NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING
DIVISION OF ELECTRIC POWER
ELECTRIC ENERGY SYSTEMS LABORATORY

A Hybrid Framework for the Cyber Resilience Enhancement of Frequency Control in Smart Grids

A dissertation submitted for the degree of Doctor of Philosophy
of

Andreas - Dorotheos Syrmakesis

Athens, July 2024

NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING
DIVISION OF ELECTRIC POWER
ELECTRIC ENERGY SYSTEMS LABORATORY

A Hybrid Framework for the Cyber Resilience Enhancement of Frequency Control in Smart Grids

A dissertation submitted for the degree of Doctor of Philosophy
of

Andreas - Dorotheos Syrmakesis

Advisory Committee: Nikolaos Hatziargyriou
Cristina Alcaraz
Georgios Korres

Approved by the seven-member committee on July 17, 2024.

.....	.....	.....
Nikolaos Hatziargyriou	Cristina Alcaraz	Georgios Korres
Professor Em., NTUA	Associate Professor, UMA	Professor, NTUA
.....	.....	.....
Charalambos Konstantinou	Haralambos Psilakis	Aris-Evangelos Dimeas
Associate Professor, KAUST	Lecturer, NTUA	Assistant Professor, NTUA

.....
Athanasios Voulodimos
Assistant Professor, NTUA

Athens
17th July 2024

.....

Andreas-Dorotheos Syrmakesis
Doctor of Philosophy in Electrical and Computer Engineering, NTUA

Ευχαριστίες

Με την παρούσα διατριβή ολοκληρώνεται ο κύκλος των διδακτορικών μου σπουδών. Ωστόσο, ο κύκλος είναι ένα σχήμα που ζει στο διδιάστατο χώρο, γεγονός που του επιτρέπει να περιγράφει επαρκώς τα επαναλαμβανόμενα γεγονότα. Ευτυχώς, η ζωή είναι πιο πολύπλοκο σχήμα. Έτσι, αν φανταστούμε ένα κύκλο να ξετυλίγεται και σε μια τρίτη διάσταση, τότε ίσως συνειδητοποιήσουμε πως οι ζωές μας μοιάζουν περισσότερο με σπείρες, οι οποίες με μικρές, πανομοιότυπες επαναλήψεις ταξιδεύουν στο άπειρο. Με την ελπίδα, λοιπόν, πως τα πέντε τελευταία χρόνια σπουδών δεν ήταν απλά ένας μονότονος κύκλος, θα αναδιατυπώσω την αρχική μου πρόταση λέγοντας πως η ολοκλήρωση της παρούσας διατριβής με φέρνει σε ένα νέο σημείο της προσωπικής μου σπειροειδούς κλίμακας.

Η μοναχικότητα του ακαδημαϊκού κόσμου ελαφραίνει όταν ο δρόμος του ερευνητή διασταυρώνεται με σπουδαίους μέντορες. Η δική μου τύχη με ευνόησε ώστε την πόρτα σε αυτό τον κόσμο να την ανοίξει διάπλατα ο επιβλέποντας Καθηγητής μου κ. Νίκος Χατζηαργυρίου. Ως ελάχιστη ένδειξη ευγνωμοσύνης, θα ήθελα να τον ευχαριστήσω διπλά, πρώτα για την εμπιστοσύνη του προς το πρόσωπο μου να αναλάβω το συγκεκριμένο ερευνητικό θέμα και έπειτα για το φως που έριχνε στα σκοτάδια που συναντούσαμε. Η στοχευμένη καθοδήγηση του σε συνδυασμό με την ελευθερία προσωπικής έκφρασης που μου παρείχε, οδήγησαν στην γέννηση επιστημονικών ιδεών για τις οποίες αισθάνομαι υπερηφάνος.

Στη συνέχεια, θα ήθελα να εκφράσω τις θερμές ευχαριστίες μου και στα υπόλοιπα μέλη της τριμελούς επιτροπής, την Αναπληρώτρια Καθηγήτρια κα. Cristina Alcaraz και τον Καθηγητή κ. Γεώργιο Κορρέ, για την πολύτιμη καθοδήγησή τους και τις σημαντικές συμβουλές που μου παρείχαν κατά την εκπόνηση της παρούσας διατριβής. Επίσης, ευχαριστώ ιδιαίτερα τον Αναπληρωτή Καθηγητή κ. Χαράλαμπο Κωνσταντίνου και τον Λέκτορα κ. Χαράλαμπο Ψυλλάκη για την αποδοχή συμμετοχής τους τόσο στην πενταμελή επιτροπή εξέτασης της ενδιάμεσης κρίσης όσο και στην επταμελή επιτροπή εξέτασης της διατριβής, καθώς επίσης και για τις χρήσιμες υποδείξεις και συμβουλές τους οι οποίες βελτίωσαν σημαντικά την ποιότητα της παρούσας διατριβής. Επιπλέον, εκφράζω την ευγνωμοσύνη μου στον Επίκουρο Καθηγητή κ. Άρη-Ευάγγελο Δημέα και στον Επίκουρο

Καθηγητή κ. Αθανάσιο Βουλόδημο για την αποδοχή συμμετοχής τους στην επταμελή επιτροπή εξέτασης της διατριβής. Συγκεκριμένα, τον κ. Δημέα τον ευχαριστώ από καρδιάς για την εμπιστοσύνη του να με κάνει μέλος της ερευνητικής ομάδας SmartRUE του εργαστηρίου Συστημάτων Ηλεκτρικής Ενέργειας (ΣΗΕ) του ΕΜΠ, για τις ηγετικές του δεξιότητες που αποτελούν πηγή έμπνευσης, καθώς και για την άριστη επαγγελματική συνεργασία που είχαμε όλα αυτά τα χρόνια.

Ξεχωριστή θέση στην καρδιά μου κατέχει ο Ερευνητής Hassan Alhelou. Ο Hassan διαδραμάτισε σπουδαίο ρόλο στην επιστημονική μου ωρίμανση, πράγμα για τον οποίο τον εκτιμώ βαθιά. Ο Hassan, μέσω της οξυδέρκειας του, αντιλαμβανόταν άμεσα τη μεγάλη εικόνα που είχε στο νου του ο επιβλέποντας Καθηγητής και την μετέτρεπε σε μικρά βήματα για εμένα. Έτσι, διευκόλυνε σε μεγάλο βαθμό το επιστημονικό μου έργο στα πλαίσια της παρούσας διατριβής.

Επιπλέον, θα ήθελα να ευχαριστήσω θερμά όλους μου τους συναδέλφους και συνεργάτες στην ερευνητική ομάδα SmartRUE του εργαστηρίου ΣΗΕ του ΕΜΠ για την άψογη συνεργασία και τις ευχάριστες αναμνήσεις που μοιραστήκαμε όλα αυτά τα χρόνια. Συγκεκριμένα, εκφράζω τις βαθιες μου ευχαριστίες στην κα. Ελένη Αυλωνίτου, στην κα. Ειρήνη Γασπαράκη και στην κα. Αλεξάνδρα Αδάμ για τη σπουδαία γραμματειακή υποστήριξη που παρείχαν κατά την εκπόνησης της παρούσας διατριβής.

Ο προσωπικός μου μόχθος στα χρόνια του διδακτορικού, παρά το μέγεθος του, δεν αρκούσε για την ολοκλήρωση αυτής της διατριβής. Χρειαζόταν κι ένα επιπλέον στοιχείο το οποίο, κατά ένα παράδοξο τρόπο, πολλαπλασιάζει μέσω της αφαίρεσης: το μοίρασμα. Σε όσους λοιπόν μοιραστήκαμε τα πέντε τελευταία (και όχι μόνο) χρόνια θα ήθελα να στείλω μια ζεστή αγκαλια και μια σεμνή υπόκλιση. Στους συνοδοιπόρους, σε όσους χάσαμε παρέα λίγο από τον εαυτό μας, μήπως και τον βρούμε στο παρακάτω του δρόμου. Και στην Κλεοπάτρα, που με επέστρεψε στο βασίλειο των χρωμάτων, όχι μόνο για να μου θυμίσει το ορατό φάσμα αλλά για να μου δείξει και νέες αποχρώσεις.

Κλείνοντας, θα ήθελα να ευχαριστήσω από καρδιάς τον Πίπη, τη Μαρία, τη Βιβή, το Σπύρο και το Γιώργο, για το ότι μου πρόσφεραν αυτό που ονομάζουμε οικογένεια. Όλοι τους μαζί γίνανε σκαλί στο ύψος που αντέχαν ώστε να φτάσω εγώ πιο κοντά στη δική μου αυτοπραγμάτωση. Ελπίζω την αγάπη που έλαβα από αυτούς να μπορέσω να την προσφέρω πολλαπλάσια τόσο στους ίδιους όσο και στους στενούς μου ανθρώπους, καθώς και σε όσους το έχουν πραγματικά ανάγκη.

Ανδρέας-Δωρόθεος Συρμακέσης,
Αθήνα, Ιούλιος 2024

"*Fear (if unmanaged) is the path to the dark side.*
Fear leads to Anger.
Anger leads to Hate.
Hate leads to Suffering."

Abstract

Modern power systems undergo a continuous digitalization for a more reliable, secure, and environmentally friendly operation. However, this advancement opens a door to a wide range of digital threats, making electrical grids vulnerable to cyberattacks. These malicious activities mainly affect the monitoring and control systems of smart power infrastructures. One of the most fundamental automation of power systems is the Load Frequency Control (LFC), which is responsible for maintaining the energy equilibrium in an electrical system by remotely adjusting the setpoints of the regulated generators. The criticality of LFC makes it a prime target for adversaries. Inspired by this threat, the present thesis introduces a novel set of active protection layers that detect, locate, estimate and mitigate the impact of cyberattacks against LFC. For each layer, both a model-based and a data-driven approach is designed, formulating a hybrid framework that increases the cyber resilience of LFC. The criteria for selecting the proper methodology at each layer are established according to the specifications of the system.

The model-based layers of the proposed hybrid framework are based on a special type of mathematical systems known as state observers. The related detection and localization methodologies use novel pairs of sliding mode (SMOs) and Luenberger observers to identify cyberattacks against LFC. The main benefit of these methodologies is their ability to distinguish cyberattacks from other types of external disturbances. Regarding the attack detection thresholds, an adaptive design has been selected to minimize false positive alarms. After determining which LFC signals have been corrupted, the introduced attack estimation technique takes place. This method approximates the characteristics of the identified cyberattacks by utilizing an innovative combination of SMO and unknown input observers. The estimated attacks are then fed to the proposed attack-resilient control to neutralize the effects of malicious activities against the considered system. The developed observer-based estimation and mitigation approaches employ an H_∞ method to minimize the effects of external disturbances on their performance.

The data-driven techniques of the introduced hybrid framework apply advanced deep learning algorithms to strengthen the cyber resilience of LFC. For the corresponding detection and localization methodologies, an autoencoder is trained on time-series that represent various

normal LFC states. After the training process, the model can replicate a given input with high accuracy under normal operation while it fails to achieve the same goal during a cyberattack. This feature makes the autoencoder a proper indicator for cyberattacks. Next, a deep neural network (DNN) is utilized for the proposed data-driven estimation and mitigation approaches. The DNN is trained on data that reflect the normal operation of LFC to estimate the healthy control signals through selected field measurements. The trained DNN is then deployed in the control center, along with backup communications channels that transfer the sensor readings and the approximated setpoints. When an attack is detected in the system, the original control loop is temporarily discarded and replaced by the proposed DNN, allowing the uninterrupted operation of LFC even under cyberattacks.

For the performance assessment of the designed cyber defense layers, a series of detailed experiments is conducted. Firstly, the effectiveness and the scalability of the proposed methodologies are tested on several use cases of growing complexity. In the LFC modeling of these use cases, several practical features have been considered, such as nonlinearities, high-voltage direct current (HVDC), thyristor controlled phase shifter-equipped (TCPS) tie-lines, disturbances due to Renewable Energy Sources (RES), etc., to emulate the operation of real-world power systems. The performance of the introduced methodologies in realistic conditions is further investigated through software/hardware-in-the-loop techniques. Next, the robustness of the presented approaches against various system uncertainties, such as system parameter miscalculations, noisy settings, time delays, etc., is numerically evaluated. Finally, the introduced cyber defense layers are compared with other, state-of-the-art works of the research field to highlight the contribution and the innovations of the present thesis.

Keywords

Smart Grids, Cybersecurity, Cyber Resilience, Cyberattacks, Load Frequency Control, Automatic Generation Control, False Data Injection Attacks, Sliding Mode Observers, Deep Neural Networks, Autoencoders.

Περίληψη Διατριβής

Τα σύγχρονα συστήματα ενέργειας υποβάλλονται σε συνεχή ψηφιοποίηση για πιο αξιόπιστη, ασφαλή και φιλική προς το περιβάλλον λειτουργία. Ωστόσο, αυτή η εξέλιξη εκθέτει τα ηλεκτρικά δίκτυα σε μια ευρεία γκάμα ψηφιακών απειλών, καθιστώντας τα ευάλωτα σε κυβερνοεπιθέσεις. Αυτές οι κακόβουλες δραστηριότητες επηρεάζουν κυρίως τα συστήματα παρακολούθησης και ελέγχου των έξυπνων ενεργειακών υποδομών. Ένας από τους πιο θεμελιώδεις αυτοματισμούς των ενεργειακών συστημάτων είναι ο έλεγχος φορτίου-συχνότητας (LFC), που είναι υπεύθυνος για τη διατήρηση του ισοζυγίου ενέργειας σε ένα ηλεκτρικό σύστημα, προσαρμόζοντας απομακρυσμένα την παραγωγή των ελεγχόμενων γεννητριών. Η χρισιμότητα του LFC τον καθιστά πρωταρχικό στόχο για τους επιτιθέμενους. Εμπνευσμένη από αυτή την απειλή, η παρούσα διατριβή παρουσιάζει ένα νέο πλαίσιο επιπέδων προστασίας που ανιχνεύουν, εντοπίζουν, εκτιμούν και μετριάζουν τον αντίκτυπο των κυβερνοεπιθέσεων εναντίον του LFC. Για κάθε επίπεδο, σχεδιάζεται τόσο μία προσέγγιση βασισμένη σε μοντέλα όσο και μία προσέγγιση βασισμένη σε δεδομένα, διαμορφώνοντας έτσι ένα υβριδικό πλαίσιο που αυξάνει την κυβερνοανθεκτικότητα του LFC. Τα κριτήρια για την επιλογή της κατάλληλης μεθοδολογίας σε κάθε επίπεδο καθορίζονται σύμφωνα με τις προδιαγραφές του συστήματος.

Το μέρος του προτεινόμενου υβριδικού πλαισίου που βασίζεται σε μοντέλα χρησιμοποιεί κυρίως έναν ειδικό τύπο μαθηματικών συστημάτων γνωστών ως παρατηρητές κατάστασης. Οι σχετικές μεθοδολογίες ανίχνευσης και εντοπισμού χρησιμοποιούν καινοτόμα ζεύγη παρατηρητών ολίσθησης (SMOs) και παρατηρητών Luenberger για την ταυτοποίηση κυβερνοεπιθέσεων εναντίον του LFC. Το κύριο πλεονέκτημα αυτών των μεθοδολογιών είναι η ικανότητά τους να διακρίνουν τις κυβερνοεπιθέσεις από άλλους τύπους εξωτερικών διαταραχών. Σχετικά με τα κατώφλια ανίχνευσης επιθέσεων, έχει επιλεγεί ένας προσαρμοστικός σχεδιασμός για την ελαχιστοποίηση των ψευδών συναγερμών. Μετά τον καθορισμό των σημάτων LFC που έχουν αλλοιωθεί, πραγματοποιείται η προτεινόμενη τεχνική εκτίμησης επίθεσης. Αυτή η μέθοδος προσεγγίζει τα χαρακτηριστικά των ταυτοποιημένων κυβερνοεπιθέσεων χρησιμοποιώντας έναν καινούργιο συνδυασμό SMO και παρατηρητών αγνώστου εισόδου. Στη συνέχεια, οι εκτιμώμενες επιθέσεις τροφοδοτούνται στον προτεινόμενο ανθεκτικό-σε-επιθέσεις έλεγχο για την εξουδετέρωση

των επιπτώσεων των κακόβουλων δραστηριοτήτων στο εξεταζόμενο σύστημα. Οι προτεινόμενες προσεγγίσεις εκτίμησης και μετριασμού των κυβερνοεπιθέσεων που βασίζονται σε παρατηρητές χρησιμοποιούν τη μέθοδο H_∞ για την ελαχιστοποίηση των επιπτώσεων των εξωτερικών διαταραχών στην απόδοσή τους.

Οι τεχνικές του προτεινόμενου υβριδικού πλαισίου που βασίζονται σε δεδομένα εφαρμόζουν προηγμένους αλγόριθμους βαθιάς μάθησης για την ενίσχυση της κυβερνοανθεκτικότητας του LFC. Για τις αντίστοιχες μεθοδολογίες ανίχνευσης και εντοπισμού, εκπαιδεύεται ένας αυτοκωδικοποιητής (autoencoder) σε χρονοσειρές που αντιπροσωπεύουν φυσιολογικές καταστάσεις του LFC. Μετά τη διαδικασία εκπαίδευσης, το μοντέλο μπορεί να αναπαράγει με υψηλή ακρίβεια τα δεδομένα εισόδου που αντιστοιχούν σε κανονική λειτουργία, ενώ αποτυγχάνει να πετύχει τον ίδιο στόχο κατά τη διάρκεια μιας κυβερνοεπίθεσης. Αυτό το χαρακτηριστικό καθιστά τον αυτοκωδικοποιητή κατάλληλο δείκτη εντοπισμού κυβερνοεπιθέσεων. Στη συνέχεια, ένα βαθύ νευρωνικό δίκτυο (DNN) χρησιμοποιείται για τις αντίστοιχες προτεινόμενες προσεγγίσεις εκτίμησης και μετριασμού κυβερνοεπιθέσεων που βασίζονται σε δεδομένα. Το DNN εκπαιδεύεται σε δεδομένα που αντικατοπτρίζουν τη φυσιολογική λειτουργία του LFC για την εκτίμηση των υγιών σημάτων ελέγχου μέσω επιλεγμένων μετρήσεων πεδίου. Το εκπαιδευμένο DNN εγκαθίσταται στη συνέχεια στο κέντρο ελέγχου, μαζί με εφεδρικά κανάλια επικοινωνίας που μεταφέρουν τις μετρήσεις αισθητήρων και τα εκτιμώμενα σήματα ρύθμισης. Όταν εντοπίζεται επίθεση στο σύστημα, ο αρχικός έλεγχος απορρίπτεται προσωρινά και αντικαθίσταται από το προτεινόμενο DNN, επιτρέποντας την αδιάλειπτη λειτουργία του LFC ακόμη και υπό κυβερνοεπιθέσεις.

Για την αξιολόγηση της απόδοσης των σχεδιασμένων επιπέδων κυβερνοάμυνας, πραγματοποιείται μια σειρά λεπτομερών πειραμάτων. Αρχικά, η αποτελεσματικότητα και η επεκτασιμότητα των προτεινόμενων μεθοδολογιών δοκιμάζονται σε διάφορα πειραματικά σενάρια αυξανόμενης πολυπλοκότητας. Για τη μοντελοποίηση του LFC σε αυτά τα σενάρια έχουν ληφθεί υπόψη αρκετά πρακτικά χαρακτηριστικά, όπως μη γραμμικότητες, γραμμές διασύνδεσης υψηλής τάσης (HVDC), γραμμές διασύνδεσης ελέγχου φάσης θυρίστορ (TCPS), διαταραχές λόγω ανανεώσιμων πηγών ενέργειας (RES) κ.λπ., για την προσομοίωση της λειτουργίας του συστήματος σε πιο πραγματικές συνθήκες. Η απόδοση των προτεινόμενων μεθοδολογιών σε ρεαλιστικά περιβάλλοντα διερευνάται περαιτέρω μέσω τεχνικών λογισμικού/υλικού-σε-βρόχο (SITL/HITL). Στη συνέχεια, η ανθεκτικότητα των παρουσιαζόμενων προσεγγίσεων έναντι διαφόρων αβεβαιοτήτων του συστήματος, όπως αστοχίες στους υπολογισμούς των παραμέτρων του συστήματος, θορυβώδεις συνθήκες, χρονοκαθυστερήσεις κ.λπ., αξιολογείται αριθμητικά. Τέλος, τα προτεινόμενα ε-

πίπεδα κυβερνοάμυνας συγκρίνονται με άλλες, σύγχρονες ερευνητικές μεθόδους για να αναδειχθούν η συνεισφορά και οι καινοτομίες της παρούσας διατριβής.

Λέξεις-κλειδια

Ευφυή ηλεκτρικά δίκτυα, Κυβερνοασφάλεια, Κυβερνοανθεκτικότητα, Κυβερνοεπιθέσεις, Έλεγχος φορτίου-συχνότητας, Αυτόματος έλεγχος παραγωγής, Παρατηρητές ολίσθησης, Επιθέσεις έγχυσης ψευδών δεδομένων, Βαθιά νευρωνικά δίκτυα, Αυτοκωδικοποιητές.

Contents

List of Figures

List of Tables

Nomenclature

Acronyms / Abbreviations

ACE Area Control Error

AD Attack Detection

AE Attack Estimation

AGC Automatic Generation Control

AI Artificial Intelligence

ALC Attack Localization

ARC Attack-resilient Control

CIA Confidentiality - Integrity - Availability

CPS Cyber-Physical Systems

CPU Central Processing Unit

DNN Deep Neural Network

DNP3 Distributed Network Protocol version 3.0

DoS Denial of Service

E-ISAC Electricity Information Sharing and Analysis Center

ENISA European Union Agency for Cybersecurity

FDIA False Data Injection Attack

GDB Governor Dead-Band

GRC Generation Rate Constraints

HITL Hardware-In-The-Loop

HVDC High-Voltage Direct Current

ICCP Inter-Control Center Communication Protocol

ICT Information and Communication Technologies

LFC Load Frequency Control

LMI Linear Matrix Inequality

LRS Load Reference Setpoint

LSTM Long Short-Term Memory

MSE Mean Squared Error

ML Machine Learning

NIST National Institute of Standards and Technology

RES Renewable Energy Source

RTAC Real-Time Automation Controller

RTDS Real Time Digital Simulator

SCADA Supervisory Control And Data Acquisition

SFR System Frequency Response

SG Smart Grid

SIL Software-In-the-Loop

SMO Sliding Mode Observer

SVM Support Vector Machine

TCPS Thyristor-Controlled Phase Shifter

TDA Time-delay Attack

TTD Transportation Time Delay

UIO Unknown Input Observer

Chapter 1

Introduction

1.1 Cybersecurity in Smart Grids

The growing demand for electrical energy at a global scale highlights the need for more reliable, secure, and environmentally friendly power systems. For this purpose, both research and industry communities in several parts of the world (e.g. U.S., E.U., China, Australia, etc.) [3, 4], focus their efforts on "smartening" the grid, in order to effectively accommodate the needs of all users, i.e., producers, consumers and prosumers. Smart Grids (SGs) are electricity networks that use advanced information and communication technologies (ICT) such as sensors, software applications, computer networks, and data analytics to provide efficient and sustainable energy services. ICT facilitates the monitoring and control of the power grid, which means that it can provide a better overview about the state of the grid and regulate its operation in an optimal manner.

While ICT offers a wide range of benefits, it also exposes SGs to several critical security challenges [5, 6]. The vulnerable spots that arise by the digital transformation of the power grid, pave the way for different types of cyberattacks. For instance, SG uses a group of heterogeneous communication technologies, such as ZigBee, wireless mesh networks, cellular network communication and powerline communication [7]. Their highly meshed structure along with the possible protocol incompatibilities can result in serious security gaps. In addition, the operation of power systems is still heavily dependent on proprietary and legacy technologies, such as conventional Supervisory Control and Data Acquisition (SCADA) systems whose design did not originally account for security measures. As a consequence, infrastructures that extensively utilize SCADA systems, such as SGs, are exposed to numerous digital risks [8]. Moreover, securing modern power systems in terms of cybersecurity is more challenging compared to the typical ICT-based infrastructures, due to their strict operational requirements and their criticality level [9].

Successful cyberattacks against Cyber-Physical Systems (CPS) have been already recorded, like the well-known case of the Ukrainian power system in December 2015. This large-scale incident is extensively reported by the SANS institute, the Electricity Information Sharing and Analysis Center (E-ISAC) and other power companies [10]. The coordinated attack consisted of malware installation via spear phishing emails, unauthorized access and SCADA system hijacking, which opened several circuit breakers remotely to interrupt the electricity supply to consumers. It also involved Denial of Service (DoS) attacks on telephone systems to prevent customers from emergency reporting to the operators. The power disruptions caused by this attack approximately affected 225,000 customers. Another notorious software, called Stuxnet, was uncovered in 2010 [11]. Stuxnet worm targeted the hosts of specific Siemens industrial control systems that were running on Windows environment and it mainly affected Iranian nuclear facilities [12]. For this reason, protecting SG systems from malicious activities is currently an active research area [13], relevant for governments [5], international organisations such as the European Union Agency for Cybersecurity (ENISA) [6] and the National Institute of Standards and Technology (NIST) [14, 13], and the academic community.

1.2 Cyber Resilience of Power Systems

Resilience is one of the most important attributes of the power grid as it ensures the uninterrupted delivery of the electrical energy. Currently, there is an extensive list of definitions for the power system resilience, provided by international institutions and organizations [15–18]. According to [19], the majority of these definitions agree that power system resilience is *the capability of a system to endure, assimilate, and promptly recuperate from an external catastrophic incident characterized by high impact but low probability*.

As electrical systems evolve rapidly over time, new types of undesired events affect their resilience, such as cyberattacks. Thus, it is critical to reconsider the typical concept of power system resilience in order to include the impact these emerging incidents. To this end, the definition of resilience provided by [19] is extended in [1] in order to include the cyber part of SGs, establishing the attribute of *cyber resilience*. Based on [1], cyber resilience is viewed as ***the ability of a system to preserve its operational state in the presence of successful cyberattacks***. More specifically, cyber resilience focuses on the minimization of the cyberattack impact on power grids and the prompt recovery from these incidents. Cyber resilience is a relatively new principle for modern power grids that has to be carefully considered by system designers.

To provide more insights on the term of cyber resilience, the typical power system resilience curve presented in [19] is modified and adjusted in [1] for the case of cyberattacks. This cyber resilience curve for SGs is depicted in Figure 1.1. In this graph, the evolution of the system performance in the event of a cyberattack is illustrated. It is a highly useful tool towards the deeper understanding of the different cyber resilience states along with their corresponding defensive measures, such as: robustness/resistance, resourcefulness/redundancy, adaptive self-organization, etc. The level of each resilience state is calculated based on selected resilience metrics, e.g. the number of customers affected or the number of residents in a population impacted, which quantitatively express the system reliability or power quality.

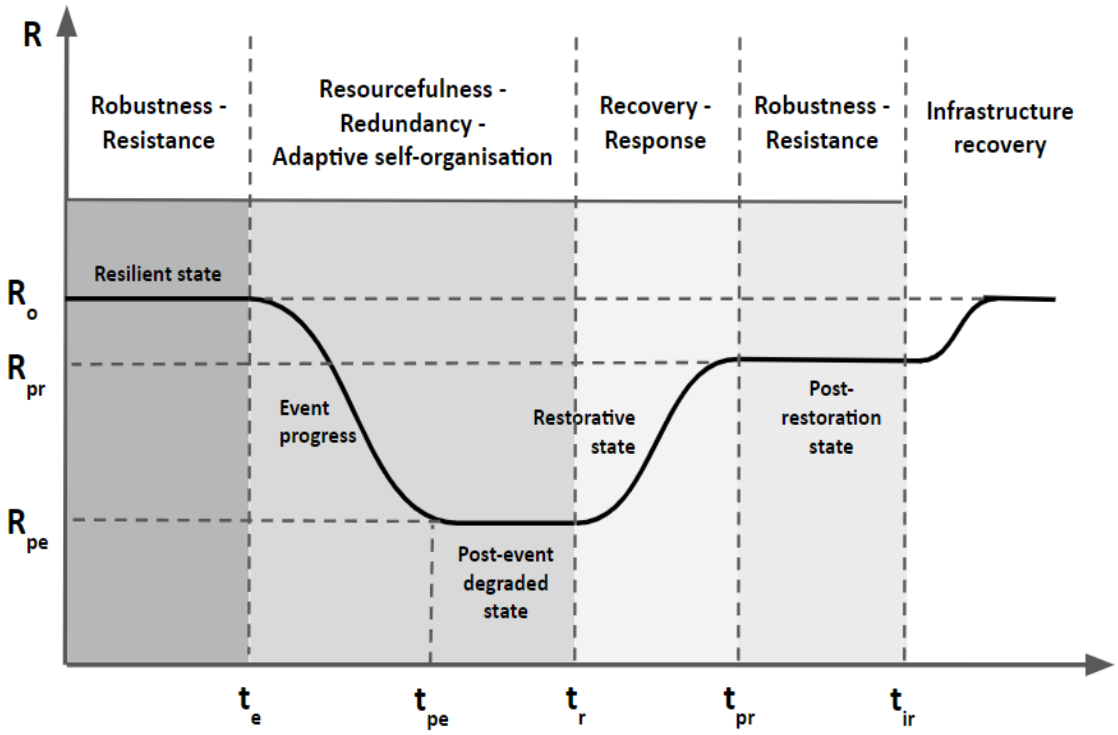


Figure 1.1 Resilience curve [1].

For a better comprehension of Figure 1.1 and the concept of cyber resilience, a detailed analysis of the different resilience states is presented in what follows:

- **Resilient state:** at this state, a well-designed power system could neutralize the impact of a launched cyberattack. Configuring a secure and intrusion tolerant grid in this phase provides a high resilience level which makes the SG capable of preventing unauthorized access and successful attacks.

- **Post-event degraded state:** in case of a successful cyberattack, the performance of the power system degrades; the percentage of this degradation depends on the impact of the attack and the preventive measures that have been applied. Key resilience techniques help reduce the impact of the attack and facilitate the progress to restoration state. For example, redundancy provides operational flexibility to the power system by offering additional resources. It should be noted that the duration of this state can be very short, thus transforming the trapezoidal shape of the resilience curve to triangular.

- **Restorative state:** at this state, the compromised power system has managed to mitigate the cyberattack and is gradually returning to its normal condition. Its recovery is almost fully completed. For example, after an accomplished attack, the power grid should modify its functionality, allocate alternative resources and optimally restore affected components or applications.

- **Post-restoration state:** this is the state where the recovery process has been completed and the power system is again operational. Nevertheless, its resilience level R_{pr} might be lower than its initial value R_o. Operational recovery refers to bringing the system back into a functional state, while infrastructure recovery refers to the restoration of the resilience level of the system to its initial value. For example, if all replicas of a SCADA master are compromised, restoring at least one of them will make the system operational again. However, all the replicas of the SCADA master have to be restored in order to reach the initial resilience level of the system.

At this point, it is important to explain the meaning of the different variables depicted in Figure 1.1:

- R_o: initial resilience value,
- R_{pe}: resilience value after a successfully completed cyberattack,
- R_{pr}: resilience value after attack mitigation,
- t_e: starting time of the cyberattack,
- t_{pe}: end of the cyberattack,
- t_r: starting time of the attack mitigation,
- t_{pr}: end time of attack mitigation and
- t_{ir}: starting time of infrastructure recovery.

1.3 Cyber Resilience of Frequency Control

As mentioned in Section 1.1, ICT enables a more efficient monitoring and control of power systems compared to conventional energy infrastructures. Among the several control mechanisms facilitated by ICT, load frequency control (LFC) is one of the most critical. The role of the LFC is the preservation of the energy balance between generation and demand in power grids to prevent any performance degradation of the system. A key indicator of energy energy equilibrium is the deviation of frequency from its nominal value, as illustrated in Fig. 1.2. To achieve the energy balance, LFC receives frequency measurements from the power plant, computes the control signal and sends the resulting setpoints to the generators.

