is exposed in the experiments from Section 4.

The problem with the latter behavior, is that the residual indices are not good indicators of the presence of cyberattacks in this configuration. It must be noticed that the attacks are being mitigated, but it still would be informative to have something indicating the presence of an attack. Reviewing Section 3.3.3, from Equations (3.18)-(3.19), it is imminent that $-Q^m_i(t)$ and $-P^m_i(t)$ are clear attack detection indicator as they tend to the attack sequences $Q^a_i(t)$ and $P^a_i(t)$ when $t \to \infty$.

Another point to highlight from this detection configuration, is that not only would $-Q^m_i(t)$ and $-P^m_i(t)$ indicate which unit(s) and consensus variable(s) is/are being attacked, but they would also match the attack sequence being injected into the system.

3.5. Definition of Attack Sequences for the Experiments

To validate the DT framework designed in this thesis, two types of attack sequences are shown in the Results (chapter 4). In the next two subsections each type will be defined and explained based on Equations (3.3)-(3.4).

3.5.1. Step Attack Sequences

The step attack, as its name implies, is a step function added to Q_i and P_i. Based on Equations (3.3)-(3.4), this would mean that attack elements $q_i(t)$ and $p_i(t)$ are constant values, which are defined as $q_i(t) = \bar{q}_i$ and $p_i(t) = \bar{p}_i$. With this definition, the attack sequences are defined as:

$$Q_i^a(t) = \begin{cases} 0 & \text{if } 0 \leq t < t_{q_i}^{\text{init}} \\ \bar{q}_i & \text{if } t_{q_i}^{\text{init}} \leq t < t_{q_i}^{\text{final}} \\ 0 & \text{if } t \geq t_{q_i}^{\text{final}} \end{cases} \tag{3.20}$$

$$P_i^a(t) = \begin{cases} 0 & \text{if } 0 \leq t < t_{p_i}^{\text{init}} \\ \bar{p}_i & \text{if } t_{p_i}^{\text{init}} \leq t < t_{p_i}^{\text{final}} \\ 0 & \text{if } t \geq t_{p_i}^{\text{final}} \end{cases} \tag{3.21}$$

In some experiments, the step attack is active during the whole simulation making $t_{q_i}^{\text{final}} = \infty$ and/or $t_{p_i}^{\text{final}} = \infty$. In others, the steps are deactivated during the simulation making more complex attack patterns.

3.5.2. Truncated Ramp Attack Sequences

The truncated attack is an attack sequence where at $t_{q_i}^{\text{init}}, t_{p_i}^{\text{init}}$ the attack takes the shape of a ramp, which is then truncated to a constant value $\bar{q}_i, \bar{p}_i$ at time $t_{q_i}^{\text{trunc}}, t_{p_i}^{\text{trunc}}$.

In all the experiments with this attack sequence, the attack is activated during the whole simulation, or in other words, $t_{q_i}^{\text{final}}, t_{p_i}^{\text{final}} = \infty$. This way the attack sequences collapse to:

$$Q_i^a(t) = \begin{cases} 0 & \text{if } 0 \leq t < t_{q_i}^{\text{init}} \\ q_i(t) & \text{if } t_{q_i}^{\text{init}} \leq t < \infty \end{cases} \tag{3.22}$$

$$P_i^a(t) = \begin{cases} 0 & \text{if } 0 \leq t < t_{p_i}^{\text{init}} \\ p_i(t) & \text{if } t_{p_i}^{\text{init}} \leq t < \infty \end{cases} \tag{3.23}$$

Where attack elements $q_i(t)$ and $p_i(t)$ are defined as:

$$q_i(t) = \begin{cases} 0 & \text{if } 0 \leq t < t_{q_i}^{\text{init}} \\ \dfrac{\bar{q}_i}{t_{q_i}^{\text{trunc}} - t_{q_i}^{\text{init}}} \left(t - t_{q_i}^{\text{init}}\right) & \text{if } t_{q_i}^{\text{init}} \leq t < t_{q_i}^{\text{trunc}} \\ \bar{q}_i & \text{if } t_{q_i}^{\text{trunc}} \leq t < \infty \end{cases} \tag{3.24}$$

$$p_i(t) = \begin{cases} 0 & \text{if } 0 \leq t < t_{p_i}^{\text{init}} \\ \dfrac{\bar{p}_i}{t_{p_i}^{\text{trunc}} - t_{p_i}^{\text{init}}} \left(t - t_{p_i}^{\text{init}}\right) & \text{if } t_{p_i}^{\text{init}} \leq t < t_{p_i}^{\text{trunc}} \\ \bar{p}_i & \text{if } t_{p_i}^{\text{trunc}} \leq t < \infty \end{cases} \tag{3.25}$$

In these definitions, it must be noted that the ramp sections of the attack elements are defined in such way that they start incrementing linearly from 0 at $t_{q_i}^{\text{init}}, t_{p_i}^{\text{init}}$ and reach exactly $\bar{q}_i, \bar{p}_i$ at $t_{q_i}^{\text{trunc}}, t_{p_i}^{\text{trunc}}$.

Chapter 4

Results

In this chapter the results of the experiments are exposed. Various experiments are shown under different scenarios, showing how the control system works, how the residual indices behave under different attack scenarios, detection under both DT configurations and how the DT mitigates the attacks. All simulations are run in Matlab®/Simulink® integrated with PLECS Blockset® using an MG with $N = 3$ DERs.

There are three principal experiments that were carried out. In Experiment 1, the system is tested in DT configuration 1 ($\lambda_i^q = 0$ and $\lambda_i^p = 0$) with no attack sequences present. More specifically, the control system is tested against load changing, showing how it is able to manage voltage and frequency levels and power-sharing. The load changing values are defined in Table 4.1, where at 0[s] a 5 [Ω] and 2[mH] load is connected to the PCC, and from 6[s] onward the load changes its values every 2[s].

Table 4.1: Load changes in the PCC.

time [s]	0	6	8	10
R [Ω]	5.0	10.0	4.0	4.0
L [mH]	2.0	3.0	3.0	7.0

Experiment 2 shows the system's behavior in DT configuration 1 with attack sequences present ($\kappa_i^q = 1$ and/or $\kappa_i^p = 1$). This experiment is separated into two sub experiments; Experiment 2.1 with step attack sequences separated into three scenarios and Experiment 2.2 with truncated ramp attack sequences separated into two scenarios.

Experiment 3 is organized in the exact same manner as Experiment 2, but using DT configuration 2 ($\lambda_i^q = 1$ and $\lambda_i^p = 1$). Scenarios in both experiments are identical and are determined by the attack sequence values. Experiment 2.1 and 3.1 are designed to be compared, as with Experiment 2.2 and 3.2. Specifically, Table 4.2 and 4.3 sums up the time intervals that define the attack sequences and elements of the scenarios of each pair of experiments.

These experiment were designed in such manner so they can be relatively easy to compare. In the outline below, each experiment is briefly summarized.

- **Experiment 2.1 and 3.1**
  - **Scenario 1**: All consensus variables are attacked at 4[s] with a step attack, showing how it reacts to the same load changes from Experiment 1 (Table 4.1).
  - **Scenario 2**: All consensus variables are attacked with step attacks in isolated intervals in time. A 10 [Ω] and 3 [mH] load is connected at 0[s] and is maintained throughout the simulation.
  - **Scenario 3**: All consensus variables are attacked with step attacks progressively, adding a new attack every 4 [s] until all consensus variables are attacked. Then the attacks are deactivated progressively and symmetrically. The same load configuration as in scenario 2 is set.
- **Experiment 2.2 and 3.2**
  - **Scenario 1**: All consensus variables are attacked at 4[s] with a truncated ramp attack, showing how it reacts to the same load changes from Experiment 1 (Table 4.1).
  - **Scenario 2**: All consensus variables are attacked with truncated ramp attacks progressively, adding a new attack every 2 [s] until all consensus variables are attacked. A 10 [Ω] and 3 [mH] load is connected at 0[s] and is maintained throughout the simulation.

Table 4.2: Attack times of the step attack sequences from Experiment 2.1 and 3.1 for each scenario.

Experiment 2.1 and 3.1			
Attack Interval [s]	Scenario 1	Scenario 2	Scenario 3
$[t_{q_1}^{init}, t_{q_1}^{final})$	[4,∞)	[4,8)	[4,28)
$[t_{q_2}^{init}, t_{q_2}^{final})$	[4,∞)	[12,16)	[12,36)
$[t_{q_3}^{init}, t_{q_3}^{final})$	[4,∞)	[20,24)	[20,44)
$[t_{p_1}^{init}, t_{p_1}^{final})$	[4,∞)	[28,32)	[8,32)
$[t_{p_2}^{init}, t_{p_2}^{final})$	[4,∞)	[36,40)	[16,40)
$[t_{p_3}^{init}, t_{p_3}^{final})$	[4,∞)	[44,48)	[24,48)

Table 4.3: Attack times of the truncated ramp attack sequences from Experiment 2.2 and 3.2 for each scenario ($t_{p_i}^{final}, t_{p_i}^{final} = \infty \; \forall i \in \{1,2,3\}$).

Experiment 2.2 and 3.2		
Attack Interval [s]	Scenario 1	Scenario 2
$[t_{q_1}^{init}, t_{q_1}^{trunc})$	[4,12)	[4,8)
$[t_{q_2}^{init}, t_{q_2}^{trunc})$	[4,12)	[8,12)
$[t_{q_3}^{init}, t_{q_3}^{trunc})$	[4,12)	[12,18)
$[t_{p_1}^{init}, t_{p_1}^{trunc})$	[4,12)	[6,10)
$[t_{p_2}^{init}, t_{p_2}^{trunc})$	[4,12)	[10,14)
$[t_{p_3}^{init}, t_{p_3}^{trunc})$	[4,12)	[14,18)

Another point to mention before observing the results in the next subsections is explain which are the values of $\bar{q}_i$ and $\bar{p}_i$ used in each experiment and scenario. This is important for understanding some of the figures ahead. In scenario 1, 2 and 3 of Experiment 2.1 and 3.1, six simulation were carried out in each, testing different values of $\bar{q}_i$ and $\bar{p}_i$ for $i = \{1, 2, 3\}$; the same Q-attack and P-attack constant values are assigned for all attacks, in other words, $\bar{q}_1 = \bar{q}_2 = \bar{q}_3$ and $\bar{p}_1 = \bar{p}_2 = \bar{p}_3$ in each scenario. The values used with their color codes can be observed in Table 4.4.

Table 4.4: Step attack constants $\bar{q}_i$ and $\bar{q}_i$ for the three scenarios of Experiment 2.1 and 3.1.

Scenarios 1, 2 and 3 of Experiment 2.1 and 3.1.						
Color code	🟦	🟧	🟨	🟪	🟩	🟦
Constants	Value	Value	Value	Value	Value	Value
$\bar{q}_i \ \forall i \in \{1, 2, 3\}$	-500	-250	-100	100	250	500
$\bar{p}_i \ \forall i \in \{1, 2, 3\}$	-100	-50	-10	10	50	100

In the corresponding figures in the results all information from Table 4.2 and 4.4 is summarized in the figure. For example, in further Figure 4.6, 4.7 and 4.8, the attack time intervals defined in Table 4.2 are illustrated on top of the graph and the constant values of the attack elements from Table 4.4 are summarized in the legend on the right, following the same color code. This same general structure persists in figures from other scenarios.

In the case of Scenario 1 and 2 of Experiment 2.2 and 3.2, only one simulation is carried out, but mantaining the same structure of having the same constant values $\bar{q}_1 = \bar{q}_2 = \bar{q}_3$ and $\bar{p}_1 = \bar{p}_2 = \bar{p}_3$ for all DERs and scenarios. These values are shown in Table 4.5 below. The information of the attack time intervals $[t_{q_i}^{init}, t_{q_i}^{final})$ and $[t_{p_i}^{init}, t_{p_i}^{final})$ for $i \in \{1, 2, 3\}$ is also illustrated on top of the figures of these experiments.

Table 4.5: Truncated ramp attack constants $\bar{q}_i$ and $\bar{q}_i$ for both scenarios of Experiment 2.2 and 3.2.

Scenarios 1 and 2 of Experiment 2.2 and 3.2.	
Constants	Value
$\bar{q}_i \ \forall i \in \{1, 2, 3\}$	250
$\bar{p}_i \ \forall i \in \{1, 2, 3\}$	-50

Finally, the output instantaneous active and reactive power is calculated in $\alpha\beta$ reference frame: $Q_i = V_i^\alpha I_i^\beta - V_i^\beta I_i^\alpha$ and $P_i = V_i^\alpha I_i^\alpha + V_i^\beta I_i^\beta$.

4.1. Experiment 1: Testing the DAPI Controller Against Load Changing

In Experiment 1 the distributed control system (DAPI controller) is tested against load changing in DT configuration 1. The load changes are describe in Table 4.1 and also shown in the figures below. In this experiment there is no presence of attack and mitigation sequences

$(\kappa_i^q, \kappa_i^p = 0 \text{ and } \lambda_i^q, \lambda_i^p = 0)$.

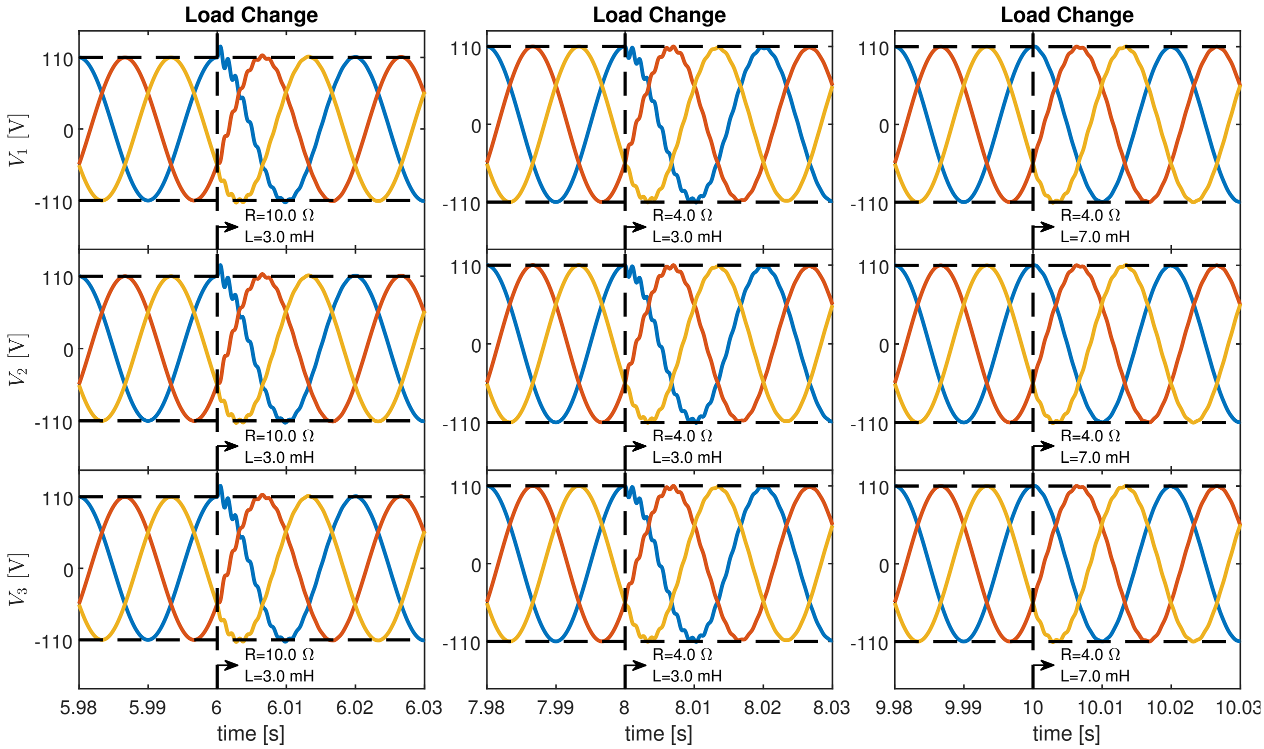

Figure 4.1: Three-phase voltage of each DER against load changes in absence of attack and mitigation sequences.

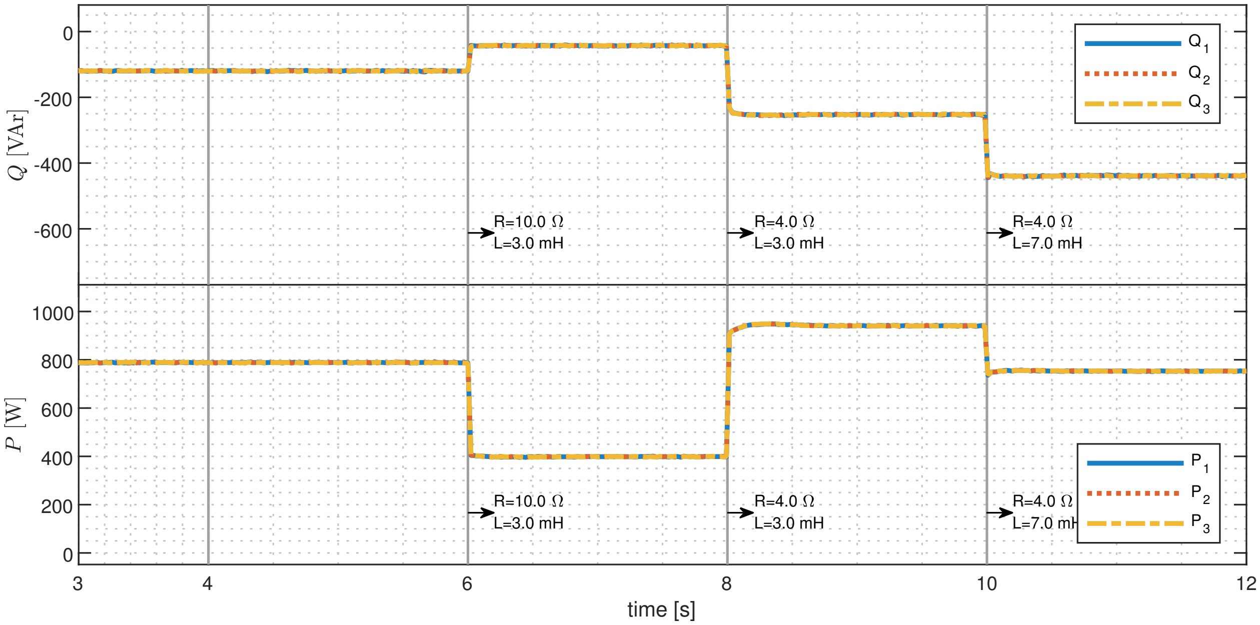

Figure 4.2: Active and reactive power generated by each DER against load changes in absence of attack and mitigation sequences.

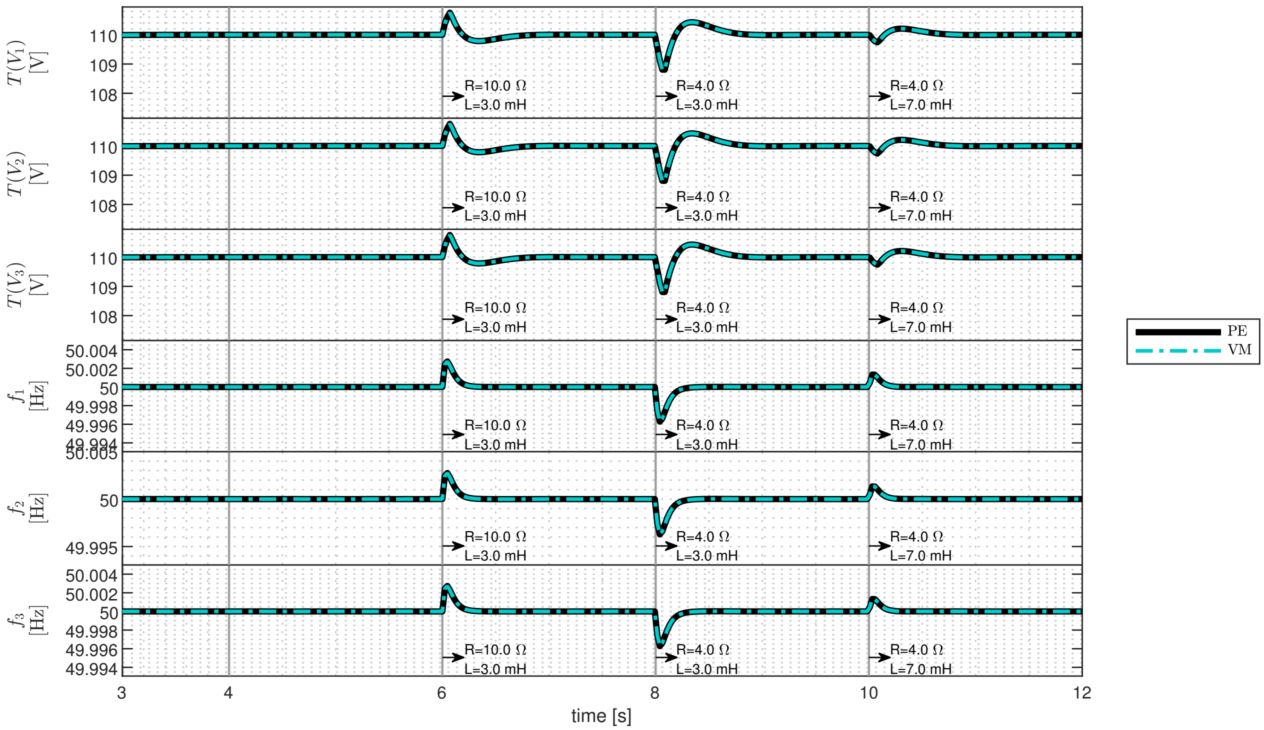

Figure 4.3: Average voltage amplitude and frequency of each DER (of both the PE and VM) against load changes in absence of attack and mitigation sequences.

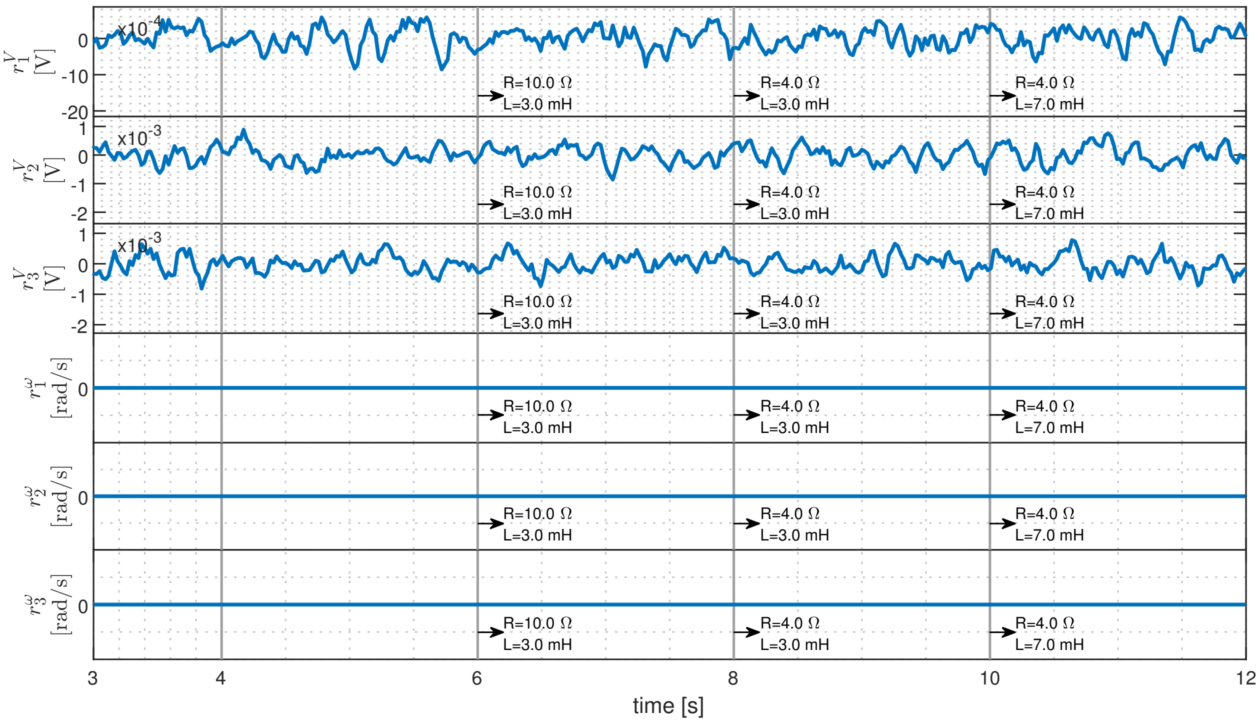

Figure 4.4: Voltage and frequency residual indices of each DER against load changes in absence of attack and mitigation sequences.

4.2. Experiment 2: Testing the Residual Indices for Detection in DT Configuration 1

In Experiment 2 the residual indices are tested as detection indicator for Dt Configuration 1. With these results the residual indices demonstrate to be indicator of a P-attack or Q-attack (which consensus variables is attacked) and which DER is being hijacked. Also, by the shape of the residual indices, a step or truncated ramp attack should be identified easily.

4.2.1. Experiment 2.1: Step Attack Sequences

In Experiment 2.1, the DT framework is tested against step attack sequences in three different scenarios. In Scenario 1 all consensus variables are attacked and attack sequences start at the same time, in Scenario 2 attack sequences are activated in isolated time intervals, and in Scenario 3 attack sequences are added progressively in time until all consensus variables are under attack.

4.2.1.1. Scenario 1

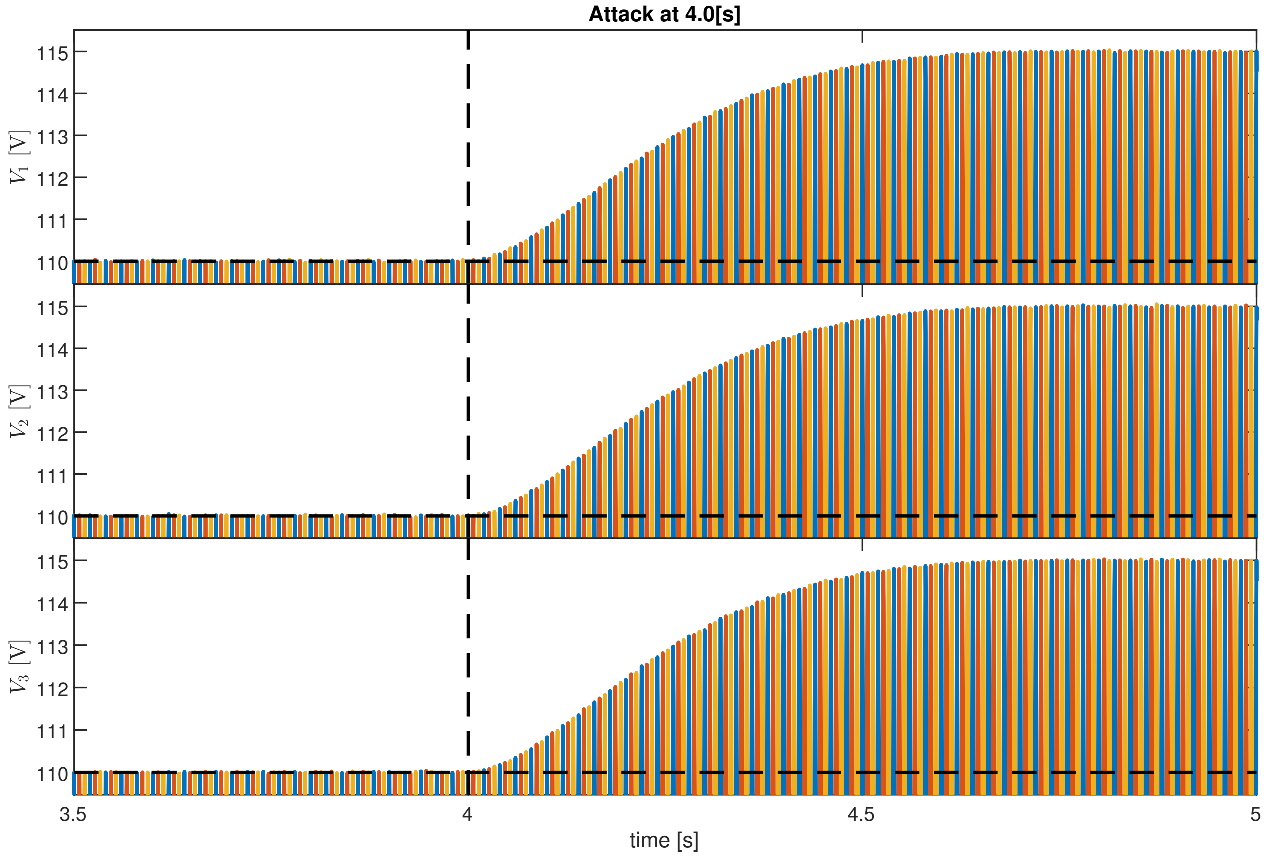

Figure 4.5: Three-phase voltage at start of the step attack on all consensus variables, specifically for $\bar{q}_i = -500$ and $\bar{p}_i = -100$. The plot is zoomed in the Y-axis showing the top peaks of the sinusoidal signals.

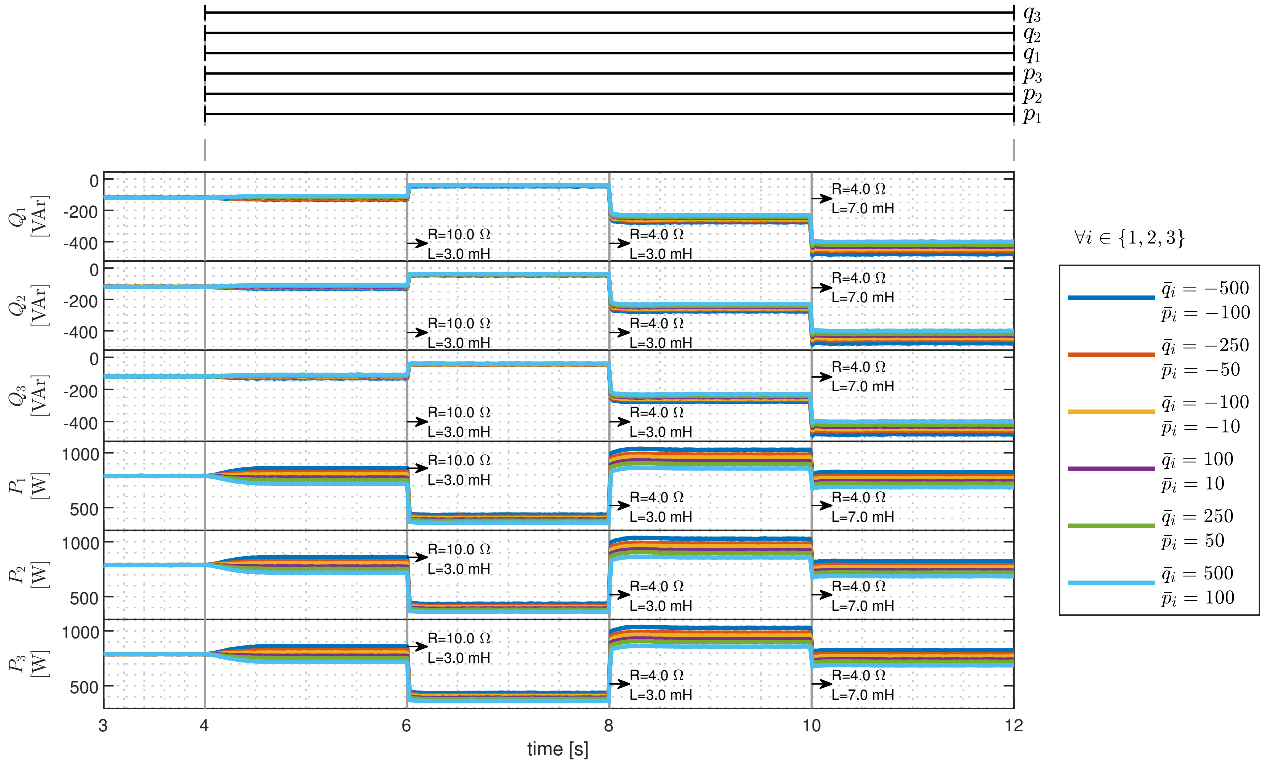

Figure 4.6: Active and reactive power generated by each DER in DT configuration 1 against step attacks and load changes.

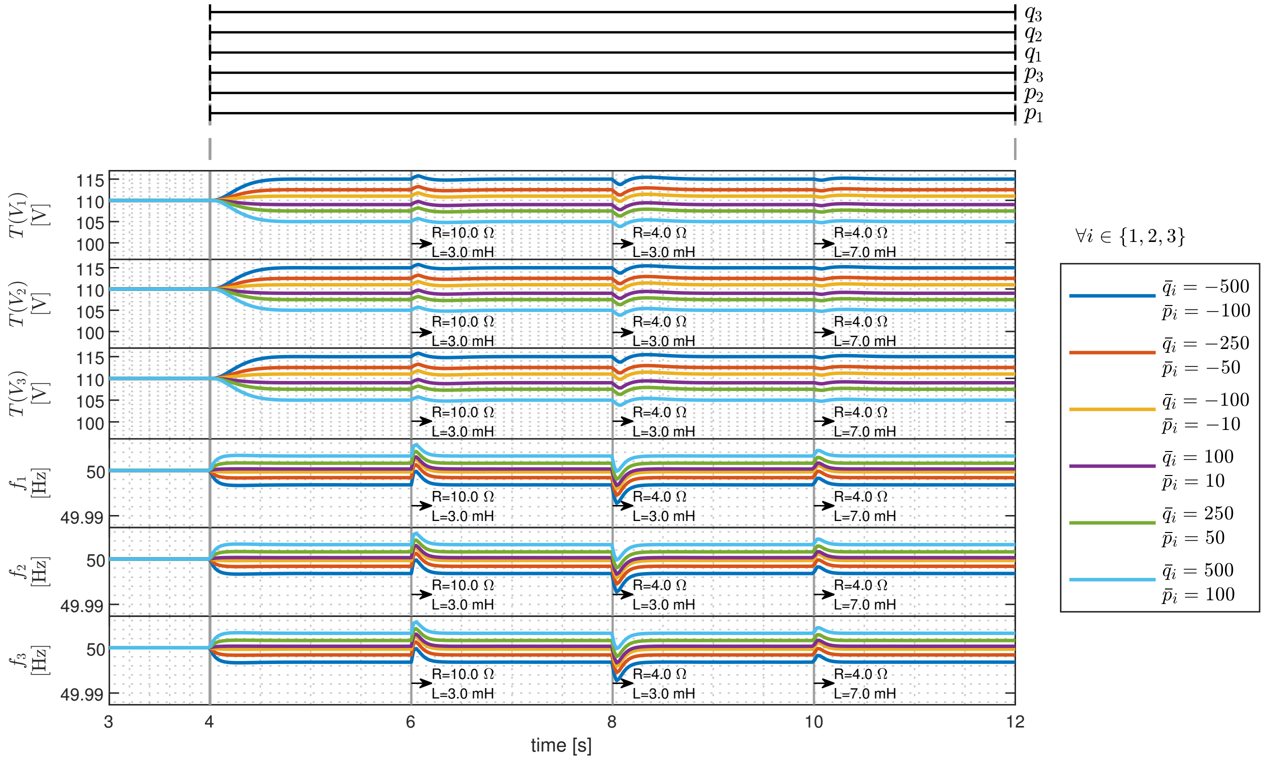

Figure 4.7: Average voltage amplitude and frequency of each DER in DT configuration 1 against step attacks and load changes.

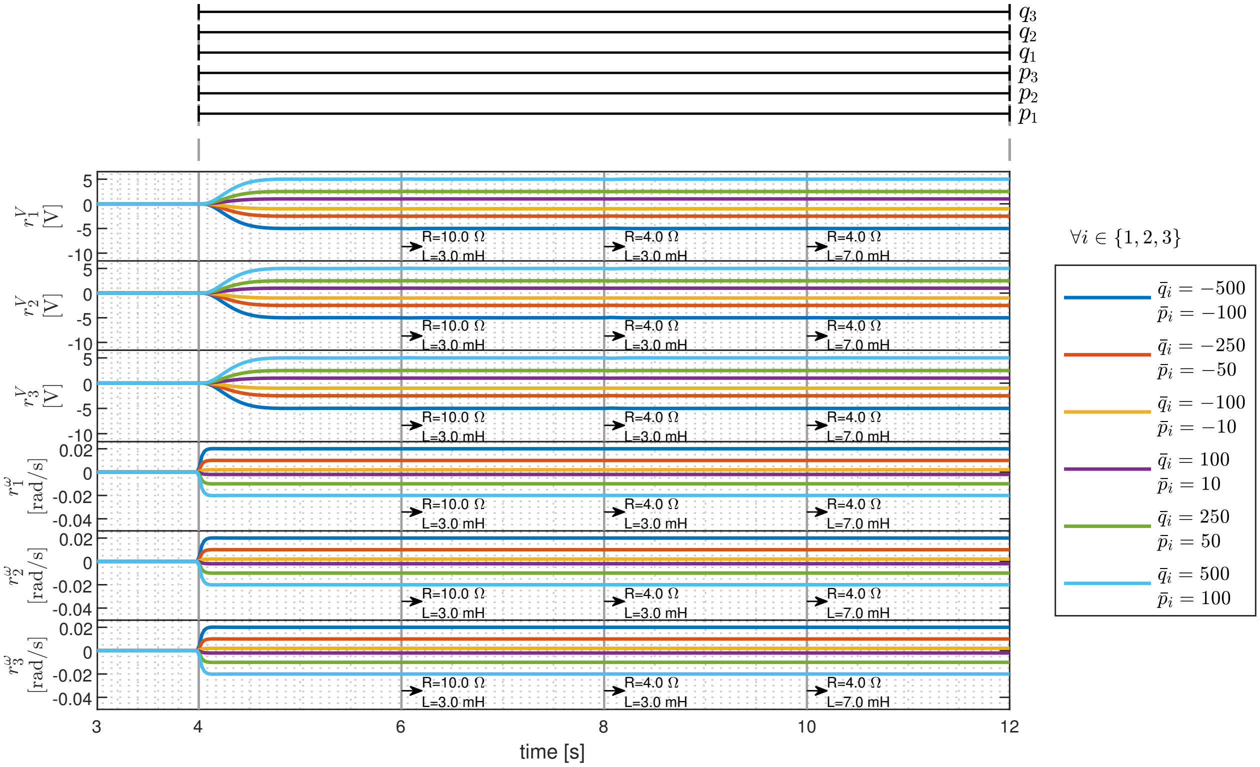

Figure 4.8: Voltage and frequency residual indices in DT configuration 1 against step attacks and load changes.

4.2.1.2. Scenario 2

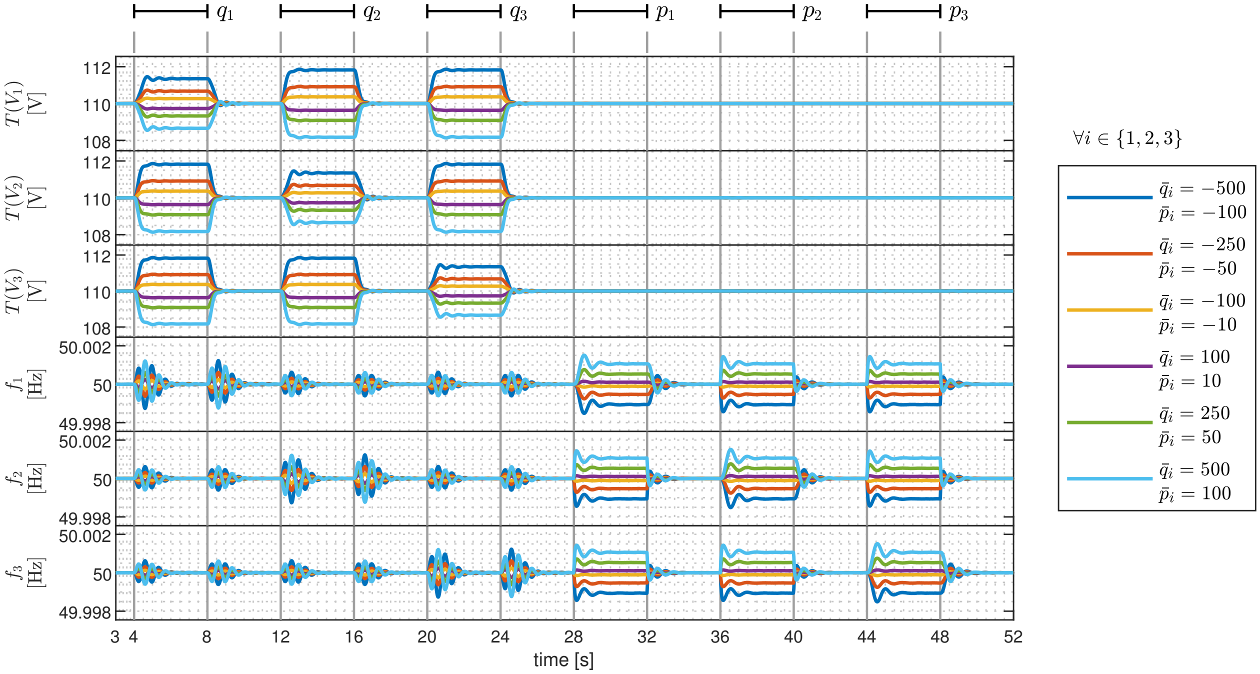

Figure 4.9: Average voltage amplitude and frequency of each DER in DT configuration 1 against step attacks in isolated time intervals.

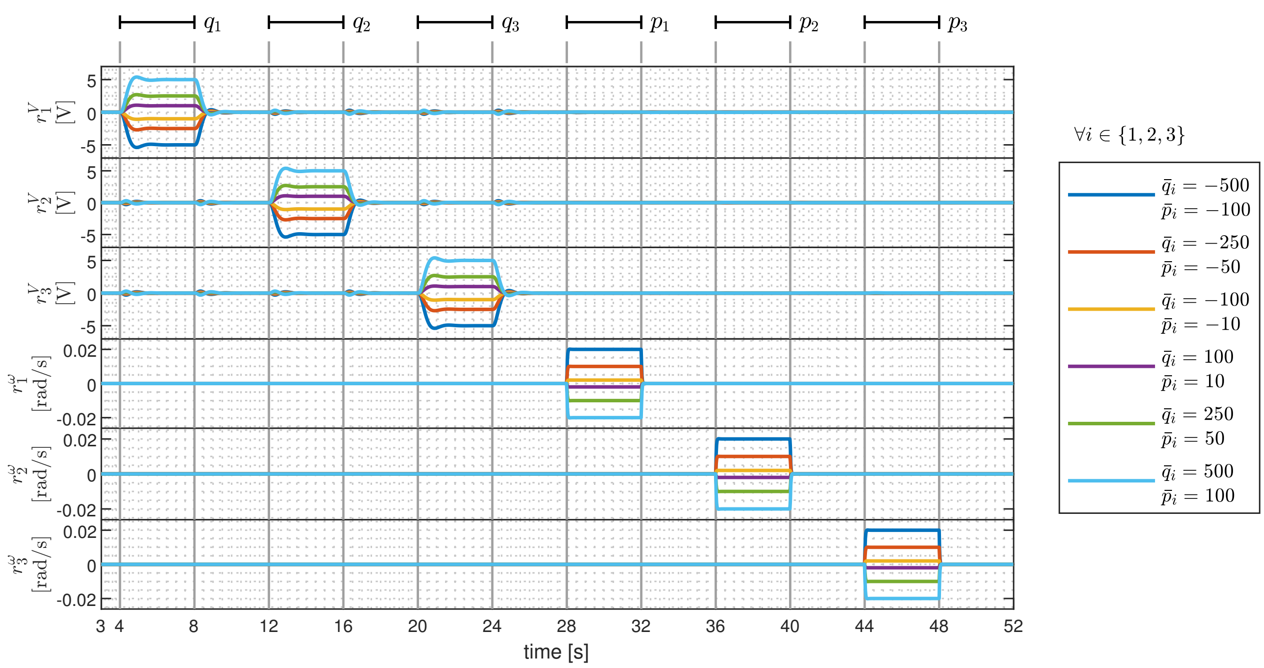

Figure 4.10: Voltage and frequency residual indices in DT configuration 1 against step attacks in isolated time intervals.

4.2.1.3. Scenario 3

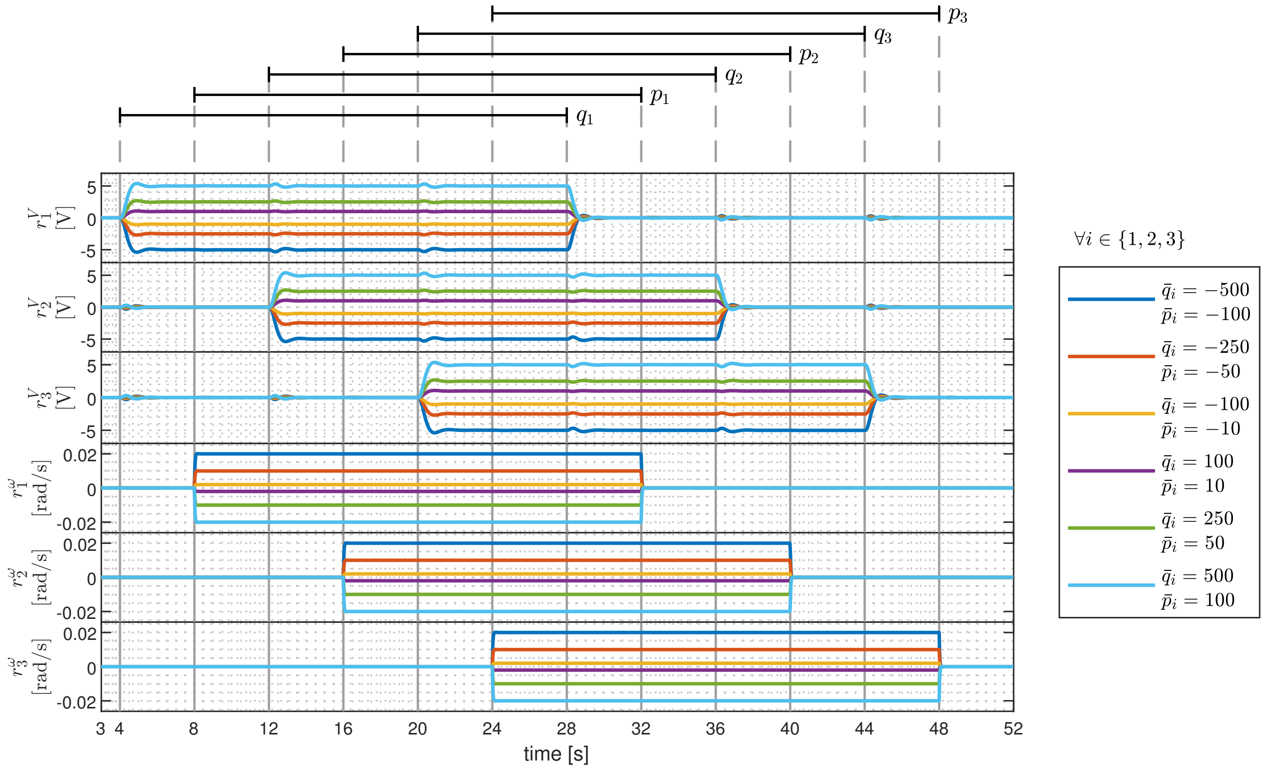

Figure 4.11: Voltage and frequency residual indices in DT configuration 1 against step attacks added progressively in time.

4.2.2. Experiment 2.2: Truncated Ramp Attack Sequences

In Experiment 2.2, the DT framework is tested against truncated ramp attacks in two scenarios: Scenario 1 attacks all consensus variables simultaneously, while Scenario 2 progressively adds attacks, truncating ramps every 4 seconds.

4.2.2.1. Scenario 1

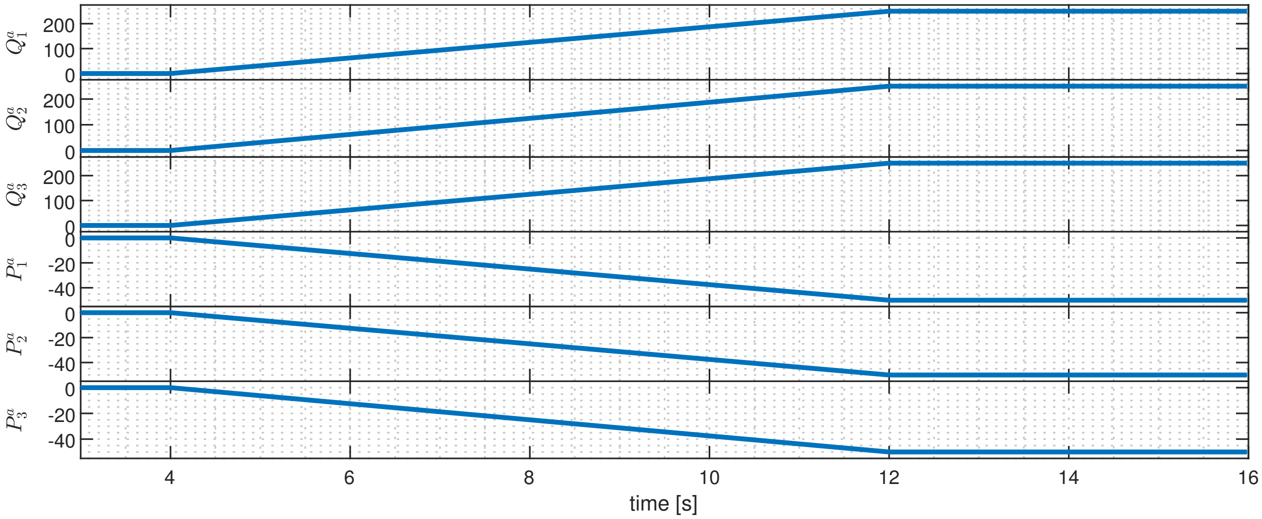

Figure 4.12: Truncated ramp attack sequences injected to consensus variables initiated and truncated at the same times.