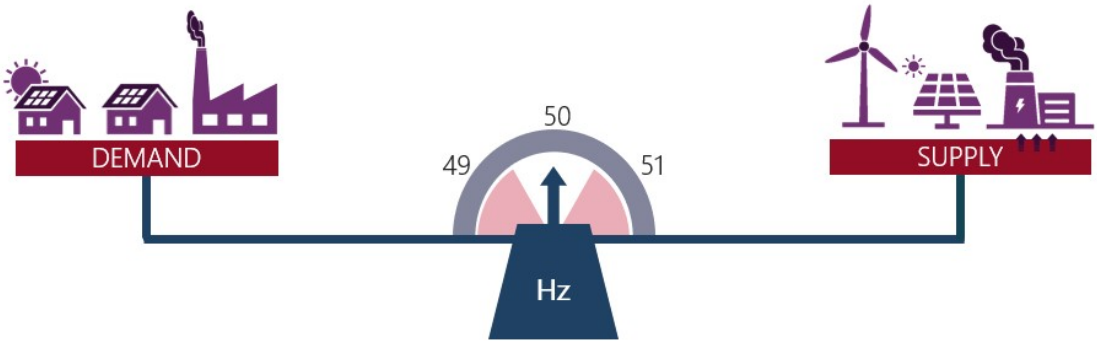


Figure 1.2 Frequency as indicator of energy equilibrium.

While there are multiple levels of frequency control, this study focuses on the two most fundamental ones, namely *primary control* and *secondary control*. Primary control is a functionality provided by the governors of generators at their local connection points to stabilize the frequency within acceptable values. Secondary control, also known as automatic generation control (AGC), is a remote communication mechanism responsible for the restoration of frequency back to its nominal value. In AGC, the communication of the physical system with its cyber layer is achieved through industrial network protocols, such as Distributed Network Protocol version 3.0 (DNP3) [20], Inter-Control Center Communication Protocol (ICCP) [21], etc. However, the vulnerabilities of these protocols [22], [23] expose the LFC to numerous cybersecurity dangers and decrease its levels of cyber resilience.

There are various types of cyber threats that affect the cyber resilience of LFC. For example, the secondary control of LFC is disabled when the system is under DoS attacks [24] and the frequency is not restored to its nominal value. Furthermore, the LFC, as a typical CPS, is also susceptible to a special kind of cyberattacks, different from those faced by standard ICT systems: data integrity attacks. Data integrity attacks include false data

injection attacks (FDIAs) [21], time-delay switching attacks (TDSAs) [25] and replay attacks [26]. FDIAs stealthily modify the data exchanged across the communication links, TDSAs strategically embed time delays into the LFC loop, and replay attacks send replicas of a particular network packet stream over a certain time period. Data integrity attacks lead the physical system to deregulation and cause financial losses to the system operator.

The significance of LFC necessitates the development of a defense-in-depth strategy to ensure its cyber resilience [27]. Since the functionality of LFC is based on ICT, the initial layer of this strategy can be provided through the standard information security practices. Such approaches include the installation and configuration of firewalls, authentication mechanisms, deployment of virtual private networks (VPNs), encryption of critical data, etc. The standard cybersecurity solutions have a proactive role and focus on preventing adversaries from launching any type of cyberattacks. Apart from the ICT systems, LFC also utilizes the control center of the power system, which allows the deployment of intelligent algorithms towards the elimination of cyberattacks. This capability forms another, active layer of cybersecurity for LFC that takes place after the successful execution of a cyberattack. The active cybersecurity approaches are typically performed using mathematical models, statistical techniques or data-driven algorithms. For a better comprehension, the synergy between the preventive and active cyber defense layers of LFC is illustrated in Fig. 1.3.

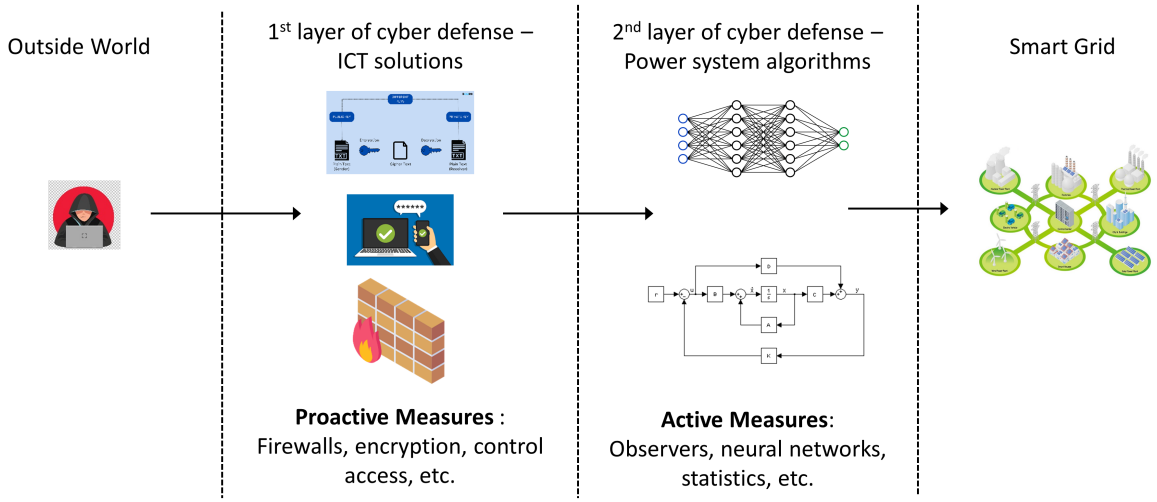


Figure 1.3 Cybersecurity layers for LFC.

The arsenal of active cybersecurity solutions for LFC is composed of the attack detection (AD), the attack localization (ALC), the attack estimation (AE) and the attack-resilient control (ARC) methods. More specifically, attack detection determines whether and when an attack has been successfully launched against the power system, while attack localization refers to the identification of the parts of the system (sensors, controllers, etc.) that are under

attack. Furthermore, attack estimation provides detailed information about the characteristics of accomplished cyber intrusions, such as shape, magnitude, duration, etc., and is essential for designing an attack-resilient control scheme for LFC. Finally, attack-resilient control is an advanced stability mechanism that preserves the normal functionality of a power system even in the presence of cyberattacks. Each of these methodologies can be applied sequentially, starting from AD and proceeding to ALC, AE and ARC, in order to form a multi-layer cybersecurity mechanism that can identify malicious activities and mitigate them.

If an active cyber resilience methodology for LFC is effectively applied to a specific power system, it is not guaranteed that it will preserve its performance in more or less complex grids. Therefore, it is necessary to examine (both theoretically and numerically) the effectiveness of a proposed cybersecurity solution across all types of power grids, from simplistic to more complicated ones. This necessity introduces the concept of scalability. Scalability is defined as the applicability of a cyber resilience methodology to various types of electrical systems, regardless of their size or complexity. Since power systems are constantly evolving infrastructures with dynamic operating points, scalability is considered a significant performance aspect. However, the works in the literature that study scalability are limited, despite its criticality. For this reason, this feature is thoroughly investigated in the present thesis.

1.4 Literature Review

1.4.1 Description & classification of related works

The importance of LFC has prompted researchers to propose various methodologies for enhancing its cyber resilience against FDIAs. For these methods, three main categories are identified: (i) *model-based*, (ii) *observer-based*, and (iii) *data-driven* approaches. In model-based methods, algorithms that process system knowledge are usually developed to tackle the effects of cyberattacks; observer-based techniques leverage the generated estimation errors to provide FDIA approximation formulas and attack-resilient LFC architectures; finally, data-driven approaches use deep learning architectures for capturing the dynamic behavior of LFC under healthy and attack conditions in order to eliminate the FDIA impact. Based on our literature review, we have identified that the related works can be classified into three main categories, which are illustrated in the following figure: The aforementioned categories are thoroughly explained in what follows:

- **model-based methods**: in this category, the proposed defense methods extract system knowledge/information and properly process them in order to identify underlying

patterns that can reveal useful insights about the attacking strategy. Some indicative examples of this category are the use of load forecasting to approximate the correct generator setpoints in case of cyberattack, the deployment of sophisticated Kalman filters that leverage the system modeling to estimate cyberattacks and the implementation of statistical methods to predict the healthy behavior of the frequency control signals.

- **observer-based methods**: this group of research methodologies leverages a special type of systems, called observers, to increase the cyber resilience of frequency control in power systems. Observers can accurately estimate the state vector of the real-world LFC systems they are designed for. From the model of the observer, a formula is extracted to describe the behavior the estimation error, i.e. the difference between the actual and the estimated state vector. This formula is utilized to tackle cyberattacks against LFC and its structure depends on the objective of the introduced methodology, e.g. detection, estimation or mitigation.

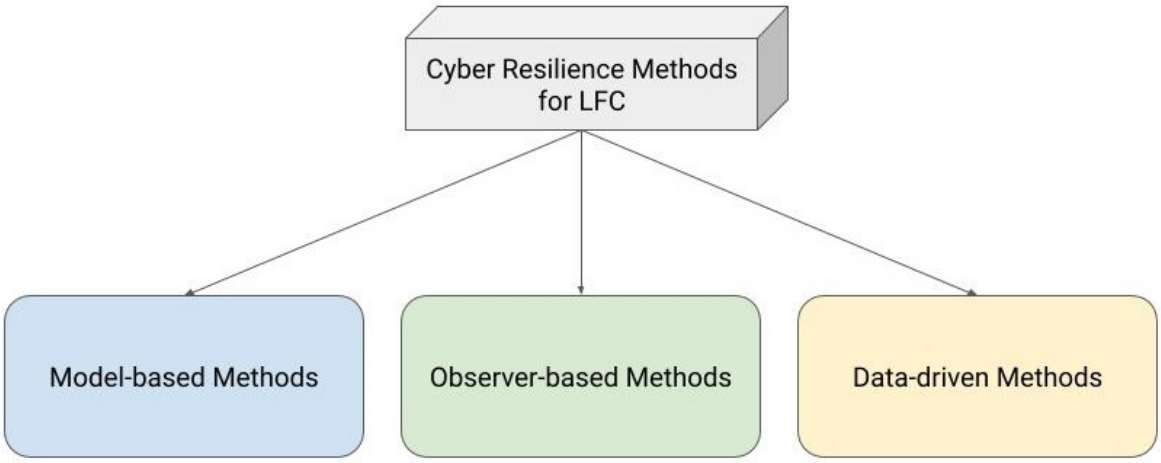


Figure 1.4 Classification of related works.

- **data-driven methods**: instead of using an analytical model of the power system frequency control, as other categories do, this type of methodologies utilizes the data generated by the actual LFC system to actively eliminate digital risks. Data-driven methodologies typically use historical databases, which keep track of past values of the LFC signals, to train their models. After this learning process, the data-driven models can determine the LFC status and acquire critical information about the compromised signals. In specific methods, these historical databases also serve as input to the developed data-driven models.

1.4.2 Related works

Various intelligent control mechanisms have been developed as active cyberattack response mechanisms for power systems and CPS in general. Particularly, the physical control model in [28] is integrated with an additional commensurate response module to tackle setpoint and actuator attacks, without considering attacks against sensors [29]. In [30], the feedback control of a standard distributed energy resources (DER) unit is integrated with a sliding mode observer to decrease the impact of cyberattacks. An attack-resilient control policy for the energy management system is presented in [31], which focuses on preserving the stability of the physical system during and after a cyberattack. The decentralized secure LFC scheme that is established in [32] can eliminate the detrimental impact of complex cyberattacks. In [33], compressed sensing techniques are applied to estimate the state of the plant during attacks. An attack-resilient state estimator is proposed in [34] which is applied to the cruise control of an electric, unmanned vehicle. Finally, a control method based on a recursive filtering algorithm is implemented in [35] to tackle specific sensor attacks. This technique estimates the states of the system by leveraging the redundant information of the controller.

Game theory is another scientific field that can provide defensive strategies for strengthening the cyber resilience of smart grids. To achieve this objective, game theory reveals the optimal responses to cyberattacks based on the activities of the adversaries. More specifically, a sequential game between an adversary and a SCADA administrator is formulated in [36–38] to analyze their interactions in case of cyberattacks. Furthermore, [39] utilizes a non-cooperative, differential game to discover the countermeasure vector against malicious activities that stealthily compromise DER actuators. In [40], a zero-sum game is modeled to represent the decision-making process between a sensor node and an adversary that launches DoS attacks. A strictly competitive game is also designed in [41] which approximates the interaction between the attacker and the defender in case of cyberattacks against power systems state estimation. Finally, a game theory-based framework is developed in [42] that analyzes the interaction between the controller and the adversary to mitigate the launched FDIAs.

Reinforcement learning is another commonly used approach for the cyber resilience enhancement of power grids. This technique is defined as the process that enables an agent to adopt the optimal behavior by interacting with a dynamic environment via trial-and-error [43]. To this end, a Q-learning technique is implemented in [44]. This strategy models the importance of the communication channels in a power system to find the optimal link recovery sequence under a limited budget. Furthermore, a Q-learning is applied in [45] to discover an optimal link/node recovery sequence in feasible time. In [46], the optimal re-closing time of power transmission lines after a successful cyber attack is investigated using a deep

reinforcement learning method. A reinforcement learning method is also proposed in [47] to maintain the cyber resilient state of an SG that uses cognitive radio network technology. The transmitter and the receiver of this methodology follow a multi-armed bandit approach to choose the most likely available and jamming-free communication channels in case of a jamming attack.

Model-based approaches are extensively used for increasing the cyber resilience of power systems and CPSs. For example, a representative linear model is developed in [48] to provide a cyberattack detection baseline and replace the tampered system data. This model is obtained by linearizing the Tennessee-Eastman process model [49] about the steady-state operating conditions. Similarly, in [50], a SCADA system with software defined networking (SDN) [51] assistance is presented, which replaces the compromised measurements with estimated ones. For evaluation, an extension of the MiniCPS [52] is developed in order to provide SDN functionalities for both supervisory and field networks. In the same context, an algorithm is proposed in [53] that estimates which sensor data links have been affected by cyberattacks. If any attack is identified, the power export deviation is accounted for the ACE computation, otherwise an attack-mitigating state estimation program is initiated. The performance of this algorithm is evaluated on a 37-bus power system model simulated in PowerWorld [54]. Regarding the model-based methods developed for LFC, a representative approach is presented in [21]. This defense mechanism uses the real-time load forecasts to approximate LFC control signals, which replace the actual ones in case of cyberattacks. In [55], a cyber-attack detection and mitigation platform (CDMP) is introduced, which utilizes the forecasted data of area control error for identification and mitigation of cyberattacks. In [56], the limitations of the Kalman filter are overcome by an input/state estimation-based algorithm which is developed to detect and approximate measurement FDIAs in the LFC system. Similarly, an attack-resilient frequency control scheme is introduced in [57] based on attack detection through state estimation. For 100% renewable energy power systems, the method designed in [58] can mitigate cyberattack impact by using a cascaded extended state filter and a robust decision-making model.

The design of effective observer structures is a well-studied research field and as a result, several observer-based techniques have been proposed for the cyber resilience enhancement of power systems. Particularly, a robust detection algorithm for SGs is developed in [59] using an adaptive observer that takes the stealthy characteristics of the bias load injection attack into account. Similarly, an unknown input interval observer-based detection and isolation scheme for FDIAs against the monitoring and control of SGs is introduced in [60]. In [61], an observer-based predictive control mechanism for grid-interactive inverters is presented with intrusion detection capabilities. Furthermore, a decentralized detection

and mitigation algorithm based on a state and attack observer is introduced in [62] for the interconnected power system subject to multi-area multichannel FDIAs. Regarding microgrids (MGs), an unknown input observer is deployed in [63] to detect and estimate cyberattacks against the secondary frequency control loop. Moreover, two typical types of observer-based schemes are proposed in [64] to tackle the attack detection problem for distributed DC MG systems. Regarding wind power systems, an observed-based dynamic event-triggered controller is presented in [65] for multi-area wind farms under dual alterable aperiodic DoS attacks. Furthermore, an adaptive observer-based resilient control method for the cyber links of wind turbines is developed in [66] to defend against time-delay attacks. Observer-based techniques have been also proposed for increasing the cyber resilience of other types of CPSs. For example, an FDIA-resilient control mechanism is designed in [67] for a networked control system using a Kalman filter as an observer. Additionally, an adaptive sliding mode observer is developed in [68] to establish a resilient control for linear CPSs under compromised measurements and control commands. Furthermore, an event-triggered, observer-based control scheme is presented in [69] to detect DoS attacks in CPSs. Since LFC is a critical part of the power systems automation, observer-based techniques have been also adopted for the strengthening of its cyber resilience. For example, a robust adaptive observer is presented in [70] for concurrent estimation of the LFC system states and FDIAs. A Luenberger observer enhanced by the extended Kalman filter is proposed in [71] and a combination of switching impulsive observer and switching state observer is introduced in [72] for cyberattack estimation and mitigation in LFC. Furthermore, an unknown input observer is designed in [73] that forms an attack-resilient control architecture for LFC.

Data-driven approaches are a potential solution when the LFC modeling is highly complex and it is difficult to find an adequate system representation. More specifically, a long short-term memory (LSTM) neural network is trained in [74], that can reconstruct the healthy LFC control signals during FDIAs, based on data extracted under normal system conditions. However, the effectiveness of this approach is not always guaranteed since the load disturbances are omitted in the theoretical system modeling. A similar approach is followed in [75]; an LSTM neural network is designed to tackle the FDIA impact on the LFC but in this case, both load disturbances and system nonlinearities are considered. In [76], a combination of a deep autoencoder and an extreme learning machine is employed to estimate the data missing by DoS attacks, preserving the operational state of LFC. This method is evaluated on the single, two and three area LFC models provided in [77] using MATLAB/Simulink. Besides LFC, data-driven methods also offer cyber resilience enhancement to other power systems and CPS applications. For example, a data clearing method based on conditional deep belief

networks is investigated in [78] as a real time cyberattack response response. This work is also expanded in [79] to detect FDIAs.

1.4.3 Limitations of related works

Due to the significance of modern power grids and their applications, there is ongoing academic work in the investigated research field. So far, several issues of this research field have been effectively addressed by existing works; each category of these related works contributes in its own, unique way to the research field. However, there are still multiple open problems to be resolved, which are either caused by the inherent characteristics of the problem or introduced by the categories of the proposed methodologies. The contributions of the existing works in the research field along with the open problems are listed per category in what follows as advantages and limitations, respectively:

- **model-based methods**: the advantages of model-based methods is that they can be easily implemented, as long as an effective model has been developed, and their low computational requirements. However, they heavily depend on the model that has been designed, which significantly determines their overall performance; defining an accurate system model is a complicated task due to simplifications and abstractions that have to be made. Furthermore, for simplicity, the methodologies of this category do not consider other types of uncertainties, besides cyberattacks. Finally, the methodologies of this category do not consider practical features of LFC and they are not validated under real-world conditions.

- **observer-based methods**: this category has the same advantages with model-based defense strategies and additionally, it can effectively distinguish cyberattacks from other types of uncertainties, such as load disturbances, RES generation, etc. Nevertheless, the performance of these methodologies depend on the modeling of the LFC system and could be potentially affected if the system is not properly defined or if it is modified. Furthermore, the methodologies of this category do not utilize practical features of LFC and they are not evaluated in a realistic environment.

- **data-driven methods**: the majority of the disadvantages of model-based and observer-based methods are overcome by the deployment of data-driven methods. Since data-driven algorithms utilize data to approximate both the normal and unhealthy behavior of the actual LFC system, they are model-agnostic and their performance is not affected by the accuracy of any developed system representation. Moreover, these algorithms can reveal the underlying system dynamics and hence, they can distinguish

cyberattacks from other types of uncertainties. However, their training procedure is typically computationally intensive and thus, they could be an infeasible solution in terms of resources. Moreover, the practicality of these methodologies is questioned because several practical features of LFC are omitted and they are not evaluated in a real-world testbed.

1.5 Proposed Hybrid Framework & Thesis Contribution

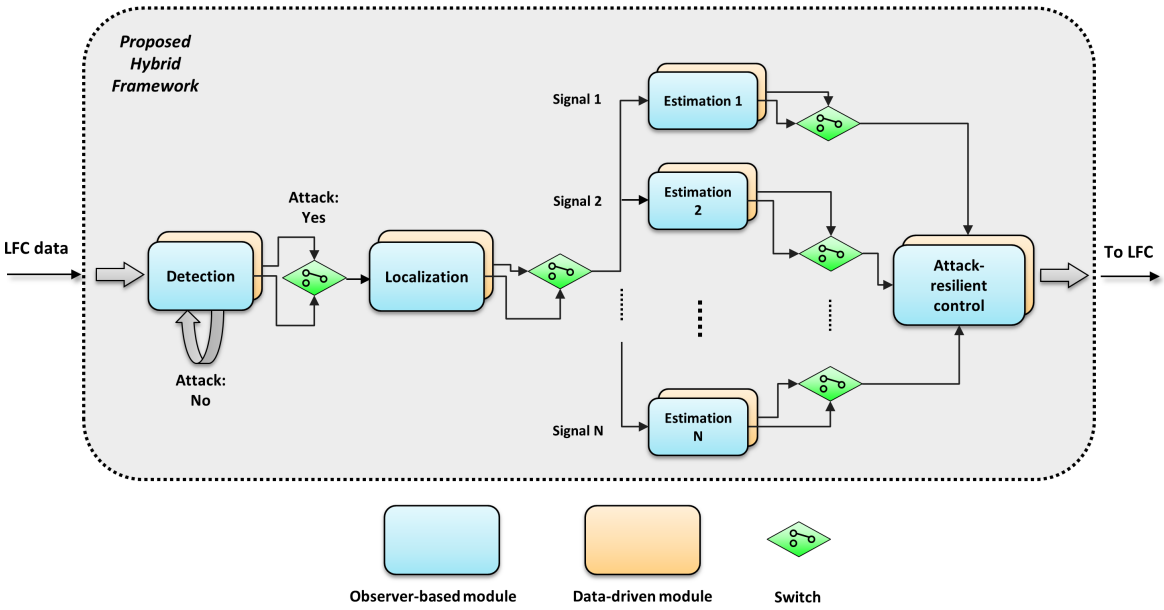


Figure 1.5 Diagram of the hybrid framework proposed in this thesis.

The previous literature review highlighted the research challenges in enhancing the cyber resilience of LFC, presented several existing solutions and determined the open problems of the field. Inspired by these problems, a hybrid framework is proposed in this thesis to effectively address the identified research gaps. As illustrated in Fig. 1.5, the introduced framework combines the four active cyber defense layers presented in Section 1.3. This diagram indicates that the internal layers of the framework are deployed in a cascading way, where the output of a specific cyber defense technique is fed to input of the next one. The synergy between the layers of the framework provides a holistic protection to LFC against successful digital intrusions. For each of these techniques, two algorithms are developed: one that utilizes a novel observer design and another that employs a deep learning model. The configuration of the green switches in Fig. 1.5 determine if the observer-based or the data-driven approach will be used at each cyber defense layer. The various combinations that

can be created allow the system operator to leverage the benefits of both cyber resilience categories. Furthermore, a customized version of the framework is assembled for each use case, taking into account the specific requirements of the LFC system to which it is applied. The functionality of each cyber resilience algorithm designed in this thesis is thoroughly discussed in Chapters 4-6.

As explained earlier, the objective of the proposed framework is to address the open problems of the investigated research field. Therefore, the novelties of this thesis can be clearly determined by studying these issues. The key contributions are summarized in what follows:

- The main contribution of the thesis is that it combines the advantages of two standard cyber resilience categories, i.e. observer-based and data-driven, to overcome their common limitations. More specifically, the introduced methodologies are model-independent due to their area-wise design, can distinguish cyberattacks from other types of external disturbances, have low computational requirements, are applicable to LFC systems with various practical features (nonlinearities, RES, HVDC/TCPS-equipped links, etc.) and are tested on Software/Hardware-in-the-Loop simulations that mimic real-world conditions.

- In the literature, there are several types of observers that have been proposed for strengthening the cyber resilience of LFC, such as unknown input observer, robust adaptive observer, etc. In this thesis though, an observer design is utilized that has never been applied to the investigated research field before, particularly the sliding-mode observer. Since it is the first time that this model is deployed to protect LFC from cyberattacks, the presented methodologies are considered innovative. This is a major contribution regarding the observer-based part of the introduced framework.

- The proposed data-driven attack detection method utilizes an autoencoder architecture based on DNNs. This variant is applied for the first time in LFC and its lightweight implementation enables the autoencoder to continuously learn new normal LFC states during its online operation, unlike similar works. Moreover, the introduced data-driven attack recovery methodology estimates the healthy control signals of LFC during cyberattacks in an innovative way. Therefore, the data-driven part of the proposed framework offers a novel set of cyber resilience techniques for LFC.

- Based on the theoretical and experimental results, the proposed framework is scalable to large power systems, unlike other methodologies that do not explore this aspect. The introduced methodologies are developed as area-wise techniques in order to be

unaffected by an increase (or decrease) in the number of power-areas (control-areas). The experimental results also indicate that the proposed framework can be successfully applied into a wide range of frequency control systems, varying from single area systems to multi-area topologies.

- The sensitivity of the introduced framework against several types of uncertainties is studied in depth, both theoretically and experimentally. Such uncertainties include the inaccuracies in the computation of the LFC parameters, the noisy environment of real-world power systems and the time delays in the data transferring caused by the deficiencies of the communication mediums. The results verify the robustness of the proposed approach against these unpredictable factors, which are difficult to be incorporated into the model of LFC.

Chapter 2

Background

This chapter provides the necessary theoretical background of the research problem addressed in this thesis and the tools that are employed to mitigate it. It begins with an in depth-analysis of state observers, which serve as the cornerstone of the observer-based part of the proposed framework. Following this, the chapter introduces the deep learning algorithms utilized by the data-driven part of the framework, providing a concise yet comprehensive overview of their role and functionality. Finally, the aspect of cybersecurity is investigated from the viewpoint of modern power systems to offer critical insights into the problem that has to be tackled.

2.1 Observers

The effective monitoring and control of power systems requires accurate information about the state variables of the grid [80]. However, measuring all the system variables is practically infeasible and highly expensive, especially for large power systems. Another way of achieving effective power system automation is by using estimated measurements instead of the actual ones. The state variables of a system can be estimated through another type of dynamical system called observers [81]. Observers utilize mathematical models and measured data to constantly provide accurate and reliable information about the internal states of the system, even in the presence of disturbances, unmeasured variables, or other types of uncertainties. They can be leveraged to enhance the performance of the system by enabling fault diagnosis and sophisticated control strategies. Observers can be designed either for continuous-time systems or discrete-time systems. Due to the similar design processes between them, this study is focused on the first ones.

The typical structure of observers is illustrated in Fig. 2.1. This figure sheds insight on their functionality, which can be summarized as follows: firstly, the observer receives the

input $u(t) \in \mathbb{R}^m$ and the output $y(t) \in \mathbb{R}^p$ of the given system as measured data. Then, these measurements are fed to a mathematical model which is developed based on the knowledge of the system dynamics. To ensure the existence of this information, it is common practice to assume that the state-space representation matrices $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{p\times n}, D \in \mathbb{R}^{p\times m}$ of the given system are known. Finally, the filter resulting from the designed observer produces an estimation of the internal state vector $\hat{x}(t) \in \mathbb{R}^n$ of the given system.

When an observer starts operating, it is reasonable to provide inaccurate estimations due to the initial conditions. However, the performance of the observer is expected to be improved over time. This behavior can be sufficiently captured by the estimation error term, which is the difference between the actual state vector and the estimated one. Formally, this error is defined as $e(t) = x(t) - \hat{x}(t)$. A designed observer exists if it is proven that its estimation error is asymptotically stable. Therefore, the estimation error determines whether an observer design process has been successful or not, typically following the next steps:

$$e(t) = x(t) - \hat{x}(t) \Rightarrow \dot{e}(t) = \dot{x}(t) - \dot{\hat{x}}(t) \Rightarrow \ldots \Rightarrow \dot{e}(t) = Qe(t),$$

where $Q \in \mathbb{R}^{n\times n}$. If Q is a Hurwitz matrix, then $e(t) \to 0$ as $t \to \infty$.

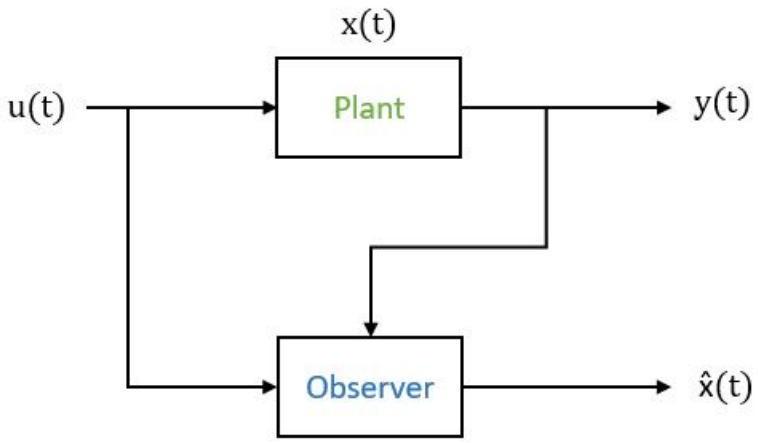


Figure 2.1 Typical observer structure.

Observers can be categorized according to their mathematical model. Each mathematical model serves a different purpose, e.g. simplicity, robustness, accuracy. In the present thesis, three major observer categories are employed for the design of the proposed cyber defense layers. A brief introduction to these observers is presented in the following subsections.

2.1.1 Luenberger observer

The simplest observer type is the Luenberger observer [82]. The mathematical model of this observer combines the matrices of the tracked system along with a correction term between the measured and the estimated output. It is typically used for linear, time invariant dynamic systems which are modeled by the next state-space representation:

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t). \end{cases} \tag{2.1}$$

The structure of the full-state Luenberger observer is described by the following dynamic system:

$$\begin{cases} \dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + L\big(y(t) - \hat{y}(t)\big) \\ \hat{y}(t) = C\hat{x}(t) + Du(t), \end{cases}$$

where $L \in \mathbb{R}^{n \times m}$ is the gain matrix of the observer.

The resulting estimation error has the following form:

$$\begin{aligned} & e(t) = x(t) - \hat{x}(t) \Rightarrow \dot{e}(t) = \dot{x}(t) - \dot{\hat{x}}(t) \Rightarrow \\ \Rightarrow\ & \dot{e}(t) = Ax(t) + Bu(t) - A\hat{x}(t) - Bu(t) - Ly(t) - L\hat{y}(t) \Rightarrow \\ \Rightarrow\ & \dot{e}(t) = A\big[x(t) - \hat{x}(t)\big] - Ly(t) + L\hat{y}(t) \Rightarrow \\ \Rightarrow\ & \dot{e}(t) = A\big[x(t) - \hat{x}(t)\big] - LCx(t) - LDu(t) + LC\hat{x}(t) + LDu(t) \Rightarrow \\ \Rightarrow\ & \dot{e}(t) = A\big[x(t) - \hat{x}(t)\big] - LC\big[x(t) - \hat{x}(t)\big] \Rightarrow \\ \Rightarrow\ & \dot{e}(t) = (A - LC)e(t). \end{aligned} \tag{2.2}$$

Eq. (2.2) yields that the estimation error can be viewed as an autonomous dynamical system described by the $A - LC$ matrix. According to system theory, if $A - LC$ is Hurwitz, then the estimation error is asymptotically stable. In other words, if the gain matrix L is chosen so that the eigenvalues of $A - LC$ are strictly in the left-half of the complex plane, then the error equation will decay to zero over time. Therefore, with a proper choice of L, the Luenberger observer can effectively estimate the states of system (2.1).

2.1.2 Unknown input observer

The model of (2.1) can adequately describe the behavior of several linear dynamic systems. However, the majority of practical systems face different kinds of disturbances or uncertainties which are not reflected by the model of system (2.1). These uncertainties can be

represented by an unknown disturbance term $Ed(t)$, which is added to system (2.1) as:

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) + Ed(t) \\ y(t) = Cx(t), \end{cases} \tag{2.3}$$

where $d(t) \in \mathbb{R}^r$ is the unknown input or disturbance vector and $E \in \mathbb{R}^{n\times r}$ is the unknown input distribution matrix.