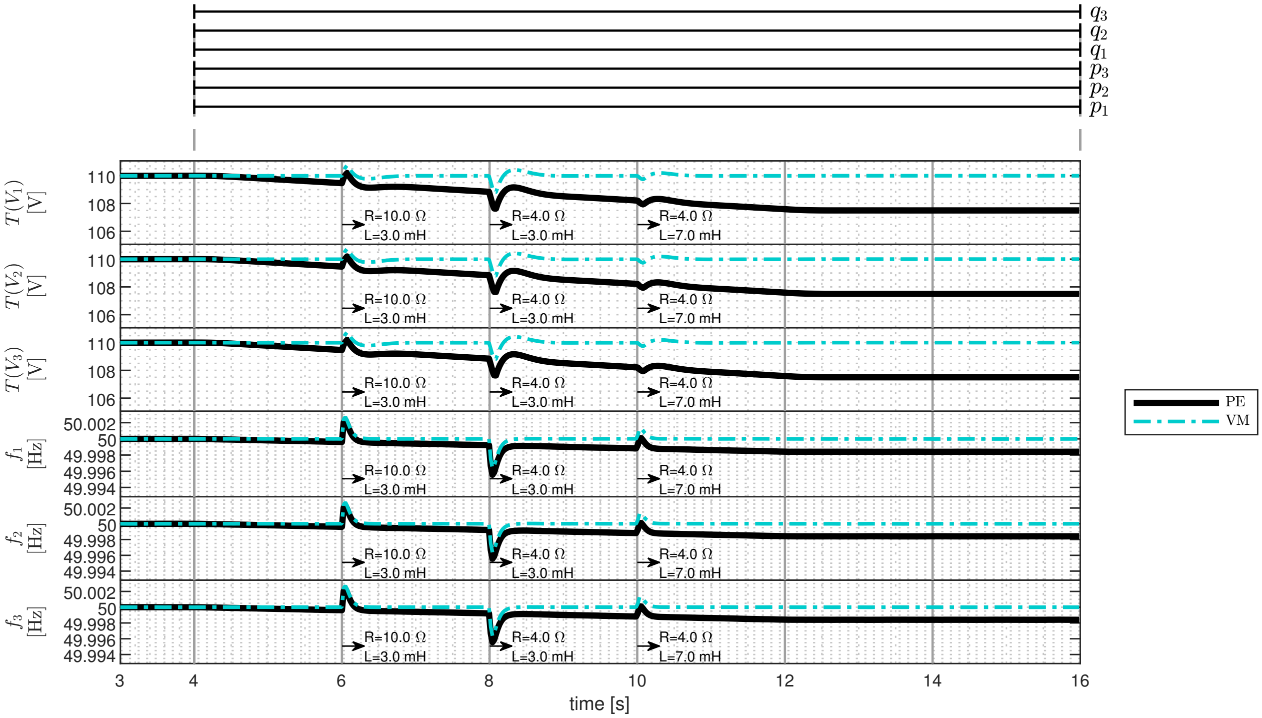

Figure 4.13: Average voltage amplitude and frequency of each DER in DT configuration 1 against truncated ramp attacks and load changes.

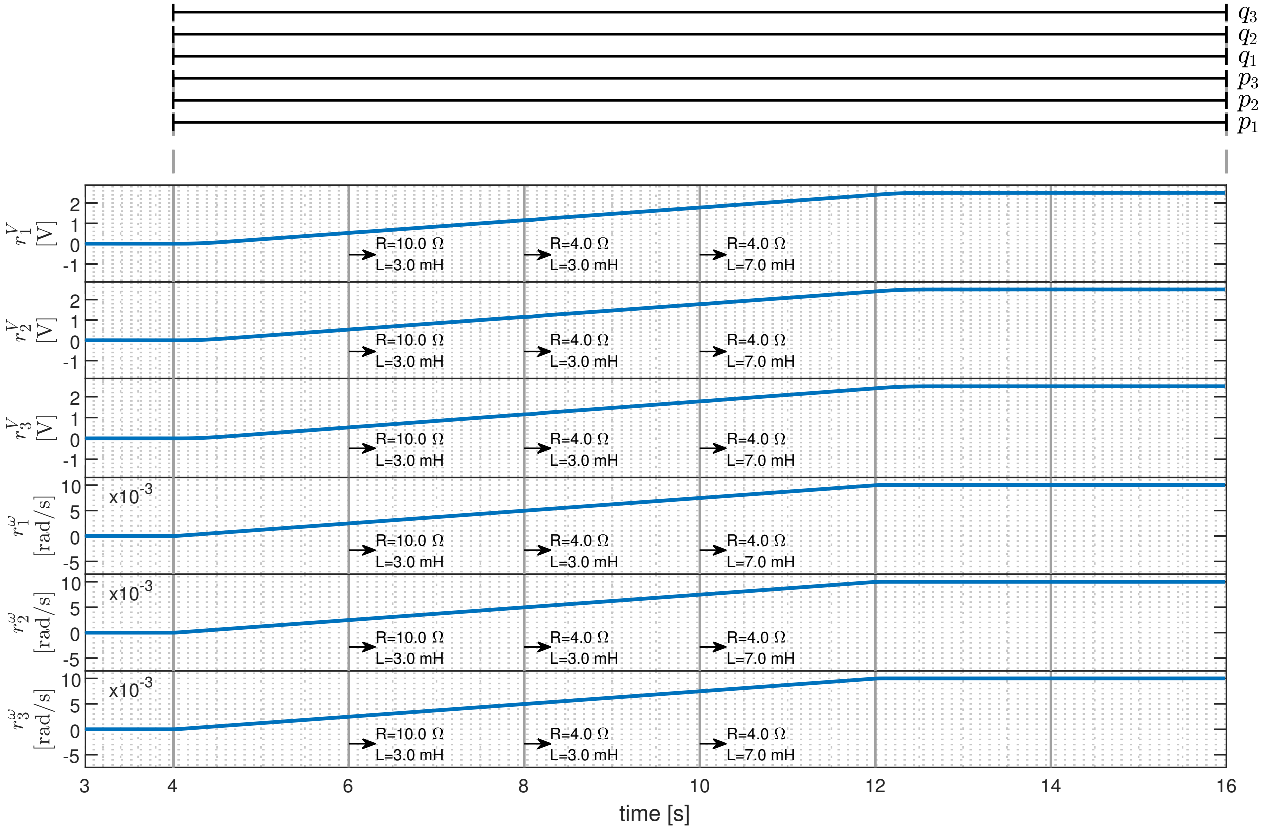

Figure 4.14: Voltage and frequency residual indices in DT configuration 1 against truncated ramp attacks and load changes.

4.2.2.2. Scenario 2

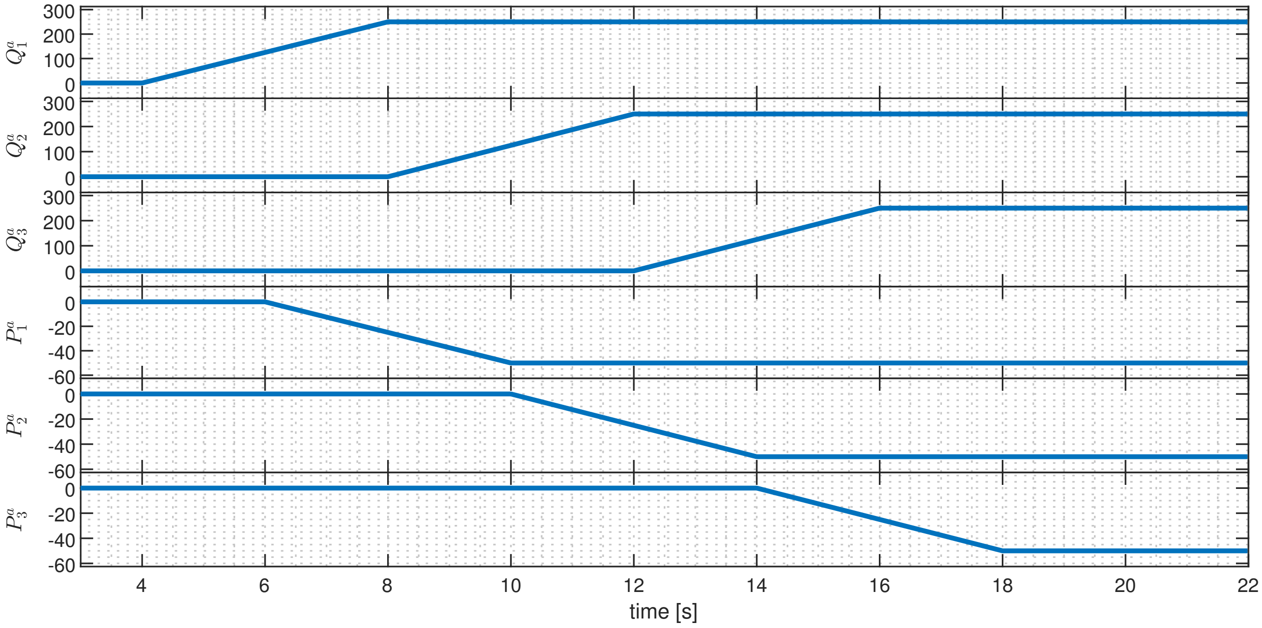

Figure 4.15: Truncated ramp attack sequences injected to consensus variables added progressively in time.

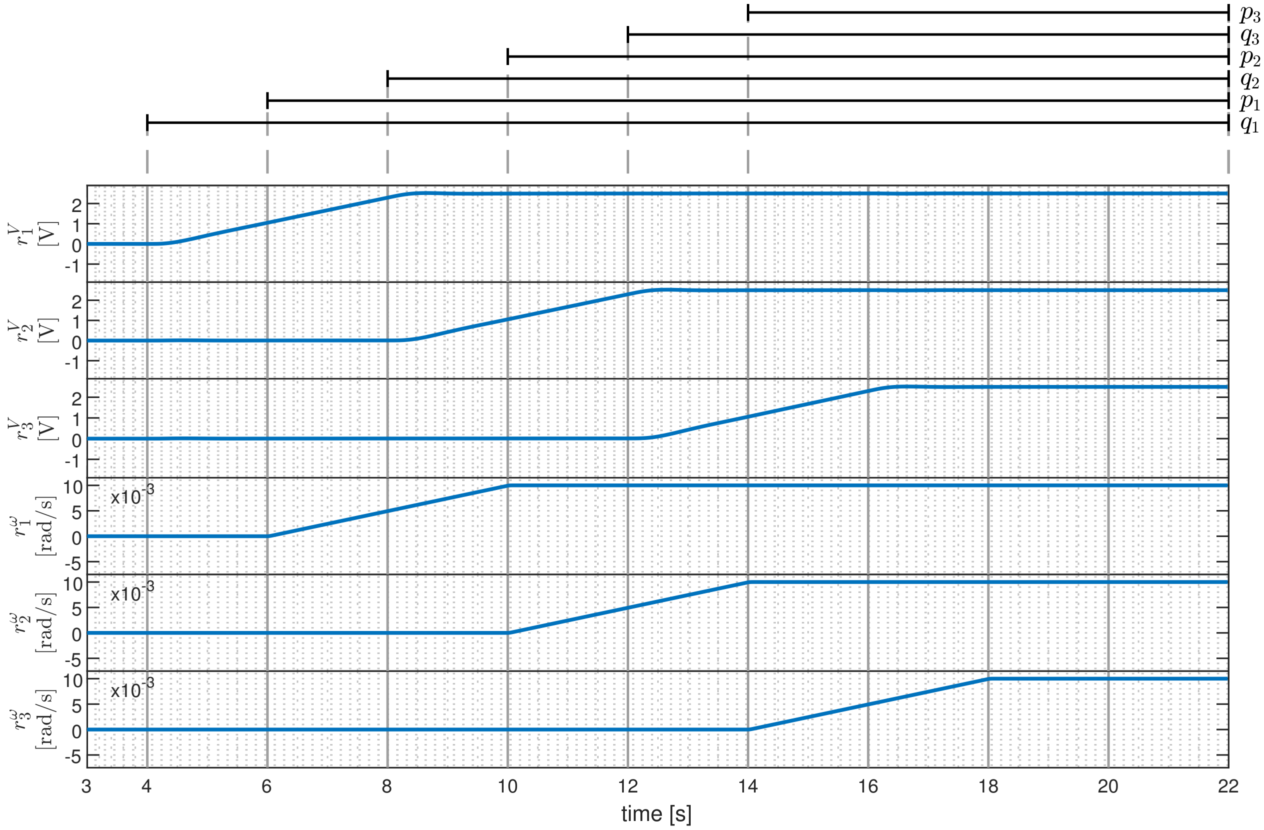

Figure 4.16: Voltage and frequency residual indices in DT configuration 1 against truncated ramp attacks added progressively in time.

4.3. Experiment 3: Testing Detection and Mitigation in DT Configuration 2

In Experiment 3 the the control block of the DT is tested for mitigation. In this experiment the residual indices are no longer good detection indicators, but the results demonstrate there effective use as input for PID mitigator. Also, $-Q_i^m(t)$ and $-P_i^m(t)$ are shown as new detection indicators for this configuration.

4.3.1. Experiment 3.1: Step Attack Sequences

In Experiment 3.1, the DT framework is tested against step attack sequences in the three exact scenarios from Experiment 2.1. In Scenario 1 all consensus variables are attacked and attack sequences start at the same time, in Scenario 2 attack sequences are activated in isolated time intervals, and in Scenario 3 attack sequences are added progressively in time until all consensus variables are under attack.

4.3.1.1. Scenario 1

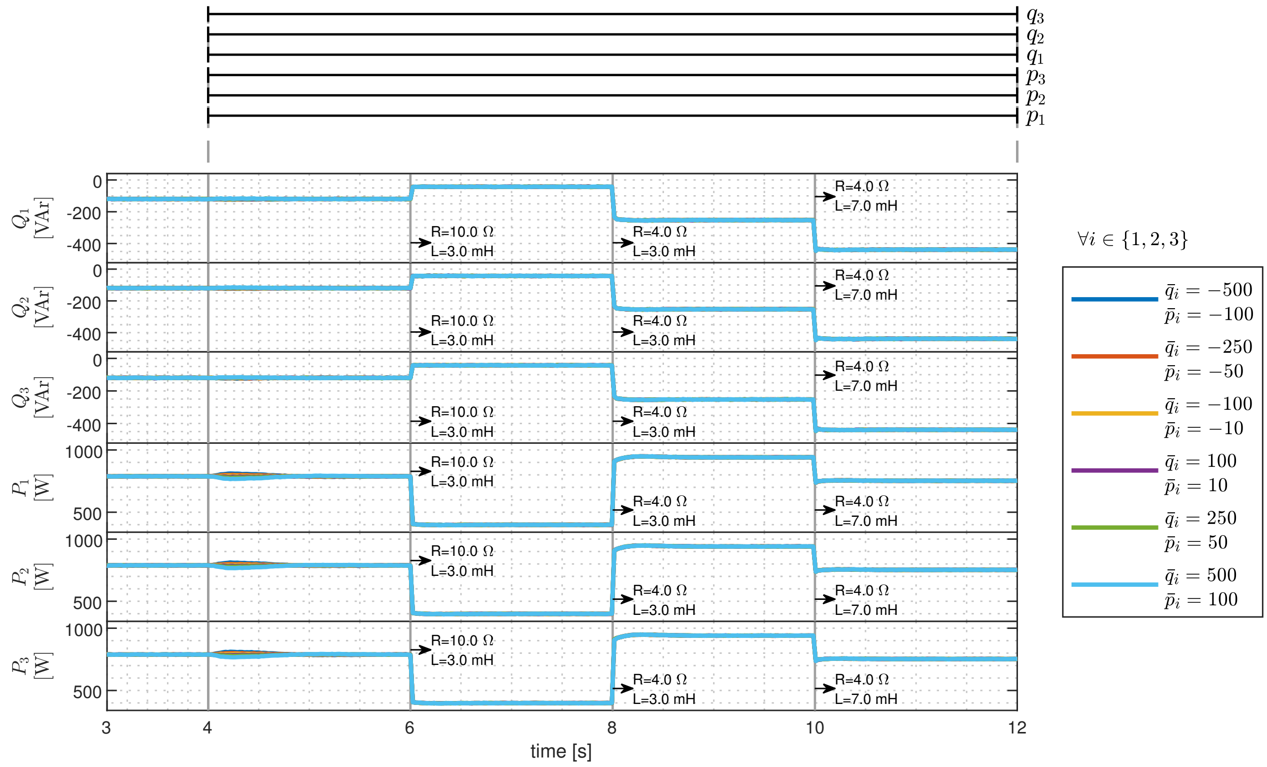

Figure 4.17: Active and reactive power generated by each DER in DT configuration 2 against step attacks and load changes.

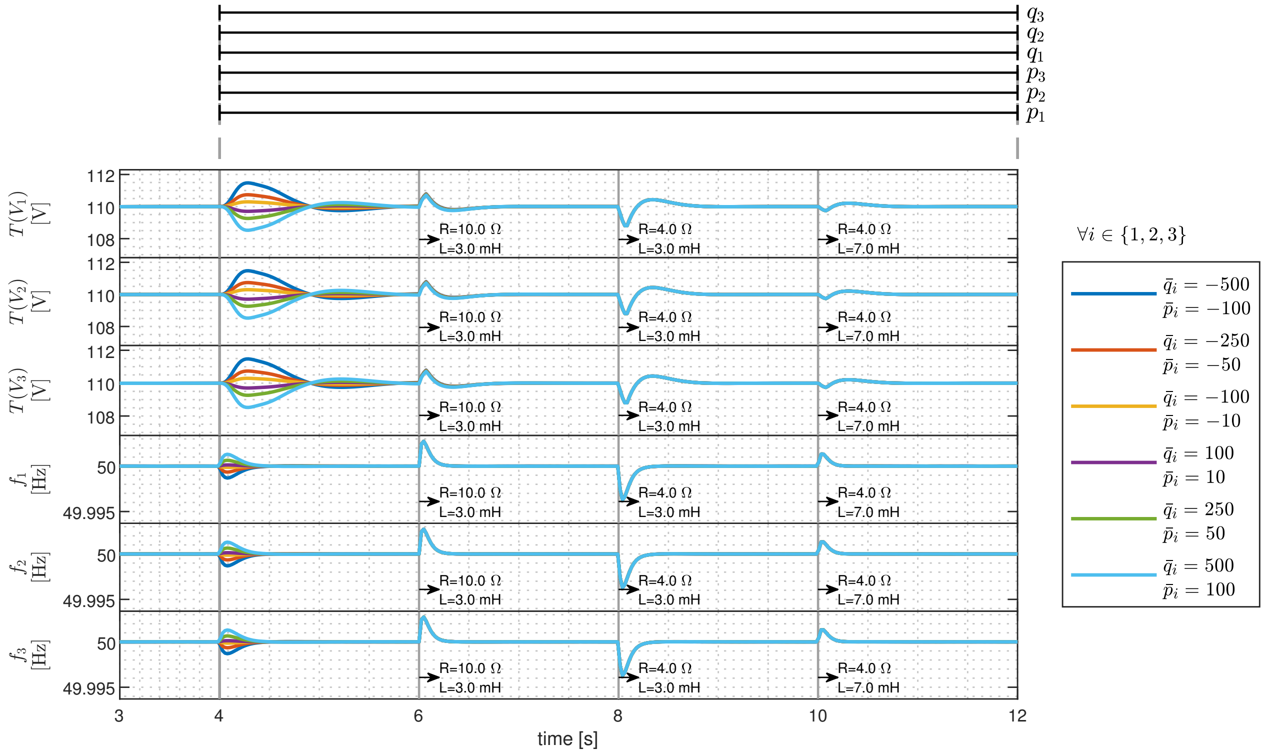

Figure 4.18: Average voltage amplitude and frequency of each DER in DT configuration 2 against step attacks and load changes.

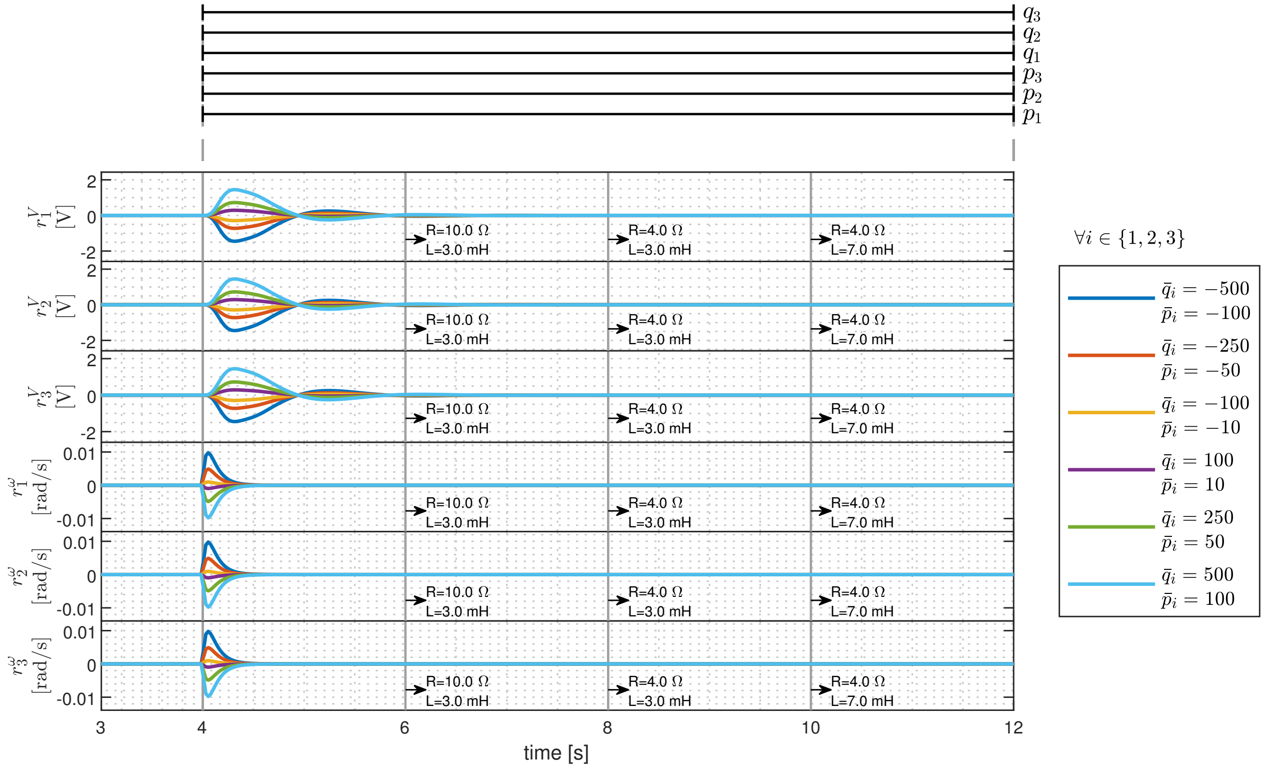

Figure 4.19: Voltage and frequency residual indices in DT configuration 2 against step attacks and load changes.

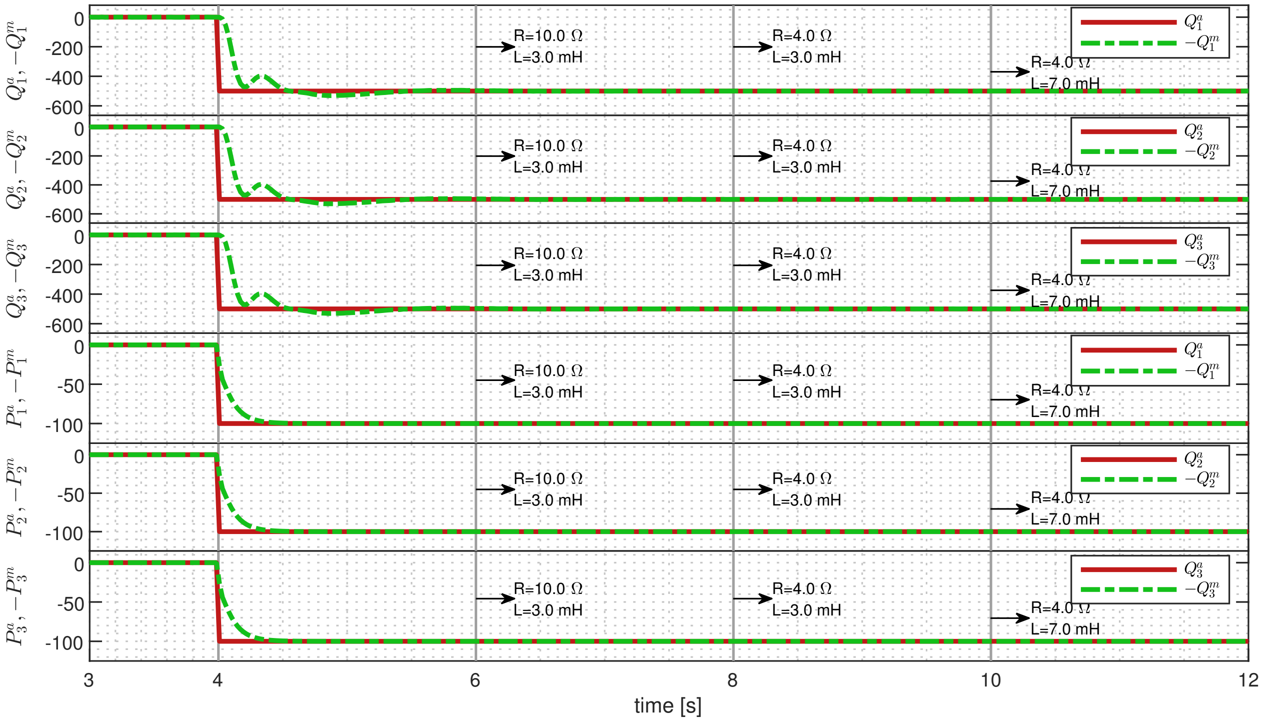

Figure 4.20: Step attack sequences injected to consensus variables initiated at the same time versus mitigation sequences.

4.3.1.2. Scenario 2

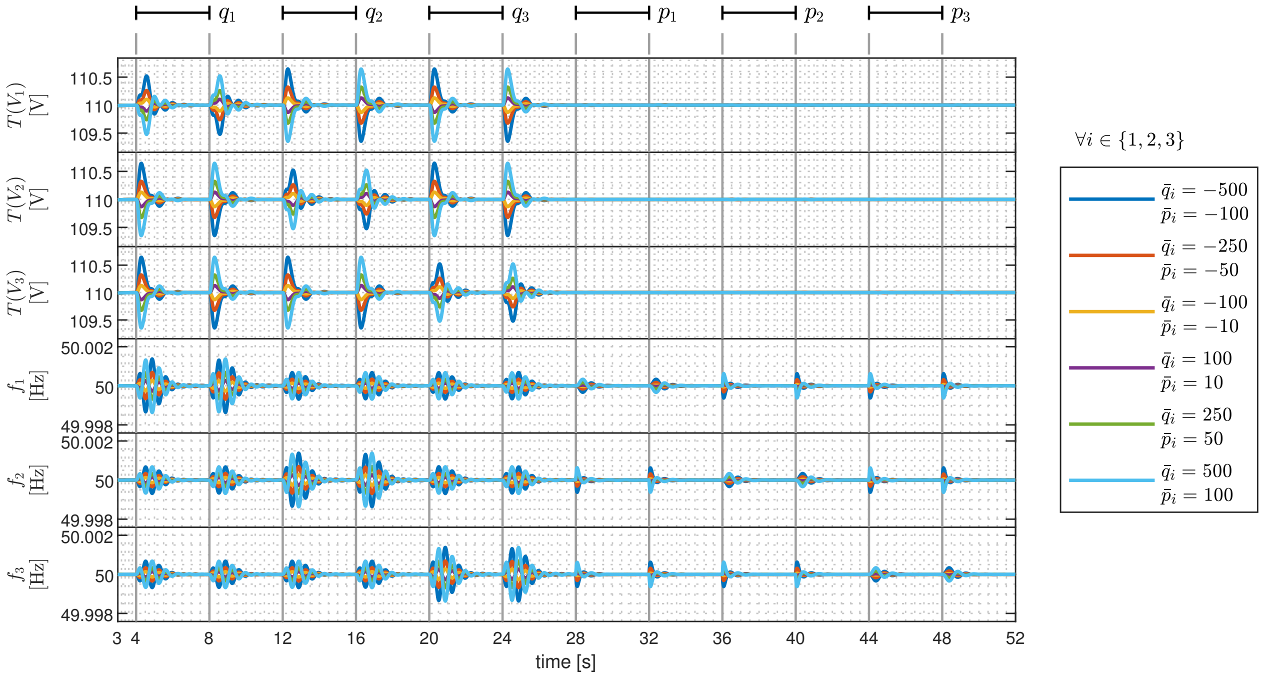

Figure 4.21: Average voltage amplitude and frequency of each DER in DT configuration 2 against step attacks in isolated time intervals.

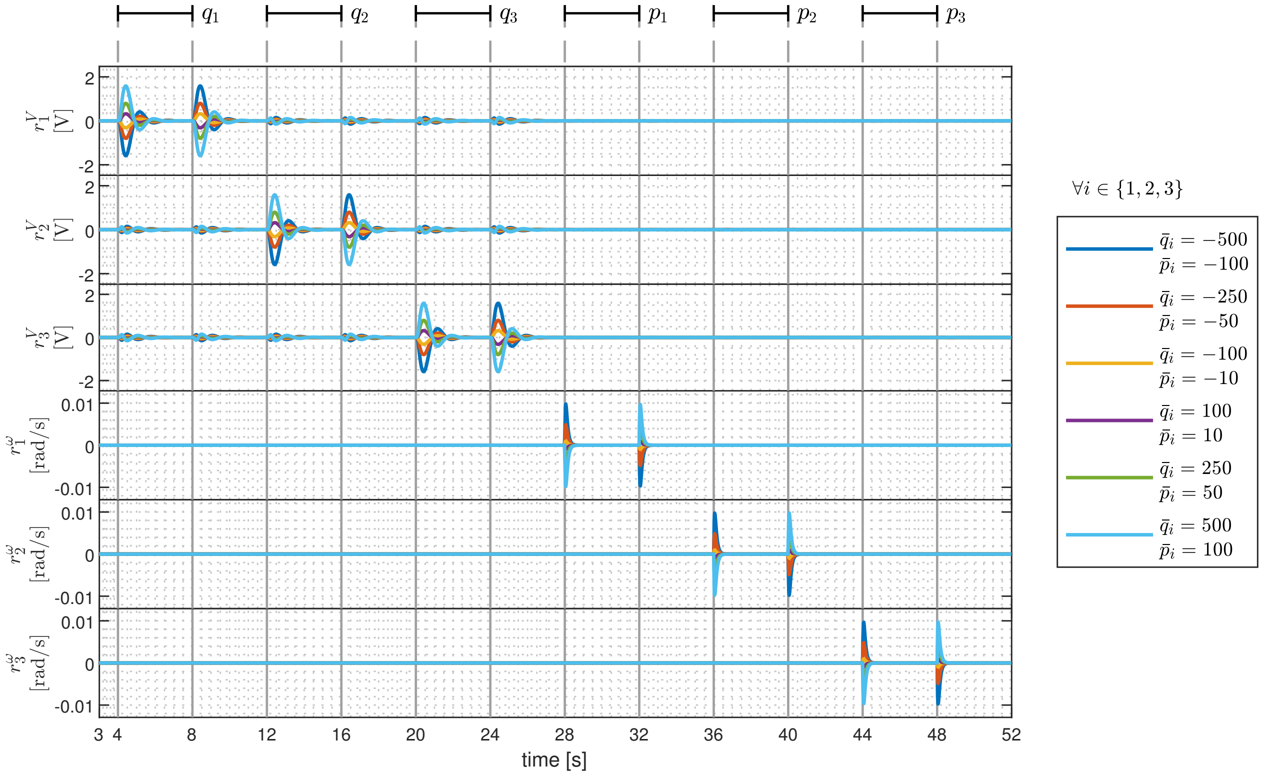

Figure 4.22: Voltage and frequency residual indices in DT configuration 2 against step attacks in isolated time intervals.

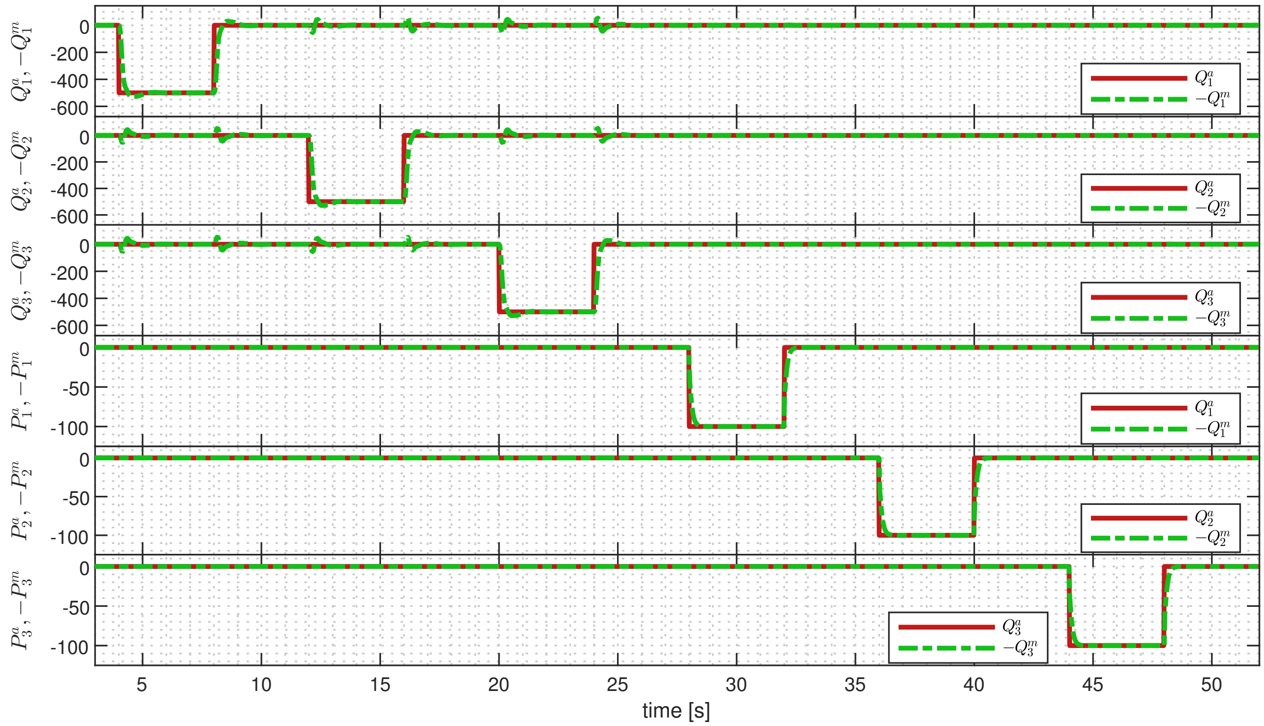

Figure 4.23: Step attack sequences injected to consensus variables in isolated time intervals versus mitigation sequences.

4.3.1.3. Scenario 3

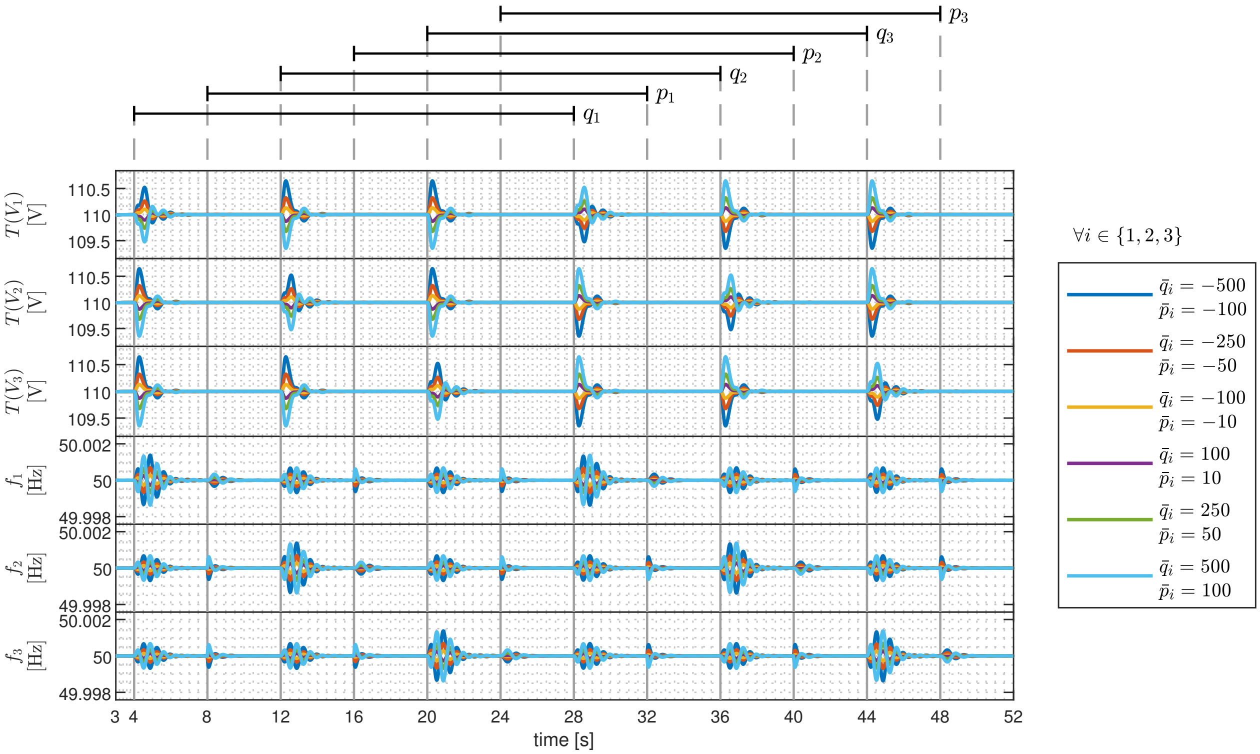

Figure 4.24: Average voltage amplitude and frequency of each DER in DT configuration 2 against step attacks added progressively in time.

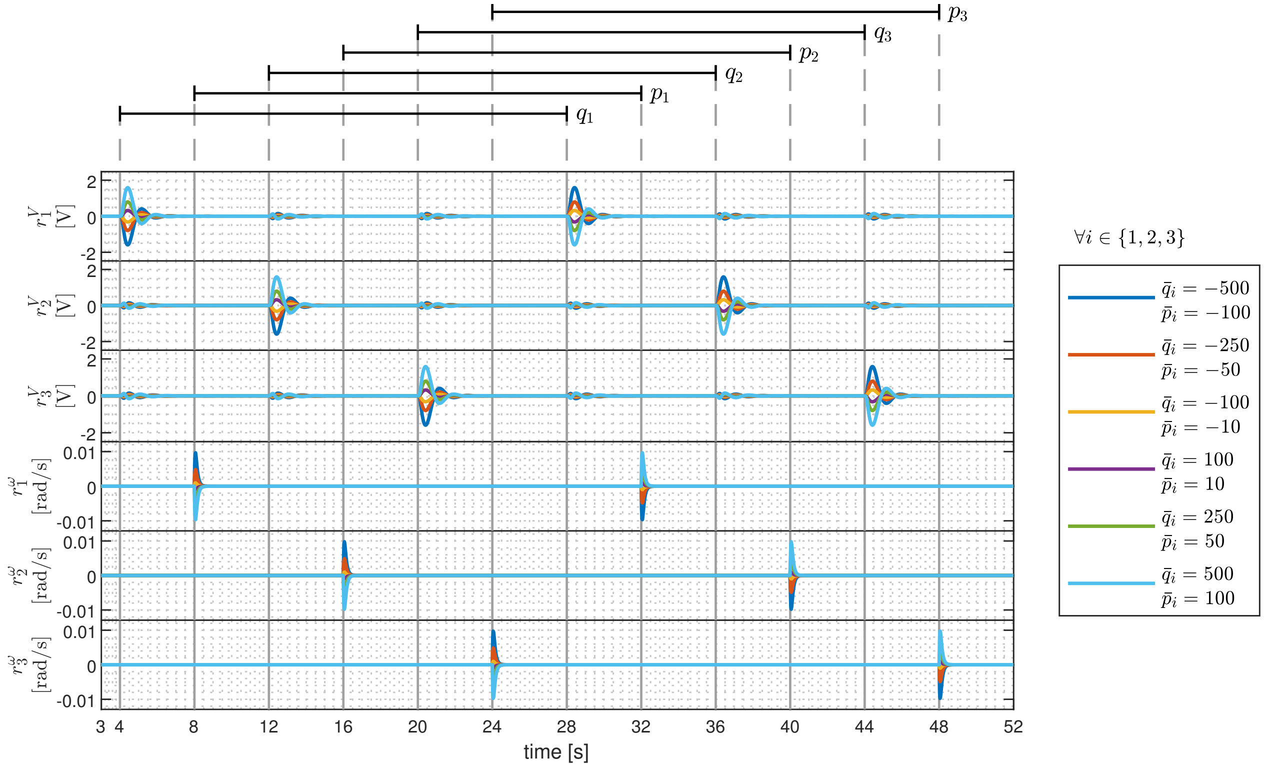

Figure 4.25: Voltage and frequency residual indices in DT configuration 2 against step attacks added progressively in time.

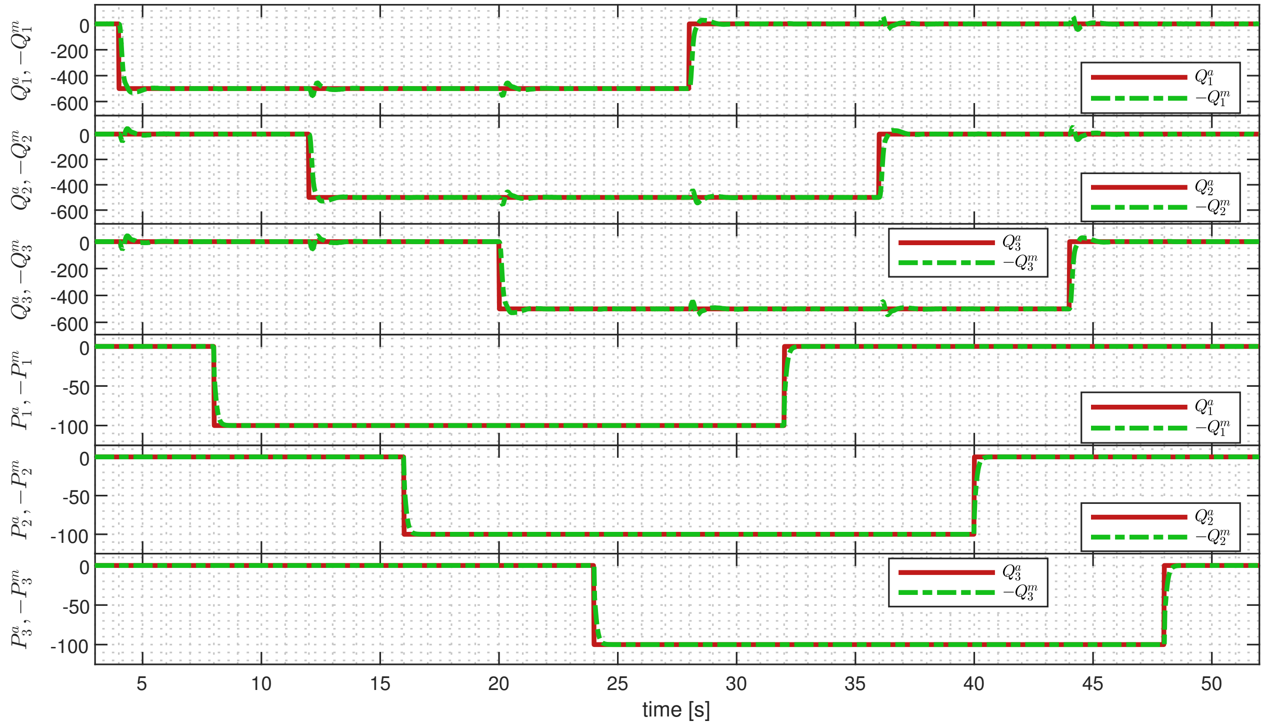

Figure 4.26: Step attack sequences injected to consensus variables added progressively in time versus mitigation sequences.

4.3.2. Experiment 3.2: Truncated Ramp Attack Sequences

In Experiment 3.2, the DT framework is tested against truncated ramp attack sequences in two different scenarios, the same scenarios from Experiment 2.2. In Scenario 1 all consensus variables are attacked and attack sequences start at the same time and in Scenario 3 attack sequences are added progressively in time until all consensus variables are under attack, where ramps are truncated every 4[s].

4.3.2.1. Scenario 1

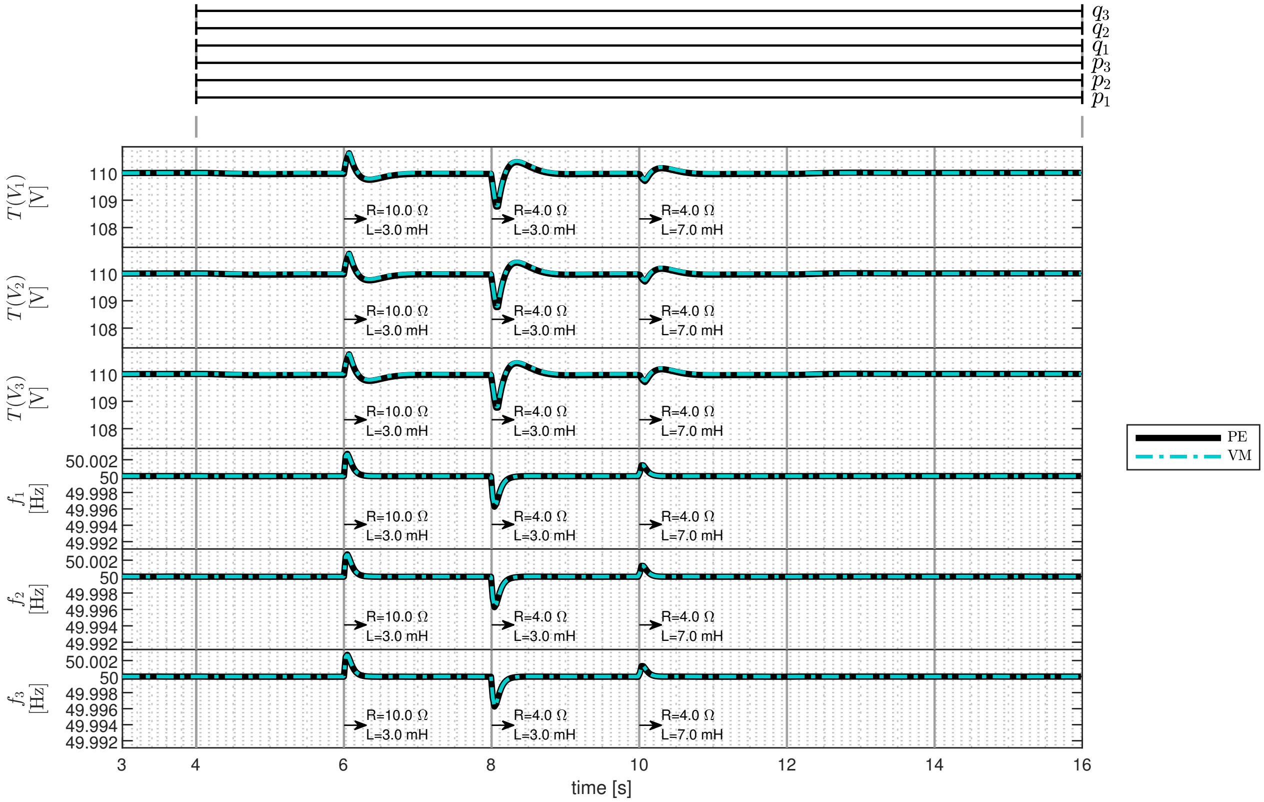

Figure 4.27: Average voltage amplitude and frequency of each DER in DT configuration 2 against truncated ramp attacks and load changes.