While the standard structure of the Luenberger observer works effectively for linear dynamic systems, it falls short for systems expressed by (2.3), which are subjected to disturbances. Particularly, the resulting estimation error between a Luenberger observer and system (2.3) is affected by the disturbance term $Ed(t)$ and hence, it cannot converge to zero. To tackle this issue, multiple observer designs have been proposed in the literature, including the unknown input observer (UIO) [83] covered in this subsection.

The UIO design process aims to decouple the resulting estimation error from the unknown input signal. This can be achieved by properly selecting the observer matrices, as explained later in this subsection. The structure for a full-order UIO is described by the following dynamic system:

$$\begin{cases} \dot{w}(t) = Fw(t) + TBu(t) + Ky(t) \\ \hat{x}(t) = w(t) + Hy(t), \end{cases}$$

where $w(t)$ is the intermediate variable and F, T, K, H are the desired observer matrices. For further comprehension, the block diagram of the UIO is shown in Fig. 2.2. The decoupling procedure of the disturbance term will be better illustrated by computing the estimation error $e(t)$, as follows:

$$\begin{aligned}
&e(t) = x(t) - \hat{x}(t) \Rightarrow \dot{e}(t) = \dot{x}(t) - \dot{\hat{x}}(t) \Rightarrow \\
&\Rightarrow \dot{e}(t) = Ax(t) + Bu(t) + Ed(t) - \dot{w}(t) - H\dot{y}(t) \Rightarrow \\
&\Rightarrow \dot{e}(t) = Ax(t) + Bu(t) + Ed(t) - Fw(t) - TBu(t) - Ky(t) - HC\big(Ax(t) + Bu(t) + Ed(t)\big) \Rightarrow \\
&\Rightarrow \dot{e}(t) = (A - HCA - K_1C)e(t) + \big[F - (A - HCA - K_1C)\big]w(t) + \\
&\qquad + \big[K_2 - (A - HCA - K_1C)\big]y(t) + \big[T - (I - HC)\big]Bu(t) + (HC - I)Ed(t).
\end{aligned}$$

If the F, T, K, H matrices are selected so that the following conditions:

$$F = A - HCA - K_1C$$
$$K_2 = FH$$
$$T = HC - I$$
$$(HC - I)E = 0$$

are satisfied, then the estimation error is simplified as

$$\dot{e}(t) = Fe(t). \tag{2.4}$$

The error dynamics that resulted from the UIO are independent of the disturbance terms E and $d(t)$, as indicated by Eq. (2.4). Therefore, the desired decoupling of the state estimator from the unknown disturbance inputs has been achieved. Furthermore, if F is a stable matrix, then $e(t)$ will approach zero asymptotically, according to the Lyapunov stability theory. This yields that the UIO can effectively estimate the states of system (2.3), despite the presence of disturbances and without any a priori knowledge about the unknown inputs.

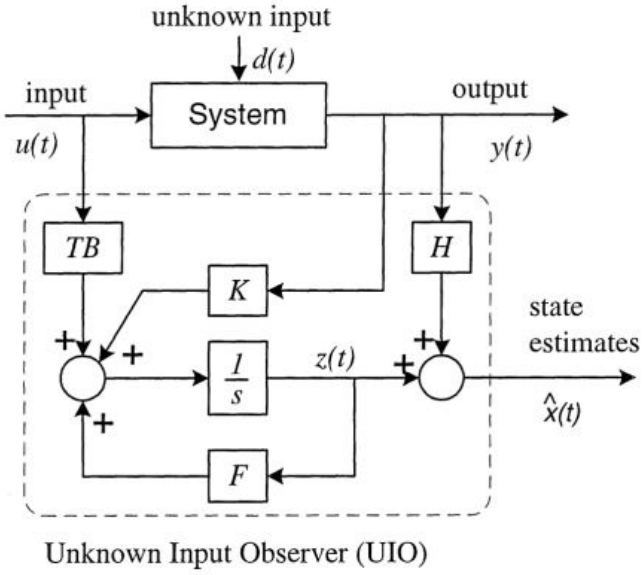


Figure 2.2 Structure of UIO [2].

2.1.3 Sliding mode observer

The types of the observers that have been discussed so far are typically designed for linear systems and may not perform optimally for systems with high uncertainties. Furthermore,

there are certain applications where the performance and robustness of the aforementioned observers do not meet the necessary requirements [84]. These limitations have inspired researchers to investigate other types of observers, such as sliding mode observers, which are presented in this subsection. The main advantage of SMOs is that they can provide accurate state estimations even in the presence of multiple uncertainties, such as nonlinearities, modeling errors, noise, etc. The majority of real-world dynamical systems includes several nonlinearities which are modeled as an extra term in (2.1).

The basic idea behind a sliding mode observer is to create a sliding surface where the estimated states will converge to, regardless of the initial conditions or uncertainties in the system [85]. The observer uses the system dynamics and the available measurements to update the estimated states and drive them towards the desired surface by sliding along it. The sliding surface is a mathematical construct defined in the state space. It is designed such that its derivative satisfies certain conditions while the updating law determines how the estimated states are updated. Additionally, the sliding surface provides formulas to reconstruct the system uncertainties.

A detailed analysis about the SMO design process exceeds the scope of this thesis, which serves only as an introductory point to this topic. For this reason, the reader may refer to [86] for more information. For the sake of completeness, the SMO design for a simple, linear dynamic system will be briefly presented. In sliding mode approaches, a coordinate transformation is typically applied before the observer design process. This technique is used for decoupling the uncertainties of the system and simplifying the corresponding model. For the simple case of system (2.1), a proper change of coordinates is the $z \mapsto Tx$ where $T = \begin{bmatrix} N_c \\ C \end{bmatrix}$ and the submatrix $N_c \in \mathbb{R}^{n\times(n-p)}$ spans the null-space of C. Assuming the $N_c x(t) = x_1(t)$, it is obtained:

$$T = \begin{bmatrix} N_c \\ C \end{bmatrix} \Rightarrow Tx(t) = \begin{bmatrix} N_c \\ C \end{bmatrix} x(t) \Rightarrow z(t) = \begin{bmatrix} x_1(t) \\ y(t) \end{bmatrix}.$$

After applying the above transformation to (2.1), the newly transformed system is derived as:

$$\begin{aligned} \dot{x}(t) = Ax(t) + Bu(t) \Rightarrow T\dot{x}(t) = TAx(t) + TBu(t) \Rightarrow \\ \Rightarrow \dot{z}(t) = TAT^{-1}z(t) + TBu(t) \Rightarrow \begin{bmatrix} \dot{x}_1(t) \\ \dot{y}(t) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ y(t) \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u(t) \Rightarrow \\ \Rightarrow \begin{cases} \dot{x}_1(t) = A_{11}x_1(t) + A_{12}y(t) + B_1u(t) \\ \dot{y}(t) = A_{21}x_1(t) + A_{22}y(t) + B_2u(t). \end{cases} \end{aligned} \tag{2.5}$$

The observer design based on the sliding mode approach that Utkin proposed [86] for the transformed system (2.5) has the following form:

$$\begin{cases} \dot{\hat{x}}_1(t) = A_{11}\hat{x}_1(t) + A_{12}\hat{y}(t) + B_1 u(t) + Lv \\ \dot{\hat{y}}(t) = A_{21}\hat{x}_1(t) + A_{22}\hat{y}(t) + B_2 u(t) - v, \end{cases}$$

where $(\hat{x}_1,\ \hat{y})$ represent the state estimations, $L \in \mathbb{R}^{(n-p)\times p}$ is a gain matrix and $v_i = Msgn(\hat{y}_i - y_i)$ where $M \in \mathbb{R}^+$.

The estimation errors are defined as $e_1(t) = \hat{x}_1 - x_1$ and $e_y(t) = \hat{y} - y$. From systems (2.1) and (2.5), the resulting estimation error dynamics are obtained as:

$$\dot{e}_1(t) = A_{11}e_1(t) + A_{12}e_y(t) + Lv,$$

$$\dot{e}_y(t) = A_{21}e_1(t) + A_{22}e_y(t) - v.$$

It can be proven [86] that, with a proper choice of L, an ideal sliding motion take places on the surface $\{(e_1, e_y) : e_y = 0\}$ after some finite time and the corresponding sliding mode dynamics represent a stable system. Therefore, the estimated states generated by the SMO can track the real states asymptotically. The form of the SMO and the type of the coordinate transformation are adjusted according to the model of the system. In this way, the effects of the system uncertainties on the estimator are eliminated and the system states can be effectively approximated.

2.2 Deep Learning Models

Machine learning (ML) is a subset of artificial intelligence (AI) that focuses on the development of algorithms that allow computers to take actions based on data [87]. Rather than being explicitly programmed to perform a specific task, ML models are designed to learn patterns and relationships from data and make decisions or predictions autonomously. Typical paradigms of ML algorithms are the decision trees, random forests, support vector machines, K-nearest neighbors, naive bayes and neural networks. The process of ML model development typically involves the following steps:

1. **Data Collection**: Gathering relevant data from various sources. This includes examples, observations or measurements of the task to be solved.

2. **Data Preprocessing**: Cleaning the collected data and preparing them for usage. This involves tasks such as removing noise, handling missing values and normalizing or scaling the data.

3. **Feature Engineering**: Selecting or extracting representative features from the pre-processed data that can help the ML algorithm learn patterns and make accurate predictions. Feature engineering is crucial for improving the performance of ML models.

4. **Model Selection**: Choosing an appropriate ML model based on the nature of the problem and the available data. Common types of ML algorithms include supervised learning, unsupervised learning and reinforcement learning.

5. **Model Training**: Training the selected ML model on the prepared data to learn patterns and relationships. During training, the model adjusts its parameters based on the input data to minimize the produced errors and maximize its performance.

6. **Evaluation**: Assessing the performance of the trained model using evaluation metrics and techniques such as cross-validation. This step helps determine how well the model generalizes to new, unseen data.

Deep learning is a subset of ML that involves algorithms and models inspired by the structure and function of the human brain [88], called artificial neural networks. What distinguishes deep learning from traditional ML algorithms is its capability to automatically learn representations of data in a hierarchical manner. The term "deep" refers to the multiple layers of neurons that are typically present in these types of neural networks. These hierarchies enable deep learning models to extract intricate features from raw data. Deep learning algorithms utilize large amounts of labeled data to train neural networks, adjusting the connections between neurons through a process called backpropagation to minimize the produced errors and improve accuracy of the model. In what follows, the deep learning architectures that are utilized in this thesis are analyzed in detail.

2.2.1 Deep neural networks

Deep neural networks (DNNs) are computing systems of artificial intelligence that are able to recognize underlying relationships in a set of data by emulating the operation of the human brain [89]. They are composed of multiple nodes, also called as *neurons*, which are grouped in multiple, parallel layers. These neurons are connected by links in a forward direction, as

shown in Fig. 2.3. The output of each neuron is computed as:

$$\phi = f\left(b + \sum_{n=1}^{n} x_i w_i\right),$$

where x_i represents an input of the neuron, w_i is the weight of the corresponding input x_i, b is the bias of the neuron, f is the activation function and n is the number of the inputs of the neuron. The goal is to compute the best weights and biases for all neurons based on the given data and a selected error function. This is achieved through the backpropagation algorithm [90], a method that calculates the gradient of the error function with respect to the weights of the neural network. So far, DNNs have successfully solved various types of problems and especially, regressions problems.

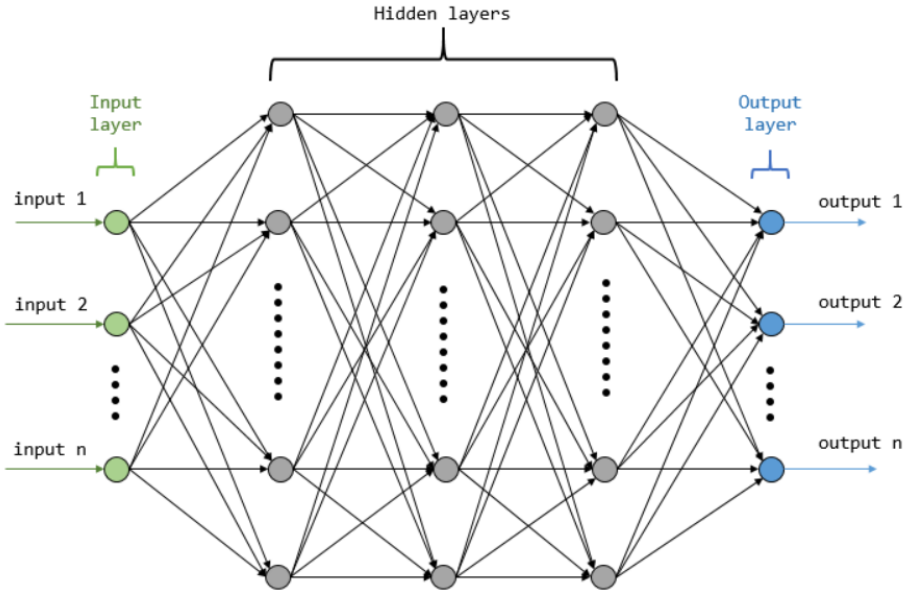


Figure 2.3 Deep feedforward neural network architecture.

2.2.2 Autoencoders

The autoencoder is a special type of neural networks whose purpose is to provide an accurate replica of the given input to its output [91]. For example, the trained autoencoder of the paradigm illustrated in Fig. 2.4 copies the input images at its output with high precision. To achieve this, the autoencoder compresses the input data into a lower-dimensional code and then uses it to reconstruct the given input. These actions are performed separately by the

three dedicated components of the autoencoder, i.e. the encoder, the code and the decoder. The aforementioned components are briefly described in what follows:

- **Encoder**: this module is composed of consecutive neural network layers of decreasing size, where each layer creates a small mapping between the input data and compressed, lower-dimensional spaces. In this way, the network is forced to learn the most representative features of the input data and store them at its output, namely the *code* of the autoencoder.
- **Code**: it is a representative summary of the input, also called as *latent-space* representation. The code encapsulates the most representative features of the input into a compressed version. The size of the code is a hyperparameter that determines the amount of the information that will be lost by the compression process.
- **Decoder**: this component is the mirror image of the encoder and thus, it performs the opposite operation. Within the decoder, the compressed information of the code is passed to a neural network with the inverse architecture of the model used in the encoder. In this way, the encoded message in the latent-space is decompressed and the original input is reconstructed with high accuracy.

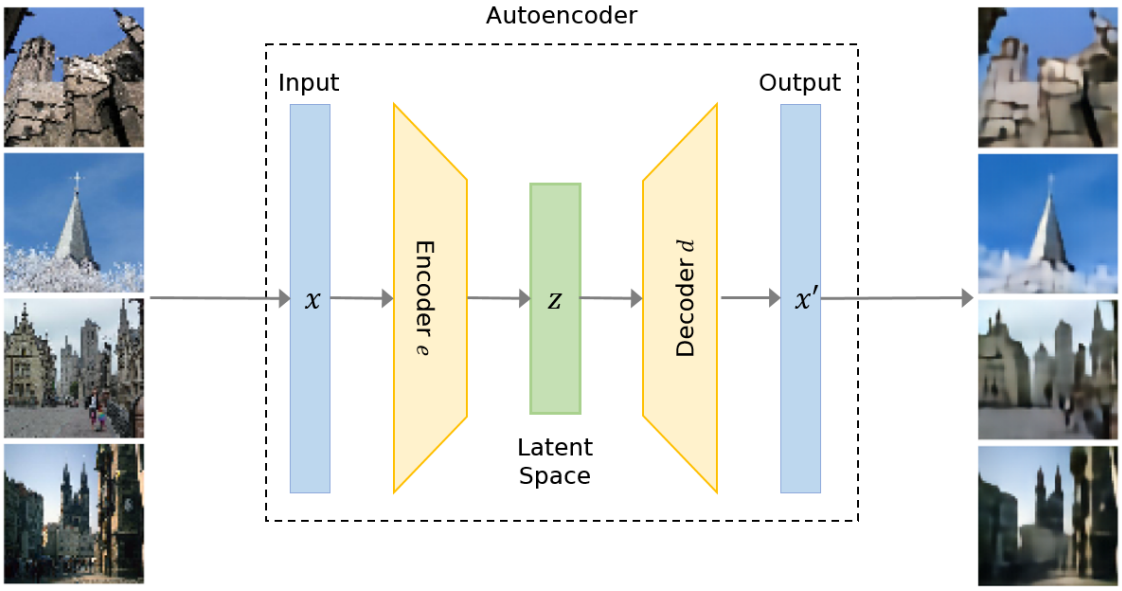


Figure 2.4 Functionality of an autoencoder.

Without loss of generality, the functionality of an autoencoder built with DNNs is mathematically formulated in the remainder of this section. More specifically, assume that a

multivariate dataset $x = \{x_1, x_2, \ldots, x_n\}$ is given, where n is the total number of input features. Then, the processes of the encoder and the decoder are formally expressed as:

$$\begin{cases} \varepsilon : x \to \mathscr{Z} : f_e(x, \theta_e) \\ \delta : \mathscr{Z} \to \tilde{x} : f_d(x, \theta_d), \end{cases}$$

where ε and δ denote the encoding and decoding mappings, respectively, $\mathscr{Z}$ is the minimum latent-space of the input features, f_e and f_d represent the nonlinear functions of the encoder and the decoder, respectively, while $\theta_e = \{W_e, b_e\}$ and $\theta_d = \{W_d, b_d\}$ reflect the weight and bias matrices of the DNNs utilized for the encoder and decoder, respectively, Considering that $W_{e_j}, b_{e_j}, W_{d_j}$ and b_{d_j} $(j = 1, 2, \ldots, k)$ refer to the weights and biases of the corresponding jth neural network layer, the nonlinear functions of the autoencoder are obtained as:

$$\begin{cases} f_e = \phi_k\Big(\ldots\ \phi_2\big(W_{e_2}\phi_1(W_{e_1}x + b_{e_1}) + b_{e_2}\big)\ldots + b_{e_k}\Big) \\ f_d = \phi_k\Big(\ldots\ \phi_2\big(W_{d_2}\phi_1(W_{d_1}x + b_{d_1}) + b_{d_2}\big)\ldots + b_{d_k}\Big), \end{cases}$$

$$\begin{cases} f_e = \phi_k\Big(\ldots\ \phi_2\big(W_{e_2}\phi_1(W_{e_1}x + b_{e_1}) + b_{e_2}\big)\ldots + b_{e_k}\Big) \\ f_d = \phi_k\Big(\ldots\ \phi_2\big(W_{d_2}\phi_1(W_{d_1}x + b_{d_1}) + b_{d_2}\big)\ldots + b_{d_k}\Big), \end{cases}$$

where $\phi_j(\cdot)$ is the activation function of the jth layer. To learn the identity function, the autoencoder computes the θ_e^* and θ_d^* sets that minimize its reconstruction error $e = x - \tilde{x}$, which is the difference between the input and output data. This is achieved by solving the next optimization problem:

$$\{\theta_e^*, \theta_d^*\} = \arg\min_{\theta_e, \theta_d} \|x - \tilde{x}\|_2 = \arg\min_{\theta_e, \theta_d} \sum_{i=1}^{n} \|x_i - f_d(f_e(x_i, \theta_e), \theta_d\|_2.$$

The accuracy of the autoencoder is evaluated through its reconstruction error. The model performance can be improved by stacking multiple layers of neurons that enable the autoencoder to learn higher-level features of the given data.

2.3 Cybersecurity Objectives

The main cybersecurity objectives when designing ICT-based systems are the confidentiality, integrity and availability. These objectives are also known as the "CIA triad" and are illustrated in Fig. 2.5. The CIA triad defines which system characteristics does a cybersecurity

mechanism enhance or oppositely, which system features are exposed to cyber risks. This triad is a foundational concept in cybersecurity which helps organizations and companies to maintain a balance between the functionality and cyber resilience of a designed system. It is also important to note that the CIA triad is interrelated and an impact on one aspect may affect the others. Therefore, there are trade-offs between the satisfaction of a cybersecurity objective over another, which is an aspect that cybersecurity designers need to carefully consider.

For a better understanding on these cybersecurity objectives, each component of the CIA triad is briefly explained in what follows:

- **Availability**: ensures that data and services are accessible when needed and focuses on preventing disruptions or downtimes. Cybersecurity measures aim to offer protection against DoS attacks and other incidents that could render data and services unavailable. Redundancy, failover mechanisms, and disaster recovery plans are commonly used to maintain availability. Without proper availability, critical services can become inaccessible, leading to productivity losses or service disruptions.

- **Integrity**: refers to the accuracy and trustworthiness of the data. It ensures that information remains unaffected by unauthorized parties and is only modified by authorized and documented processes. Maintaining data integrity is essential to prevent data corruption, tampering and manipulation. Techniques like data hashing, checksums, and digital signatures are employed to ensure the integrity of data. A breach of data integrity can result in the dissemination of inaccurate information or system malfunctions.

- **Confidentiality**: focuses on protecting the exchanged information from unauthorized access. It ensures that data is only accessible to those who have the appropriate permissions or privileges. This triad aspect is crucial for protecting sensitive and private information. In practice, measures like access controls, encryption, and authentication are used to maintain confidentiality. Breaches in confidentiality can lead to data leaks, privacy violations, and security incidents.

2.4 Location of Cyberattacks

In typical computer networks, such as corporate database systems, web servers, etc., the primary cybersecurity concerns are related to maintaining the privacy of data and ensuring uninterrupted access. For example, adversaries usually attempt to steal the information stored in the networked database system of a bank or disrupt the normal operation of a web server to demand a ransom. Therefore, the cybersecurity objectives that are threatened in this case

are the confidentiality and availability. In case of power grids, their parts that are exposed to cyber risks are the monitoring and control systems, since they utilize communication infrastructures along with software/hardware applications. Cyberattackers aim for either modifying or blocking the normal data transfer of the automation systems to degrade the stability of the power system. Hence, the main cybersecurity objectives that are in jeopardy, regarding power systems, are the integrity and availability.

Figure 2.5 Cybersecurity objectives

Analyzing the distinct components of a remote automation system facilitates the identification of the vulnerable points across a power grid in terms of cybersecurity [92]. For this reason, the standard control loop of a power system is depicted in Figure 2.6. Building upon this illustration, the next paragraphs provide a detailed breakdown of the power system components susceptible to digital threats:

- **Sensors**: they are field devices that periodically measure critical variables of the physical system. Typically, they deployed in a dedicated hardware and utilize a lightweight software environment for configuration.
- **Measurement Channels**: they are communications channels that are responsible for the transfer of the measurements from the field devices to the control center. Their implementation depends on the application that are deigned for and the architecture of the utilized communication protocol.
- **Control Center**: it is the cornerstone of an automation system. The control center receives the field measurements and process the accordingly in order to generate. The

applications that receive and the control center input are software applications that run a deisgned algorithm.

- **Control Command Channels**: they are communications channels that are responsible for the transfer of the control command from the control center to the power plant. Their implementation is similar to the measurement channels.
- **Actuators**: they are devices that convert control signals or commands into physical actions or movements within the power system. Actuators are typically implemented as mechanical, hydraulic or electronic devices.

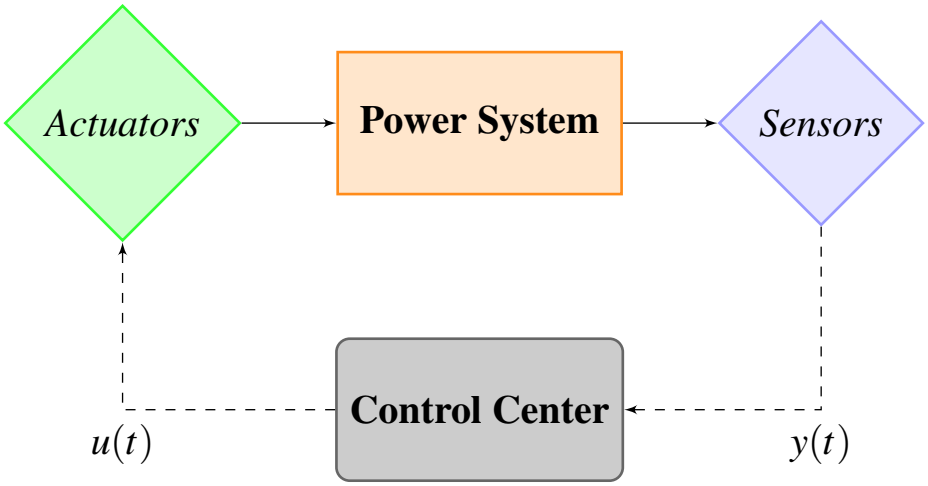


Figure 2.6 Automation in power systems.

2.5 Types of Cyberattacks

In this section, the most critical types of cyberattacks against analyzed in detail and practically modelled.

2.5.1 Denial-of-Service attacks

In a computer network, the primary objective of a Denial-of-Service (DoS) attack is to make the delivered data or service unavailable to its legitimate users [14]. This is typically achieved by exploiting the cyber vulnerabilities of a computer or network system in order to gain access to a critical infrastructure. Then, the different parts of the computer network are flooded with an excessive amount of data, traffic or requests to saturate all the available resources of the

system [93]. In this way, the overall performance of the system is deteriorated, becoming unable to provide the intended data or services. This results in a severe operational and financial impact on both organizations and individuals.

Since LFC system uses a remote communication network to regulate the generation of a power system, it is directly threatened by DoS cyberattacks. DoS attacks can be launched against all the communication parts of the LFC, namely sensors, interconnection links and control center applications [24]. When the sensors of the LFC or their communications channels suffer from a DoS attack, the control center cannot receive the frequency and the tie-line power flow measurements to calculate the generator setpoint. Similarly, if a DoS attack has been successfully launched against the control center, the actuators or their interconnection links, the calculated setpoints cannot be transferred to the their generators. Therefore, the secondary control of LFC is practically deactivated and the power system operates only with its primary frequency control [94]. As a result, the frequency of the system cannot be stabilized to its nominal value, becoming completely dependent to the external disturbances, such as load or Renewable Energy Sources (RES) variations.

During DoS attacks, the standard state-space representation of the LFC system becomes as follows:

$$\begin{cases} \dot{x}(t) = Ax(t) + J\phi(x,t) + Bu_p(t) + Ed(t) \\ y(t) = Cx(t), \end{cases} \tag{2.6}$$

where u_p is the input of the LFC that is regulated only by the primary frequency control mechanism of the power system. The form of the u_p can be easily derived by the dynamics of the LFC.

2.5.2 Time-delay attacks

To preserve the stability of the grid, the measurements required for the operation of the LFC and the control commands issued by this mechanism have to be exchanged in a timely manner. The occurrence of small time delays due to the limitations in the computer and network resources deployed for the LFC, is a natural event and their impact can be easily eliminated by the typical control schemes. However, when substantial amounts of time delays are deliberately injected across the LFC loop by adversaries, the stability of the system is significantly degraded [95]. If the setpoints of the generators are not computed or transferred promptly, the system frequency fails to converge to its scheduled value and demonstrates large fluctuations. Therefore, a new type of cyber threat arises for LFC (and cyber-physical systems (CPSs) in general), namely time-delay attacks (TDAs) [96]. TDAs have severe

consequences on the performance of LFC as they can stealthily affect its normal operation or lead to power disruptions [97].

When the LFC system suffers from TDAs, its original state-space representation is converted into the following form [25, 98, 99]:

$$\begin{cases} \dot{x}(t) = Ax(t) + A^{*}x\big(t - \tau_d(t)\big) + J\phi(x,t) + Bu(t) + Ed(t) \\ y(t) = Cx(t), \end{cases} \tag{2.7}$$

where τ_d is the time-delay function that is used by the adversary. τ_d can have the form of a constant, linear, random, etc. function.

2.5.3 False data injections attacks

Another digital threat against the integrity of the data exchanged across the LFC loop are the false data injection attacks (FDIAs) [100]. If adversaries have gained unauthorized access into a part of the communication system of LFC, they can manipulate the data encapsulated within the transmitted network packets [101]. FDIAs can maliciously modify the content of the network packets in numerous ways and they typically modeled by mathematical functions of varying complexity. Depending on the intentions of the adversaries, the FDIAs can be constructed either to stealthily damage to the attacked electrical grid or to cause a complete power outage [102]. As a results, FDIAs can heavily degrade the stability of a power system and lead to substantial financial losses.

Without loss of generality, FDIAs are modeled in this study as a term added to the measurements and/or the control signals of LFC. Hence, when the LFC system faces FDIAs, its standard state-space representation is transformed into [103]:

$$\begin{cases} \dot{x}(t) = Ax(t) + J\phi(x,t) + B\big(u(t) + a_c(t)\big) + Ed(t) \\ y(t) = Cx(t) + Da_m(t), \end{cases} \tag{2.8}$$

where $a_c(t)$ is the corruption term that is injected by the adversary into the control signals and $a_m(t)$ is the same variable the measurements. This study investigates three types of FDIAs, namely step attack, ramp attack and sine attack [104], which are modeled in what follows:

i) Step FDIA: it adds a constant value to the actual measurements. When a step FDIA is launched, the attack vector is modeled as:

$$a_m(t) = \begin{cases} 0 & t \notin \tau_m \\ 1 & t \in \tau_m. \end{cases}$$

ii) Ramp FDIA: it alters the actual values of the measurements linearly with time. In case of a ramp FDIA, the attack vector is:

$$a_m(t) = \begin{cases} 0 & t \notin \tau_m \\ t & t \in \tau_m. \end{cases}$$

iii) Sine FDIA: it oscillates the actual values of the measurements. When a step FDIA is launched, the attack vector has the following form:

$$a_m(t) = \begin{cases} 0 & t \notin \tau_m \\ sin(t) & t \in \tau_m. \end{cases}$$

In the above definitions, t represents time and τ_m is the attack interval, while the scale of the attacks is determined by D matrix.

Chapter 3

Generation Control System

In this chapter, the principal aspects of the LFC system are analyzed in detail. Firstly, the generator speed governing system is presented, which is the backbone of the power systems frequency control. Then, the operation of LFC is explained through its resulting block diagram and its hierarchical control levels. Finally, the state-space representation of LFC is formulated using the differential-algebraic equations that describe the dynamic behavior of this system. The specific modeling type forms the basis of the developed cyber defense mechanisms, as it will be shown in the following chapters.

3.1 Fundamentals of Speed Governing

The core of the LFC functionality is the speed governing system and therefore, it is necessary to introduce its fundamental concepts. This section is dedicated to the modeling of the basic components that comprise the speed governing system in the frequency domain. These components are analyzed based on the low-order system frequency response (SFR) model that was introduced in [105] to approximate the average frequency behavior of a large power system in response to sudden load changes. While the SFR model is a simplification of other similar representations, it effectively captures the essential system dynamics and can be conveniently handled. In the analysis that follows, it is considered that the different quantities deviate about their steady-state values. The steady-state or nominal values are designated by a "0" subscript, e.g. ω_0, and the deviations from the nominal values are annotated with a "Δ", e.g., $\Delta\omega$.

3.1.1 Model of generator

Assume an isolated generator regulated by a turbine that supplies power to a single load. The operation of this simple power grid produces two opposing torques which act on the rotation of the considered electrical machine: the mechanical torque T_m, caused by the turbine, and the electrical torque T_e, produced by the electromagnetic field of the generator. If T_m and T_e are equal in magnitude, then the rotational speed of the generator ω is constant. According to [106], the balance between T_m and T_e is affected when T_e is increased or decreased due to an increase or decrease in the electrical load, respectively. When $T_e > T_m$, then the entire rotating system starts decelerating and vice versa.

When the generator is accelerating, the temporal evolution of its speed is described by:

$$\omega = \omega_0 + \alpha t \Rightarrow \omega - \omega_0 = \alpha t \Rightarrow \Delta\omega = \alpha t, \tag{3.1}$$

where α is the rotational acceleration. The generator phase angle δ is defined as:

$$\Delta\delta = \int \Delta\omega \, dt. \tag{3.2}$$

The relationship between the net accelerating torque T_{net}, which is the combination of multiple torques acting on a system, and ω is derived by the classical mechanics as:

$$T_{net} = I\alpha = I\frac{d}{dt}(\Delta\omega) = I\frac{d^2}{dt^2}(\Delta\delta), \tag{3.3}$$

where I denotes the moment of inertia of the generator.