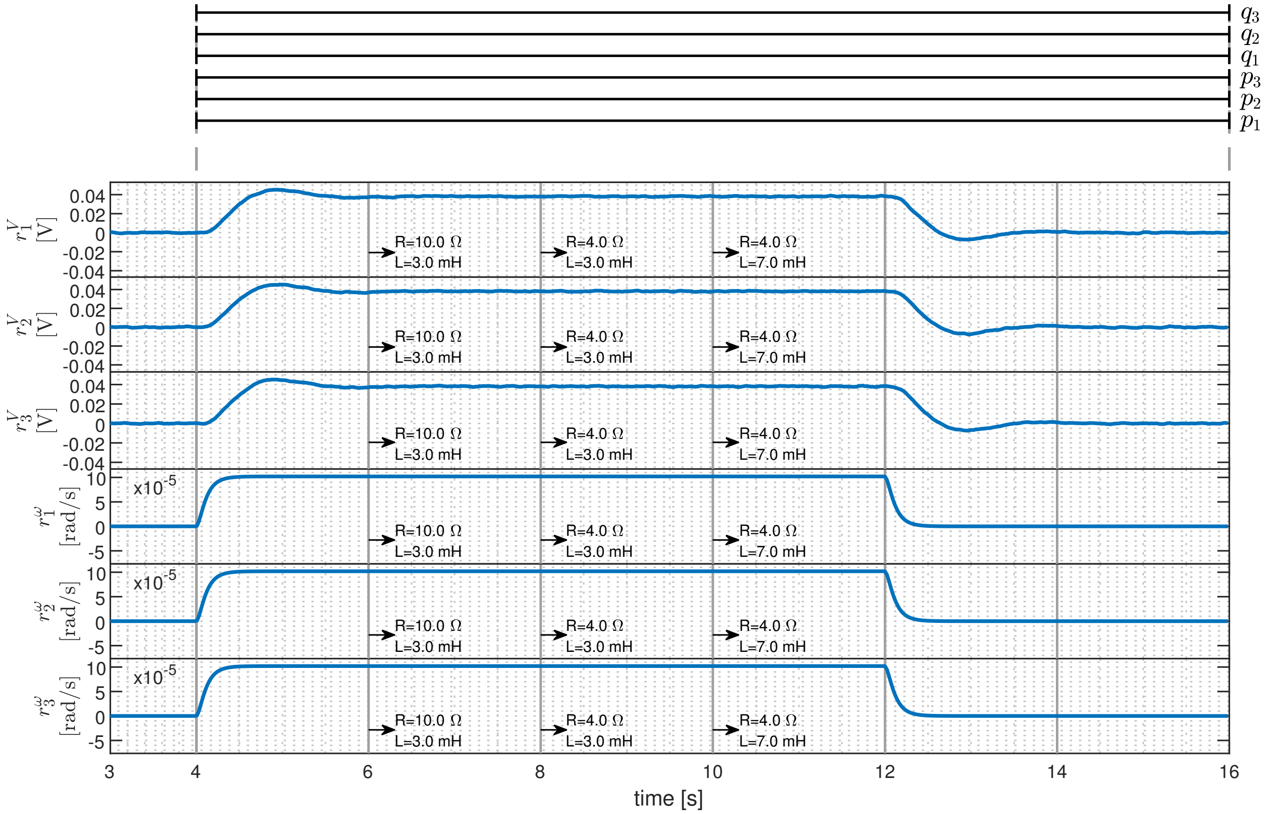

Figure 4.28: Voltage and frequency residual indices in DT configuration 12 against truncated ramp attacks and load changes.

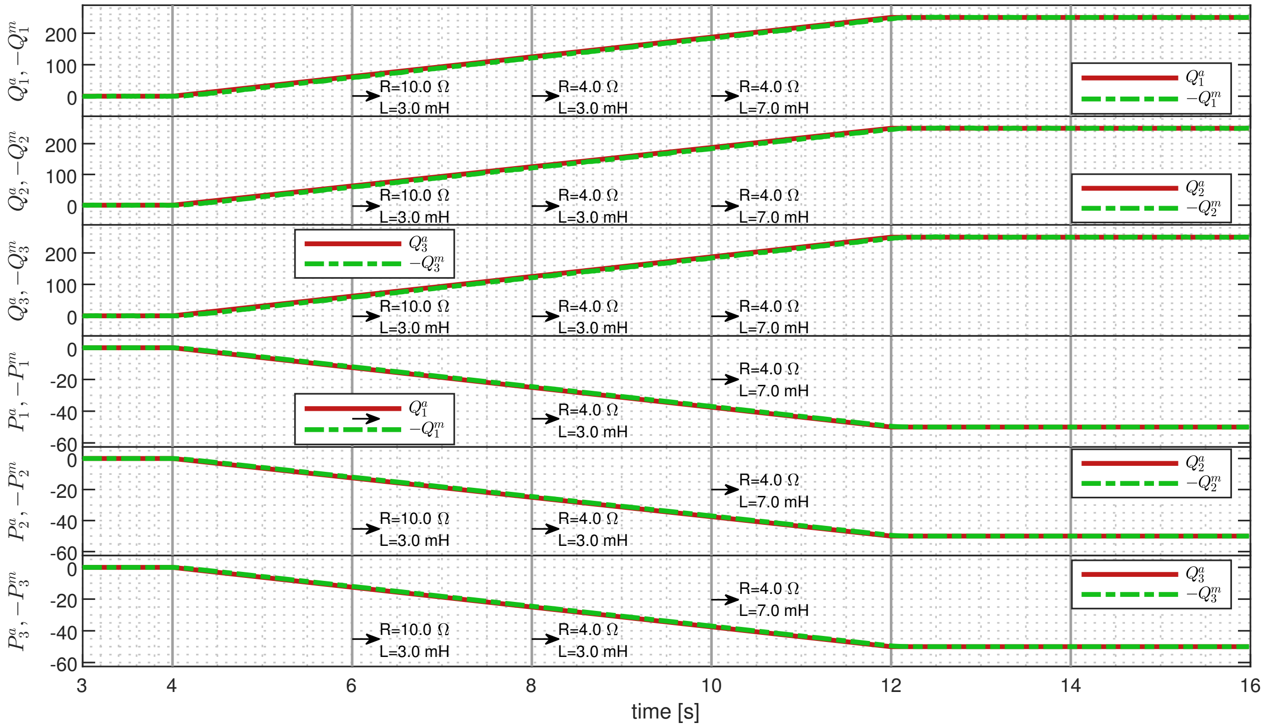

Figure 4.29: Truncated ramp attack sequences injected to consensus variables initiated and truncated at the same times versus mitigation sequences.

4.3.2.2. Scenario 2

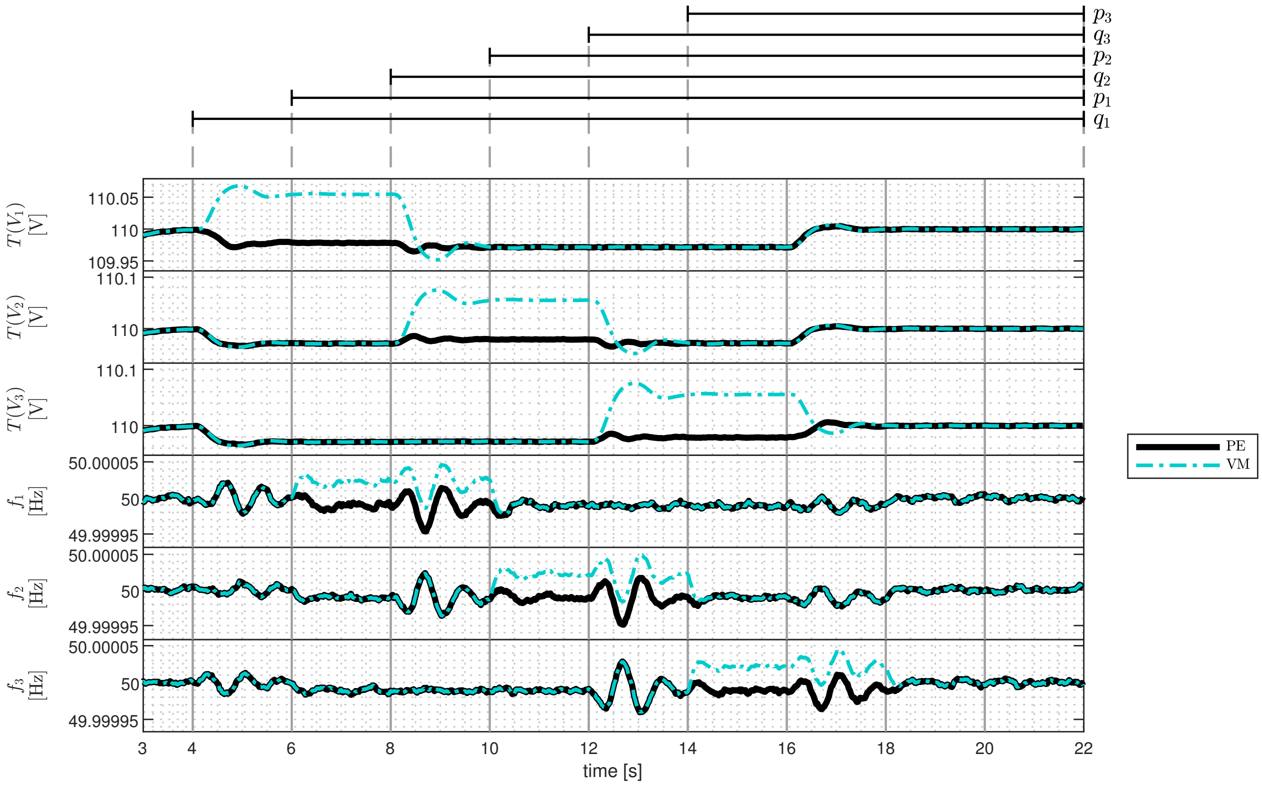

Figure 4.30: Average voltage amplitude and frequency of each DER in DT configuration 2 against truncated ramp attacks added progressively in time.

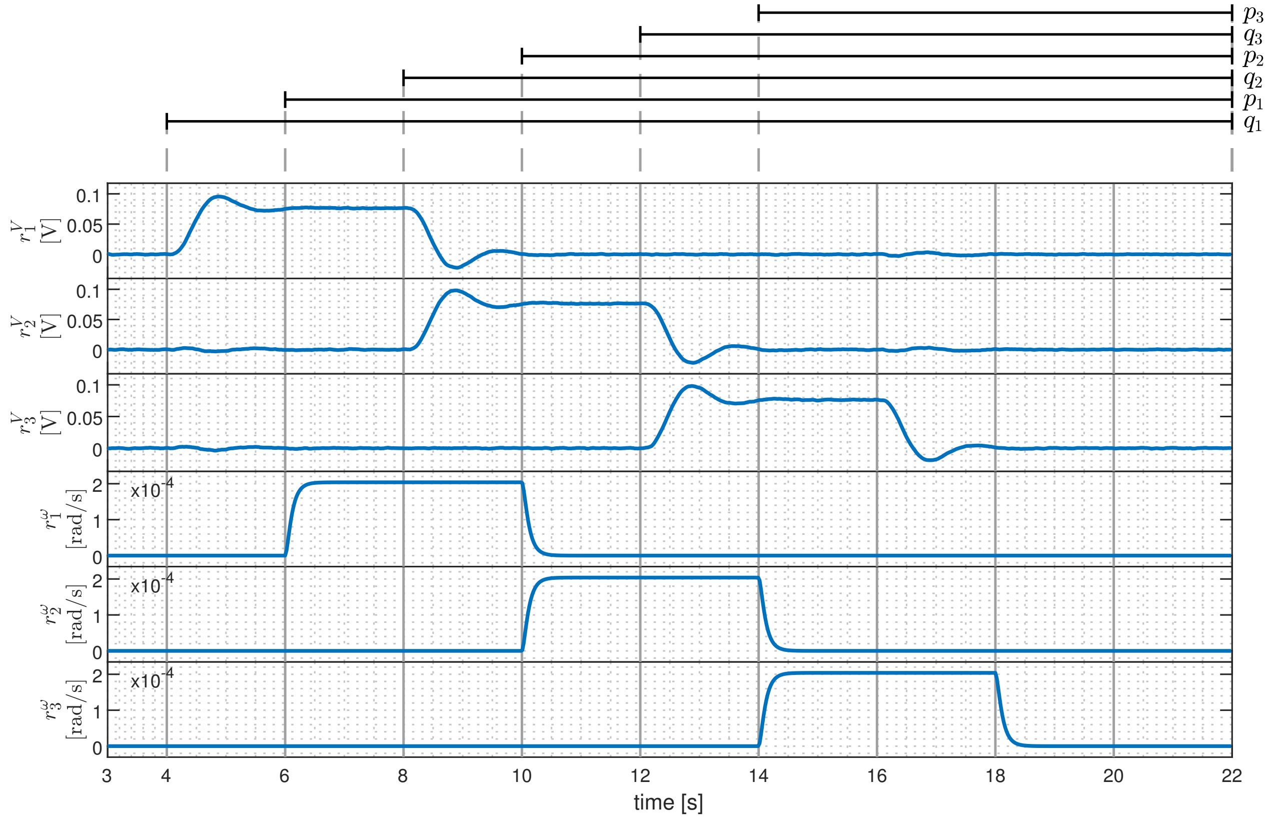

Figure 4.31: Voltage and frequency residual indices in DT configuration 2 against truncated ramp attacks added progressively in time.

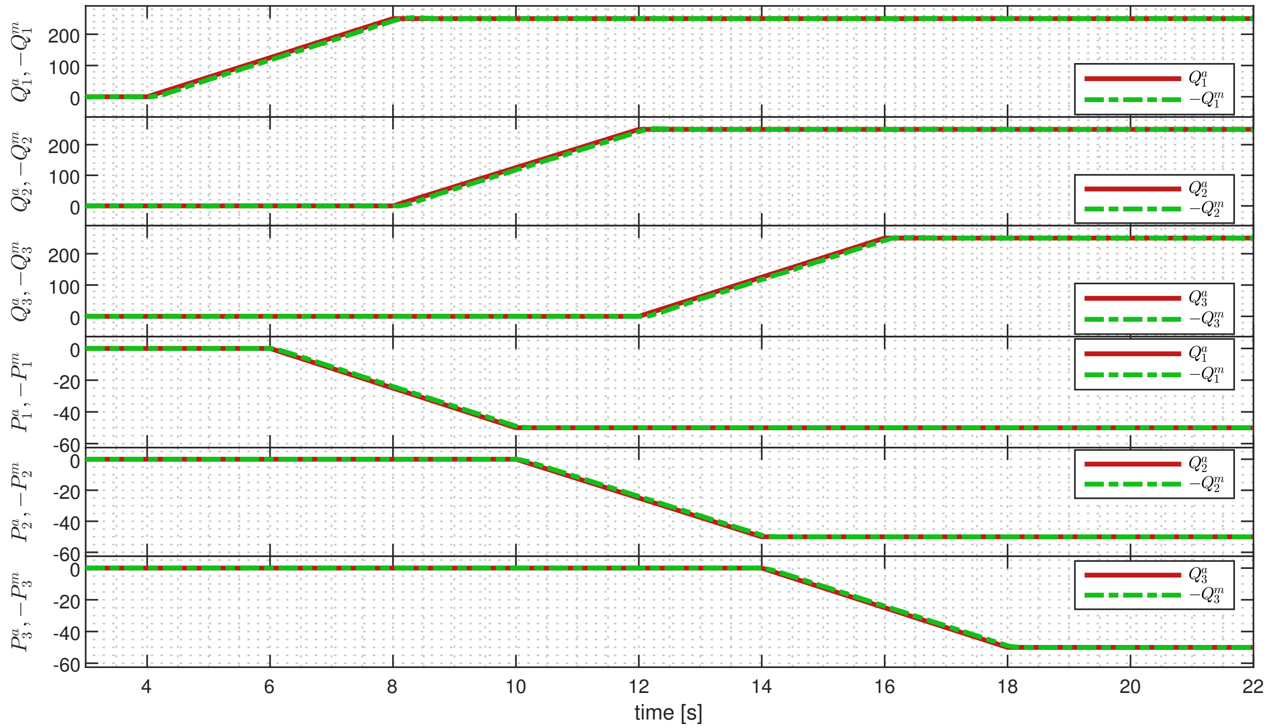

Figure 4.32: Truncated ramp attack sequences injected to consensus variables added progressively in time versus mitigation variables.

Chapter 5

Analysis and Discussion

This chapter presents a detailed analysis and discussion of the results obtained in Chapter 4, validating the DT framework as a novel approach for detecting and mitigating False Data Injection Attacks in MGs.

5.1. Validation of the DAPI controller

The validation of the DAPI controller was conducted using the results from Experiment 1 (section 4.1). As depicted in Fig. 4.1, the disturbance in the three-phase voltage of the DERs caused by changes in the resistance (R) and inductance (L) of the load at 6 seconds was rapidly corrected. A similar disturbance occurred at 8 seconds when only the resistance changed. At 10 seconds, a change in inductance caused minimal disturbance to the voltage amplitude.

Furthermore, as shown in Fig. 4.2, both the active and reactive instantaneous output power of the DERs adjusted in response to the load changes. Notably, the output power levels of all three DERs were consistent and exhibited identical transient behaviors in response to these load variations.

Fig. 4.3 also demonstrates that the voltage amplitude and frequency were at nominal values of 110V and 50Hz before the first load change. After each load adjustment, the values in the PE experienced transients that returned to nominal values within 0.5 to 1 second.

This analysis confirms that the DAPI controller works as expected, ensuring that three-phase voltage is accurately restored, power-sharing among the DERs is maintained, and both voltage amplitude and frequency are returned to their nominal values following load changes.

5.2. VM and Residual Indices in Absence of Attack and Mitigation Sequences

Experiment 1 (Section 4.1) provides insights into the operation of the VM and the functionality of the residual indices in DT Configuration 1, particularly in the absence of attack sequences. As illustrated in Fig. 4.3, the average voltage amplitude and frequency for all DERs in the VM closely mirrored the behavior of the PE, both during steady-state conditions and load changes. This observation validates that, in the absence of attacks, the VM nearly perfectly replicates the state of the PE.

Moreover, when analyzing Fig. 4.4, the voltage residuals were observed to fluctuate very close to zero. Given that these residuals represent the differences between the VM and PE, this result further corroborates the VM's accuracy. The frequency residuals remained precisely at zero, reinforcing the VM's ability to replicate the PE with high fidelity.

The slight variations in voltage residuals is attributed to the white noise associated with the sensors. In the case of the PE, voltage measurements include this noise, whereas the VM's virtual measurements do not. However, the frequency residuals remained exactly zero because the frequency is not a measurement but a control variable of the primary controller.

In conclusion, Fig. 4.4 demonstrates that the residuals fulfill their intended function: in the absence of attack sequences, they remain at zero. Furthermore, these residuals were not affected by load changes, effectively showing that the VM replicates the PE even during transients, thereby preventing false detections due to load variations.

5.3. DT Configuration 1: Residual Indices for FDIA Detection

In DT Configuration 1, where mitigation sequences are deactivated, the residual indices play a crucial role in detecting FDIAs. This section discusses the results of Experiment 2 (Section 4.2) involving step and truncated ramp attack sequences, demonstrating the effectiveness of the DT framework in identifying these attacks.

5.3.1. Step Attack Sequences

As presented in Experiment 2.1, the step attack sequences were introduced across three scenarios to evaluate the DT's detection capability. In Scenario 1, all consensus variables were attacked simultaneously at 4 seconds, while the system underwent load changes similar to Experiment 1.

In Scenario 1, as shown in Fig. 4.5, a $\bar{q}_i = -500$ and $\bar{p}_i = -100$ step attack for all $i \in \{1, 2, 3\}$ begins at 4 seconds. Prior to this time, the sinusoidal signal amplitudes are stable at the nominal value of 110V. Once the attack is initiated, the voltage amplitudes deviate smoothly, eventually reaching 115V.

Fig. 4.6 displays the instantaneous output power of the DERs. At the start of the attack on consensus variables at 4 seconds, the power generation is visibly affected. Despite the same load changes as in Experiment 1, the power values respond differently under attack conditions, with the impact varying according to the load's magnitude.

As observed in Fig. 4.7, when the attack starts at 4 seconds, both voltage amplitudes and frequency deviate from the nominal values. The amplitude of these deviations is directly influenced by the magnitude of the step attack. Specifically, when $\bar{q}_i$ and $\bar{p}_i$ are negative, the voltage increases, while positive values result in a decrease. The frequency, conversely, exhibits the opposite behavior.

Additionally, during load changes, the voltage and frequency display transients with a consistent shape but shifted by a constant value on the Y-axis, regardless of the step attack's magnitude. These transients closely resemble those observed under normal conditions in Experiment 1.

The residual indices depicted in Fig. 4.8 behave as expected. At 4 seconds, when all attacks are introduced, each residual index is activated. Similar to Experiment 1, these indices do not exhibit transients due to load changes, even in the presence of attack sequences. This further validates the residual indices as reliable indicators for attack detection.

Similar to the voltage amplitude and frequency, the sign of the attack magnitude affects the residual indices' amplitude. When $\bar{q}_i$ and $\bar{p}_i$ are negative, the voltage residues are also negative; when they are positive, the residues are positive. For the frequency residues, the opposite occurs. Additionally, the magnitude of the step attack proportionally influences the amplitude of the resulting residues: the voltage amplitude residue is directly proportional to the attack magnitude, whereas for frequency, the relationship is inversely proportional.

To this point, the full potential of the residual indices has not been fully demonstrated. Scenarios 2 and 3 of Experiment 2.1 involve more complex attack sequences to further validate these residual indices as effective indicators within DT Configuration 1.

In Scenario 2, the attack intervals are non-overlapping (refer to Table 4.2). As shown in Fig. 4.7, when $q_1(t)$ is injected, the voltage of all three DERs deviates from the nominal 110V. For DER$_1$, the deviation is slightly less pronounced compared to DER$_2$ and DER$_3$ for all attack magnitudes. Some transients are observed in the frequency, which eventually converge back to 50Hz. This pattern repeats for $q_2(t)$ and $q_3(t)$. In contrast, when $p_1(t)$ is activated, the frequency of all three DERs deviates from 50Hz, with no transients observed in the voltage amplitudes.

Regarding the the residual indices from Fig. 4.10, two crucial observations can be made. Firstly, when each attack sequence is activated, a specific residual signal is triggered. When q_i is initiated, the residual index r_i^V is activated; similarly, when p_i is triggered, r_i^ω is activated. Conversely, if r_i^V is activated, it indicates a Q-attack on Q_i in DER_i. Analogously, if r_i^ω is activated, it indicates a P-attack on P_i in DER_i. This correlation aligns with the secondary control equations: a P-attack directly influences the frequency restoration variable Ω_i (Equation (3.10)), which in turn affects the frequency control variable ω_i via primary control. Similarly, a Q-attack directly impacts the voltage restoration variable δV_i (Equation (3.8)), which subsequently affects the voltage control variable V_i^{ref} through primary control.

Secondly, the magnitude of the attack elements proportionally influences the residual indices' amplitude, as observed in Scenario 1, with a clear distinction between q_i and p_i. For Q-attacks, comparing the q_i values with the residual indices r_i^V reveals a direct proportionality. When $q_i = -500$, the three dark blue curves for voltage indices show values near -5. As q_i increases, the voltage residue values rise, changing sign and reaching approximately 5 when $q_i = 500$. In contrast, for P-attacks, there is an inverse proportionality between p_i and

r_i^{ω}, analogous to the Q-attacks. This difference in proportionality is due to the signs of the control parameters η_i and γ_i listed in Table 3.2, where η_i is negative, and γ_i is positive, thus affecting the sign of the consensus variables sum in Equations (3.10) and (3.8).

From this experiment, a significant preliminary conclusion can be drawn. In scenarios where P-attacks and Q-attacks occur in isolated time intervals, if the index $r_i^V(t)$ is activated, a Q-attack is occurring on the consensus variable Q_i in DER_i. Conversely, if $r_i^{\omega}(t)$ is activated, a P-attack is taking place on the consensus variable P_i in DER_i. Furthermore, the proportionality between the residual indices and the attack elements suggests that it may be possible to estimate the attack element's value.

Scenario 3 further reinforces the validity of the proposed residual indices. In this scenario, the attack pattern is more complex, with attack time intervals appearing progressively until all attacks overlap (refer to Table 4.2). As shown in Figure 4.11, Q-attacks and P-attacks are activated in an interleaved manner, yet the pattern observed in the previous experiments persists. It is possible to accurately identify whether the attack is a P-attack or Q-attack, as well as the specific variable and DER being targeted.