By definition, the net accelerating power P_{net} of a generator is:

$$P_{net} = \omega T_{net}. \tag{3.4}$$

Furthermore, P_{net} can be expressed by the electrical P_e and mechanical P_m powers as:

$$P_{net} = P_m - P_e, \tag{3.5}$$

and can be also represented as the sum of its nominal value and its deviation term:

$$P_{net} = P_{net_0} + \Delta P_{net} = (P_{m_0} - P_{e_0}) + (\Delta P_m - \Delta P_e). \tag{3.6}$$

Following a similar analysis for T_{net}, we have:

$$T_{net} = T_{net_0} + \Delta T_{net} = (T_{m_0} - T_{e_0}) + (\Delta T_m - \Delta T_e). \tag{3.7}$$

Combining Eq. (3.4), (3.5), (3.6) and (3.7), it is obtained:

$$\begin{aligned} P_{net} &= (\omega_0 + \Delta\omega)(T_{net_0} + \Delta T_{net}) \Rightarrow \\ \Rightarrow (P_{m_0} - P_{e_0}) + (\Delta P_m - \Delta P_e) &= (\omega_0 + \Delta\omega)\big[(T_{m_0} - T_{e_0}) + (\Delta T_m - \Delta T_e)\big]. \end{aligned} \tag{3.8}$$

In the steady-state, ω is constant and thus, $P_{m_0} = P_{e_0}$ and $T_{m_0} = T_{e_0}$. Furthermore, the second-order terms in Eq. (3.8), which include products of the $\Delta\omega$, ΔT_m and ΔT_e deviations, are negligible and therefore, they are omitted . Consequently, Eq. (3.8) is converted into:

$$\Delta P_m - \Delta P_e = \omega_0(\Delta T_m - \Delta T_e). \tag{3.9}$$

The combination of Eq. (3.3) and (3.9) provides the relationship between ΔP_{net} and ω as follows:

$$\Delta P_{net} = \Delta P_m - \Delta P_e = \omega_0 I \frac{d}{dt}(\Delta\omega), \tag{3.10}$$

By applying the Laplace transformation to Eq. (3.10), the generator response to the net power deviation in the frequency domain is derived:

$$\Delta P_m - \Delta P_e = Ms\Delta\omega \Rightarrow \frac{\Delta\omega}{\Delta P_{net}} = \frac{1}{2Hs}, \tag{3.11}$$

where M is the angular momentum of the generator.

The resulting formula (3.11) describes the transfer function of the generator and is illustrated in Fig. 3.1.

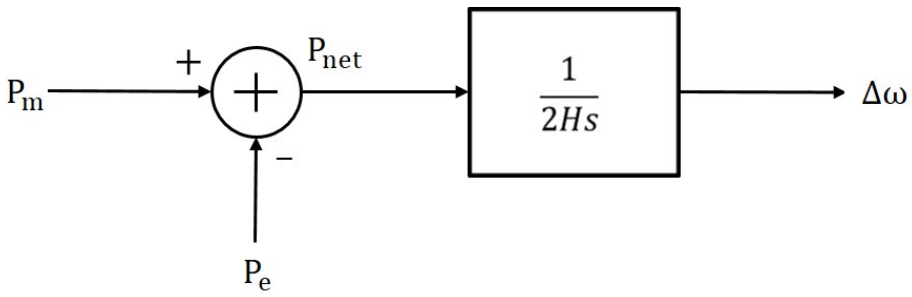


Figure 3.1 Generator transfer function.

3.1.2 Model of load

The role of an electrical machine is to supply power to the connected electrical load. The total electrical load within a power system is collectively composed of a wide range of electrical devices. These devices can be categorized based on their dependence to frequency.

The resulting categories include resistive loads, such as lightning and heating, which are independent of frequency, motor loads, e.g. pumps and fans, where the electrical power is affected by changes in the motor speed and other types of loads that demonstrate more complex frequency characteristics. When the power system is dominated by motor loads, the equation that relates the changes in frequency due to the load variations is:

$$\Delta P_L^f = D\,\Delta\omega,$$

where ΔP_L^f frequency-sensitive load change and D is the damping constant that describes the percent change in load for a given percent change in frequency. Therefore, the variations in the electrical power of a machine can be expressed as:

$$\Delta P_e = \Delta P_L + \Delta P_L^f = \Delta P_L + D\,\Delta\omega,$$

where ΔP_L is the non frequency-sensitive load change. Now, the transfer function of the motion equation for a machine connected to a load can be acquired, which is demonstrated in the Fig. 3.2 that follows:

Figure 3.2 Generator transfer function considering load changes.

3.1.3 Model of turbine

The prime mover of a generator is the engine that constantly provides mechanical energy to the machine in order to convert it into electrical power. Turbines are a widely adopted type of engines for this task because they can convert the energy of an element, e.g. water, steam, diesel, etc. into mechanical energy. The mechanical energy output of a turbine is determined by the position of its valve or gate, depending on the type of the turbine. Typically, the prime mover that drives an electrical machine can be a steam turbine or a hydroturbine. To model the prime mover of a generator, the steam supply and boiler control system characteristics have to be reflected in case of a steam turbine, or the penstock characteristics for a hydro turbine. Without loss of generality, the simplest prime-mover model will be considered for this study, namely the nonreheat turbine.

The transfer function of machine-load model that utilizes a nonreheat turbine as its prime mover is shown in the Fig. 3.3 that follows:

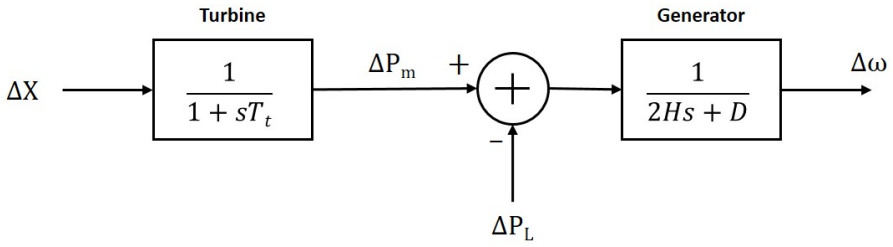


Figure 3.3 Machine-load model driven by a nonreheat turbine.

3.1.4 Model of governor

Consider a generating unit with a fixed position of its turbine valve/gate that is connected to a load. Due to the fixed mechanical power output of the turbine, the electrical energy provided by the generator will be constant. In this case, any load change would affect the energy equilibrium of the power system. This imbalance is reflected as a deviation of the speed machine from its nominal or scheduled value, according to Fig. 3.3. Under such circumstances, the system frequency will eventually be driven far beyond its acceptable operational limits, degrading the stability of the system. This issue can be resolved by adding a control mechanism that identifies machine speed changes and regulates the position of the turbine valve/gate accordingly; in this way, the generated power is properly adjusted to compensate for the demand-side changes and system frequency is restored to its nominal value.

In the following paragraphs, the different speed governing techniques of the electrical generators are presented, depending of the type of the power system.

3.1.4.1 Isochronous control mode

The simplest type of speed governing systems is the isochronous control. In isochronous control mode, an integral controller is typically used to stabilize the speed of the generator. The integrator is a basic control mechanism that a drives a system variable to a predefined setpoint through the reset action. For the isochronous speed governing of a generator, the reset action involves driving the speed error, which is the difference between the actual and the desired or reference speed, to zero by continuously integrating it. The model of a isochronous governor is illustrated in Fig. 3.4.

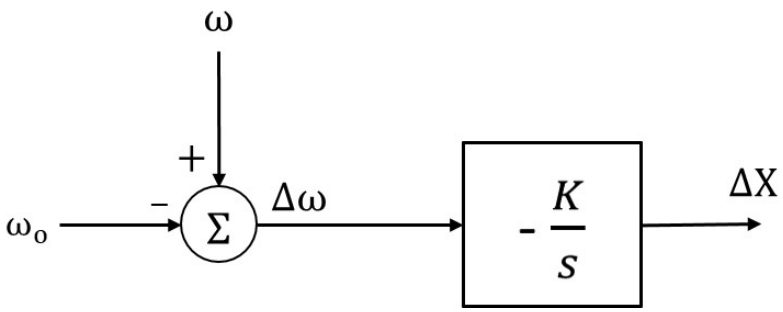


Figure 3.4 Transfer function of isochronous governor.

Determining the relationship between the speed and power output of a generator that uses isochronous governor provides a better insight on this control mode. To achieve this, the value of the governor transfer function, depicted in Fig. 3.4, is calculated for the steady-state (where $t \to \infty \Rightarrow s \to 0$) as:

$$\frac{\Delta X}{\Delta \omega} = -\frac{K}{s} \Rightarrow \frac{\Delta \omega}{\Delta X} = -\frac{s}{K} \xrightarrow{s \to 0} \frac{\Delta \omega}{\Delta X} = 0 \Rightarrow \omega = \omega_0. \tag{3.12}$$

Eq. (3.12) yields that the speed is independent of the governor output. Since the governor output is proportional to the generator power output, it is concluded that frequency is independent of the generated electrical power. The relationship between the speed and the power output in isochronous control mode is plotted in Fig. 3.5.

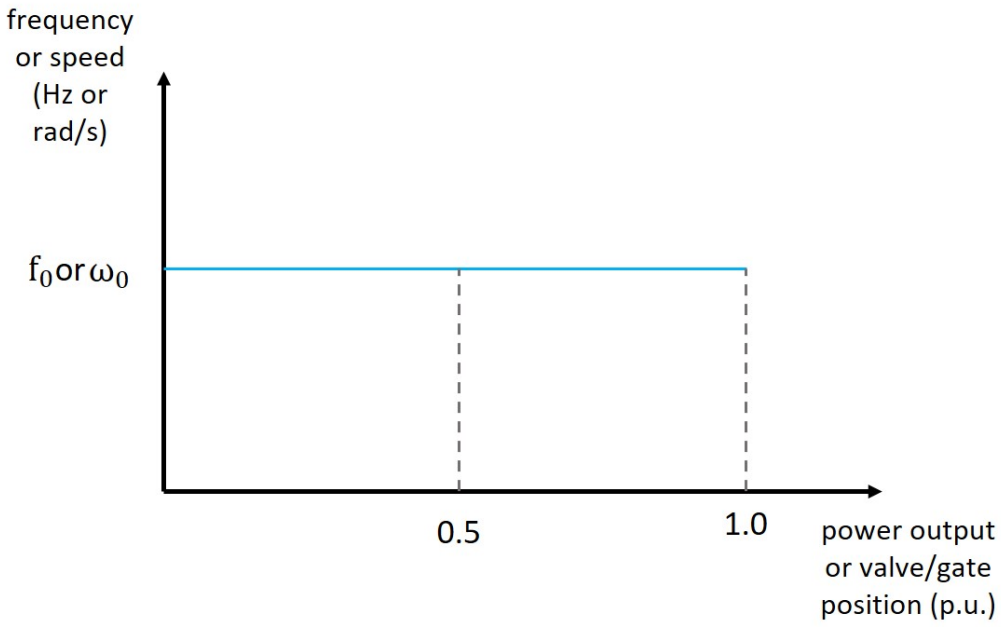


Figure 3.5 Isochronous governor characteristic.

3.1.4.2 Droop control mode

The isochronous control mode is typically used when a single generator operates in the power system. For multi-machine power systems, where two or more generating units work in parallel, the load must be properly shared between the connected generators. If all generators use isochronous governors, their synchronization is heavily affected by the load changes and eventually fails. This happens because the generators compete with each other on forcing their own speed setting to the system, based on their status before the synchronization process. Poor parallel operation leads to degradation of the system performance and damage in the equipment.

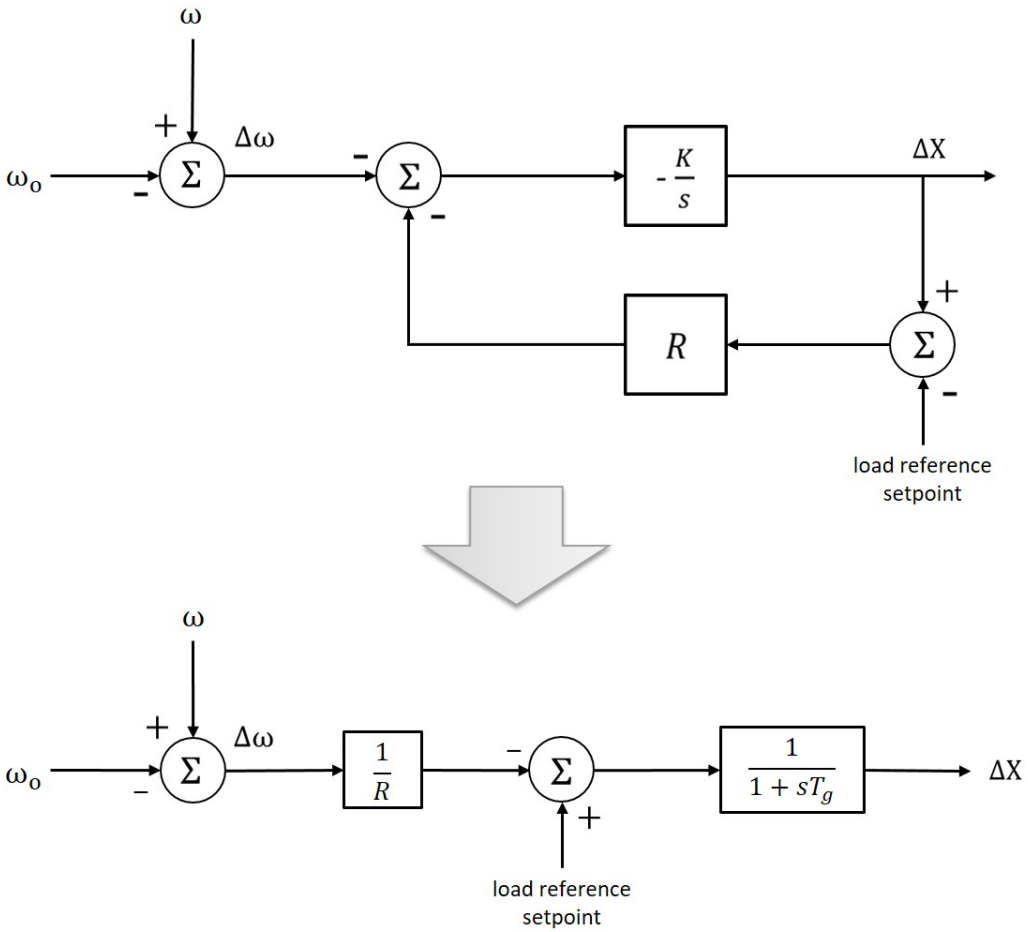


Figure 3.6 Transfer function of droop-equipped governor.

A practical solution towards the efficient load sharing among multiple generators is the transformation of the relationship between their speed and power output. This can be achieved by introducing a droop characteristic to the governors, as shown in Fig. 3.6. In

this type of speed governing, the relationship between the speed and the power output of the generator is inversely proportional. This is verified by computing the value of the governor transfer function, demonstrated in Fig. 3.6, for the steady-state (where $t \to \infty \Rightarrow s \to 0$) as:

$$\frac{\Delta X}{-\frac{\Delta\omega}{R}+lr} = \frac{1}{1+sT_g} \xrightarrow{lr=0} \frac{\Delta X}{\Delta\omega} = \frac{-\frac{1}{R}}{1+sT_g} \xrightarrow{s\to 0} \frac{\Delta\omega}{\Delta X} = -R. \tag{3.13}$$

With this configuration, the generator speed shifts to a certain value for a specific change in the power output. The relationship between the speed and the power output in droop control mode is plotted in Fig. 3.7.

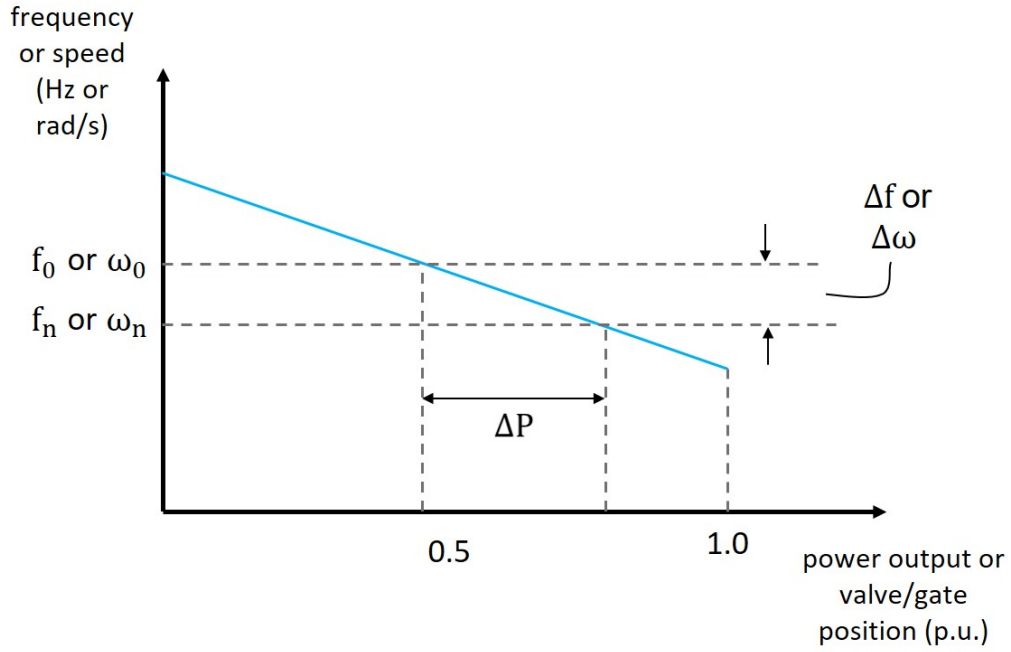


Figure 3.7 Droop governor characteristic.

The effectiveness of the droop control mode on the load sharing process is highlighted in the following paradigm. Assume two parallel units that use droop governors, where Fig. 3.8 shows their characteristics. Initially, these generators operate at nominal speed ω_0 and their power outputs are P_{11} and P_{21}. A load increase that happens after a time period, causes the generating units to slow down, leading to a decrease in their speeds. However, their speeds are driven to a new, common value ω_n. That is because the droop control provides the same frequency change for different amounts of power output, as Fig. 3.8 indicates. The load amount distributed in each unit is determined by the gradients R_1 and R_2 of their droop

characteristics, as follows:

$$R_1 = \frac{\Delta\omega}{\Delta P_1} \Rightarrow \Delta P_1 = P_{12} - P_{11} = \frac{\Delta\omega}{R_1},$$

$$R_2 = \frac{\Delta\omega}{\Delta P_2} \Rightarrow \Delta P_2 = P_{22} - P_{21} = \frac{\Delta\omega}{R_2}.$$

By properly selecting the droop value of each machine, it is possible to perform an efficient load division to the connected generators.

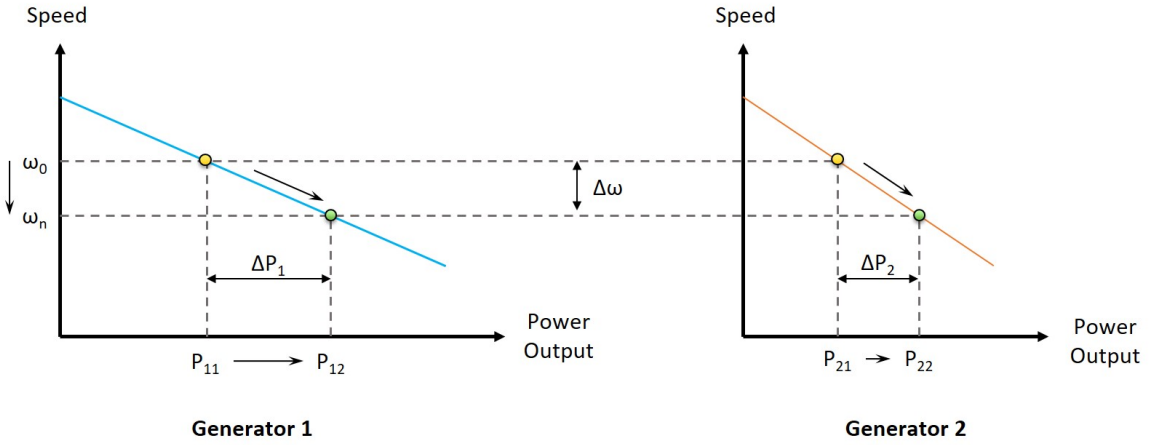


Figure 3.8 Load distribution to parallel units using droop control.

3.1.4.3 Load reference setpoint

In the load sharing paradigm that was previously described, it is implied that the system frequency converges to a non-nominal value after a load change. This droop control issue can be addressed by adjusting the input of the *load reference setpoint* (LRS), shown in 3.6. LRS allows the modification of the vertical intercept of the droop characteristic. This means that the droop characteristic can be freely shifted across the vertical axis, as shown in Fig. 3.9. For example, assume that the blue curve of Fig. 3.9 represents the droop characteristic of a generating unit. When this unit provides 0.5 p.u. of its power output, it operates at nominal speed. To make the unit operate at nominal speed for its full power output, the LRS has to be increased until the blue curve aligns with the yellow one. Similarly, to make the unit operate at nominal speed without any load, the LRS has to be decreased until the blue curve reaches the green one.

By properly configuring the value of the LRS, it is possible to drive a generator to its nominal frequency for any desired value of its power output. Consider the paradigm of the previous section but with a configurable LRS for generating unit 1. After a load increase,

unit 1 slow downs and its operating point is temporarily moved from (P_{11}, ω_0) to (P_{12}, ω_n), indicated by the yellow and fainted green dots in Fig. 3.10, respectively. Then, the LRS is immediately increased, causing the droop characteristics of the generators to vertically slide upwards, as shown in Fig. 3.10. In this way, the desired operating point (P_{12}, ω_0), represented by the green dot in Fig. 3.10, is reached and the power system frequency is stabilized to its nominal value. LRS is the primary control input of a generating unit that is regulated by AGC.

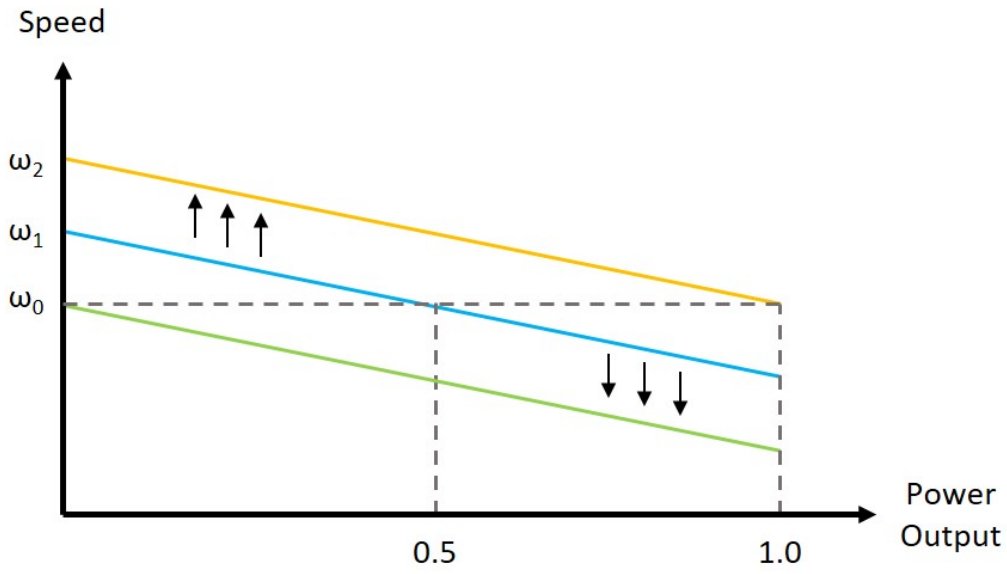


Figure 3.9 Effects of the load reference setpoint to the droop characteristic.

3.1.5 Model of tie-line

It is common practice to separate power systems into multiple, distinct areas as a strategic approach to enhance their efficiency and reliability, while meeting the diverse energy needs of different regions. These power areas are interconnected through transmission lines, which are referred to as tie-lines. The modeling of the power that flows across a tie-line between areas i and j follows the DC load flow method presented in [107], described as:

$$P_{tie} = \frac{1}{X_{tie}}(\theta_i - \theta_j), \tag{3.14}$$

where X_{tie} is the tie-line reactance, θ_i is the tie-line phase angle from the side of area i and θ_j tie-line phase angle from the side of area j. For a small deviation from the initial values of

(3.14), the tie-line power flow deviation is acquired as:

$$P_{tie}+\Delta P_{tie}=\frac{1}{X_{tie}}\left[(\theta_i+\Delta\theta_i)-(\theta_j+\Delta\theta_j)\right]=\frac{1}{X_{tie}}(\theta_i-\theta_j)+\frac{1}{X_{tie}}(\Delta\theta_i-\Delta\theta_j). \tag{3.15}$$

Eq. (3.15) yields that:

$$\Delta P_{tie}=\frac{1}{X_{tie}}(\Delta\theta_i-\Delta\theta_j).$$

Without loss of generality, it is assumed that the tie-line phase angle is equal to the rotor angle of the equivalent area generator. Therefore:

$$\Delta P_{tie}=\frac{1}{X_{tie}}(\Delta\theta_i-\Delta\theta_j)=\frac{1}{X_{tie}}(\Delta\delta_i-\Delta\delta_j). \tag{3.16}$$

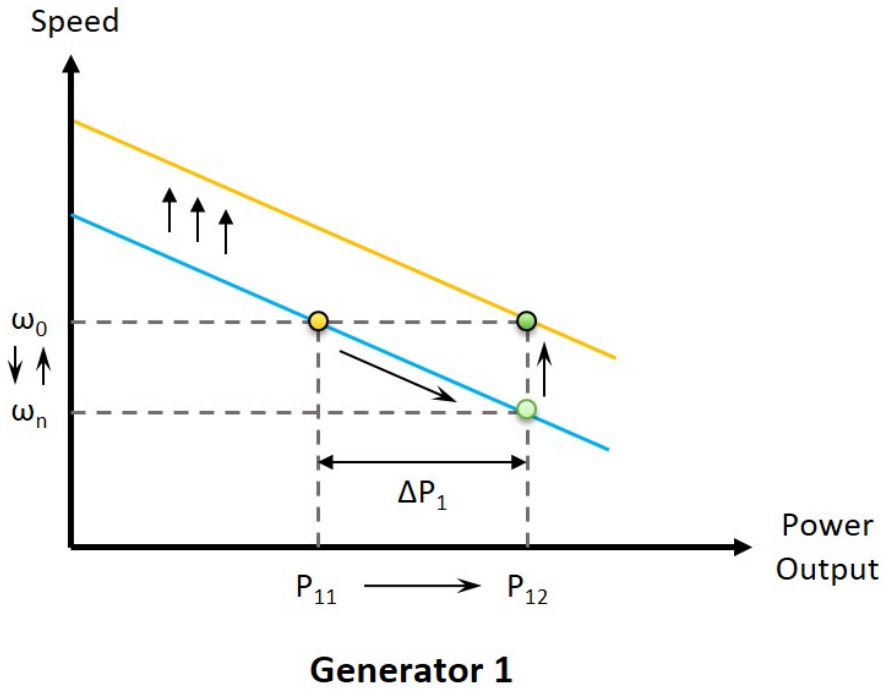


Figure 3.10 Load distribution to parallel units using droop control.

From the definition of generator phase angle (3.2) and Eq. (3.16), it is obtained:

$$\Delta P_{tie}=\frac{T_{ij}}{s}(\Delta\omega_i-\Delta\omega_j)\Rightarrow\Delta P_{tie}=\frac{2\pi T_{ij}}{s}(\Delta f_i-\Delta f_j),$$

where $T_{ij}=\frac{1}{X_{tie}}$ represents the stiffness coefficient of the tie-line.

3.2 Load Frequency Control System

The normal operation of power systems requires the continuous preservation of the energy equilibrium within acceptable limits. This balance is usually affected by the load variations that constantly occur in the grid. Therefore, the generated power must be always adjusted according to the levels of the energy demand. A key indicator of the energy imbalances is the power system frequency: any deviation of frequency from its nominal value implies that there is a mismatch between generation and demand. The *load frequency control* system is a mechanism that utilizes frequency measurements to achieve the aforementioned balance. The LFC receives frequency measurements as its input and senses any deviation from their nominal value. Then, it properly regulates the output of the system generators to compensate for the energy mismatches.

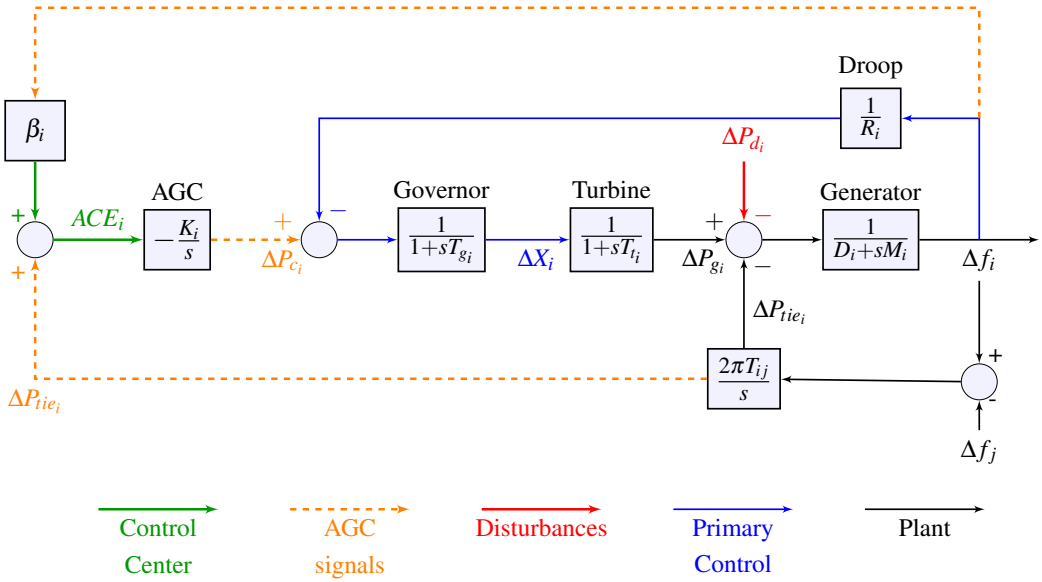


Figure 3.11 Block diagram of LFC for the i*th* control area.

The assembly of the different speed governing components that were presented in Section 3.1 forms the typical block diagram of the LFC system. This LFC diagram for the i*th* power system area is illustrated in Fig. 3.11. In this representation, the frequency domain response of each speed governing component (generating units, governor-turbine systems, controllers, etc.) to power imbalances is modeled by a system block. These blocks are mathematically formulated based on the physical characteristics and the behavior of each speed governing component. The transmission system performance and the intermachine oscillations are disregarded in LFC analysis while the overall dynamic performance of the area generators

is represented by an equivalent unit model. This model is extensively used in the literature and by industries, such as the European Network of Transmission System Operators for Electricity (ENTSO-E) [108], and therefore, it is considered suitable for frequency control studies.

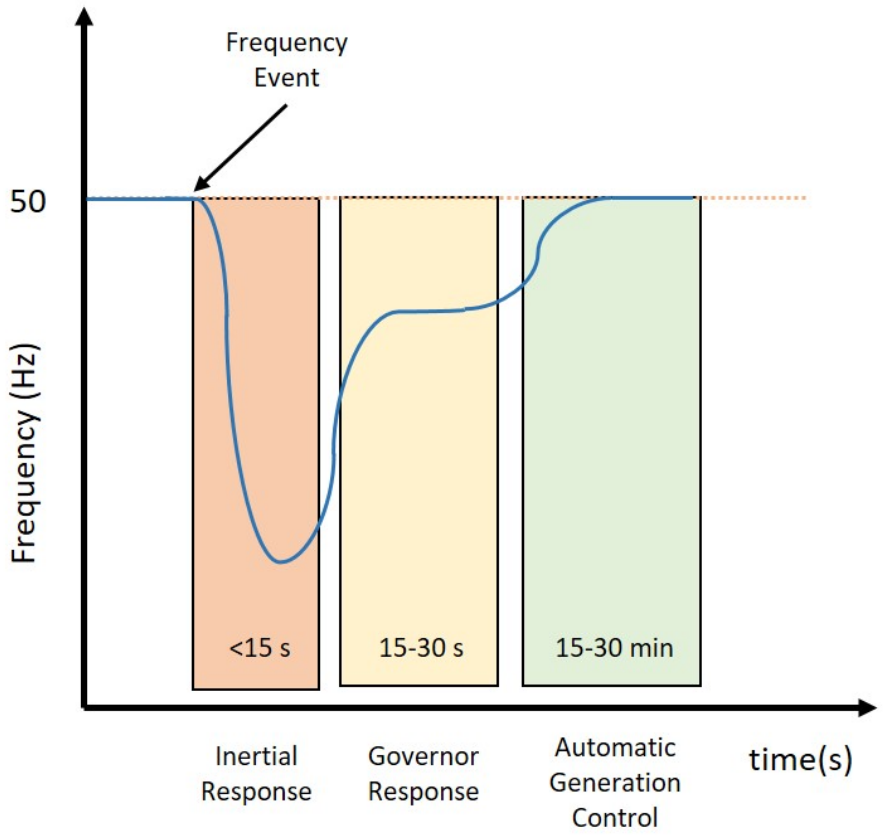


Figure 3.12 Frequency response paradigm of a power system utilizing LFC to a load disturbance event.

As mentioned in Section 1.3, the LFC is composed of multiple control levels, arranged in a hierarchical manner. Each of these control levels performs a specific LFC operation and plays a unique role in the frequency stabilization. This study focuses on the primary and the secondary levels of the frequency control, due to their importance in the LFC functionality. A detailed analysis of these frequency control levels follows:

- **Primary control:** it is the initial frequency control level and is responsible for stabilizing the power system frequency within acceptable values. Its operation can be described as: the local governors of the generators automatically sense a power imbalance event as a deviation of frequency from its nominal value. Then, their outputs are properly adjusted to regulate the position of the turbine valve/gate. In this way,

governors force the produced power to follow the energy demand levels. While this type of control stabilizes the system frequency, it is unable to restore it to its nominal value and thus, a frequency error remains after its operation. In the block diagram of LFC (Fig. 3.11), primary control corresponds to the blue arrows and its impact on frequency is demonstrated in the "Governor Response" window of the Fig. 3.12 paradigm.

- **Secondary/Supplementary control:** it is a control center application, also known as *Automatic Generation Control* (AGC), that takes place after the primary frequency control. Its purpose is to compensate for the limitations of the primary control; namely, to restore the frequency back to its nominal value and keep the tie-line flows at their scheduled levels (in case of multi-area power systems). To achieve this, it forms the area control error (ACE) by the received frequency and tie-line power flow measurements, calculates the command signal and sends it as an input to the load reference setpoint of the governors that participate in it. In the block diagram of LFC (Fig. 3.11), AGC corresponds to the orange and green arrows and its impact on frequency is demonstrated in the "Automatic Generation Control" window of the Fig. 3.12 paradigm.

Table 3.1 Main features of the frequency control levels.

	Response Time	**Duration Time**	**Operation**
Primary Control	15-30 seconds	15 minutes	Automatic
Secondary Control	200 seconds	120 minutes	Communication-based
Tertiary Control	15 minutes	Indicated by TSO	Upon request

For a better insight into the functionality of LFC, the main features of the different frequency control levels are illustrated in Table 3.1. According to this table, the primary control takes place 15-30 seconds after a power imbalance event and lasts for approximately 15 minutes. Its operation is automatic since the local governors are installed in the generating units that drive; this yields that the governors are interconnected with the turbines and they can immediately sense frequency deviations by measuring the generator speed. Regarding the AGC, it is applied around 200 seconds after a power imbalance event and its duration is about 15 minutes. AGC is a software application within the control center and therefore, its implementation is communication-based. This indicates that it uses telemetry to receive the necessary frequency and tie-line power flow measurements from the field devices and remote communication channels to send the control signals to the load reference setpoints of the participating generators. Finally, while the tertiary control is out of scope for the

present study, its main features are included in Table 3.1, for the sake of completeness. This type of frequency control aims to restore the reserve margin used for the AGC and occurs approximately 15 minutes after a power imbalance event upon request to the transmission system operator (TSO).

3.3 State-space Representation of LFC

For the development of the proposed cyber defense layers, the state-space representation of LFC is required, as it will be explained in the following chapters. This type of modeling expresses the dynamical behavior of the LFC system as a set of its input, output and state variables related by first-order differential equations. These equations are obtained by the transfer function of each LFC component in the frequency domain, demonstrated in Fig. 3.11. In what follows, i corresponds to a power system area and for multi-area power systems, each area i is represented by an equivalent generating unit.

Note: *For feasibility purposes of the introduced cyber defense methodologies, several practical features are considered in the LFC modeling of this study, such as nonlinearities, HVDC tie-lines, disturbances due to RES generation, etc. While these practical features are not included in the core components of the speed governing system, their dynamical equations, which reveal their contribution to the LFC system, are analyzed in this section.*

From the generator dynamics of area i, it is acquired:

$$\Delta \dot{f}_i = \frac{1}{2H_i}(\Delta P_{g_i} - D_i \Delta f_i - \Delta P_{tie_i} - \Delta P_{d_i}), \tag{3.17}$$

where the frequency deviation is denoted as Δf_i, the damping (load frequency relief) and inertia constants are represented by D_i and H_i, respectively, and the deviations in the tie-line power interchange and external disturbances are denoted as ΔP_{tie_i} and ΔP_{d_i}, respectively.

From the governor-turbine dynamics of area i, it is obtained:

$$\Delta \dot{P}_{g_i} = \frac{1}{T_{t_i}}(\Delta X_i - \Delta P_{g_i}), \tag{3.18}$$

$$\Delta \dot{X}_i = \frac{1}{T_{g_i}}(\Delta P_{c_i} - \Delta X_i - \frac{\Delta f_i}{R_i}), \tag{3.19}$$

where ΔP_{g_i} and ΔX_i denote the deviations in the turbine power output and governor valve position, respectively; T_{t_i} and T_{g_i} represent the turbine and governor time constants, respectively; R_i is the droop characteristic and ΔP_{c_i}, which also serves as the area control signal, denotes the deviation in the load reference setpoint of the governor.

In a power system composed of N areas, the total tie-line power deviation of the ith area is determined by:

$$\Delta P_{tie_i} = \Delta P_{ac_i} + \Delta P_{dc_i}, \tag{3.20}$$

where ΔP_{ac_i} denotes the AC tie-line power deviation and ΔP_{dc_i} stands for the high-voltage direct current (HVDC) tie-line power variation. When the AC tie-line does not have a thyristor-controlled phase shifter (TCPS), ΔP_{ac_i} can be calculated as:

$$\Delta P_{ac_i} = \frac{2\pi}{s} \sum_{j=1, j\neq j}^{N} T_{ij}(\Delta f_i - \Delta f_j), \tag{3.21}$$

where T_{ij} is the tie-line synchronizing coefficient between areas i and j, $j = 1, 2, \cdots, N$, $j \neq i$. When the AC tie-line is equipped with a TCPS, the value of ΔP_{ac_i} can be obtained from the following equation [109–111]:

$$\Delta P_{ac_i} = \frac{2\pi}{s} \sum_{j=1, j\neq j}^{N} T_{ij}(\Delta f_i - \Delta f_j) + \sum_{j=1, j\neq j}^{N} T_{ij} \frac{K_{s_{ij}}}{1 + sT_{s_{ij}}}(\Delta f_i - \Delta f_j), \tag{3.22}$$

where $T_{s_{ij}}$ and $K_{s_{ij}}$ are the time and gain constants of TCPS links between areas i and j, respectively. When areas i and j are connected with a HVDC link, the value of ΔP_{dc_i} is calculated based on the difference between the Δf_i and the rest of the Δf_j [110, 112, 113], as shown below:

$$\Delta P_{dc_i} = \frac{1}{1 + sT_{dc_i}} \sum_{j=1, j\neq j}^{N} K_{ij}(\Delta f_i - \Delta f_j), \tag{3.23}$$

where T_{dc_i} and K_{ij} represent the HVDC time and gain constants between areas i and j, respectively. If areas i and j are not coupled with an HVDC link, then $K_{ij} = 0$.

The external disturbances ΔP_{d_i}, considering the variations due to RES generation, are modeled as:

$$\Delta P_{d_i} = \Delta P_{L_i} - \Delta P_{RES_i}, \tag{3.24}$$

where ΔP_{L_i} and ΔP_{RES_i} express the load and RES disturbances, respectively.

The RES disturbances ΔP_{RES_i} consist of the generation deviations due to photovoltaics ΔP_{PV_i} and generation variations caused by wind farms ΔP_{W_i}. Formally, this is expressed by:

$$\Delta P_{RES_i} = \Delta P_{W_i} + \Delta P_{PV_i}. \tag{3.25}$$

The variations due to solar energy generation follow the next model:

$$\Delta P_{PV_i} = 0.6\sqrt{\Delta P_{solar}}, \tag{3.26}$$

where ΔP_{solar} reflects the solar power deviation from its initial value. The wind turbine output power adopts the following model:

$$\Delta P_{W_i} = \frac{1}{2}\rho_W A_T \Delta V_{W_i}^3 C_P(\lambda_W, \beta_W), \tag{3.27}$$

where ρ_W is the air density, A_T denotes the rotor swept area, ΔV_{W_i} represents the wind speed deviation from its nominal value for area i, C_P reflects the rotor blade parameter, λ_W is the optimal tip speed ratio and β_W is the pitch angle of the blade. The detailed models of the RES developed in this work and their parameter values are available in [114].

The load reference setpoints ΔP_{c_i} of the governors that are not part of the AGC have a fixed value. The governors driven by the AGC adjust their setpoints according to the output of this controller. The ACE is used as input to the AGC. For each control area i, ACE_i is defined as:

$$ACE_i = \beta_i \Delta f_i + \Delta P_{tie_i}, \tag{3.28}$$

where the frequency bias is represented by β_i. Therefore, the load reference setpoint of a governor that contributes to AGC, driven by an integral controller, is modeled as:

$$\Delta P_{c_i} = -K_{I_i} \int ACE_i dt, \tag{3.29}$$

where K_{I_i} is the integral gain.

This study considers the standard nonlinearities of LFC, which are modeled based on [115, 116]. More specifically:

- *Generation Rate Constraints (GRC)*: the rate of change of the turbine output is limited by a fixed value, denoted as P_{grc}. This yields that:

$$\Delta \dot{P}_{g_i} < P_{grc}. \tag{3.30}$$

- *Governor Dead-Band (GDB)*: the generator governor does not respond to minor fluctuations in the active power. Mathematically:

$$\Delta X_i(\Delta P_{d_i}) = \begin{cases} 0, & \Delta P_{d_i} < |P_{gdb}| \\ \Delta X_i(\Delta P_{d_i}) & \Delta P_{d_i} > |P_{gdb}|, \end{cases} \tag{3.31}$$

  where P_{gdb} represents the threshold power change at which the governor starts to respond.

- *Transportation Time Delay (TTD)*: the signals that are remotely exchanged with the control center face delays due to communication and mechanical system responses, as:

$$\Delta P_{c_i}(t) = -K_{I_i} \int ACE_i(t-\tau_d)dt, \tag{3.32}$$

where τ_d is the aggregated communication and mechanical delay.

The values of P_{grc}, P_{gdb} and τ_d used in the present study, are provided in the Appendix C.

The local state vector used in the state-space representation of the ith area, i.e. $x_i \in \mathbb{R}^O$, where O is the dynamic order in the ith area, is defined as:

$$\Delta x_i = \begin{bmatrix} \Delta f_i & \Delta P_{g_i} & \Delta X_{g_i} & \int ACE_i & \Delta P_{ac_i} & \Delta P_{dc_i} \end{bmatrix}^T. \tag{3.33}$$

From Eq. (3.17) – (3.33), the differential-algebraic expression of the total N-area power system in its compact form is:

$$\begin{cases} \dot{x}(t) = Ax(t) + F\phi(x,t) + Bu(t) + Ed(t) \\ y(t) = Cx(t), \end{cases} \tag{3.34}$$

where $x(t) = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \end{bmatrix}^T \in \mathbb{R}^{O \cdot N}$ is the global state vector ($O \cdot N = n$), $u(t) = \begin{bmatrix} \Delta P_{c_1} & \Delta P_{c_2} & \cdots & \Delta P_{c_m} \end{bmatrix}^T \in \mathbb{R}^m$ is the input vector, $d(t) = \begin{bmatrix} \Delta P_{d_1} & \Delta P_{d_2} & \cdots & \Delta P_{d_r} \end{bmatrix}^T \in \mathbb{R}^r$ is the disturbance vector (modeled as unknown input), $y(t) \in \mathbb{R}^p$ is the output vector and $\phi(x,t) \in \mathbb{R}^\upsilon$ models LFC nonlinearities. The matrices $A \in \mathbb{R}^{n \times n}$, $F \in \mathbb{R}^{n \times \upsilon}$, $B \in \mathbb{R}^{n \times m}$, $E \in \mathbb{R}^{n \times r}$ and $C \in \mathbb{R}^{p \times n}$ are known.

Chapter 4

SMO-based Attack Detection & Localization for LFC

This chapter is dedicated to the presentation of the proposed observer-based attack detection and localization methods for the frequency control of power systems. It has been already stated in Section 1.3 that AD approaches determine whether a cyberattack has been launched against the LFC and when it occurred, while ALC methodologies aim to identify which signals within the LFC loop have been affected by digital threats. In what follows, the state-space representation of LFC during FDIAs is initially is established, forming the basis for developing the introduced AD and ALC methodologies. Then, the observer design procedure for the proposed AD and ALC is demonstrated, along with the observer stability conditions. The formulation of an adaptive threshold selection follows, which minimizes the false positive alarm rate of the presented AD and ALC strategies. Finally, the performance of the proposed cybersecurity techniques is evaluated on a series of experimental testbeds to demonstrate its effectiveness.

4.1 Modelling FDIAs against LFC

When system (3.34) is under FDIAs against measurements and control signals, its state-space representation is transformed into the following form:

$$\begin{cases} \dot{x}(t) = Ax(t) + F\phi(x,t) + Bu(t) + Ed(t) \\ y(t) = Cx(t) + Da_m(t) \end{cases} \tag{4.1}$$

where $a_m(t) \in \mathbb{R}^q$ is the attack vector and $D \in \mathbb{R}^{p\times q}$. $p \geq q+r$ and it is assumed that C, D, and E have full column rank. For the adversaries, we assume that they have access to the communication channels of LFC and can modify the exchanged data. Moreover, they have full knowledge of the A, B, C, D and E matrices.

4.2 Observer Design Preliminaries

The core concept behind observed-based approaches for the detection and localization of cyberattacks against LFC is to design observers in a way that the resulting estimation errors are asymptotically stable only under attack-free conditions. This ensures that the estimation errors will always be zero unless a cyberattack occurs, making them reliable cyber threat indicators. The main challenge of these methodologies is to distinguish cyberattacks from other types of external disturbances. In the proposed approach, this decoupling is achieved through system coordinate transformation, inspired by advanced observation techniques [117]. Initially, the original system is virtually split into *subsystem-I*, which carries only the system disturbances, and *subsystem-II*, which carries only the cyberattacks. Then, we develop *observer-I* for *subsystem-I* which is susceptible only to system disturbances and *observer-II* for *subsystem-II* which is sensitive only to cyberattacks.

Before proceeding to the presentation of the proposed methodology, a series of mathematical conditions have to be satisfied regarding the considered LFC system (4.1). These conditions are required for the existence of the introduced observers and are included in the assumptions that follow. After each assumption, a brief explanation is provided to shed more insight into the observer design process.

Assumption 1. $rank(E) = rank(CE)$.

Assumption 1 is the necessary condition for the existence of the state and output transformation (G, H), as:

$$g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = Gx \text{ and } h = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = Hy.$$

where $g_1 \in \mathbb{R}^r$ and $h_1 \in \mathbb{R}^r$. After applying the above transformation technique to (4.1), it is acquired:

$$\begin{cases} \dot{x} = Ax + F\phi + Bu + Ed \\ y = Cx + Da_m \end{cases} \xrightarrow[\text{mult. by H}]{\text{mult. by G}} \begin{cases} G\dot{x} = G\big(Ax + F\phi(x,t) + Bu + Ed\big) \\ Hy = HCx + HDa_m \end{cases} \xrightarrow[y=H^{-1}h]{x=G^{-1}g}$$

$$\Rightarrow \begin{cases} \dot{g} = GAG^{-1}g + GF\phi(G^{-1}g,t) + GBu + GEd \\ h = HCG^{-1}h + HDa_m. \end{cases} . \tag{4.2}$$

The (G,H) transformation is selected so that the following properties are satisfied for the matrices of system (4.2):

$$GAG^{-1} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, GF = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}, GB = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, GE = \begin{bmatrix} E_1 \\ \underline{0} \end{bmatrix}, HCG^{-1} = \begin{bmatrix} C_1 & \underline{0} \\ \underline{0} & C_4 \end{bmatrix} \text{ and } HD = \begin{bmatrix} \underline{0} \\ D_2 \end{bmatrix},$$

where $G = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \in \mathbb{R}^{n\times n}$, $H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} \in \mathbb{R}^{p\times p}$, $G_1 \in \mathbb{R}^{r\times n}$, $H_1 \in \mathbb{R}^{r\times p}$, $A_1 \in \mathbb{R}^{r\times r}$, $A_4 \in \mathbb{R}^{(n-r)\times(n-r)}$, $F_1 \in \mathbb{R}^{r\times k}$, $B_1 \in \mathbb{R}^{r\times m}$, $E_1 \in \mathbb{R}^{r\times r}$, $D_2 \in \mathbb{R}^{(p-r)\times q}$, $C_1 \in \mathbb{R}^{r\times r}$, $C_4 \in \mathbb{R}^{(p-r)\times(n-r)}$ and C_1 is invertible. Therefore, system (4.2) can be written as:

$$\begin{cases} \begin{bmatrix} \dot{g}_1 \\ \dot{g}_2 \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \phi + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u + \begin{bmatrix} E_1 \\ \underline{0} \end{bmatrix} d \\ \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} C_1 & \underline{0} \\ \underline{0} & C_4 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} \underline{0} \\ D_2 \end{bmatrix} a_m. \end{cases} \tag{4.3}$$

By splitting system (4.3) into its upper and lower part, the transformed system (4.2) can be separated into the next subsystems:

$$\begin{cases} \dot{g}_1 = A_1 g_1 + A_2 g_2 + F_1 \phi(G^{-1}g,t) + B_1 u + E_1 d \\ h_1 = C_1 g_1, \end{cases} \tag{4.4}$$

$$\begin{cases} \dot{g}_2 = A_3 g_1 + A_4 g_2 + F_2 \phi(G^{-1}g,t) + B_2 u \\ h_2 = C_4 g_2 + D_2 a_m. \end{cases} \tag{4.5}$$

For the observer design procedure of the present study, it is more convenient to transfer the a_m vector from the output equation of system (4.5) to its state equation. To this end, a new state $g_3 = \int_0^t h_2(\tau)\, d\tau$ is defined so that:

$$\dot{g}_3 = C_4 g_2 + D_2 a_m.$$

With the new g_3 state, subsystem (4.5) can be converted into the following augmented form of $n+p-2r$ order:

$$\begin{cases} \begin{bmatrix} \dot{g}_2 \\ \dot{g}_3 \end{bmatrix} = \begin{bmatrix} A_4 & \underline{0} \\ C_4 & \underline{0} \end{bmatrix} \begin{bmatrix} g_2 \\ g_3 \end{bmatrix} + \begin{bmatrix} A_3 \\ \underline{0} \end{bmatrix} g_1 + \begin{bmatrix} F_2 \\ \underline{0} \end{bmatrix} \phi(G^{-1}g,t) + \begin{bmatrix} B_2 \\ \underline{0} \end{bmatrix} u + \begin{bmatrix} \underline{0} \\ D_2 \end{bmatrix} a_m \\ h_3 = g_3 \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} \dot{g}_0 = A_0 g_0 + \bar{A}_3 g_1 + \bar{F}_2 \phi(G^{-1}g,t) + B_0 u + D_0 a_m \\ g_3 = C_0 g_0, \end{cases} \tag{4.6}$$

where $g_0 = \begin{bmatrix} g_2 \\ g_3 \end{bmatrix} \in \mathbb{R}^{n+p-2r}$, $h_3 \in \mathbb{R}^{p-r}$, $A_0 = \begin{bmatrix} A_4 & \underline{0} \\ C_4 & \underline{0} \end{bmatrix} \in \mathbb{R}^{(n+p-2r)\times(n+p-2r)}$, $\bar{A}_3 = \begin{bmatrix} A_3 \\ \underline{0} \end{bmatrix} \in \mathbb{R}^{(n+p-2r)\times r}$, $\bar{F}_2 = \begin{bmatrix} F_2 \\ \underline{0} \end{bmatrix}$, $B_0 = \begin{bmatrix} B_2 \\ \underline{0} \end{bmatrix} \in \mathbb{R}^{(n+p-2r)\times m}$, $D_0 = \begin{bmatrix} \underline{0} \\ D_2 \end{bmatrix} \in \mathbb{R}^{(n+p-2r)\times q}$ and $C_0 = \begin{bmatrix} \underline{0} & I_{p-r} \end{bmatrix} \in \mathbb{R}^{(p-r)\times(n+p-2r)}$.

In a similar manner, subsystem (4.4) is restructured as:

$$\begin{cases} \dot{g}_1 = A_1 g_1 + \bar{A}_2 g_0 + F_1 \phi(G^{-1}g,t) + B_1 u + E_1 d \\ h_1 = C_1 g_1, \end{cases} \tag{4.7}$$

where $\bar{A}_2 = \begin{bmatrix} A_2 & 0_{r\times(p-r)} \end{bmatrix}$.

Assumption 2. *For every complex number $s = z + wi$, where $z \geq 0$, it is true that:*

$$rank \begin{bmatrix} sI - A & E \\ C & \underline{0} \end{bmatrix} = rank(E) + n.$$

From Assumption 2, the following Lemmas are derived:

Lemma 1. *If and only if Assumption 2 is satisfied, then the pair (A_4, C_4) is detectable.*

Proof. See [118] , [119]. □

Lemma 2. *If Assumption 2 is satisfied, then the pair (A_0, C_0) is observable.*

Proof. Refer to Appendix B. □

According to Lemma 2, there is a $L_0 \in \mathbb{R}^{(n+p-2r)\times(p-r)}$ matrix that guarantees the stability of $A_0 - L_0 C_0$.

Assumption 3. *The nonlinear term of system (4.1) is Lipschitz about x with $\mathscr{L}_\phi$ as Lipschitz constant, hence:*

$$\|\phi(x,t)-\phi(\hat{x},t)\| \leq \mathscr{L}_\phi \quad \forall\, x,\hat{x} \in \mathbb{R}^n.$$

Assumption 4. *The attack vector a_m and disturbance vector d are constrained by the known constants $\rho > 0$ and $\xi > 0$ respectively, thus:*

$$\|a_m\| \leq \rho \text{ and } \|d\| \leq \xi.$$

Assumptions 3 and 4 provide a set of inequalities that are necessary for proving the existence of the proposed observers and the stability of their error dynamics, as it will be shown in the following section.

4.3 Observer Design for Attack Detection

The development of SMOs directly for the compromised system (4.1) is not indicated because the impact of the FDIAs on state estimation errors might be affected by the variable structure term [120]. This issue can be addressed by the virtual separation of (4.1) into subsystems (4.6) and (4.7), as mentioned previously in Section 4.2. The theoretical verification of this statement is demonstrated in this section. More specifically, it is mathematically proven that the estimation errors of the observers designed for subsystems (4.6) and (4.7) will converge to zero under attack-free conditions if the necessary requirements are satisfied (Theorem 4.3.1) while the error dynamics (4.16) and (4.17) are only affected by the occurrence of FDIAs. Therefore, these estimation errors are proper indicators (or residuals) of whether the LFC system faces cyberattacks or not.

For subsystem (4.7), the following SMO is designed:

$$\begin{cases} \dot{\hat{g}}_1 = A_1\hat{g}_1 + \bar{A}_2\hat{g}_0 + F_1\phi(G^{-1}\hat{g},t) + B_1u + (A_1 - A_1^s)C_1^{-1}(h_1 - \hat{h}_1) + v_1 \\ \hat{h}_1 = C_1\hat{g}_1 \end{cases} \tag{4.8}$$

where $A_1^s \in \mathbb{R}^{r\times r}$ is a stable matrix that has to be computed and $\hat{g} := col(C_1^{-1}h_1, \hat{g}_2)$. The discontinuous output error injection term v_1 of SMO (4.8) is given by:

$$v_1 = \begin{cases} (\|E_1\|\xi + \eta_1)\frac{P_1(C_1^{-1}h_1 - \hat{g}_1)}{\|P_1(C_1^{-1}h_1 - \hat{g}_1)\|} & \text{if } C_1^{-1}h_1 - \hat{g}_1 \neq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{4.9}$$

where $P_1 > 0 \in \mathbb{R}^{r\times r}$ is a symmetric definite matrix and $\eta_1 > 0$ is a scalar to be calculated.

For subsystem (4.6), the next Luenberger observer with $L_0 \in \mathbb{R}^{(n+p-2r)\times(p-r)}$ gain is developed:

$$\begin{cases} \dot{\hat{g}}_0 = A_0\hat{g}_0 + \bar{A}_3 C_1^{-1} h_1 + \bar{F}_2\phi(G^{-1}\hat{g},t) + B_0 u + L_0(h_3 - \hat{h}_3) \\ \hat{h}_3 = C_0\hat{g}_0. \end{cases} \tag{4.10}$$

After designing observers (4.8) and (4.10), the dynamics of the resulting estimation errors can be obtained. The estimation errors of the developed SMO (4.8) and the proposed Luenberger observer (4.10) are defined as $e_1 = g_1 - \hat{g}_1$ and $e_0 = g_0 - \hat{g}_0$, respectively. The differentiation of the estimation errors provides the error dynamics under attack-free conditions as:

$$\dot{e}_1 = A_1^s e_1 + \bar{A}_2 e_0 + F_1\phi(G^{-1}g,t) - F_1\phi(G^{-1}\hat{g},t) + E_1 d - v_1 \tag{4.11}$$

$$\dot{e}_0 = (A_0 - L_0C_0)e_0 + \bar{F}_2\phi(G^{-1}g,t) - \bar{F}_2\phi(G^{-1}\hat{g},t). \tag{4.12}$$

The existence of the developed observers is proven by the theorem that follows. This theorem paves also the way for determining the values of A_1, P_1, L_0 and η_1.

Theorem 4.3.1. *Let system (4.1) along with Assumptions 1-4. In the absence of cyberattacks, if there are matrices $A_1^s < 0$, L_0, $P_1 = P_1^T > 0 \in \mathbb{R}$ and $P_0 = P_0^T > 0$ and scalars $\alpha_1 > 0$ and $\alpha_0 > 0$ that satisfy the following Inequality:*

$$\Lambda := \begin{bmatrix} A_1^{sT}P_1 + P_1A_1^s + \frac{1}{\alpha_1}P_1P_1 & P_1\bar{A}_2 \\ \bar{A}_2^T P_1 & (A_0 - L_0C_0)^T P_0 + P_0(A_0 - L_0C_0) + \frac{1}{\alpha_0}P_0P_0 + aI_{n+p-2r} \end{bmatrix} < 0, \tag{4.13}$$

where $a = \alpha_1(\|F_1\|\mathscr{L}_\phi\|G^{-1}\|)^2 + \alpha_0(\|F_2\|\mathscr{L}_\phi\|G^{-1}\|)^2$, the error dynamics (4.11) and (4.12) are asymptotically stable.

Proof. Refer to Appendix B. □

Theorem 4.3.1 establishes the necessary conditions for the existence of the designed observers (4.8) and (4.10). Nevertheless, it does not provide a systematic way to calculate the necessary matrices that satisfy Inequality (4.13). This is achieved by using the Schur complement, which converts the problem of determining matrices that satisfy Inequality (4.13) into the next Linear Matrix Inequality (LMI) feasibility problem.

Remark 1. *There are matrices* $A_1^s < 0$, L_0, , $P_0 = P_0^T > 0$ *and* $P_1 = P_1^T > 0$ *and scalars* $\alpha_0 > 0$ *and* $\alpha_1 > 0$ *so that:*

$$\begin{bmatrix} P_1A_1^s + (P_1A_1^s)^T & P_1 & P_1\bar{A}_2 & \underline{0} \\ P_1 & -\alpha_1 I & \underline{0} & \underline{0} \\ \bar{A}_2^T P_1 & \underline{0} & A_0^T P_0 + P_0A_0 - C_0^T(P_0L_0)^T - P_0L_0C_0 + aI & P_0 \\ \underline{0} & \underline{0} & P_0 & -\alpha_0 I \end{bmatrix} < 0. \quad (4.14)$$

After determining the values of A_1^s, L_0, P_1 and P_0, the next step is to configure the parameter η_1 of (4.9) so that the error dynamics are driven to the following sliding surface:

$$\mathscr{S} = \{(e_1, e_0) | e_1 = 0\} \quad (4.15)$$

within a finite time frame and ensure a continuous sliding motion is sustained on $\mathscr{S}$ thereafter. This objective can be accomplished by using the next theorem.

Theorem 4.3.2. *Let system (4.1), Assumptions 1-4 and the observers (4.8) and (4.10). The error dynamics (4.11) and (4.12) can be directed to the sliding surface (4.15) within finite time frame, provided that* η_1 *of (4.9) meets the following condition:*

$$\eta_1 \geq (\|\bar{A}_2\| + \|F_1\|\mathscr{L}_\phi\|T^{-1}\|)\varepsilon + \eta_2,$$

where ε *is the upper limit of* $\|e\|$ *and* $\eta_2 > 0$ *is a scalar, and the LMI feasibility problem (4.14) can be solved.*

Proof. Refer to Appendix B. □

When the LFC system faces FDIAs, the error dynamics (4.11) and (4.12) are converted into the following form:

$$\dot{e}_1 = A_1^s e_1 + \bar{A}_2 e_0 + F_1\phi(G^{-1}g,t) - F_1\phi(G^{-1}\hat{g},t) + E_1 d - v_1 \quad (4.16)$$

$$\dot{e}_0 = (A_0 - L_0C_0)e_0 + \bar{F}_2\phi(G^{-1}g,t) - \bar{F}_2\phi(G^{-1}\hat{g},t) + D_0a_m. \quad (4.17)$$

Eq. (4.17) includes only the attack vector a_m while it does not contain the system disturbances vector d or the discontinuous output error injection term v_1 of the SMO (4.8). The additional D_0a_m term of (4.17) definitely affects the asymptotic stability of e_0, yielding that e_0 is only prone to FDIAs and immune to the system disturbances or the error injection term. Theorem (4.3.1) and the form of e_0 during FDIAs clearly evidence that if the estimation error e_0 has converged to zero, the system is in its normal state; otherwise, if the e_0 is non-zero, the system is under cyberattacks. This property of estimation error e_0 make it a suitable

indicator for identifying FDIAs against LFC, commonly referred to as *residual*. Before selecting the proper form of the residual for the proposed AD strategy, it should be noticed that a_m affects only the last $p-r$ components of e_0, namely $e_{g_3} = g_3 - \hat{g}_3$, since $D_0 = \begin{bmatrix} 0 \\ \hline D_2 \end{bmatrix}$. Consequently, it is a rational choice to use $\|e_{h_3}\| = \|C_0 e_0\| = \|e_{g_3}\|$ as the residual for detecting FDIAs against LFC. The attack detection method that is proposed in the present study can be summarized as:

Proposed Attack Detection Strategy: *Assume that $\|e_{h_3}\|$ is the detection residual, ς_d represents an adaptive threshold, t_d denotes the exact time of the attack detection and t_e is the time elapsed from the previous attack detection until t_d. If $\|e_{h_3}\| \geq \varsigma_d$, then a FDIA is detected at time t_d; otherwise, the system is considered to be in healthy state for the t_e.*

4.4 Observer Design for Attack Localization

The attack detector described in Section 4.3 can only determine whether and when the LFC faces a cyberattack or not. However, it can not identify which LFC signal has been affected by a cyberattack, especially in scenarios of multiple FDIAs. This can be achieved by the proposed attack localization method that is presented in the remainder of this section. The core idea of the introduced ALC scheme is the following: the attack vector can be expressed as $a_m = [a_m^1,\ a_m^2, \ldots\ a_m^q]^T$. After determining whether the attack vector elements $a_m^\iota = 0$ $(\iota = 1, 2, \ldots, q)$ or not, the identification of the compromised state variables can be performed by multiplying a_m with the known attack distribution matrix D. In case of a successful FDIA, an alarm informs the system operator about the location of the attack through a digital logic system.