At 4 seconds, $q_1(t)$ is activated, resulting in the activation of r_1^V. At 8 seconds, $p_1(t)$ is activated, causing r_1^{ω} to activate, while r_1^V remains active without affecting other residual indices. This alternating pattern continues until 24 seconds, by which time all DERs are under attack, leading to the activation of all indices without interference among them. After 28 seconds, the attacks begin to deactivate in the same interleaved manner, with the same identifiability maintained. The proportionality between the attack elements and the indices remains consistent, even when all DERs are under attack.

In this last scenario of Experiment 2.1, it is concluded that the DT framework utilizing residual indices for step attacks is highly effective in accurately identifying whether a Q-attack or P-attack has been executed, as well as determining which consensus variables and DERs are being targeted at any given moment, even when multiple, overlapping attacks are occurring.

5.3.2. Truncated Ramp Attack Sequences

For the truncated ramp attack sequences, the results closely resemble those observed with step attacks, with the principal difference in the ramp-like progression of the attack. In Scenario 1, as illustrated in Fig. 4.12, the attack sequence injected begins at 4 seconds and gradually increases until 12 seconds, where Q_i^a reaches $\bar{q}_i = 250$ and P_i^a reaches $\bar{p}_i = 50$ for all $i \in U = 1, 2, 3$ (refer to Tables 4.3 and 4.5).

Examining Fig. 4.13, it is evident that the voltage amplitude decreases steadily until 12 seconds, at which point it stabilizes at a constant value. The same behavior is observed for the frequency. Similar to the observations in Experiment 1 and Scenario 1 of Experiment 2.1, both voltage and frequency exhibit transients during load changes.

Fig. 4.14 shows that the residual indices increase linearly, mirroring the ramp-like progression of the attack sequences, and then stabilize at a constant value at 12 seconds. As with previous load change experiments, these load changes do not influence the residuals.

While this outcome could have been anticipated from the results of the step attacks, it is nonetheless an intriguing result to demonstrate. The linear increase in the residuals' shape allows for the identification of the attack as a truncated ramp, highlighting the specific information these indices can provide.

In Scenario 2, a more complex attack pattern is tested. As depicted in Fig. 4.15, the attack sequences are introduced progressively, with the ramps being truncated every 4 seconds,exactly as Scenario 3 of Experiment 2.1. By analyzing the residual indices in Fig. 4.16, it is apparent that they, too, follow the shape of the attack sequences. At 4 seconds, $q_1(t)$ is injected, prompting $r_1^V(t)$ to begin increasing linearly in a ramp-like form. At 6 seconds, $p_1(t)$ is injected, causing $r_1^\omega(t)$ to increase similarly, without interfering with $r_1^V(t)$. This consistent pattern across all residual indices once again confirms that the residual indices operate independently of one another.

5.4. DT Configuration 2: Mitigation and Alternative Detection Approach

In this section, the results from Experiment 3 are analyzed to validate the mitigation of step and truncated ramp attacks within DT Configuration 2. The discussion is divided into two subsections: the first focuses on the mitigation strategy, and the second explores an alternative detection approach.

5.4.1. Mitigation

This subsection is further divided into two parts: the first discusses the mitigation of step attacks, and the second addresses the mitigation of truncated ramp attacks.

5.4.1.1. Step Attack Sequences

In this discussion, the effectiveness of the PID controller as a mitigator for step attack sequences is validated. Starting with Scenario 1 of Experiment 3.1, which includes load changes, Fig. 4.17 shows that when step attacks are introduced at 4 seconds, the active power exhibits minor disturbances. These disturbances are rapidly managed across all step attack amplitudes. A similar response is observed in the reactive power, where the disturbance is so minimal that it is barely noticeable in the plot. This indicates that before the first load change, the system successfully returns to normal operation, recovering the behavior observed in Experiment 1.

Regarding voltage amplitude and frequency, Fig. 4.18 displays these values across all six simulations. At 4 seconds, deviations from the nominal values are evident, with the amount of deviation depending on the attack's amplitude. Notably, by the time of the first load change at 6 seconds, both voltage and frequency are restored to their nominal values, exhibiting the same response to load changes as shown in Figure 4.3 from Experiment 1. Additionally, when compared to Scenario 1 of Experiment 2.1 (specifically Fig.4.7), it is clear that the deviations in voltage and frequency do not reach the same levels, highlighting the effectiveness of the mitigator.

To conclude the analysis of Scenario 1, the residual indices presented in Fig. 4.19 are examined. Here, the voltage and frequency residues demonstrate behavior similar to that shown in Fig. 4.18. At 4 seconds, the residues deviate from zero, indicating that an attack has occurred, but they quickly converge back to zero. As discussed throughout this chapter, these residues do not exhibit disturbances due to load changes.

In Scenario 2, a similar behavior is observed for isolated attack time intervals. In Fig. 4.21, when $q_1(t)$ is injected into the communication network, the voltage and frequency of all DERs are affected but are restored to their nominal values within approximately 2 seconds. The same effect is seen with $q_2(t)$ and $q_3(t)$. For $p_i(t)$, only the frequency is disturbed, but it too is quickly restored.

The residual indices shown in Fig. 4.22 behave as expected. When q_i is injected, the corresponding ith voltage residual index $r_i^V(t)$ deviates but returns to zero. A similar response occurs when p_i is injected: the ith frequency residual index $r_i^\omega(t)$ deviates but also returns to zero, with the frequency indices restoring slightly faster.

For Scenario 3, being redundant to this point, shows the same behavior in Fig. 4.25: when q_i is injected, $r_i^V(t)$ is altered and restored back to zero, and when p_i is injected, $r_i^\omega(t)$ is altered and restored back to zero too. In this case the attack injections are overlapped, but as discussed before, the residual indices are independent, so neither the attack or mitigation sequence affects other residual indices.

Scenario 3, though somewhat redundant at this point, exhibits the same behavior as seen in Fig. 4.25. When q_i is injected, $r_i^V(t)$ is altered and then restored to zero. Likewise, when p_i is injected, $r_i^\omega(t)$ is altered and restored to zero. Even though the attack injections overlap in this scenario, the residual indices remain independent, so neither the attacks nor the mitigation sequences interfere with other residual indices.

These results clearly demonstrate that the PID controller functions effectively as a mitigator for step attack sequences. Power sharing is restored, voltage amplitude and frequency return to their nominal values, and residual indices converge to zero. This is a significant result, as it indicates that even if the attack persists, the system can rapidly and safely return to normal operation. The voltage and frequency values do not deviate far from nominal before being restored. As discussed in Experiment 2.1, when $\bar{q}_i = \pm 500$, the voltage deviated by approximately ±5V from nominal values. However, the mitigation is so effective that, for the same attack amplitude, the voltage deviation does not even reach ±2V.

5.4.1.2. Truncated Ramp Attack Sequences

In this discussion, the PID controller's effectiveness as a mitigator for truncated ramp attack sequences is validated. Beginning with Scenario 1 of Experiment 3.2, which includes load changes, Fig. 4.27 shows that when ramp attacks are injected at 4 seconds, both the PE and VM deviate slightly from their nominal values, though these deviations are barely noticeable in the plot. From the first load change at 6 seconds onward, the system operates normally, as illustrated in Fig. 4.3 from Experiment 1.

Although Figure Fig. 4.27 suggests that the voltage amplitude and frequency are perfectly restored, the residual indices in Figure Fig. 4.28 indicate that from 4 to 12 seconds, when the ramp is increasing, the residual indices reach a constant value. This value is relatively small and can be considered negligible, meaning the system is operating normally despite the small error. Observing Figure 4.29, this behavior can be explained by the mitigation variables $Q_i^m(t)$ and $P_i^m(t)$. The PID controller is designed to follow a reference, so when $Q_i^a(t)$ and $P_i^a(t)$ are increasing (or decreasing) before being truncated, the corresponding $-Q_i^m(t)$ and $-P_i^m(t)$ generated by the PID controller attempt to follow them closely but cannot match them exactly while they are still varying. At 12 seconds, when the attack is constant, $-Q_i^m(t) = Q_i^a(t)$ and $-P_i^m(t) = P_i^a(t)$. This satisfies the restrictions of Equations (3.18)-(3.19).

In Scenario 2, the behavior of the PE and VM can be better visualized and analyzed. In Figure Fig. 4.30, when the ramp-like $q_1(t)$ is injected, the voltage amplitude of all three PEs deviates slightly below 110V. However, the VM's $T(V_1)$ does not follow the PE's value exactly, reaching approximately 110.05V, while $T(V_2)$ and $T(V_3)$ follow the PE closely. When $q_1(t)$ is truncated at 8 seconds, the VM converges with the PE to a value deviating less than 0.05V from the nominal value. This behavior is consistent across all other attacks.

As in Scenario 1, the residual indices in Scenario 2, depicted in Fig. 4.31, show a constant value while the ramp attack is varying, converging back to zero at truncation. This constant value is explained by the behavior of the attack and mitigation sequences, as shown in Fig. 4.32. The PID controller attempts to follow $-Q_i^a(t)$ and $-P_i^a(t)$ closely but cannot exactly match them while they are still varying.

These experiments validate the effectiveness of the DT's control block in mitigating the attacks. Although a small error persists during the linear increase of the ramp, the PID control mitigator successfully maintains this error constant and relatively small, resulting in voltage deviations of less than 0.05V and frequency deviations of less than 50×10^{-6} Hz.

5.4.2. Alternative Detection Approach

As discussed in the previous subsection, the mitigator successfully brings the residual indices back to zero in response to FDIAs. However, this creates a small stepback, as the residual indices are no longer effective indicators of ongoing attacks. Nevertheless, as mentioned in Section 3.4.4, in DT Configuration 2, the mitigation variables $-Q_i^m(t)$ and $-P_i^m(t)$ become the new indicators for attack detection. This can be validated through Figs. 4.20, 4.23, 4.26, 4.29 and 4.32. In these figures, a consistent pattern is observed: as $t \to \infty$, $-Q_i^m(t) \to Q_i^a(t)$ and $-P_i^m(t) \to P_i^a(t)$.

For step attack sequences, the mitigation variables closely follow the attack sequences, as shown in Figs. 4.20, 4.23 and 4.26. When the step is activated, within less than 2 seconds, $-Q_i^m(t) \approx Q_i^a(t)$ and $-P_i^m(t) \approx P_i^a(t)$. On the other hand, for truncated ramp attack sequences, the mitigation sequences cannot exactly match the attack sequence values during the ramp's growth phase (as discussed in Section 5.4.1.2). However, this does not pose a problem for detection purposes, as the role previously held by the residual indices is effectively replaced by the mitigation sequences in this DT configuration.

5.5. Summary of The Most Important Points Discussed

From this chapter, the most important points discussed are outlined below:

- The DAPI controller functions perfectly, restoring voltage amplitude and frequency and balancing power-sharing. It works smoothly in response load changes.
- DT Configuration 1 works as intended. Residual indices are reliable indicator for FDIA detection, with r_i^V signaling a Q-attack on DER_i and r_i^ω indicating a P-attack on DER_i. Additionally, the shape of the residuals can distinguish between step and truncated ramp attacks.
- Residual indices are unaffected by load changes, improving their robustness for attack detection.
- Q-attack amplitude $\bar{q}_i$ is directly proportional to r_i^V, and P-attack amplitude $\bar{p}_i$ is inversely proportional to r_i^ω.
- The independence between residual indices against Q-attack and P-attacks makes them ideal candidates for being processed for mitigation.
- The PID control mitigator is effective against step and truncated ramp attacks.
- Mitigation variables $-Q_i^m$ and $-P_i^m$ are excellent attack indicators in DT Configuration 2 against step and truncated ramp attacks.

Chapter 6

Conclussions and Future Work

The research presented in this thesis has successfully demonstrated the efficacy of a Digital Twin (DT) based framework for the detection and mitigation of False Data Injection Attacks (FDIAs) in isolated AC microgrids (MGs). This work addresses the need for better cybersecurity measures in MGs, particularly as these systems become increasingly integral to the modern energy landscape. The objectives outlined at the beginning of this research have been met with success, and this conclusion summarizes these achievements, justify the fulfillment of the research objectives, and propose directions for future work.

The general objective of this thesis was to develop a robust digital twin framework capable of detecting and mitigating FDIAs within a consensus-based distributed control system of an isolated inverter-based AC microgrid. This objective has been successfully fulfilled. The DT framework designed and implemented in this study proved to be highly effective in identifying ongoing cyberattacks by utilizing proposed residual indices, which represent discrepancies between the real MG and its digital counterpart. Furthermore, the incorporation of a PID controller to process these residual indices and generate mitigation sequences was shown to be effective in counteracting the effects of the FDIAs, thereby improving the stability and reliability of the MG under attack.

The first specific objective was to develop a virtual model that replicates the physical behavior of an isolated inverter-based AC microgrid, including its control systems and communication networks. This objective was successfully achieved. The DT framework was designed to emulate the real-time operations of the microgrid, capturing the dynamics of the control systems and communication processes. By comparing the real-time data from the physical microgrid with the virtual model, the DT was able to generate residual indices that served as reliable indicators of ongoing FDIAs. These indices were instrumental in identifying the presence and nature of attacks.

The second objective required the creation of a distinct and meaningful difference between the physical microgrid and its digital replica to generate residual indices that would make sense and be useful in detecting attacks. This was accomplished by introducing subtle yet critical distinctions in the VM's design, allowing it to respond differently under attack conditions compared to the physical microgrid. This distinction was proved crucial in creating residual indices that were sensitive to FDIAs, thus enabling precise detection and effective mitigation.

The third objective involved conducting extensive simulations of various FDIA scenarios, including simultaneous and overlapping attacks, as well as step-like and ramp-like attacks. This objective was met through a comprehensive series of simulation experiments that tested the DT framework under diverse conditions. The simulations demonstrated that the DT could reliably detect and mitigate FDIAs even when multiple DERs were attacked simultaneously or in overlapping time intervals. The framework's robustness was proven across different attack patterns, validating its potential for more research in this direction.

The final objective was to validate the DT-based detection and mitigation methods by analyzing the residual indices and ensuring that the mitigation system could restore normal operation. This objective was thoroughly fulfilled. The validation process involved rigorous testing under various attack scenarios, where the DT framework's performance was assessed in terms of its ability to maintain voltage and frequency stability while mitigating the effects of FDIAs. The results confirmed that the DT not only detected the attacks but also effectively neutralized their impact, thereby restoring the MG to its nominal operating conditions.

Given the successful fulfillment of both the general and specific objectives, the hypothesis of this thesis has been thoroughly validated. The research demonstrated that the DT framework can effectively detect and mitigate FDIAs in isolated inverter-based AC MGs by utilizing residual indices of voltage amplitude and frequency. These indices proved to be reliable indicators of cyberattacks, even under challenging conditions such as simultaneous and overlapping attacks. The integration of a PID controller further enhanced the framework's capability to maintain stability and restore normal operation after an attack. Thus, the hypothesis that the DT framework can secure MGs against FDIAs by using residual indices has been confirmed, establishing the DT as a robust solution for enhancing cybersecurity in modern energy systems.

While this research has made significant advances in cybersecurity of AC MGs, there remain several directions for future exploration and improvement. The following suggestions outline potential areas of future work that could further strengthen the DT framework and expand its applicability:

1. **Expanding the Microgrid Model**: Future research should explore the application of the DT framework to microgrids with a larger number of DERs. A more complex microgrid topology would present additional challenges, such as increased communication delays and more intricate control interactions, which would test the scalability and robustness of the DT approach. Moreover, incorporating various types of DERs, including renewable energy sources and energy storage systems, would provide a more comprehensive validation of the framework.

2. **Hardware-in-the-Loop (HIL) Validation**: To bridge the gap between simulation and real-world implementation, future research should focus on validating the DT framework using Hardware-in-the-Loop (HIL) testing. HIL would allow for the integration of real-time hardware components with the DT framework, providing a more realistic assessment of its performance. This approach would help identify any potential discrepancies between the simulated environment and actual hardware behavior, thereby improving the reliability of the DT framework for field deployment.

3. **Advanced FDIA Detection Using Reinforcement Learning (RL)**: While the

current study focused on predefined FDIAs, future work could involve the development of more sophisticated attack scenarios using reinforcement learning (RL) agents. An RL agent could be trained to identify vulnerabilities in the MG and DT framework and execute more complex and adaptive FDIAs. This would provide a more rigorous test of the DT framework's detection and mitigation capabilities, ensuring that it can handle even the most advanced cyber threats.

4. **Machine Learning-Based Mitigation Strategies**: The PID-based mitigation approach used in this study, while effective, is relatively straightforward. Future research could explore the use of machine learning algorithms to design more adaptive and intelligent mitigators. Machine learning models, such as neural networks, could be trained to predict the optimal mitigation actions based on historical attack data and real-time residual indices. This would enable a more dynamic response to cyberattacks, potentially improving the speed and accuracy of the mitigation process.

5. **Exploration of More Complex Attack Types**: Finally, future research should investigate other types of cyberattacks beyond FDIAs, such as denial-of-service (DoS) attacks, replay attacks, and coordinated multi-point attacks. Understanding how the DT framework responds to these different threats would provide a more comprehensive assessment of its robustness and versatility.

In conclusion, this thesis has made significant contributions to the field of microgrid cybersecurity by developing and validating a novel DT-based approach for detecting and mitigating FDIAs. The framework's success in simulation suggests that it shows potential to resume the research in this direction, particularly as microgrids become more prevalent in modern energy systems. The proposed future work outlines several opportunities to build on this research, potentially leading to even more robust and adaptable cybersecurity solutions for microgrids. As the energy sector continues to evolve, ensuring the security and stability of MGs will remain a priority, and the advancements presented in this thesis provide a solid foundation for ongoing innovation in this area.

Bibliography

[1] Albright, D., Brannan, P., y Walrond, C., "Did stuxnet take out 1,000 centrifuges at the natanz enrichment plant?", 2010, http://isis-online.org/isis-reports/detail/supplement-to-irans-gas-.

[2] Pasqualetti, F., Dorfler, F., y Bullo, F., "Attack detection and identification in cyber-physical systems", IEEE Transactions on Automatic Control, vol. 58, pp. 2715–2729, 2013, doi:10.1109/TAC.2013.2266831.

[3] Huang, Y., Tang, J., Cheng, Y., Li, H., Campbell, K. A., y Han, Z., "Real-time detection of false data injection in smart grid networks: An adaptive cusum method and analysis", IEEE Systems Journal, vol. 10, pp. 532–543, 2016, doi:10.1109/JSYST.2014.2323266.

[4] Abhinav, S., Modares, H., Lewis, F. L., Ferrese, F., y Davoudi, A., "Synchrony in networked microgrids under attacks", IEEE Transactions on Smart Grid, vol. 9, pp. 6731–6741, 2018, doi:10.1109/TSG.2017.2721382.

[5] Sahoo, S., Peng, J. C. H., Devakumar, A., Mishra, S., y Dragičević, T., "On detection of false data in cooperative dc microgrids - a discordant element approach", IEEE Transactions on Industrial Electronics, vol. 67, pp. 6562–6571, 2020, doi:10.1109/TIE.2019.2938497.

[6] Liu, L., Esmalifalak, M., Ding, Q., Emesih, V. A., y Han, Z., "Detecting false data injection attacks on power grid by sparse optimization", IEEE Transactions on Smart Grid, vol. 5, pp. 612–621, 2014, doi:10.1109/TSG.2013.2284438.

[7] Burgos-Mellado, C., Donoso, F., Dragicevic, T., Cardenas-Dobson, R., Wheeler, P., Clare, J., y Watson, A., "Cyber-attacks in modular multilevel converters", IEEE Transactions on Power Electronics, vol. 37, pp. 8488–8501, 2022, doi:10.1109/TPEL.2022.3147466.

[8] Liu, Y., Ning, P., y Reiter, M. K., "False data injection attacks against state estimation in electric power grids", en ACM Transactions on Information and System Security, vol. 14, 2011, doi:10.1145/1952982.1952995.

[9] Liu, Y., Du, Z., Chen, Y., y Zhan, H., "Resilient distributed control of islanded microgrids under hybrid attacks", Frontiers in Energy Research, vol. 11, 2023, doi:10.3389/fenrg.2023.1320968.

[10] Shafei, H., Li, L., y Aguilera, R. P., A Comprehensive Review on Cyber-Attack Detection and Control of Microgrid Systems, pp. 1–45. Springer Nature, 2023, doi:10.1007/978-3-031-20360-2_1.

[11] Jones, D., Snider, C., Nassehi, A., Yon, J., y Hicks, B., "Characterising the digital twin: A systematic literature review", CIRP Journal of Manufacturing Science and Technology,

vol. 29, pp. 36–52, 2020, doi:10.1016/j.cirpj.2020.02.002.

[12] Danilczyk, W., Sun, Y., y He, H., “Angel: An intelligent digital twin framework for microgrid security”, en 2019 North American Power Symposium (NAPS), pp. 1–6, 2019, doi:10.1109/NAPS46351.2019.9000371.

[13] Saad, A., Faddel, S., Youssef, T., y Mohammed, O. A., “On the implementation of iot-based digital twin for networked microgrids resiliency against cyber attacks”, IEEE Transactions on Smart Grid, vol. 11, pp. 5138–5150, 2020, doi:10.1109/TSG.2020.3000958.

[14] Farrokhabadi, M., Lagos, D., Wies, R. W., Paolone, M., Liserre, M., Meegahapola, L., Kabalan, M., Hajimiragha, A. H., Peralta, D., Elizondo, M. A., Schneider, K. P., Canizares, C. A., Tuffner, F. K., Reilly, J., Simpson-Porco, J. W., Nasr, E., Fan, L., Mendoza-Araya, P. A., Tonkoski, R., Tamrakar, U., y Hatziargyriou, N., “Microgrid stability definitions, analysis, and examples”, IEEE Transactions on Power Systems, vol. 35, pp. 13–29, 2020, doi:10.1109/TPWRS.2019.2925703.

[15] Shahzad, S., Abbasi, M. A., Ali, H., Iqbal, M., Munir, R., y Kilic, H., “Possibilities, challenges, and future opportunities of microgrids: A review”, 2023, doi:10.3390/su15086366.

[16] Sami, M. S., Abrar, M., Akram, R., Hussain, M. M., Nazir, M. H., Khan, M. S., y Raza, S., “Energy management of microgrids for smart cities: A review”, Energies, vol. 14, 2021, doi:10.3390/en14185976.

[17] Beg, N., Armstorfer, A., Rosin, A., y Biechl, H., “Mathematical modeling and stability analysis of a microgrid in island operation”, en 2018 International Conference on Smart Energy Systems and Technologies (SEST) Conference Proceedings, p. 1 – 6, 2018, doi:10.1109/SEST.2018.8495694.

[18] He, J., Wu, X., Wu, X., Xu, Y., y Guerrero, J. M., “Small-signal stability analysis and optimal parameters design of microgrid clusters”, IEEE Access, vol. 7, pp. 36896–36909, 2019, doi:10.1109/ACCESS.2019.2900728.

[19] Nandanoori, S. P., Kundu, S., Du, W., Tuffner, F. K., y Schneider, K. P., “Distributed small-signal stability conditions for inverter-based unbalanced microgrids”, IEEE Transactions on Power Systems, vol. 35, pp. 3981–3990, 2020, doi:10.1109/TPWRS.2020.2982795.

[20] Pires, V. F., Pires, A., y Cordeiro, A., “Dc microgrids: Benefits, architectures, perspectives and challenges”, Energies, vol. 16, no. 3, 2023, doi:10.3390/en16031217.

[21] Espina, E., Llanos, J., Burgos-Mellado, C., Cárdenas-Dobson, R., Martínez-Gómez, M., y Sáez, D., “Distributed control strategies for microgrids: An overview”, IEEE Access, vol. 8, pp. 193412–193448, 2020, doi:10.1109/ACCESS.2020.3032378.

[22] Simpson-Porco, J. W., Shafiee, Q., Dorfler, F., Vasquez, J. C., Guerrero, J. M., y Bullo, F., “Secondary frequency and voltage control of islanded microgrids via distributed averaging”, IEEE Transactions on Industrial Electronics, vol. 62, pp. 7025–7038, 2015, doi:10.1109/TIE.2015.2436879.

[23] Wang, F., Shan, Q., Teng, F., He, Z., Xiao, Y., y Wang, Z., “Distributed secondary control strategy against bounded fdi attacks for microgrid with layered communication network”, Frontiers in Energy Research, vol. 10, 2022, doi:10.3389/fenrg.2022.914132.

[24] Griffor, E. R., Greer, C., Wollman, D. A., y Burns, M. J., “Framework for cyber-physical systems: volume 1, overview”, 2017, doi:10.6028/NIST.SP.1500-201.

[25] “Guide for conducting risk assessments”, 2012, doi:10.6028/NIST.SP.800-30r1.

[26] Grieves, M., “Digital twin: Manufacturing excellence through virtual factory replication”, 2015.