To this end, a dedicated pair of SMOs is designed for each a_m^ι, forming a bank $2q$ observers. Each pair of SMOs constitutes a bank slot, where its SMO-I is designed to be susceptible to system disturbances and robust against cyberattacks, while its SMO-II is designed to be prone to cyberattacks and resilient against system disturbances. The special feature of SMO-II is that its discontinuous output error injection term v_s^ι is designed to be equivalent to $\bar{a}_m^\iota$, which is the vector of all attack elements except a_m^ι. In this way, the resulting error dynamics of SMO-II can neglect the impact of $\bar{a}_m^\iota$ under specific conditions. As a result, SMO-II is sensitive only to its corresponding attack vector element a_m^ι. This yields that the estimation error of SMO-II deviates from zero only when a_m^ι is non-zero, making it a suitable attack localization residual.

For subsystem (4.7), the following SMO-I is designed to isolate the ιth attack a_m^ι:

$$\begin{cases} \dot{\hat{g}}_1^\iota = A_1\hat{g}_1^\iota + \tilde{A}_2\hat{g}_0^\iota + F_1\phi(G^{-1}\hat{g}^\iota,t) + B_1u \ + (A_1 - A_1^s)C_1^{-1}(h_1^\iota - \hat{h}_1^\iota) + v_1^\iota \\ \hat{h}_1^\iota = C_1\hat{g}_1^\iota \end{cases} \tag{4.18}$$

where $\hat{g}^\iota$ and $\hat{h}^\iota$ represent the estimated state and output vectors obtained by the proposed ALC method, respectively and $\hat{g}^\iota := col(C_1^{-1}h_1, [I_{n-r}\ 0]\hat{g}_0^\iota)$. Regarding the output error injection term v_1^ι and considering $P_1 > 0 \in \mathbb{R}^{r\times r}$ as a symmetric definite matrix, we have:

$$v_1^\iota = \begin{cases} (\|E_1\|\xi + \eta_1)\frac{P_1(C_1^{-1}h_1 - \hat{g}_1^\iota)}{\|P_1(C_1^{-1}h_1 - \hat{g}_1^\iota)\|} & \text{if } C_1^{-1}h_1 - \hat{g}_1^\iota \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The following SMO-II with a L_0 gain is constructed for subsystem (4.6):

$$\begin{cases} \dot{\hat{g}}_0^\iota = A_0\hat{g}_0^\iota + \bar{A}_3C_1^{-1}h_1^\iota + \bar{F}_2\phi(G^{-1}\hat{g}^\iota,t) + B_0u \ + L_0(h_3^\iota - \hat{h}_3^\iota) + \bar{D}_0^\iota v_2^\iota \\ \hat{h}_3^\iota = C_0\hat{g}_0^\iota. \end{cases} \tag{4.19}$$

If D_0 is written as $D_0 = [D_0^1, ..., D_0^q]$, then D_0^ι denotes the ιth column of D_0 and $\bar{D}_0^\iota$ refers to the rest of them. For the discontinuous output error injection term v_2^ι, we have:

$$v_2^\iota = \begin{cases} (\rho + \eta_3)\frac{\bar{F}_0^\iota e_{h_3}^\iota}{\|\bar{F}_0^\iota e_{h_3}^\iota\|} & \text{if } e_{h_3} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $e_{h_3}^\iota = h_3^\iota - \hat{h}_3^\iota$, η_3 is a positive scalar and $F_0 \in \mathbb{R}^{q\times(p-r)}$ is a matrix to be calculated. F_0^ι represents the ιth row of F_0 and $\bar{F}_0^\iota$ denotes the rest of them.

The state estimation errors that result from the aforementioned SMOs for each possible $a_m^\iota \neq 0$ are defined as $e_1^\iota = g_1^\iota - \hat{g}_1^\iota$ and $e_0^\iota = g_0^\iota - \hat{g}_0^\iota$. From the differentiation of e_1 and e_0, their error dynamics after the occurrence of FDIAs can be obtained as:

$$\dot{e}_1^\iota = A_1^s e_1^\iota + \bar{A}_2 e_0^\iota + F_1\phi(G^{-1}g,t) - F_1\phi(G^{-1}\hat{g}^\iota,t) \ + E_1d - v_1^\iota \tag{4.20}$$

$$\begin{aligned} \dot{e}_0^\iota &= (A_0 - L_0C_0)e_0^\iota + \bar{F}_2\phi(G^{-1}g,t) - \bar{F}_2\phi(G^{-1}\hat{g}^\iota,\ t) \ + D_0a_m - \bar{D}_0^\iota v_2^\iota = \\ &= (A_0 - L_0C_0)e_0^\iota + \bar{F}_2\phi(G^{-1}g,t) - \bar{F}_2\phi(G^{-1}\hat{g}^\iota,t) + D_0^\iota a_m^\iota + \bar{D}_0^\iota(\bar{a}_m^\iota - v_2^\iota). \end{aligned} \tag{4.21}$$

By noticing the structure of the error dynamics (4.20) and (4.21), it is possible to establish the stability conditions that are required for achieving the desired observer behavior. These conditions are presented in Theorem 4.4.1 that follows.

Theorem 4.4.1. *Given system (4.1) with Assumptions 1-4. If there are* L_0, $A_1^s < 0$, $P_0 = P_0^T > 0$, $P_1 = P_1^T > 0$ *and* F_0, *and scalars* $\alpha_0 > 0$ *and* $\alpha_1 > 0$ *so that:*

$$D_0^T P_0 = F_0 C_0 \tag{4.22}$$

$$\begin{bmatrix} \Pi_1 + \frac{1}{\alpha_1} P_1 P_1 & P_1 \bar{A}_2 \\ \bar{A}_2^T P_1 & \Pi_0 + \frac{1}{\alpha_0} P_0 P_0 + a I_{n+p-2r} \end{bmatrix} < 0 \tag{4.23}$$

where $\Pi_1 = A_1^{sT} P_1 + P_1 A_1^s$, $\Pi_0 = (A_0 - L_0 C_0)^T P_0 + P_0 (A_0 - L_0 C_0)$ *and* $a = \alpha_1 \mathscr{L}_{\phi_1}^2 \|G^{-1}\|^2 + \alpha_0 \mathscr{L}_{\phi_2}^2 \|G^{-1}\|^2$, *then the state estimation error* e_0^ι *will exponentially converge to zero when* $a_m^\iota = 0$; *when* $a_m^\iota \neq 0$, e_0^ι *satisfies* $\dot{e}_0^\iota = (A_0 - L_0 C_0) e_0^\iota + \bar{F}_2 \phi(G^{-1} g, t) - \bar{F}_2 \phi(G^{-1} \hat{g}^\iota, t) + D_0^\iota a_m^\iota + \bar{D}_0^\iota (\bar{a}_m^\iota - v_2^\iota)$.

Proof. Refer to Appendix B. □

If P_0 has not a particular structure, matrices P_0, F_0 can be determined to satisfy both (4.22) and (4.23) by solving the next LMI optimization problem:

$$\text{minimize } \gamma \text{ s.t.}$$

$$P_0 > 0,\ P_1 > 0 \text{ and}$$

$$\begin{bmatrix} -\gamma I_{n+p-2r} & (D_0^T P_0 - F_0 C_0)^T \\ D_0^T P_0 - F_0 C_0 & -\gamma I_q \end{bmatrix} < 0$$

where $X = P_1 A_1^s$ and $Y_0 = P_0 L_0$.

Theorem 4.4.1 specifies the behavior of the designed observers and also establishes the core concept of the proposed ALC strategy. Particularly, when the proposed AD scheme identifies a FDIA at time step t_d, the proposed AI mechanism is activated to locate the affected signals. For each a_m^ι, two dedicated observers, designed according to (4.18) and (4.19), estimate the state and the output vectors. When $a_m^\iota = 0$, the generated state estimation error e_0^i will tend to zero. Otherwise, if $a_m^\iota \neq 0$, e_0^ι will exceed an adaptive threshold at time step $t_{is} > t_d$. According to this approach, the $\|e_{h_3}^\iota\| = \|C_0 e_0^\iota\|$ is selected as a proper ALC residual and the proposed attack localization scheme is defined as:

Proposed Attack Localization Scheme: *If the selected residual* $\|e_{h_3}^\iota\|$, $(\iota = 1, 2, \ldots, q)$, *exceeds its corresponding adaptive threshold* ς_{is}^ι, *then* $a_m^\iota \neq 0$, *otherwise* $a_m^\iota = 0$. *When the values of every* a_m^ι *have been determined, the affected signals are identified based on the structure of* D_0.

4.5 Threshold Selection

According to the previous analysis, the selected residuals for the proposed AD and ALC methods are non-zero in case of a cyberattack and zero, otherwise. In practice, these residuals can be non-zero due to various reasons, e.g. initial estimation error, approximate linearization error, system nonlinearities, etc. Therefore, if the boundaries of the selected residuals are investigated under attack-free conditions, it is possible to design adaptive thresholds for each power system area in order to minimize the amount of false positive alarms.

Without loss of generality, the adaptive threshold design proposed in this section is described for the AD method and can be easily expanded to the ALC scheme. Let $r(t)$ be the selected residual and ς denote the upper bound of $r(t)$ in normal conditions. From the error dynamics (4.12), it is clear that ς is comprised of the initial estimation error threshold ς_{ie} and the system nonlinearities threshold ς_{nl} as:

$$\varsigma = \varsigma_{ie} + \varsigma_{nl}.$$

Regarding ς_{ie}, it is obvious that:

$$\varsigma_{ie} = \|C_0 e^{(A_0 - L_0 C_0)(t-t_0)} e_0(t_0)\|. \tag{4.24}$$

For ς_{nl}, we consider the residual equation, which is:

$$r(t) = C_0 e_0(t). \tag{4.25}$$

By applying Laplace transformation to (4.12) and (4.25), the following formula is obtained:

$$r(s) = C_0(sI_{n+p-2r} - A_0 + L_0 C_0)^{-1}(\bar{F}_2\phi(G^{-1}g,s) - \bar{F}_2\phi(G^{-1}\hat{g},s)). \tag{4.26}$$

From Eq. (4.26), the upper bound of the residual can be obtained as:

$$\|r(t)\| \leq |sup_{\omega\geq 0} W(C_0(j\omega I_{n+p-2r} - A_0 + L_0 C_0)^{-1}| \times \mathscr{L}_{\phi_2}\|G^{-1}\|\|e_0(t)\|, \tag{4.27}$$

where $W(Z) = |\lambda_{max}(Z^T * Z)|^{1/2}$. Combining (4.24) and (4.27), the adaptive threshold ς can be computed as:

$$\begin{aligned} &\varsigma = \varsigma_{ie} + \varsigma_{nl} \Rightarrow \\ \Rightarrow\ &\varsigma = \|C_0 e^{(A_0 - L_0 C_0)(t-t_0)} e_0(t_0)\| + \\ &\quad + |sup_{\omega\geq 0} W(C_0(j\omega I_{n+p-2r} - A_0 + L_0 C_0)^{-1}| \times \mathscr{L}_{\phi_2}\|G^{-1}\|\|e_0(t)\|. \end{aligned} \tag{4.28}$$

Using Eq. (4.28), the adaptive AD threshold can be computed over time and compared with the AD residual to determine whether the system under investigation is under attack or not. The same process is followed for the computation of the ALC adaptive thresholds.

To make the present work more comprehensible, the proposed FDIA detection and localization method is briefly illustrated in Algorithm 1.

Algorithm 1 Algorithm of the proposed attack detection and localization methodology

Require:
- $rank(CE) = rank(E)$
- $rank\begin{bmatrix} sI-A & E \\ C & \underline{0} \end{bmatrix} = n + rank(E),\ \forall s \in \mathbb{C}$ with nonnegative real part
- $\|a_m\| \leq \rho$ and $\|d\| \leq \xi$
- $\|\phi(x,t) - \phi(\hat{x},t)\| \leq \mathscr{L}_\phi \|x - \hat{x}\|, \quad \forall x, \hat{x} \in \mathbb{R}^n$

Ensure: System matrices A, B, C, E, F and D are known

$t \leftarrow 0$; ▷ time initialization
while $t \geq 0$ **do** ▷ Proposed mechanism is enabled
  Compute system output $y(t) = Cx(t)$;
  Compute transformations $h(t) = Hy(t)$ & $g(t) = Gx(t)$;
  Provide $h(t)$ and system input $u(t)$ to the AD SMOs and ALC SMOs;
  Compute AD residual $\|e_{h_3}\| = \|h_3 - \hat{g}_3\|$;
  Compute ALC residuals $\|e^1_{h_3}\|$, $\|e^2_{h_3}\|$, ... , $\|e^N_{h_3}\|$; ▷ bank of SMOs
  Compute AD adaptive threshold ς_d;
  Compute ALC adaptive thresholds ς^1_{is}, ς^2_{is}, ... , ς^N_{is};
  if $\|e_{h_3}\| > \varsigma_d$ **then**
    System is under attack;
    if $\|e^1_{h_3}\| > \varsigma^1_{is}$ or $\|e^2_{h_3}\| > \varsigma^2_{is}$ or ... $\|e^N_{h_3}\| > \varsigma^N_{is}$ **then**
      Signal of area $i \in \{1, 2, \ldots, N\}$ is under attack;
    end if
  else
    System is in normal condition;
  end if
  $t \leftarrow t + 1$; ▷ Next time step
end while ▷ Proposed mechanism is disabled

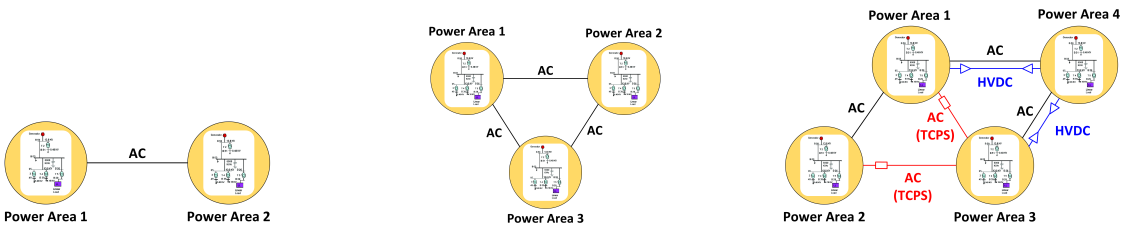
4.6 Experimental Results

This section includes the results from the experiments that were conducted to evaluate the performance of the introduced AD and ALC approaches. Particularly, the section starts with a detailed description of the implemented testbeds and then, the performance assessment of the presented methodologies follow. For further validation, the robustness of the proposed methods against various system parameters is investigated and along with a comparative analysis against other, related works from the literature to highlight its superior points. For the simulations, the MATLAB/Simulink platform was utilized on a desktop computer, equipped with a 64-bit Intel Core i7 CPU of 2.7 GHz.

4.6.1 Use case analysis

For testing the proposed AD and ALC techniques, a series of case studies is implemented. These case studies include various LFC systems of growing complexity, subjected to diverse types of FDIAs and disturbances. In this way, the effectiveness and scalability (as defined in Section 1.3, it is the effective application of a methodology to various power systems, regardless of their size) of the introduced methodologies are demonstrated. The implemented case studies are:

- **Case Study 1:** 2-area power system, interconnected via an AC tie-line, where 1% p.u. step load disturbances occur in both areas at $t = 4$ sec. A 1‰ p.u step FDIA is considered against the control signal of area 1 at $t = 10$ sec and a time delay attack of 1 second is launched against the control signal of area 2 at $t = 20$ sec;
- **Case Study 2:** 3-area power system, interconnected via AC tie-lines where a 1‰ p.u. step load disturbance occurs in area 1 at $t = 3$ sec, a 10% p.u. step load disturbance occurs in area 2 at $t = 11$ sec and a 1% p.u. step load disturbance occurs in area 3 at $t = 20$ sec. Regarding the cyberattacks, a 1‰ p.u step FDIA is considered against the frequency measurement of area 2 at $t = 5$ sec, followed by a DoS attack against the frequency measurement of area 3 at $t = 13$ sec;
- **Case Study 3:** 4-area power system, connected via AC (either equipped with TCPS or not) and HVDC tie-lines, where a 1% p.u. step load disturbance occurs in area 1 at $t = 10$ sec, a 1% p.u. step load disturbance occurs in area 2 at $t = 20$ sec, a 5‰ p.u. step load disturbance occurs in area 3 at $t = 20$ sec and a 1% p.u. step load disturbance occurs in area 4 at $t = 35$ sec. The FDIAs considered in this case study are a 1‰ p.u. step attack in area 2 at $t = 33$ sec and a 1‰ p.u. step attack in area 3 at $t = 17$ sec against the $\int$ACE signals.



(a) Case study 1 (b) Case study 2 (c) Case study 3

Figure 4.1 Use case topologies for evaluation of the proposed AD and ALC schemes.

In each case study, solar and wind power disturbances occur throughout the simulation, along with measurement time delays that range from 1 to 2 seconds. The topology, i.e.

number of power areas and types of the interconnections, of each case study is depicted in Fig. 4.1. The power system parameters of each area are given in the Appendix C. The topology complexity of each case study $k \in \{1,2,3\}$ is greater than its previous one(s) to verify the scalability of the proposed AD and ALC methods. Finally, to guarantee the normal operation of the implemented LFC systems, the frequency response of selected areas under 10% load disturbance for each case study is illustrated in Fig. 4.2.

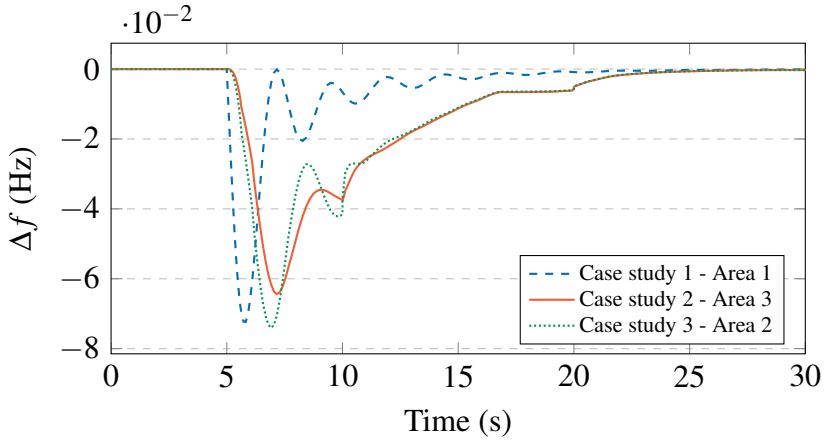


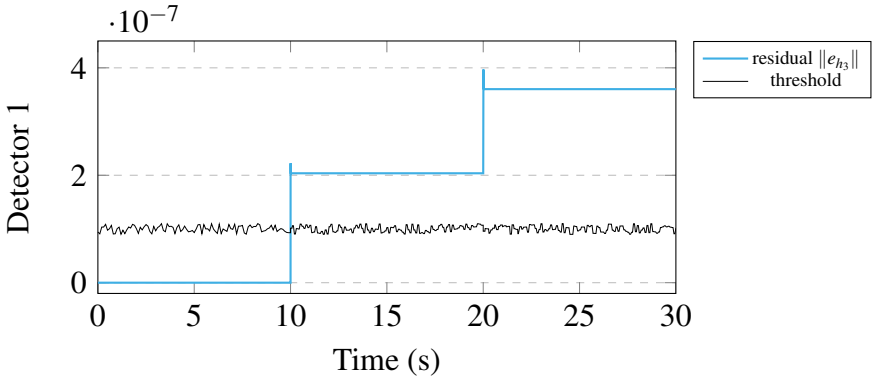
Figure 4.2 Frequency response of selected power areas for each AD & ALC case study.

4.6.2 Performance analysis

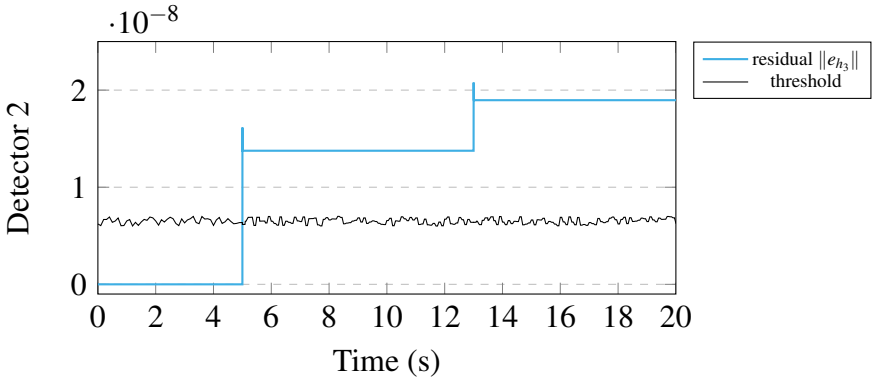
The developed attack detector is applied to the case studies described in 4.6.1 for the evaluation of its performance. The results of this experiment are illustrated in Fig. 4.3. According to this figure, the output of the attack detector (blue lines in Fig. 4.3) exceeds the designed adaptive threshold (black lines in Fig. 4.3) almost immediately after the launch of a cyberattack in every case study. This indicates that all of the simulated cyberattacks have been successfully identified in a real-time manner. The results confirm the effectiveness of the particular defense mechanism and its scalability over various power systems. Moreover, the output of the attack detector does not deviate from zero before the occurrence of any attack, which implies that the introduced defense scheme is not prone to false positive alarms.

The experiments also demonstrate that the proposed AD scheme can accurately distinguish cyberattacks from other types of disturbances, e.g. load and RES. More specifically, Fig. 4.3 illustrates that the attack detector is triggered in the event of a cyberattack, while being insensitive to the load disturbances that occur at $t = 4$ sec or to the RES disturbances that happen throughout the simulation. This is a critical feature for power system operators in order to design real-world cyberattack detectors, that other efforts [121], [122] do not consider. Fig. 4.3 also shows the benefits of the adaptive threshold over predefined ones. The

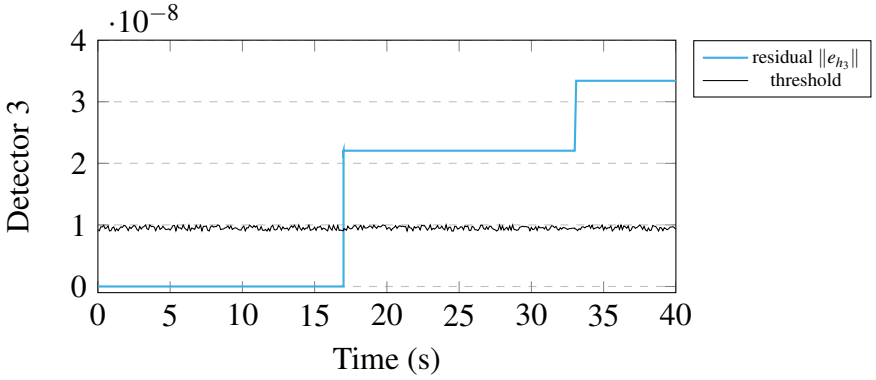
detection thresholds of each case study are dynamically evolved over time and automatically adjusted, based on the sensitivity of every system to cyberattacks. In this way, the selection of the detection thresholds is not based on arbitrary assumptions, leading to the minimization of the false positive attack alarms.



(a) Case Study 1



(b) Case Study 2



(c) Case Study 3

Figure 4.3 Performance of the proposed AD scheme.

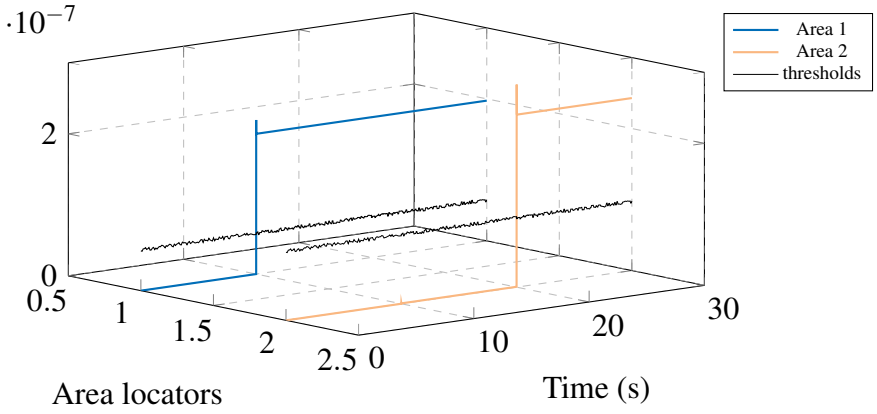
Similarly to the attack detector, the developed attack locator is applied to the case studies described in 4.6.1 for the assessment of its performance. Fig. 4.4 depicts the results of this experiment, which highlight the effectiveness and the scalability of the proposed cyber defense layer. More specifically, the sub-observer designed for $\int \text{ACE}_1$ in case study 1 is activated when the FDIA against the specific signal occurs at $t_1 = 10s$, based on Fig. 4.4a. On the other hand, when the cyberattack against $\int \text{ACE}_2$ takes place at $t_2 = 20s$, the output of the aforementioned sub-observer does not deviate, verifying that it is only sensitive to attacks against $\int \text{ACE}_1$, as designed. Similar is the case for the sub-observer of $\int \text{ACE}_2$, depicted in Fig. 4.4a. The same conclusions can be drawn for the rest of the case studies, based on Fig. 4.4b and 4.4c.

An interesting remark is that the residuals of the sub-observers neither deviate in the event of a load disturbance, e.g. at $t = 4s$ in case 1, nor during RES disturbances that happen throughout the simulation. Thus, similarly to the designed AD method, the introduced ALC scheme is insensitive to load and RES disturbances and can successfully distinguish whether an event is a cyberattack or not. It is also important to mention that the locator of area 1 for case study 2 and the locators of area 1 and 4 for case study 3 do not fluctuate, according to Fig. 4.4b and 4.4b, respectively. This is reasonable considering that these areas do not suffer from cyberattacks and indicates that the designed ALC scheme is not prone to false positive alarms. Finally, the selected localization thresholds are adaptive, following the design that is described in Section 4.5; the benefits of this feature are already discussed in the present section.

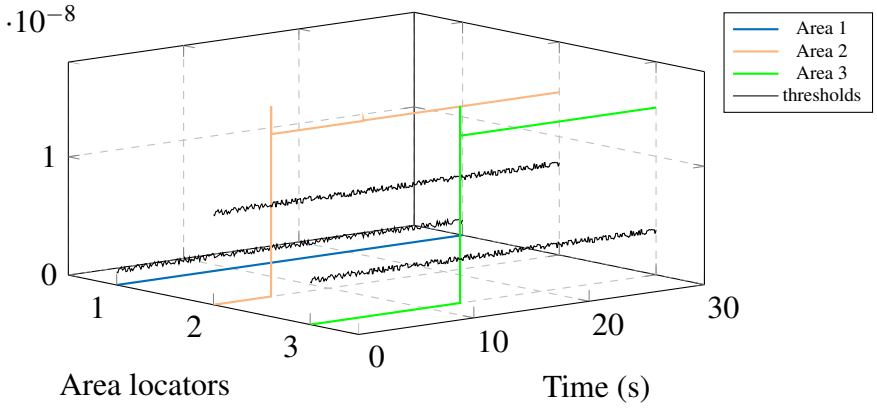
4.6.3 Sensitivity analysis on power system parameters

To effectively capture the behavior of LFC system, its modeling requires the accurate computation of several power system parameters, such as turbine and governor time constants, tie-line synchronizing coefficients, load frequency relief, regional inertia constants, etc. However, this is a challenging task due to the necessary model linearizations, system approximations or other simplifications that must be taken into account. Therefore, the resulting matrix values of the designed observers might deviate from the actual values of the system parameters. In this context, it is necessary to investigate the sensitivity of the proposed AD & ALC techniques to these parameters in order to validate their robustness and their applicability to realistic conditions.

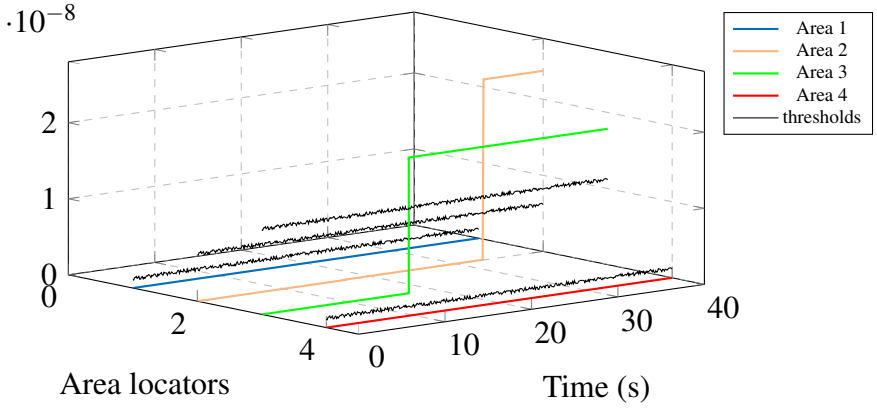
To perform the particular sensitivity analysis, the proposed AD & ALC schemes are consecutively applied to case study 1 (without loss of generality) for different system parameter values in each iteration. For this purpose, the following scenarios are investigated:



(a) Case Study 1



(b) Case Study 2



(c) Case Study 3

Figure 4.4 Performance of the proposed ALC scheme.

- a 10%, 20%, and 30% increase in the turbine and governor time constants of areas 1 and 2 ($T_{t_1}, T_{g_1}, T_{g_2}, T_{t_2}$, respectively),

- a 10%, 20%, and 30% decrease in the turbine and governor time constants of areas 1 and 2 ($T_{t_1}, T_{g_1}, T_{g_2}, T_{t_2}$, respectively),
- a 10%, 30% and 50% increase in the tie-line synchronizing coefficient between areas 1 and 2 (T_{12}),
- a 10%, 30% and 50% decrease in the tie-line synchronizing coefficient between areas 1 and 2 (T_{12}),

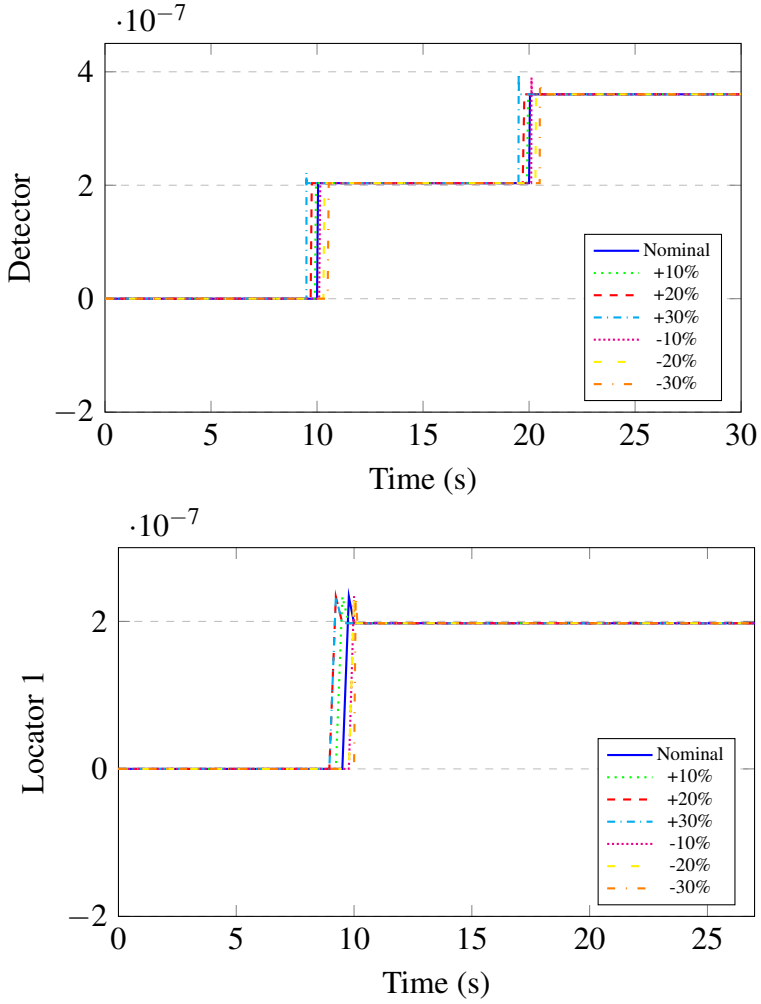


Figure 4.5 Sensitivity analysis of the proposed AD & ALC schemes on T_{g_i} and T_{t_i}, $i = 1, 2$.