[27] Madni, A. M., Madni, C. C., y Lucero, S. D., “Leveraging digital twin technology in model-based systems engineering”, Systems, vol. 7, 2019, doi:10.3390/systems7010007.

[28] Rojas, F., Cardenas, R., Clare, J., Diaz, M., Pereda, J., y Kennel, R., “A design methodology of multiresonant controllers for high performance 400 hz ground power units”, IEEE Transactions on Industrial Electronics, vol. 66, pp. 6549–6559, 2019, doi:10.1109/TIE.2019.2898610. Publisher Copyright: © 1982-2012 IEEE.

[29] Diaz, M., Cardenas, R., Wheeler, P., Clare, J., y Rojas, F., “Resonant control system for low-voltage ride-through in wind energy conversion systems”, IET Power Electronics, vol. 9, 2016, doi:10.1049/iet-pel.2015.0488.

[30] Cardenas, R. y Pena, R., “Sensorless vector control of induction machines for variable-speed wind energy applications”, Energy Conversion, IEEE Transactions on, vol. 19, pp. 196 – 205, 2004, doi:10.1109/TEC.2003.821863.

[31] Cardenas, R., Pena, R., Perez, M., Clare, J., Asher, G., y Vargas, F., “Vector control of front-end converters for variable-speed wind–diesel systems”, IEEE Transactions on Industrial Electronics, vol. 53, pp. 1127–1136, 2006, doi:10.1109/TIE.2006.878321.

[32] Mora, A., Cardenas, R., Aguilera, R., Angulo, A., Donoso, F., y Rodriguez, J., “Computationally efficient cascaded optimal switching sequence mpc for grid-connected three-level npc converters”, IEEE Transactions on Power Electronics, vol. PP, 2019, doi:10.1109/TPEL.2019.2906805.

[33] Donoso, F., Mora, A., Cardenas, R., Angulo, A., Saez, D., y Rivera, M., “Finite-set model-predictive control strategies for a 3l-npc inverter operating with fixed switching frequency”, IEEE Transactions on Industrial Electronics, vol. PP, 2017, doi:10.1109/TIE.2017.2760840.

ANNEXES

Annex A. Experiment 2: Testing Residual Indices for Detection in DT Configuration 1

A.1. Experiment 2.1: Step Attack Sequences

A.1.1. Scenario 1

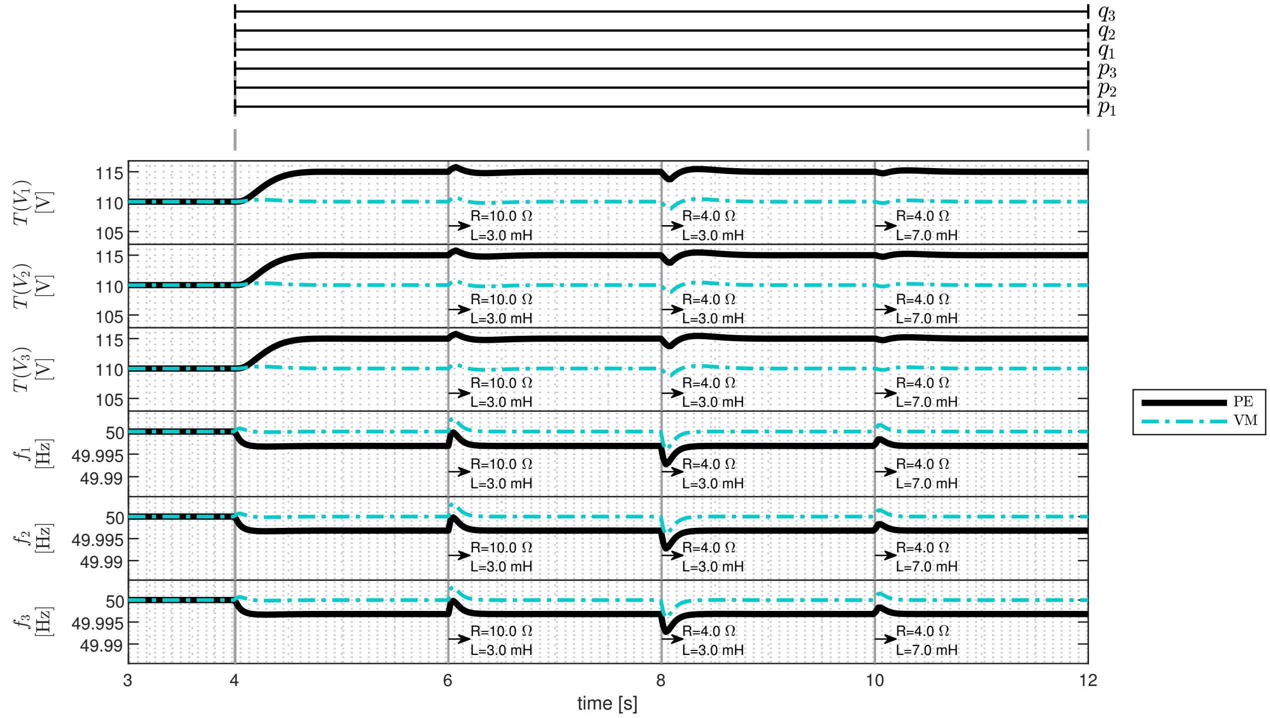

Figure A.1: Average voltage amplitude and frequency of each DER (of both the PE and VM) in DT configuration 1 against step attacks and load changes, specifically for $\bar{q}_i = -500$ and $\bar{p}_i = -100$.

A.1.2. Scenario 2

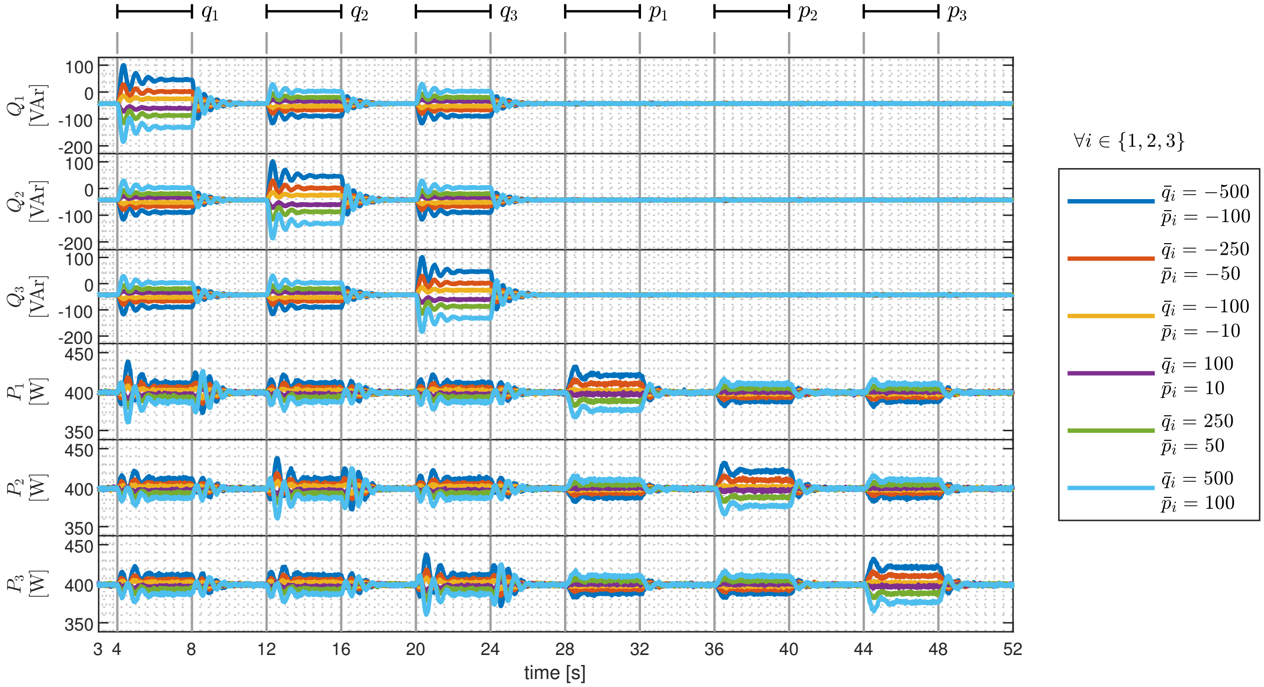

Figure A.2: Active and reactive power generated by each DER in DT configuration 1 against step attacks in isolated time intervals.

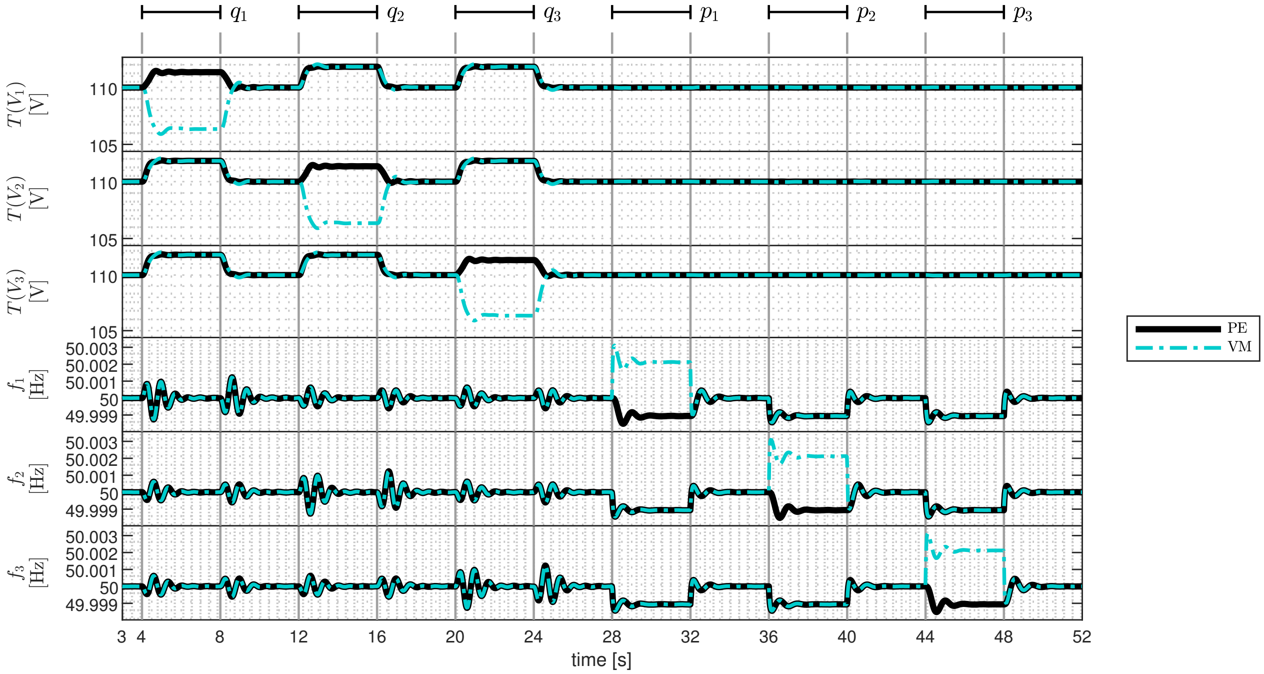

Figure A.3: Average voltage amplitude and frequency of each DER (of both the PE and VM) in DT configuration 1 against step attacks in isolated time intervals, specifically for $\bar{q}_i = -500$ and $\bar{p}_i = -100$.

A.1.3. Scenario 3

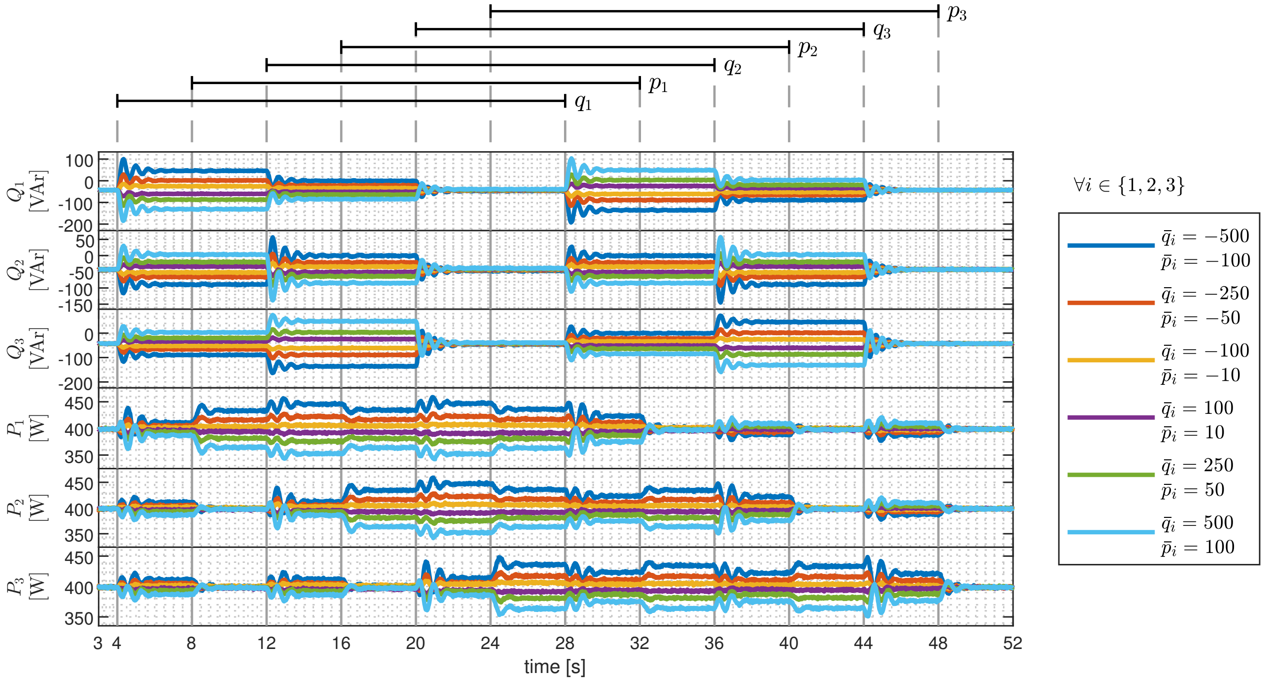

Figure A.4: Active and reactive power generated by each DER in DT configuration 1 against step attacks added progressively in time.

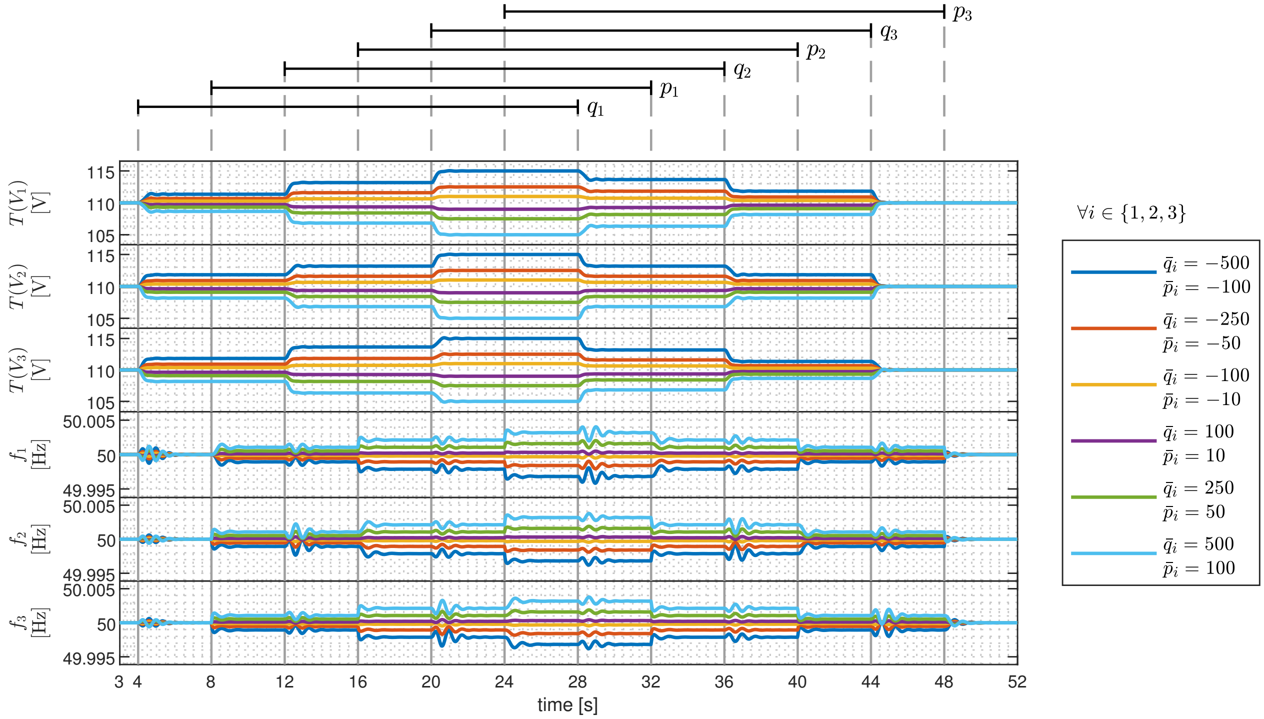

Figure A.5: Average voltage amplitude and frequency of each DER in DT configuration 1 against step attacks added progressively in time.

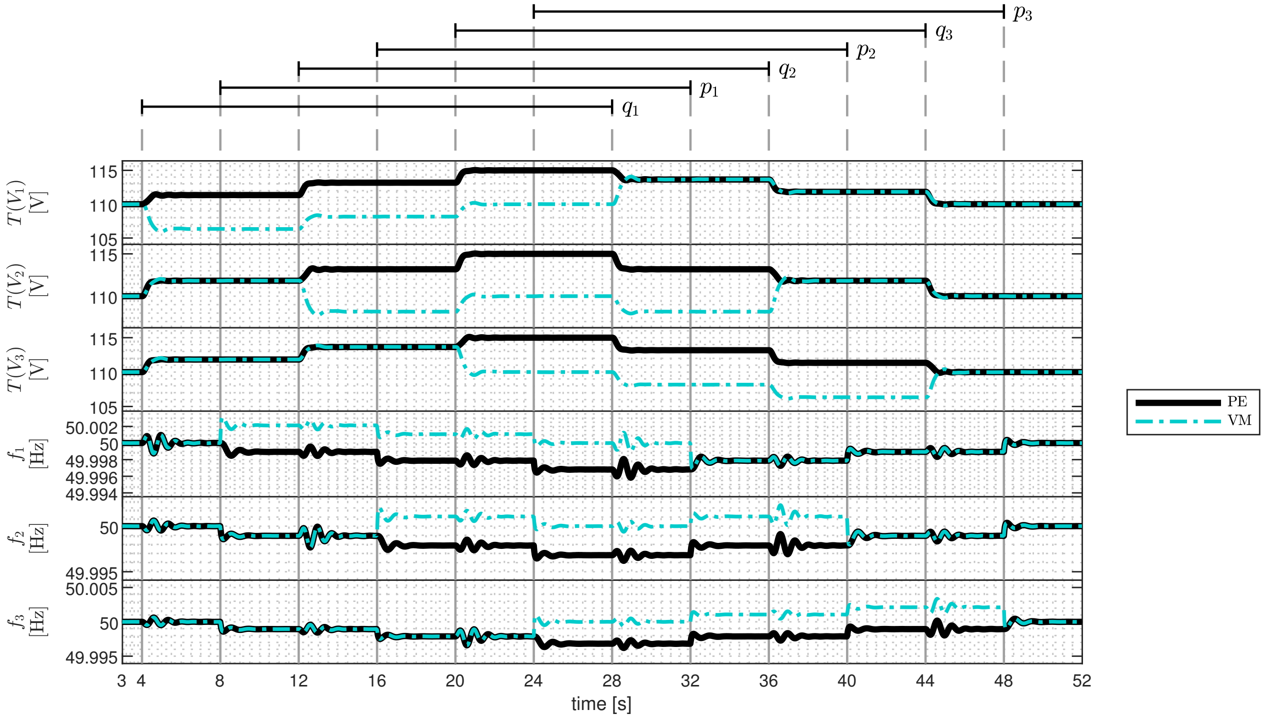

Figure A.6: Average voltage amplitude and frequency of each DER (of both the PE and VM) in DT configuration 1 against step attacks added progressively in time, specifically for $\bar{q}_i = -500$ and $\bar{p}_i = -100$.

A.2. Experiment 2.2: Truncated Ramp Attack Sequences

A.2.1. Scenario 1

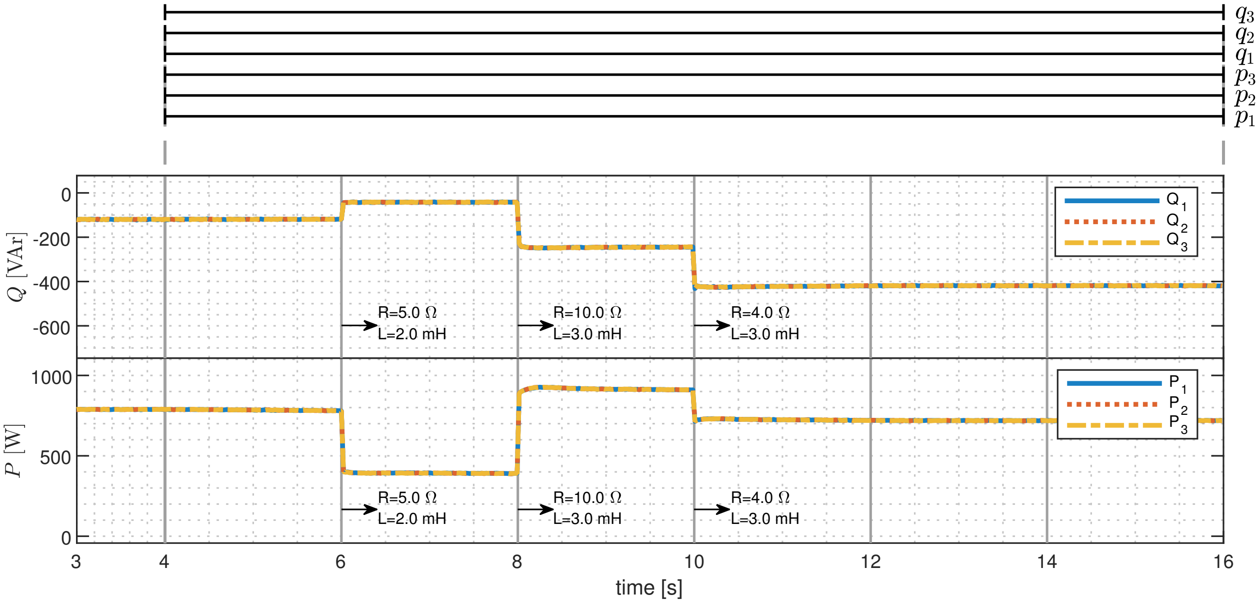

Figure A.7: Active and reactive power generated by each DER in DT configuration 1 against truncated ramp attacks and load changes.

A.2.2. Scenario 2

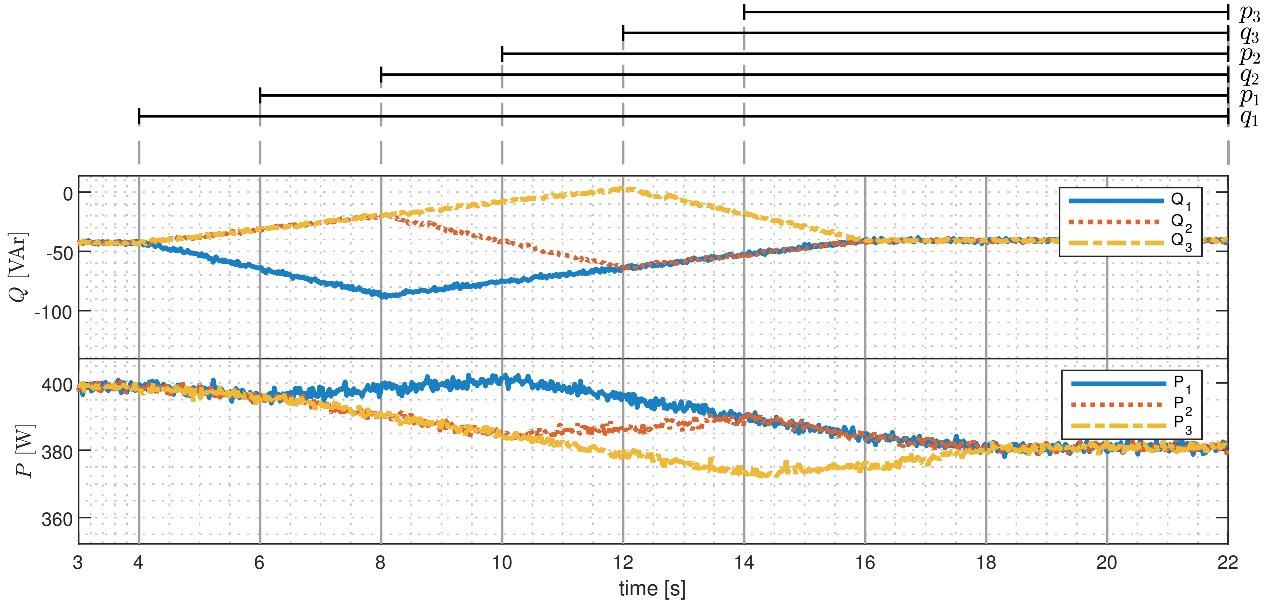

Figure A.8: Active and reactive power generated by each DER in DT configuration 1 against truncated ramp attacks added progressively in time.

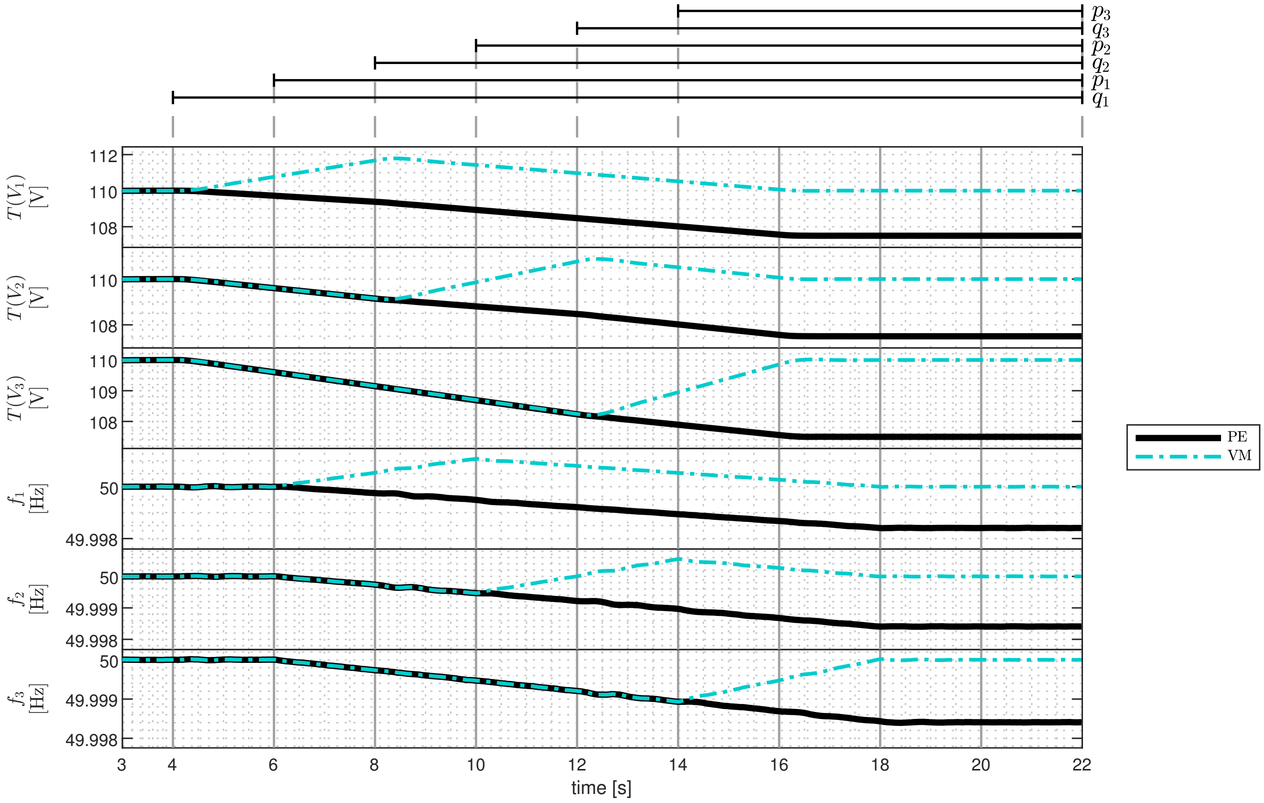

Figure A.9: Average voltage amplitude and frequency of each DER in DT configuration 1 against truncated ramp attacks added progressively in time.

Annex B. Experiment 3: Testing Detection and Mitigation in DT Configuration 2

B.1. Experiment 3.1: Step Attack Sequences

B.1.1. Scenario 1

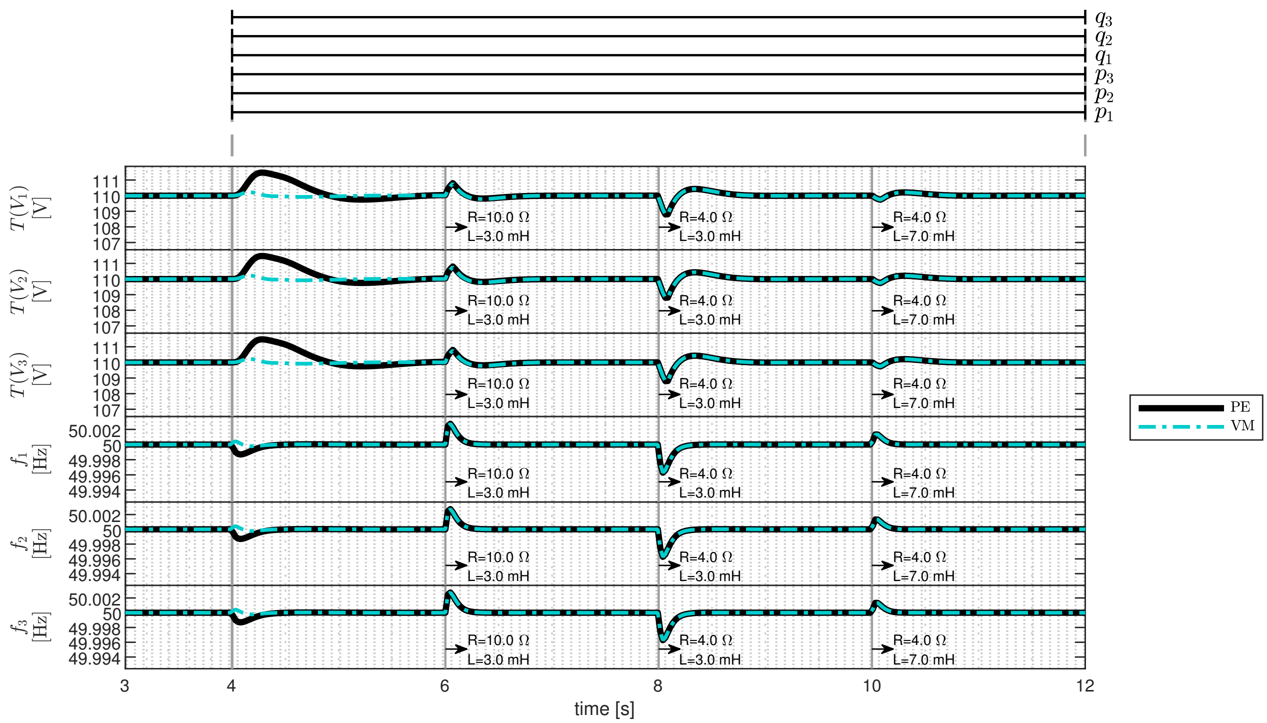

Figure B.1: Average voltage amplitude and frequency of each DER (of both the PE and VM) in DT configuration 2 against step attacks and load changes, specifically for $\bar{q}_i = -500$ and $\bar{p}_i = -100$.

B.1.2. Scenario 2

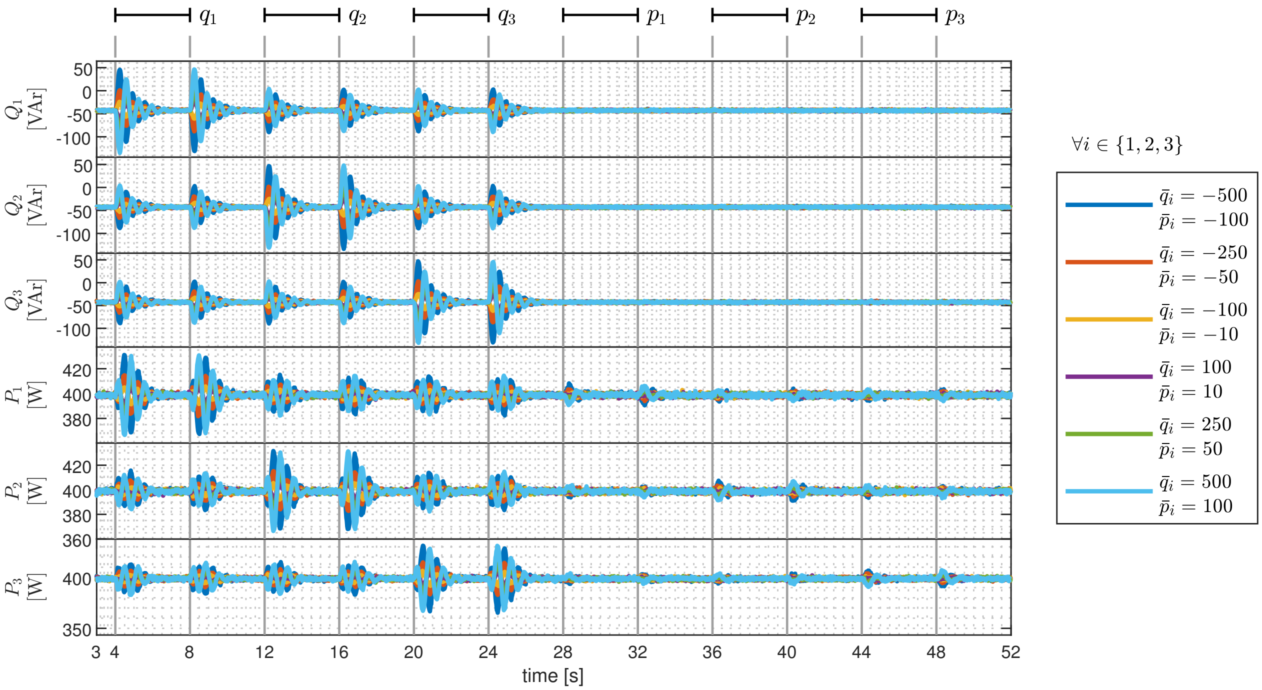

Figure B.2: Active and reactive power generated by each DER in DT configuration 2 against step attacks in isolated time intervals.

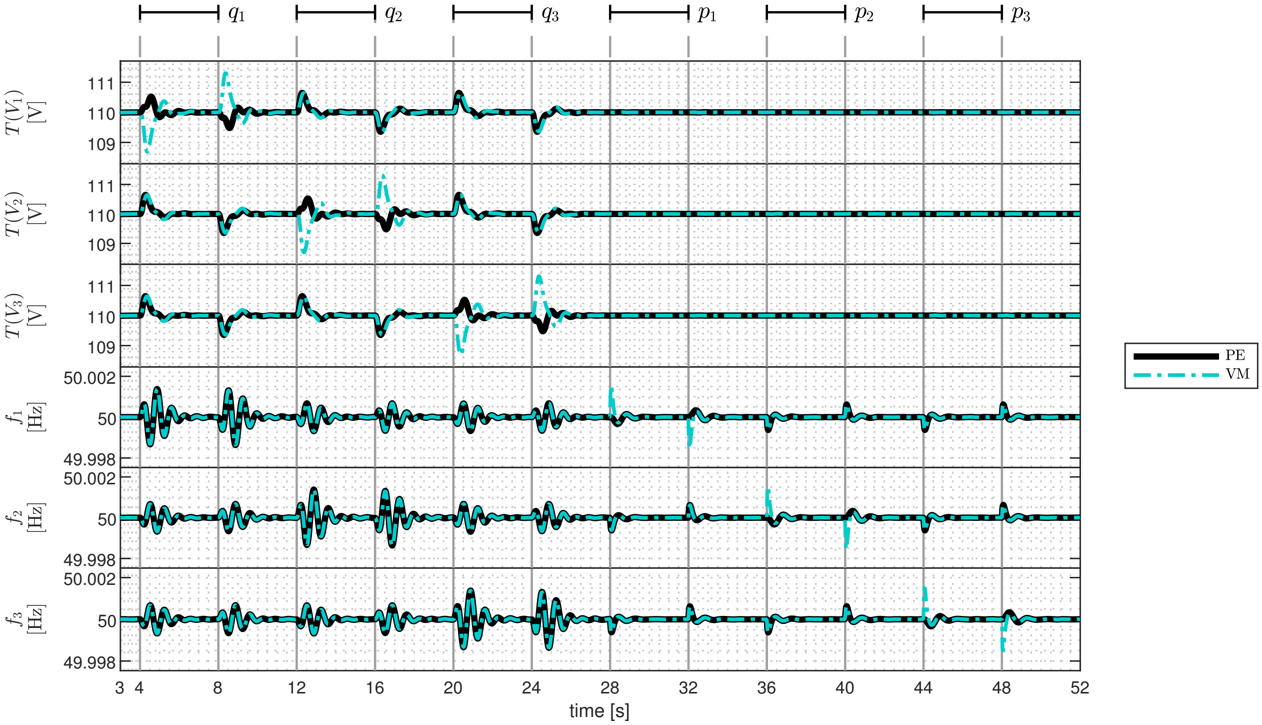

Figure B.3: Average voltage amplitude and frequency of each DER (of both the PE and VM) in DT configuration 2 against step attacks in isolated time intervals, specifically for $\bar{q}_i = -500$ and $\bar{p}_i = -100$.

B.1.3. Scenario 3

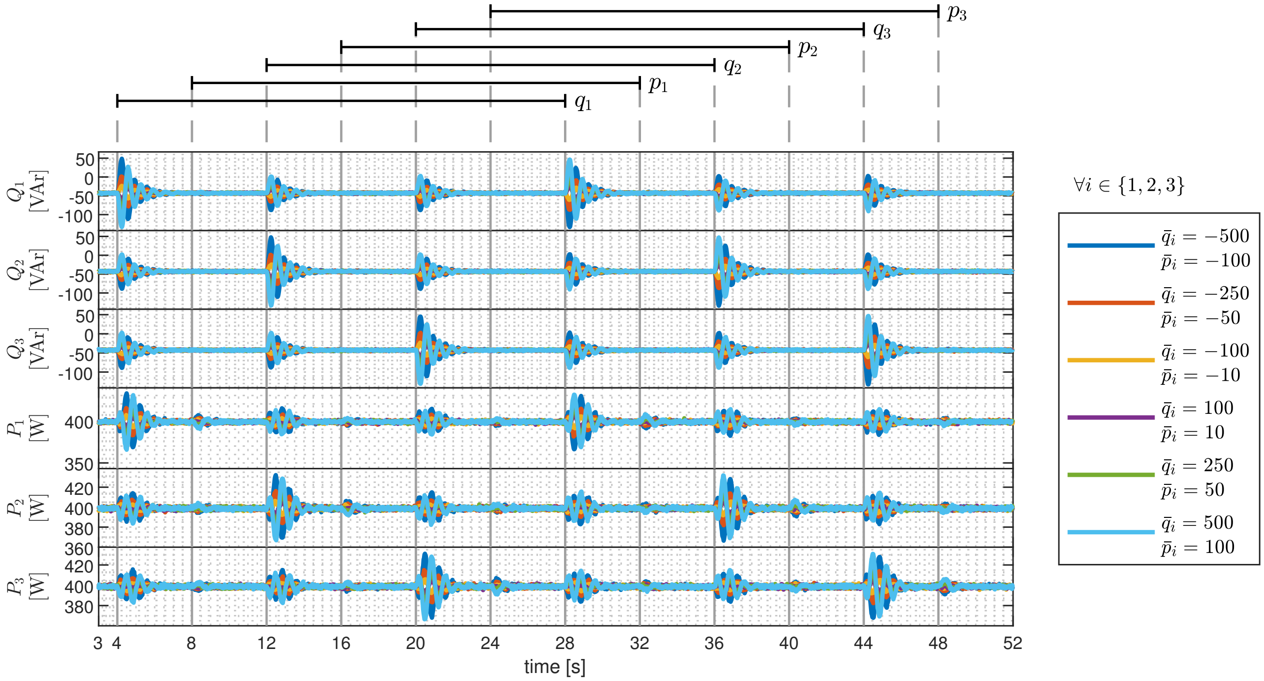

Figure B.4: Active and reactive power generated by each DER in DT configuration 2 against step attacks added progressively in time.

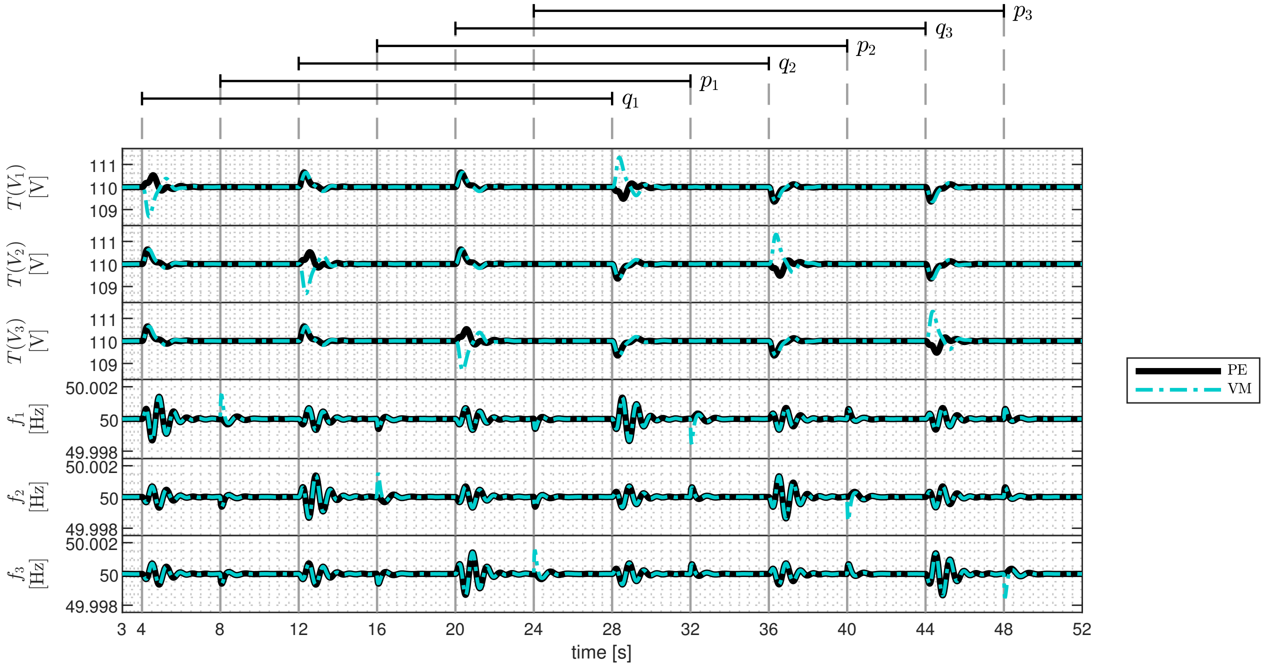

Figure B.5: Average voltage amplitude and frequency of each DER (of both the PE and VM) in DT configuration 2 against step attacks added progressively in time, specifically for $\bar{q}_i = -500$ and $\bar{p}_i = -100$.

B.2. Experiment 3.2: Truncated Ramp Attack Sequences

B.2.1. Scenario 1

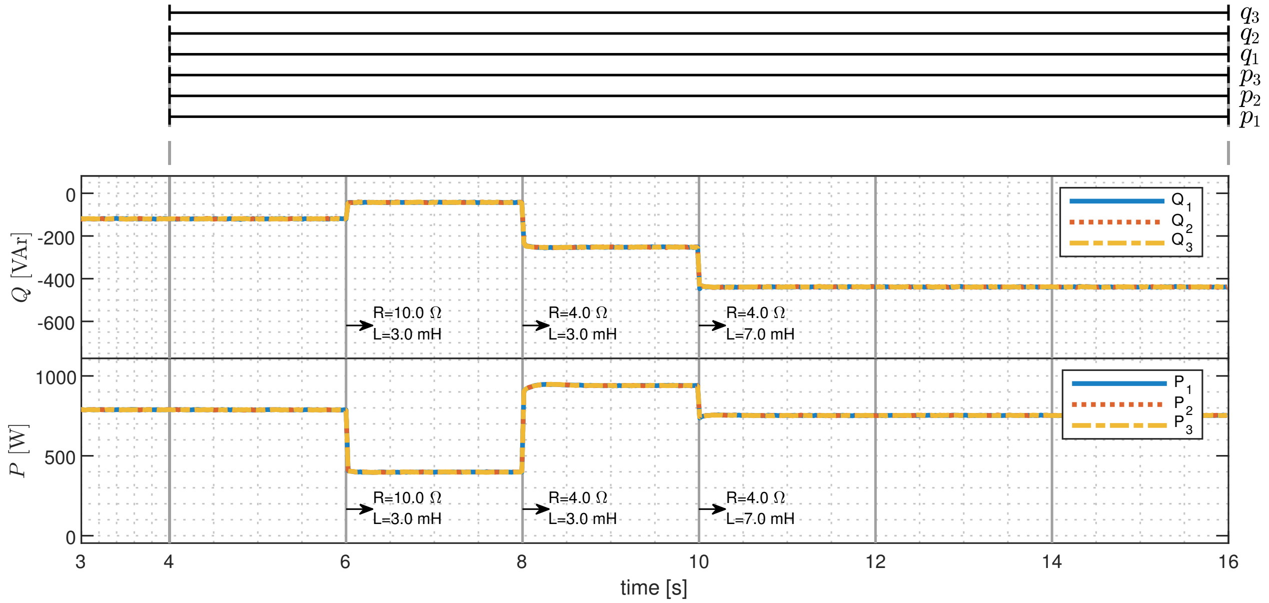

Figure B.6: Active and reactive power generated by each DER in DT configuration 2 against truncated ramp attacks and load changes.

B.2.2. Scenario 2

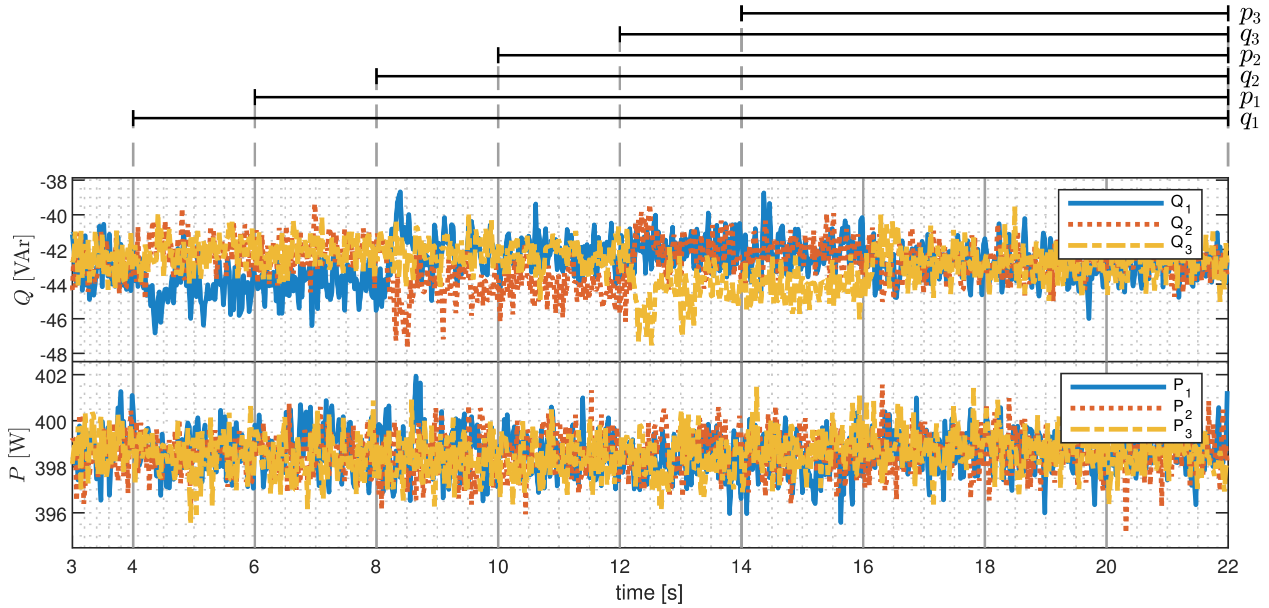

Figure B.7: Active and reactive power generated by each DER in DT configuration 2 against truncated ramp attacks added progressively in time.