The results of this sensitivity analysis are illustrated in Fig. 4.5 and 4.6, where the outputs of the attack detectors and locators for different system parameter values are depicted. It is apparent that despite the changes in the power system parameters, the effectiveness of the proposed method and schemes is preserved; the outputs of the designed AD & ALC schemes

for the modified parameters demonstrate negligible deviations from the corresponding output for the nominal parameters. Therefore, it can be safely concluded that the suggested AD & ALC methods are robust against uncertainties in power system parameters.

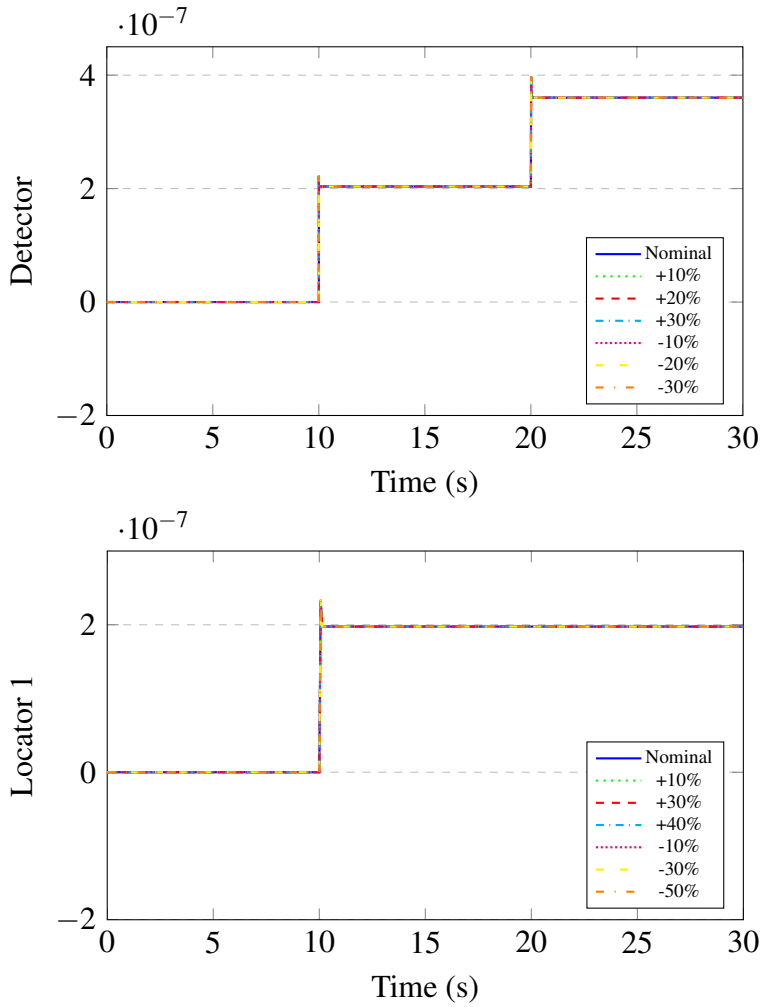


Figure 4.6 Sensitivity analysis of the proposed AD & ALC schemes on T_{12}.

4.6.4 Sensitivity analysis on noisy measurements

In real-world power systems, the noise of instruments and telemetry channels interferes with the grid measurements and corrupts the actual information [123]. To further verify the feasibility of the presented AD & ALC methods in realistic conditions, it is necessary to investigate their robustness against noisy measurements. For this numerical assessment, the introduced AD & ALC methods are applied to a noiseless and noisy setting of case study 1 (without loss of generality). Then, their outputs in each environment are obtained and

compared, similar to [124]. If the performance of the proposed cyber defense layers is similar in both settings, then they are assumed to be robust against noisy measurements.

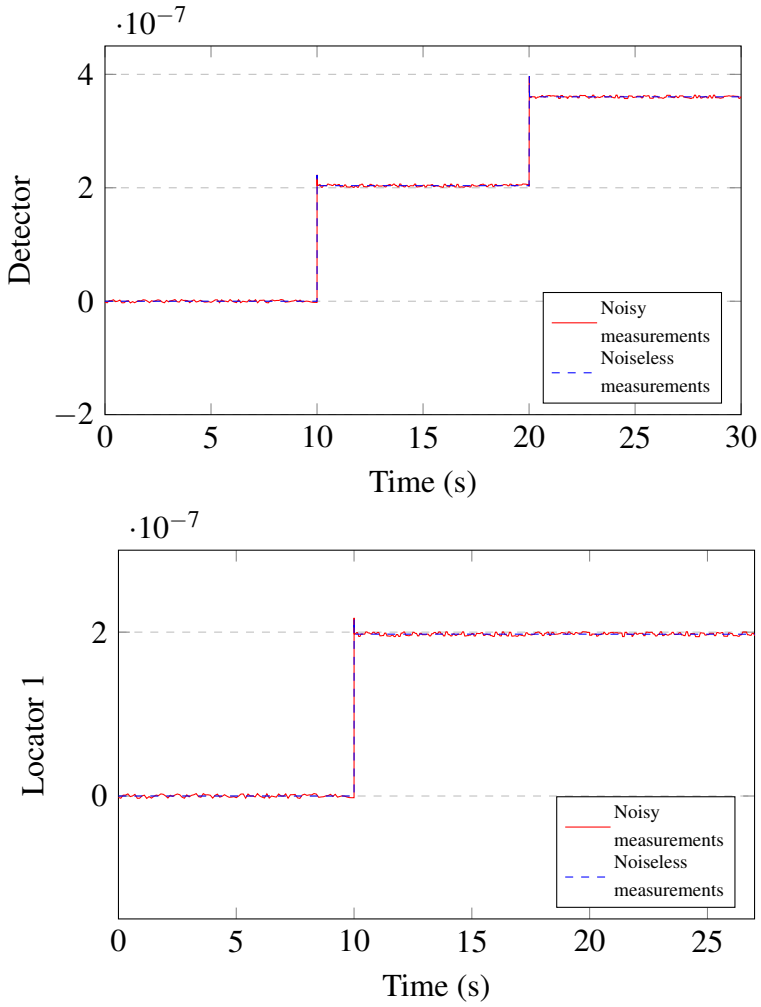


Figure 4.7 Sensitivity analysis of the proposed AD & ALC schemes against noisy measurements.

For this analysis, each y_m^{κ} element ($\kappa = 1, 2, ..., p$) of the case study 1 measurement vector y_m is modeled as:

$$y_m^{\kappa} = y_a^{\kappa} + e_{ns}^{\kappa}$$

where y_a^{κ} represents the actual value of the κ-th measured variable and e_{ns}^{κ} denotes its measurement noise. It is assumed that each e_{ns}^{κ} follows the normal distribution with a mean value of μ_{κ} and a standard deviation of δ_{κ}. For the noiseless environment, it is considered that $\mu_{\kappa} = 0$ and $\delta_{\kappa} = 0\%$ and for the noisy environment, the values of $\mu_{\kappa} = 0$ and $\delta_{\kappa} = 1\%$ are selected, $\forall\ 1 \leq \kappa \leq p$.

Fig. 4.7 illustrates the outputs of the introduced AD & ALC schemes when they are applied to case study 1 with noiseless and noisy measurements. The performance of the AD & ALC scheme in the noiseless setting (blue lines in Fig. 4.7) is the ideal one, as expected. Regarding the noisy environment, the outputs of the attack detector and the attack locator (red lines in Fig. 4.7) slightly deviate from the ideal curves. However, these fluctuations are very small and practically negligible (with approximately 0.1% order of magnitude). Therefore, the robustness of the proposed AD & ALC methods against noisy measurements is ensured.

4.6.5 Comparative study

In this subsection, a comparative study is performed in order to identify the innovations of the proposed cyberattack detection and localization methods in comparison to other relevant techniques from the literature. The first part of this analysis investigates whether an attack detection method for LFC meets a necessary set of quality features or not. This quality comparison is illustrated in Table 4.1, where "✓" denotes that the specific feature is considered in the corresponding method and "×" means that it does not. The selected quality features are the following:

1. **Localization**: the attribute of a method to locate the signals of the system that have been affected by cyberattacks;
2. **Decoupling**: the property of a method to distinguish attacks from other types of external disturbances;
3. **Adaptive threshold**: determines if a method uses adaptive detection thresholds or not;
4. **Nonlinearities**: indicates whether a method considers LFC nonlinearities or not;
5. **Diverse tie-lines**: determines whether a method includes diverse types of tie-lines among power system areas or not;
6. **RES**: states if a method includes RES disturbances or not;
7. **Parameter uncertainties**: indicates if a method is sensitive to power system parameter uncertainties;
8. **Noisy measurements**: determines if a method is robust against noisy measurements;
9. **Scalability**: the attribute of a method to be applicable to various systems irrespective of their size.

Table 4.1 Quality comparative analysis of the proposed AD scheme.

Features \ Methods	[125]	[21]	[26]	[20], [126]	**Proposed**
Localization	×	×	×	✓	✓
Decoupling	✓	×	×	✓	✓
Adaptive threshold	×	×	×	×	✓
Nonlinearities	×	×	×	×	✓
Diverse tie-lines	×	×	×	×	✓
RES	×	×	×	×	✓
Parameter uncertainties	×	×	×	×	✓
Noisy measurements	✓	×	×	×	✓
Scalability	✓	×	×	×	✓

There are several interesting conclusions drawn from Table 4.1. Firstly, while [125], [21] and [26] are able to successfully identify cyberattacks, they cannot locate which signals have been affected, like the presented filter does. Also, the proposed method is capable of handling the LFC nonlinearities in both theory and practice, while the rest of the techniques disregard them. Likewise, the introduced AD & ALC methods can successfully decouple cyberattacks from disturbances due to RES and load variations, contrary to [21] and [26]. Regarding the attack detection residuals, the presented approach uses adaptive thresholds, while the compared techniques employ predefined ones. Furthermore, the robustness of the proposed methodology against power system parameter uncertainties and noisy measurements is verified, unlike the rest of the methods. For the experiments of the suggested method, several realistic features of power grids are considered, such as diverse types of tie-lines (HVDC, TCPS-equipped) and RES generation, in contrast with the other compared works. Moreover, the stability of the proposed algorithm is mathematically established, unlike data-driven [125] and statistical [21] methods. Finally, the presented approach has been applied to different power systems of varying complexity to ensure that is scalable, while [21], [26], [20] and [126] have not.

The second part of this analysis involves a quantity comparison where the proposed AD & ALC mechanisms for LFC are compared to other sophisticated research methods based on selected detection metrics. The difference from the first part of this analysis is that the metrics of the quantity comparison can be measured, unlike the features of the quality comparative

analysis. This quantity comparison is illustrated in Table 4.2 and the selected metrics are the following:

1. **Detection time**: the elapsed time between the launch and the identification of the attack;
2. **Precision**: the percentage of the total triggered alarms that are actual cyberattacks;
3. **Recall**: the percentage of the total cyberattacks that are actually identified.

Table 4.2 Quantity comparative analysis of the proposed AD scheme.

Metrics \ **Methods**	[73]	[127]	**Proposed**
Detection time (sec)	0.058	1.006	0.012
Precision	100%	86%	100%
Recall	100%	79%	100%

For the quantity comparison, the attack detection methods [73], [127] have been implemented and applied to the case study 1 (without loss of generality) described in Section 4.6.1. For simplicity, only the step FDIA at $t = 10$ sec has been considered. Precision and recall metrics are measured by running multiple simulations of case study 1. From Table 4.2, it is concluded that the proposed work can detect cyberattacks faster than the rest of the compared techniques. Moreover, the introduced defense method and [73] can identify all the simulated attacks without any false positive alarms, unlike [127]; this is a reasonable outcome for model-based AD approaches compared to the data-driven ones. The above discussion highlights the merit of the proposed technique over other related works and therefore, it verifies its research contribution.

Chapter 5

SMO-based Attack Estimation & Attack-resilient LFC

In this chapter, the basic principles of the proposed observer-based attack estimation & attack-resilient control mechanism for the frequency control of power systems are presented. As mentioned in Section 1.3, AE methods provide an approximation of cyberattacks against LFC while ARC approaches can mitigate their impact against this system. In the remainder of this section, the model of LFC during FDIAs is initially introduced, upon which the proposed AE & ARC methodologies are established. Then, the design procedure of the proposed observers is developed. The attack estimation formulas that are derived from the introduced observers are demonstrated in the next section. To verify the effectiveness of the designed AE scheme, a discussion about the experimental results follows Finally, the ARC mechanism based on the suggested AE method is presented along with its stability analysis, which proves its existence. Similar to the others cybersecurity techniques, the introduced ARC mechanism is accompanied by the experiments that validate its performance.

5.1 LFC modeling under FDIAs

As already mentioned, the development of cyber defense methods that are based on observers requires a compact algebraic-differential model that takes cyber threats into account. To this end, the original state-space representation (3.34) of LFC is re-modeled considering FDIAs against the measurements and control signals, as follows:

$$\begin{cases} \dot{x}(t) = Ax(t) + F\phi(x,t) + B\big(u(t) + a_c(t)\big) + Ed(t) \\ y(t) = Cx(t) + Da_m(t) \end{cases} \tag{5.1}$$

where $a_c(t) \in \mathbb{R}^m$ represents the vector of attacks against control signals, $a_m(t) \in \mathbb{R}^q$ is the vector of attacks against measurements and $D \in \mathbb{R}^{p \times q}$. Without loss of generality, it is assumed that the FDIAs against the control signals target every input channel, hence $u(t)$ and $a_c(t)$ share the same matrix B. Moreover, $p - m \geq q$ while B and D have full column rank.

At this point, it should be noted that the measurement and control signal attacks are represented by two distinct terms $a_c(t)$ and $a_m(t)$, respectively. This unified framework is a more generalized modeling of LFC under FDIAs, compared to the corresponding representation demonstrated in Section 4.1. While both of these models effectively capture the impact of cyberattacks against measurements and control signals, the more specialized representation of this section may lead to less misinterpretations regarding the location of the attack. To include both of these representations in this study for the sake of completeness, each of them is used for a different part of the proposed cyber defense strategy.

5.2 Observer Design Preliminaries

The introduced attack estimation and attack-resilient control methods are based on a pair of specially designed observers. To guarantee the existence of these observers, a solid mathematical foundation has to be established before their design process [117]. This mathematical background is composed of a series of assumptions which are demonstrated in what follows. The results from these assumptions are briefly analyzed to shed more insight into the observer design process.

Assumption 5. $rank(B) = rank(CB)$.

Assumption 5 is required to virtually split system (5.1) into *subsystem-I* and *subsystem-II*; *subsystem-I* is susceptible to the FDIAs against control signals but free from measurement FDIAs and *subsystem-II* is prone to FDIAs against measurements but free from control signal FDIAs. This partition facilitates the observer design process, as it separates the original system into simpler, equivalent subsystems with less terms. Each of these subsystems is dedicated to a particular type of cyberattacks, either measurement or control signal. Formally, Assumption 5 is the necessary condition for the existence of the following state and output transformations:

$$\zeta = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} = Tx = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} x \text{ and } \omega = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} = Sy = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} y,$$

respectively, where $T \in \mathbb{R}^{n\times n}$, $S \in \mathbb{R}^{p\times p}$, $T_1 \in \mathbb{R}^{m\times n}$, $S_1 \in \mathbb{R}^{m\times p}$, $\zeta_1 \in \mathbb{R}^m$ and $\omega_1 \in \mathbb{R}^m$. After applying the above transformation technique to (5.1), it is acquired:

$$\begin{cases} \dot{x} = Ax + F\phi(x,t) + B\big(u + a_c\big) + Ed \\ y = Cx + Da_m \end{cases} \xrightarrow[\text{mult. by S}]{\text{mult. by T}}$$

$$\Rightarrow \begin{cases} T\dot{x} = TAx + T\big(F\phi(x,t) + Bu + Ba_c + Ed\big) \\ Sy = SCx + SDa_m \end{cases} \xrightarrow[y=S^{-1}\omega]{x=T^{-1}\zeta}$$

$$\Rightarrow \begin{cases} \dot{\zeta} = TAT^{-1}\zeta + TF\phi(T^{-1}\zeta,t) + TBu + TBa_c + TEd \\ \omega = SCT^{-1}\zeta + SDa_m \end{cases}$$

and the new system matrices are:

$$TAT^{-1} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, TB = \begin{bmatrix} B_1 \\ \underline{0} \end{bmatrix}, TE = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}, SCT^{-1} = \begin{bmatrix} C_1 & \underline{0} \\ \underline{0} & C_4 \end{bmatrix}, TF = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}$$

$$\text{and } SD = \begin{bmatrix} \underline{0} \\ D_2 \end{bmatrix},$$

where $A_1 \in \mathbb{R}^{m\times m}, A_4 \in \mathbb{R}^{(n-m)\times(n-m)}, B_1 \in \mathbb{R}^{m\times m}, E_1 \in \mathbb{R}^{m\times r}, F_1 \in \mathbb{R}^{m\times \iota}, C_1 \in \mathbb{R}^{m\times m}, C_4 \in \mathbb{R}^{(p-m)\times(n-m)}$ and $D_2 \in \mathbb{R}^{(p-m)\times q}$. C_1 and B_1 are invertible.

The newly transformed system can be separated into the next two virtual subsystems:

$$\begin{cases} \dot{\zeta}_1 = A_1\zeta_1 + A_2\zeta_2 + F_1\phi(T^{-1}\zeta,t) + B_1(u + a_c) + E_1 d \\ \omega_1 = C_1\zeta_1 \end{cases} \tag{5.2}$$

$$\begin{cases} \dot{\zeta}_2 = A_3\zeta_1 + A_4\zeta_2 + F_2\phi(T^{-1}\zeta,t) + E_2 d \\ \omega_2 = C_4\zeta_2 + D_2 a_m. \end{cases} \tag{5.3}$$

By considering measurement attacks a_m as auxiliary states, the augmented form of subsystem (5.3) is obtained as:

$$\begin{cases} \dot{\bar{\zeta}}_2 = \bar{A}_3\zeta_1 + \bar{A}_4\bar{\zeta}_2 + \bar{F}_2\phi(T^{-1}\zeta,t) + \bar{E}_2 d + \bar{E}\dot{a}_m \\ \omega_2 = \bar{C}_4\bar{\zeta}_2 \end{cases} \tag{5.4}$$

where $\bar{\zeta}_2 = \begin{bmatrix} \zeta_2 \\ a_m \end{bmatrix} \in \mathbb{R}^{n+q-m}$, $\bar{A}_4 = \begin{bmatrix} A_4 & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix} \in \mathbb{R}^{(n+q-m)\times(n+q-m)}$, $\bar{A}_3 = \begin{bmatrix} A_3 \\ \underline{0} \end{bmatrix} \in \mathbb{R}^{(n+q-m)\times m}$, $\bar{F}_2 = \begin{bmatrix} F_2 \\ \underline{0} \end{bmatrix} \in \mathbb{R}^{(n+q-m)\times \iota}$, $\bar{E}_2 = \begin{bmatrix} E_2 \\ \underline{0} \end{bmatrix} \in \mathbb{R}^{(n+q-m)\times r}$, $\bar{E} = \begin{bmatrix} \underline{0} \\ I_q \end{bmatrix} \in \mathbb{R}^{(n+q-m)\times q}$ and $\bar{C}_4 = \begin{bmatrix} C_4 & D_2 \end{bmatrix} \in \mathbb{R}^{(p-m)\times(n+q-m)}$.

In a similar way, system (5.2) can be expressed as:

$$\begin{cases} \dot{\zeta}_1 = A_1\zeta_1 + \bar{A}_2\bar{\zeta}_2 + F_1\phi(T^{-1}\zeta, t) + B_1(u + a_c) + E_1 d \\ \omega_1 = C_1\zeta_1, \end{cases} \tag{5.5}$$

where $\bar{A}_2 = \begin{bmatrix} A_2 & \underline{0} \end{bmatrix}$.

Assumption 6. *The nonlinear term of system (5.1) is a Lipschitz continuous function about x, with a Lipschitz constant of $\mathscr{L}_\phi$. Formally:*

$$\|\phi(x,t) - \phi(\hat{x},t)\| \leq \mathscr{L}_\phi \|x - \hat{x}\| \quad \forall\, x, \hat{x} \in \mathbb{R}^n.$$

Assumption 7. *The control signal FDIA vector a_c and the disturbance vector d are bounded by the known, positive constants ρ and ξ, respectively, as $\|a_c\| \leq \rho$ and $\|d\| \leq \xi$. Furthermore, the first derivative of measurement FDIA $\dot{a}_m$ exists and $\dot{a}_m \in \mathscr{L}_2[0,\infty)$.*

Assumptions 6 and 7 are the necessary conditions to prove that the error dynamics of the proposed observers are asymptotically stable. Particularly, the boundedness of different terms, e.g. nonlinearities cyberattacks, etc., provides the system designer with useful inequalities which lead to the proof that the derivative of the selected Lyapunov function is negative, as shown in the next section.

5.3 Observer Design Process

According to Section 2.1, the purpose of the observers is to produce an estimation of the state of the system that they are designed for. Mathematically, the observer of a system exists if the error between the actual state vector and the estimated state vector, namely the estimation error, converges to zero. The goal of this work is to design observers for LFC in such a way so that each estimation error is sensitive only to a single attack vector. Therefore, when the estimation error converges to zero, a set of formulas is derived that can approximate the related attack vectors. The rest of this section is dedicated to the development process of the proposed observers.

The SMO described by Eq. (5.6) is designed for subsystem (5.5) to estimate ζ_1 and ω_1 as $\hat{\zeta}_1$ and $\hat{\omega}_1$, respectively:

$$\begin{cases} \dot{\hat{\zeta}}_1 = A_1\hat{\zeta}_1 + \bar{A}_2\hat{\bar{\zeta}}_2 + F_1\phi(T^{-1}\hat{\zeta},t) + B_1(u+v) \ + (A_1 - A_1^s)C_1^{-1}(\omega_1 - \hat{\omega}_1) + \\ \quad + \frac{1}{2}\hat{k}_1F_1F_1^TP_1C_1^{-1}(\omega_1 - \hat{\omega}_1) \\ \hat{\omega}_1 = C_1\hat{\zeta}_1, \end{cases} \tag{5.6}$$

where $A_1^s \in \mathbb{R}^{m\times m}$ is a Hurwiz matrix to be calculated, $P_1 \in \mathbb{R}^{m\times m}$ is the definite symmetric Lyapunov matrix of A_1^s and $\hat{\zeta} := col(C_1^{-1}S_1y, \begin{bmatrix} I_{n-m} & 0 \end{bmatrix} \hat{\bar{\zeta}}_2)$. The estimated $\bar{\zeta}_2$, denoted as $\hat{\bar{\zeta}}_2$, will be determined by observer (5.8). Regarding $\hat{k}_1$, the following adaptation law $\dot{\hat{k}}_1 = l_{k_1}\|F_1^TP_1(C_1^{-1}\omega_1 - \hat{\zeta}_1)\|^2$ is satisfied, where l_{k_1} represents a positive scalar. For the discontinuous output error injection term v, we have:

$$v = \begin{cases} (\rho+\eta)\frac{B_1^TP_1(C_1^{-1}\omega_1 - \hat{\zeta}_1)}{\|B_1^TP_1(C_1^{-1}\omega_1 - \hat{\zeta}_1)\|} & \text{if } C_1^{-1}\omega_1 - \hat{\zeta}_1 \neq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{5.7}$$

where η is a positive scalar to be calculated.

For subsystem (5.4), the UIO described by Eq. (5.8) is constructed to estimate ζ_2 and ω_2 as $\hat{\zeta}_2$ and $\hat{\omega}_2$, respectively:

$$\begin{cases} \dot{h} = F_0h + M_0\bar{F}_2\phi(T^{-1}\hat{\zeta},t) + L_0\omega_2 + M_0\bar{A}_3C_1^{-1}\omega_1 \ + \frac{1}{2}\hat{k}_2M_0\bar{F}_2H_0(\omega_2 - \hat{\omega}_2) \\ \hat{\bar{\zeta}}_2 = h + N_0\omega_2 \\ \hat{\omega}_2 = \bar{C}_4\hat{\bar{\zeta}}_2. \end{cases} \tag{5.8}$$

where $h \in \mathbb{R}^{n+q-m}$ is the middle variable, $N_0 \in \mathbb{R}^{(n+q-m)\times(p-m)}$, $H_0 \in \mathbb{R}^{\iota\times(p-m)}$, $M_0 \in \mathbb{R}^{(n+q-m)\times(n+q-m)}$, $L_0 \in \mathbb{R}^{(n+q-m)\times(p-m)}$ and $F_0 \in \mathbb{R}^{(n+q-m)\times(n+q-m)}$ are matrices to be computed. Regarding $\hat{k}_2$, the $\dot{\hat{k}}_2 = l_{k_2}\|H_0(\omega_2 - \hat{\omega}_2)\|^2$ adaptation law is satisfied, where l_{k_2} denotes a positive scalar.

After designing the proposed observers (5.6) and (5.8), the estimation errors and their dynamics can be obtained. Let $e_1 = \zeta_1 - \hat{\zeta}_1$ and $\bar{e}_2 = \bar{\zeta}_2 - \hat{\bar{\zeta}}_2$ be the estimation errors generated by the observers (5.6) and (5.8), respectively. The error dynamics are modeled as first order differential equations between the estimation errors. By differentiating (5.8), it

follows that:

$$\begin{aligned}\dot{\hat{\bar{\zeta}}}_2 &= \dot{h} + N_0\dot{\omega}_2 = \\ &= F_0\hat{\bar{\zeta}}_2 + (L_0\bar{C}_4 + N_0\bar{C}_4\bar{A}_4 - F_0N_0\bar{C}_4)\bar{\zeta}_2 + M_0\bar{F}_2\phi(T^{-1}\hat{\zeta},t) + N_0\bar{C}_4\bar{F}_2\phi(T^{-1}\zeta,t) + \\ &\quad + (M_0 + N_0\bar{C}_4)\bar{A}_3\zeta_1 + N_0\bar{C}_4\bar{E}_2 d + N_0\bar{C}_4\bar{E}\dot{a}_m + \frac{1}{2}\hat{k}_2M_0\bar{F}_2H_0\bar{C}_4(\omega_2 - \hat{\omega}_2).\end{aligned}$$

Then, the error dynamics after the occurrence of cyberattacks are:

$$\dot{e}_1 = A_1^s e_1 + \bar{A}_2\bar{e}_2 + F_1\big(\phi(T^{-1}\zeta,t) - \phi(T^{-1}\hat{\zeta},t)\big) + B_1(a_c - v) + E_1 d - \frac{1}{2}\hat{k}_1F_1F_1^TP_1e_1 \tag{5.9}$$

$$\begin{aligned}\dot{\bar{e}}_2 &= (\bar{A}_4 + F_0N_0\bar{C}_4 - L_0\bar{C}_4 - N_0\bar{C}_4\bar{A}_4)\bar{\zeta}_2 - F_0\hat{\bar{\zeta}}_2 + (I_{n+q-m} - N_0\bar{C}_4)\bar{F}_2\phi(T^{-1}\zeta,t) - \\ &\quad - M_0\bar{F}_2\phi(T^{-1}\hat{\zeta},t) + (I_{n+q-m} - N_0\bar{C}_4)\bar{E}_2 d + (I_{n+q-m} - N_0\bar{C}_4)\bar{E}\dot{a}_m - \frac{1}{2}\hat{k}_2M_0\bar{F}_2H_0\bar{C}_4\bar{e}_2.\end{aligned} \tag{5.10}$$

To further simplify the error dynamics (5.10), we need to find matrices M_0, N_0, F_0, and L_0 such that:

$$M_0 = I_{n+q-m} - N_0\bar{C}_4 \tag{5.11}$$

$$F_0 = M_0\bar{A}_4 + (F_0N_0 - L_0)\bar{C}_4 \tag{5.12}$$

$$M_0\bar{E} = 0. \tag{5.13}$$

Then, (5.10) becomes:

$$\dot{\bar{e}}_2 = F_0\bar{e}_2 + M_0\bar{F}_2\big(\phi(T^{-1}\zeta,t) - \phi(T^{-1}\hat{\zeta},t)\big) + M_0\bar{E}_2 d - \frac{1}{2}\hat{k}_2M_0\bar{F}_2H_0\bar{C}_4\bar{e}_2. \tag{5.14}$$

The structure of the resulting error dynamics (5.9) and (5.14) indicates that the goal of the observer design process has been achieved: e_1 is susceptible only to control signal attacks and $\bar{e}_2$ is susceptible only to measurement attacks. However, (5.9) and (5.14) are still not completely decoupled from external system disturbances. To tackle this, a prescribed H_∞ disturbance attenuation level is integrated into the proposed observers. Formally, this newly introduced feature guarantees that the estimation errors are bounded by system disturbances.

Let $r = He = H\begin{bmatrix} e_1 \\ \bar{e}_2 \end{bmatrix}$ be the controlled estimation error where H is a predefined weight matrix in the form of $\begin{bmatrix} H_1 & \underline{0} \\ \underline{0} & H_2 \end{bmatrix}$, with $H_1 \in \mathbb{R}^{m\times m}$ and $H_2 \in \mathbb{R}^{(n+q-m)\times(n+q-m)}$. The next

theorem establishes the necessary conditions for the existence of the proposed observers with the prescribed H_∞ performance $\|r\|_{\mathscr{L}_2} \leq \sqrt{\mu}\|d\|_{\mathscr{L}_2}$.

Theorem 5.3.1. *Consider system (5.1), Assumptions 5-7 and a positive scalar μ. If matrices L_0, F_0, M_0, and N_0 satisfy conditions (5.11)-(5.13) and there are matrices $P_1 = P_1^T > 0$, $P_2 = P_2^T > 0$ and H_0 such that:*

$$H_0\bar{C}_4 = \bar{F}_2 M_0^T P_2, \tag{5.15}$$

$$\Lambda := \begin{bmatrix} \Pi_1 + H_1^T H_1 & P_1\bar{A}_2 & P_1E_1 \\ \bar{A}_2^T P_1 & \Pi_2 + H_2^T H_2 & \Pi_2 M_0 \bar{E}_2 \\ E_1^T P_1 & \bar{E}_2^T M_0^T P_2 & -\mu I_r \end{bmatrix} < 0, \tag{5.16}$$

where $\Pi_1 = A_1^{s^T} P_1 + P_1 A_1^s$ and $\Pi_2 = P_2F_0 + F_0^T P_2 + 2I_{n+q-m}$, then the estimation error dynamics are asymptotically stable with the prescribed H_∞ tracking performance.

Proof. Refer to Appendix B. □

If matrices L_0, F_0, M_0, and N_0 satisfy the conditions that Theorem 5.3.1 founds, then the design of the proposed observers (5.6) and (5.8) is feasible. Finally, Theorem 5.3.2 provides a way of selecting the value of η in order to drive e_1 and $\bar{e}_2$ to the sliding surface:

$$\mathscr{S} = \{(e_1, \bar{e}_2) | e_1 = 0\} \tag{5.17}$$

in finite time while simultaneously preserving their sliding movement.

Theorem 5.3.2. *Consider system (5.1), Assumptions 5-7 and observers (5.6) and (5.8). The error dynamics (5.9) and (5.14) can be driven to the sliding surface (5.17) in finite time when the following inequality:*

$$\eta \geq \|B_1^{-T}\|(\|\bar{A}_2\|\varepsilon + \mathscr{L}_\phi\|F_1\|\|T^{-1}\|\varepsilon + \|E_1\|\xi) + \eta_1,$$

where $\|e\| < \varepsilon$ and $\eta_1 > 0$ is a scalar, holds true and the LMI feasibility problem (5.16) has at least one solution.

Proof. Refer to Appendix B. □

5.4 Estimation of FDIAs

The measurement attack vector a_m can be easily estimated using the proposed UIO (5.8). Observer (5.8) can produce an estimation $\hat{\bar{\zeta}}_2$ of the augmented state vector $\bar{\zeta}_2$ with the

prescribed performance. According to Section 5.2, $\bar{\zeta}_2$ is a superset of a_m, and thus:

$$\hat{a}_m \approx \begin{bmatrix} 0 & I_q \end{bmatrix} \hat{\bar{\zeta}}_2. \tag{5.18}$$

The estimation of the control signal attack vector a_c will be achieved through the error dynamics (5.9) of e_1. According to Theorem 5.3.2, when e_1 and $\bar{e}_2$ are driven to the sliding surface $\mathscr{S}$ (5.17), it is true that $e_1 = 0$. Thus:

$$\bar{A}_2\bar{e}_2 + F_1\big(\phi(T^{-1}\zeta,t) - \phi(T^{-1}\hat{\zeta},t)\big) + B_1(a_c - v_{eq}) + E_1 d = 0, \tag{5.19}$$

where v_{eq} is the equivalent output error injection signal during the sliding motion [128], which expresses the average behavior v. The v_{eq} term can be accurately approximated by [129], [130]:

$$v_{eq} \approx (\rho+\eta)\frac{B_1^T P_1 (C_1^{-1}\omega_1 - \hat{\zeta}_1)}{\|B_1^T P_1 (C_1^{-1}\omega_1 - \hat{\zeta}_1)\| + \delta},$$

where $\delta > 0$ is a small scalar added to the denominator of (5.7) to tackle the chattering effect. Since B_1^{-1} exists, Eq. (5.19) becomes:

$$a_c - v_{eq} = -B_1^{-1}\big(\bar{A}_2\bar{e}_2 + F_1(\phi(T^{-1}\zeta,t) - \phi(T^{-1}\hat{\zeta},t)) + E_1 d\big). \tag{5.20}$$

From the L_2 norm of (5.20), we obtain:

$$\begin{aligned} \|a_c - v_{eq}\|_{\mathscr{L}_2} &= \|B_1^{-1}\Big(\bar{A}_2\bar{e}_2 + F_1\big(\phi(T^{-1}\zeta,t) - \phi(T^{-1}\hat{\zeta},t)\big) + E_1 d\Big)\|_{\mathscr{L}_2} \leq \\ \leq \big(\sigma_{max}(B_1^{-1}\bar{A}_2) &+ \sigma_{max}(B_1^{-1}F_1)\mathscr{L}_\phi\|T^{-1}\|\big)\|\bar{e}_2\|_{\mathscr{L}_2} + \sigma_{max}(B_1^{-1}E_1)\|d\|_{\mathscr{L}_2}, \end{aligned}$$

where $\sigma_{max}(\cdot)$ is the maximum singular value of the considered matrix. Theorem 5.3.1 implies that $\|e\|_{\mathscr{L}_2} \leq \sigma_{max}(H^{-1})\sqrt{\mu}\|d\|_{\mathscr{L}_2}$ and thus, we have:

$$\begin{aligned} &\|a_c - v_{eq}\|_{\mathscr{L}_2} \leq \\ \leq \Big(\sqrt{\mu}\big(\sigma_{max}(B_1^{-1}\bar{A}_2) &+ \sigma_{max}(B_1^{-1}F_1)\mathscr{L}_\phi\|T^{-1}\|\big)\sigma_{max}(H^{-1}) + \sigma_{max}(B_1^{-1}E_1)\Big)\|d\|_{\mathscr{L}_2} \Rightarrow \\ &\Rightarrow \sup_{\|d\|_{\mathscr{L}_2}\neq 0} \frac{\|a_c - v_{eq}\|_{\mathscr{L}_2}}{\|d\|_{\mathscr{L}_2}} = \beta_1 + \sqrt{\mu}\beta_2, \end{aligned}$$

where $\beta_1 = \sigma_{max}(B_1^{-1}E_1)$ and $\beta_2 = \big(\sigma_{max}(B_1^{-1}\bar{A}_2) + \sigma_{max}(B_1^{-1}F_1)\mathscr{L}_\phi\|T^{-1}\|\big)\sigma_{max}(H^{-1})$. Therefore, if $\beta_1 + \sqrt{\mu}\beta_2$ is close to zero, attacks against control signals can be approx-

imated as:

$$\hat{a}_c \approx (\rho + \eta)\frac{B_1^T P_1 (C_1^{-1}\omega_1 - \hat{\zeta}_1)}{\|B_1^T P_1 (C_1^{-1}\omega_1 - \hat{\zeta}_1)\| + \delta}. \tag{5.21}$$

5.4.1 Experimental results

The experimental results from the performance assessment of the presented AE method are shown in this subsection. Its structure is similar to the previous parts of the manuscript that describe numerical results. More insight regarding the subsection structure and the simulation platform can be found in the introduction of Section 4.6.

5.4.1.1 Use case analysis

As explained in Section 4.6.1, a series of case studies is required for evaluating the performance of a proposed cyber defense layer. Regarding the designed AE scheme, the implementation of the necessary case studies follows the specifications that were established for the introduced AD and ALC methodologies, as presented in Section 4.6.1. These specifications require the consideration of various topologies, different types of FDIAs and multiple disturbances to evidence the effectiveness and the scalability of the proposed AE mechanism. An in-depth analysis of these case studies follows:

(a) Case study 1

(b) Case studies 2

Figure 5.1 Topologies of the use cases implemented for the proposed AE and ARC schemes.

- **Case study 1:** 2-area power system, connected via an AC tie-line. The simulated load disturbance is modeled as 1% p.u. step function in area 1 at $t = 5$ sec. Regarding the FDIAs, a 0.01 p.u. sine attack is launched against the control signal of area 2 at $t = 25$ sec and a 10% p.u. step attack is launched against the frequency measurement of area 1 at $t = 35$ sec;

- **Case study 2:** 4-area power system, interconnected via AC (either equipped with TCPS or not) and HVDC tie-lines. In this case, a 5% p.u. step load disturbance occurs

in area 1 at $t = 5$ sec. The simulated FDIAs include a [-0.1, 0.1] p.u. random attack against the control signal of area 2 at $t = 25$ sec and a 1% p.u. ramp attack against the frequency measurement of area 3 at $t = 35$ sec.

The topology of each simulation scenario for the presented AE scheme is depicted in Fig. 5.1. The power system parameter values of each area can be found in the Appendix C. Solar and wind generation disturbances occur throughout the simulations of every case study and their implementation is based on the modeling of Section 3.3 Furthermore, measurement time delays, that vary between 1 to 2 seconds, are simulated in all case studies.

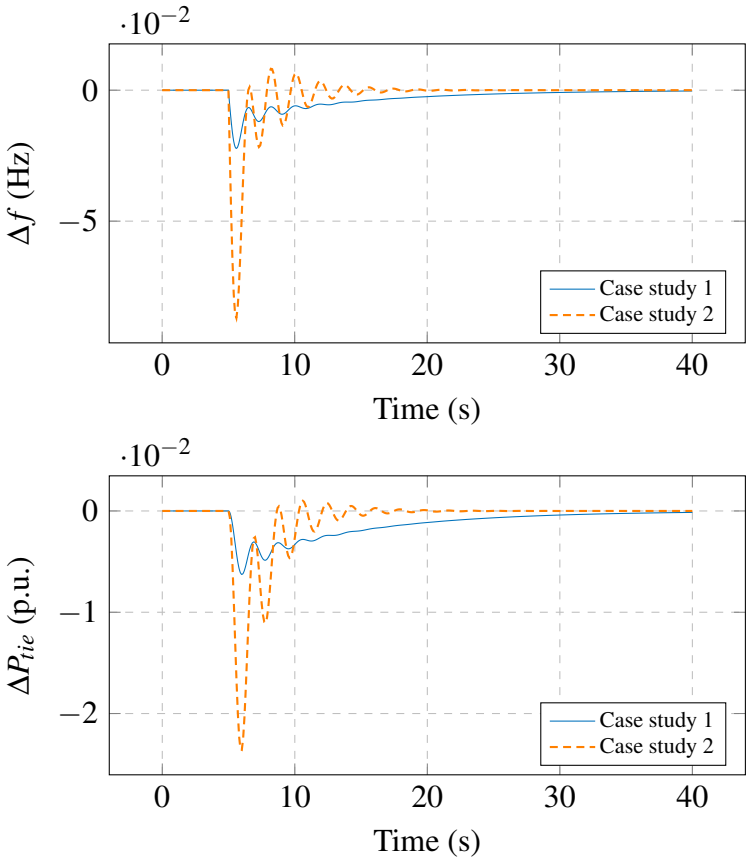


Figure 5.2 Frequency & tie-line power flow responses to the disturbances of each case study implemented for the proposed AE scheme.

Fig. 5.2 portrays the frequency and tie-line power flow responses for case study 1 under 1% p.u. step load disturbance at $t = 5$ sec in area 1 and variations due to RES, and for case study 2 under 5% p.u. step load disturbance at $t = 5$ sec in area 1 and RES variations. In this way, the proper operation of the implemented LFC systems is ensured. Moreover, the waveforms of the aggregated external disturbances, caused by both load and RES variations, are plotted in Fig. 5.3, for a better insight on the simulated scenarios.

5.4.1.2 Performance analysis

The designed AE mechanism is applied to case studies 1 and 2 defined in 5.4.1.1 for performance assessment and the results are portrayed in Fig. 5.4 and 5.5. The blue lines of Fig. 5.4 represent the actual FDIAs launched in each case study while the red lines depict the corresponding attack estimations generated by the proposed AE scheme. The upper graph of Fig. 5.4a demonstrates how the control signal attack estimator of case study 1 performs and the lower graph of 5.4a illustrates the performance of the measurement attack estimator of case study 1. Fig. 5.4b contains the same information but for case study 2. For a better insight into the performance of the proposed AE technique, the resulting attack estimation errors are plotted in Fig. 5.5. These errors are defined as the difference between the actual and the approximated attack signals and annotated as $e_a^c = a_c - \hat{a}_c$, when referred to control signal attacks, and $e_a^m = a_m - \hat{a}_m$, when referred to measurement attacks.

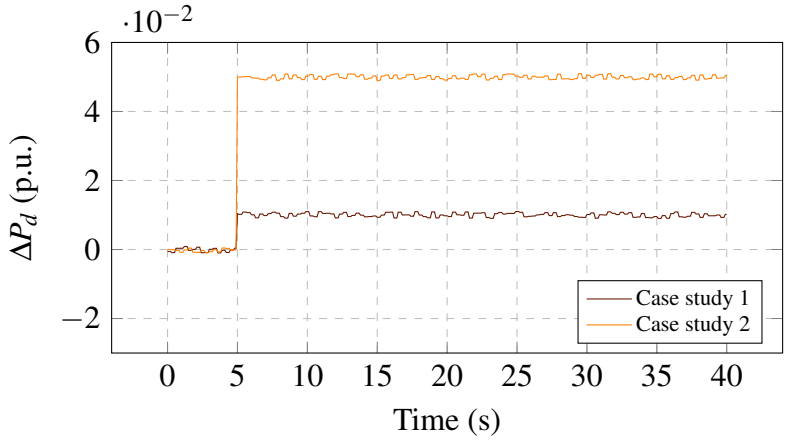
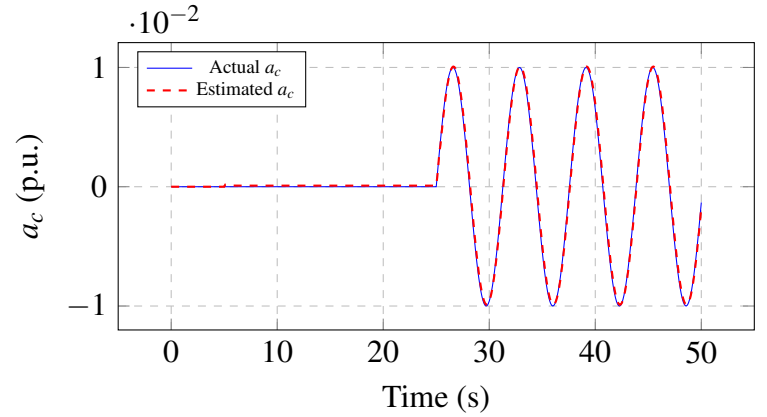
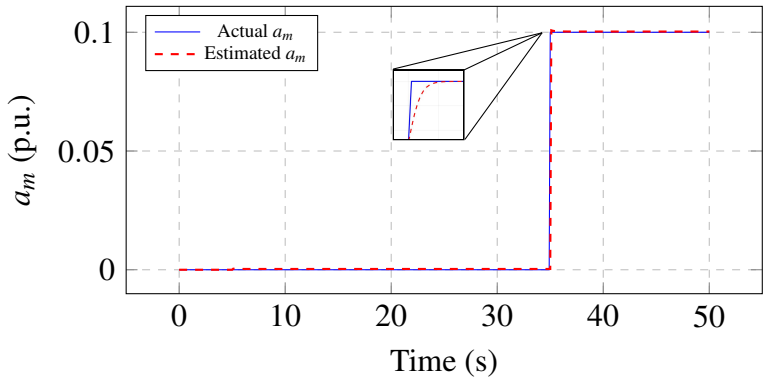


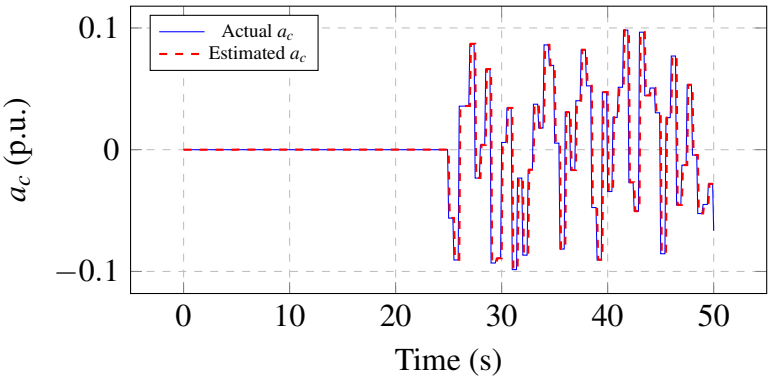
Figure 5.3 Simulated disturbances for each case study of the proposed AE scheme.

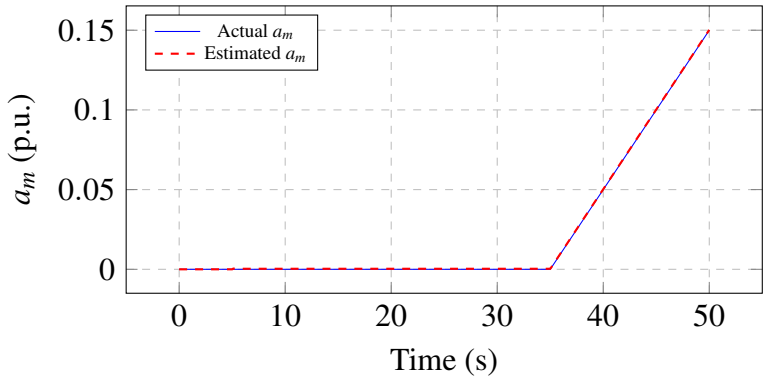
Fig. 5.4 reveals that the actual and the estimated FDIAs are almost identical, which validates the effectiveness of the presented method. This is further verified by e_a^c and e_a^m in Fig. 5.5; the differences between the actual and the approximated attack signals are always close to zero, despite some negligible spikes. Fig. 5.4 and 5.5 also indicate that the suggested AE strategy is resilient against the scheduled load disturbances at $t = 5$ sec, the RES disturbances that occur throughout the simulation and the varying time delays. Moreover, the upper graph of Fig. 5.4a shows that the a_c estimator module of case study 1 is unaffected by the a_m attack at $t = 35$ sec. The same applies to the rest of the simulations, confirming that each estimator module is sensitive only to the attack that is designed for. Finally, the presented AE methodology performs successfully to power systems of various sizes, which proves its scalability.





(a) Case study 1





(b) Case study 2

Figure 5.4 Performance of the proposed AE scheme.

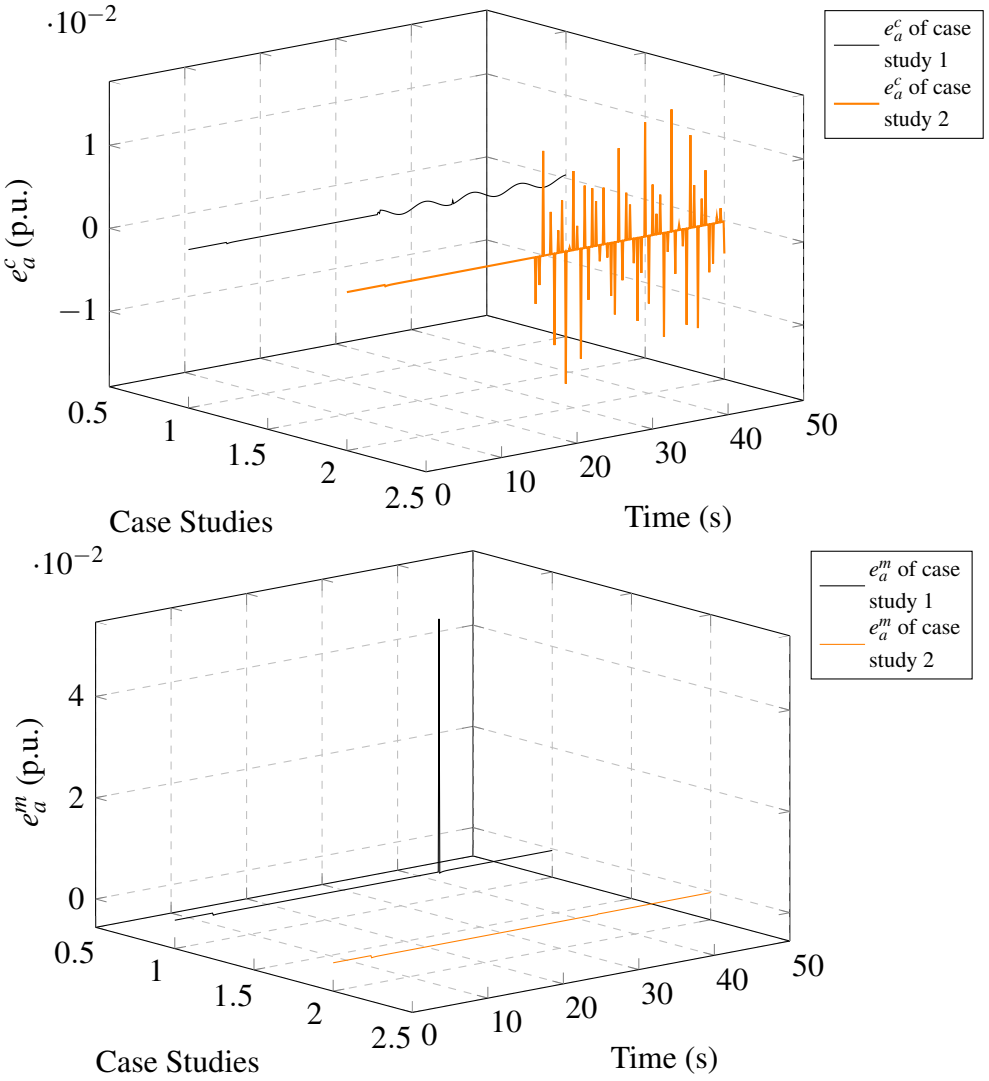


Figure 5.5 Resulting attack estimation errors of the proposed AE scheme.

5.4.1.3 Sensitivity analysis in noisy environments

As already explained in Section 4.6.4, it is highly important to test the effectiveness of a cyber defense layer in the presence of noise. To this end, this subsection is dedicated to the sensitivity analysis of the proposed AE scheme in noisy environments. The methodology and the modeling of this experiment are similar to the ones followed in Section 4.6.4, adjusted to the case study 2 environment of the proposed AE scheme. For more information on this topic, the reader may refer to the aforementioned part of the present study. A similar analysis for the proposed ARC is deemed redundant; if the introduced AE method is effective, the same holds true for the ARC strategy as well, according to both theoretical and numerical evidence.

The results of this experiment are depicted in Fig. 5.6: the orange line refers to the generated attack estimation when applied to the noiseless case study 2 while the green, dotted line corresponds to the same signal for the noisy case study 2. The upper graph Fig. 5.6 shows the launched control signal FDIA and the lower graph portrays the launched measurement FDIA. In the presence of noise, the produced FDIA approximations demonstrate similar behavior with the original waveforms of the corresponding FDIAs, despite some minor fluctuations (approximately 0.1%). Since the impact of these deviations is minimal and the overall performance of the defense mechanism is not downgraded, it is implied that the introduced methodology is robust against noisy measurements.

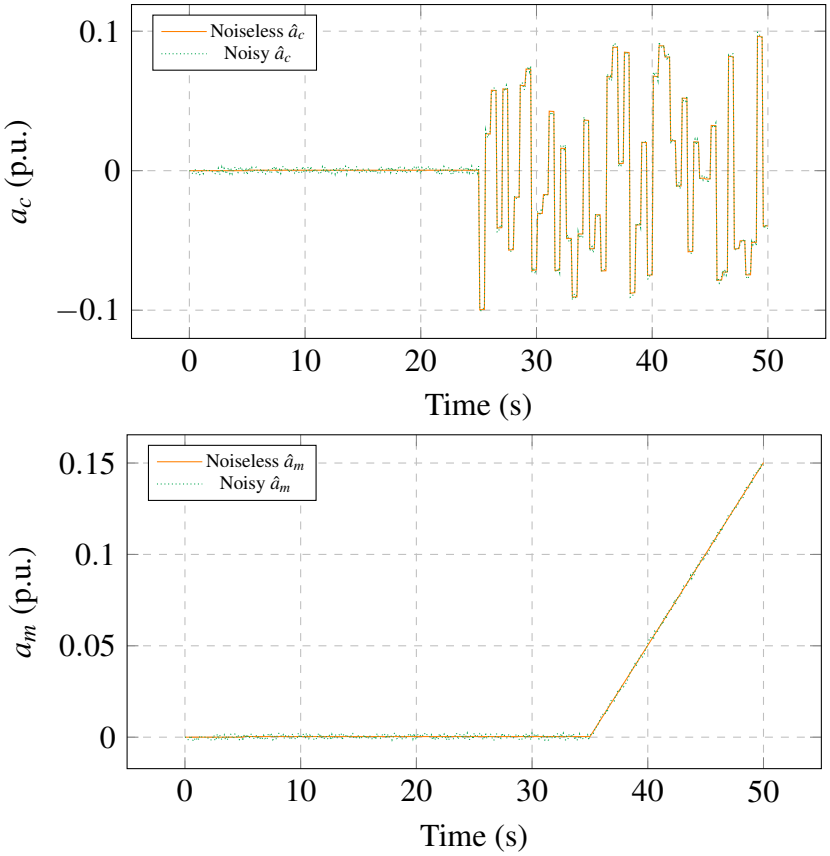


Figure 5.6 Performance of AE in the presence of noise.

5.5 Observer-based Attack-resilient Control Strategy

Eq. (5.18) and (5.21) provide the necessary formulas for the estimation of the measurement and control signal attack vectors, respectively. The estimated signals $\hat{a}_m$ and $\hat{a}_c$ can be used to form an attack-resilient frequency control strategy by properly inserting them into

the compromised control loop (5.1). The attack mitigation method proposed in this study establishes an ARC mechanism like this and is described in what follows. A stability analysis for the introduced ARC strategy is also presented to theoretically verify its effectiveness in stabilizing the power system frequency in the presence of both external disturbances and FDIAs.

For the case of control signals attacks, the integration of the $\hat{a}_c$ term into the LFC loop for the FDIA elimination is straightforward. This can be achieved by adding the $\hat{a}_c$ as a supplementary control input to the compromised system (5.1) in order to compensate for the a_c. The addition of this control input does not modify the input term $u(t)$ which in turn, would affect the original frequency control and thus, the system stability is preserved. The compromised system (5.1) integrated with the new control input $\hat{a}_c$, is converted into the following form:

$$\begin{cases} \dot{x}(t) = Ax(t) + F\phi(x,t) + B\big(u(t) + a_c(t) - \hat{a}_c(t)\big) + Ed(t) \\ y(t) = Cx(t) + Da_m(t). \end{cases} \tag{5.22}$$

For simplicity, in the remainder of this section it is assumed that control signals attacks are immediately mitigated in case of system (5.22) and the difference between the actual and the estimated control signal attack vectors is negligible. Hence, system (5.22) is transformed into:

$$\begin{cases} \dot{x}(t) = Ax(t) + F\phi(x,t) + Bu(t) + Ed(t) \\ y(t) = Cx(t) + Da_m(t). \end{cases} \tag{5.23}$$

By noticing system (5.23), it can be concluded that the compensation for the a_m impact can be achieved by adding the $\hat{a}_m$ term to the system output $y(t)$. However, this integration of a_m into the LFC system modifies the original frequency controller, as explained below. More specifically, Eq. (3.29) and (3.33) indicate the the original control law is a static output-feedback controller described by:

$$u = -Ky,$$

where $K \in \mathbb{R}^{m\times p}$ is the matrix of the feedback gains (its optimal design exceeds the scope of this study, as discussed in Section 3.3). Therefore, it is necessary to prove the effectiveness of the new, attack-resilient control law against both external disturbances and cyberattacks.

For the redesign of the original controller, the actual measurement vector y will be modified by subtracting the estimated attack vector $\hat{a}_m$ from it, since y already includes the

actual measurement attack vector a_m. Thus, y is converted into y_{cr} as:

$$y_{cr} = y - D\hat{a}_m = Cx + Da_m - D\hat{a}_m \Rightarrow y_{cr} = Cx + D\left(a_m - \hat{a}_m\right) \Rightarrow y_{cr} = Cx - De_a^m,$$

where $e_a^m = \hat{a}_m - a_m$ denotes the attack estimation error which is the difference between the actual and estimated attack vectors. From the previous analysis, the form of the updated attack-resilient control law can be obtained as:

$$u_{ar} = -Ky_{cr}.$$

The objective now is to provide a stability analysis for the updated control law which proves that the LFC operates as expected and is resilient to cyberattacks. The remainder of this section is dedicated to this proof.

Stability analysis

By replacing the actual measurement vector y with the corrected one y_{cr} into system (5.23), it is obtained:

$$\begin{cases} \dot{x} = (A - BKC)x + BKDe_a^m + F\phi(x,t) + Ed \\ y_{cr} = Cx - De_a^m. \end{cases} \tag{5.24}$$

The following Lyapunov function is selected to prove the stability of the resulting system (5.24):

$$V_{ar} = V_x(x) + \gamma V_e(e_m^a) = x^T P_x x + \gamma (e_a^m)^T P_e e_a^m, \tag{5.25}$$

where $\gamma > 0$ is scalar and $P_x > 0$ and $P_e > 0$ are definite symmetric matrices with proper dimensions.

By differentiating V_x, it is acquired:

$$\dot{V}_x = x^T \left[(A - BKC)^T P_x + P_x(A - BKC)\right]x + 2x^T P_x\left(F\phi(x,t) + Ed + BKDe_a^m\right).$$

The LFC considered in this study is asymptotically stable under attack-free conditions, according to Eq. (3.34). This implies the existence of a scalar $\eta_x > 0$, such that:

$$x^T \left[(A - BKC)^T P_x + P_x(A - BKC)\right]x + x^T P_x\left(F\phi(x,t) + Ed\right) \leq -\eta_x \|x\|^2. \tag{5.26}$$

From the analysis of Section 5.4, it is concluded that the attack estimation error $e_m^a \rightarrow 0$ as $t \rightarrow \infty$. This yields that a scalar $\eta_e > 0$ exists, such that:

$$\dot{V}_e(e_m^a) \leq -\eta_e \|e_a^m\|^2. \tag{5.27}$$

From (5.27), (5.26) and the differentiation of (5.25), it follows that:

$$\dot{V}_{ar} \leq -\eta_x \|x\|^2 + 2\|P_x BKD\| \|x\| \|e_a^m\| - \gamma \eta_e \|e_a^m\|^2. \tag{5.28}$$

By selecting:

$$\gamma \geq \frac{\left(2\|P_x BKD\|\right)^2}{\eta_x \eta_e},$$

inequality (5.28) becomes:

$$\dot{V}_{ar} \leq -\eta_x \|x\|^2 + \sqrt{\gamma \eta_x \eta_e} \|x\| \|e_a^m\| - \gamma \eta_e \|e_a^m\|^2 \leq -\frac{\eta_x}{2} \|x\|^2 - \gamma \frac{\eta_e}{2} \|e_a^m\|^2.$$

which proves that $e(t) \rightarrow 0$ and $x(t) \rightarrow 0$ as $t \rightarrow \infty$. Thus, the LFC system that is integrated with the proposed attack-resilient control law is asymptotically stable. This means that the normal operation of LFC is preserved with the introduced cyber defense strategy even in the presence of measurement attacks. The new, attack-resilient LFC state-space representation is expressed as:

$$\begin{cases} \dot{x}(t) = Ax(t) + F\phi(x,t) + Bu_{ar}(t) + Ed(t) \\ y_{cr}(t) = Cx(t) + D\big(a_m(t) - \hat{a}_m(t)\big). \end{cases}$$

The basic concepts of the introduced AE and ARC methods are briefly summarized in Algorithm 2 that follows. Algorithm 2 provides the operational flow of the suggested methodology to make the present work more comprehensible.

5.5.1 Experimental results

In this subsection, the experimental results from the performance evaluation of the proposed ARC mechanism are demonstrated. The structure of this subsection is similar to the previous experimental sections. More information about this topic and the simulation platform can be found in the introduction of Section 4.6.

Algorithm 2 Summary of the proposed FDIA defense strategy

Require: • $rank(B) = rank(CB)$,
• $\|\phi(x,t) - \phi(\hat{x},t)\| \leq \mathscr{L}_\phi \|x - \hat{x}\| \;\; \forall x, \hat{x} \in \mathbb{R}^n$,
• $\|a_c\| \leq \rho$, $\|d\| \leq \xi$ and a_m is differentiable.
Ensure: A, F, B, E, C and D are known.
1: Find proper T and S;
2: Construct the designed SMO and UIO;
3: $t \leftarrow 0$; ▷ AE & ARC mechanisms are enabled
4: **while** $t \geq 0$ **do**
5: Apply the T and S to the original system;
6: Compute $\omega(t) = Sy(t)$;
7: Provide $\omega(t)$ and $u(t)$ to the attack estimator module;
8: Calculate $\hat{\zeta}_1(t)$ and $\hat{\bar{\zeta}}_2(t)$ through the SMO and UIO, respectively;
9: Compute $\hat{a}_m \approx \begin{bmatrix} 0 & I_q \end{bmatrix} \hat{\bar{\zeta}}_2$;
10: Compute $\hat{a}_c \approx (\rho + \eta) \frac{B_1^T P_1 (C_1^{-1} \omega_1 - \hat{\zeta}_1)}{\|B_1^T P_1 (C_1^{-1} \omega_1 - \hat{\zeta}_1)\| + \delta}$;
11: Correct $y(t)$ as $y_{cr}(t) = Cx(t) + D\big(a_m(t) - \hat{a}_m(t)\big)$;
12: Provide the new LFC control input signal as $u_{ar}(t) + a_c(t) - \hat{a}_c(t)$;
13: $t \leftarrow t + 1$; ▷ Next time step
14: **end while** ▷ AE & ARC mechanisms are disabled

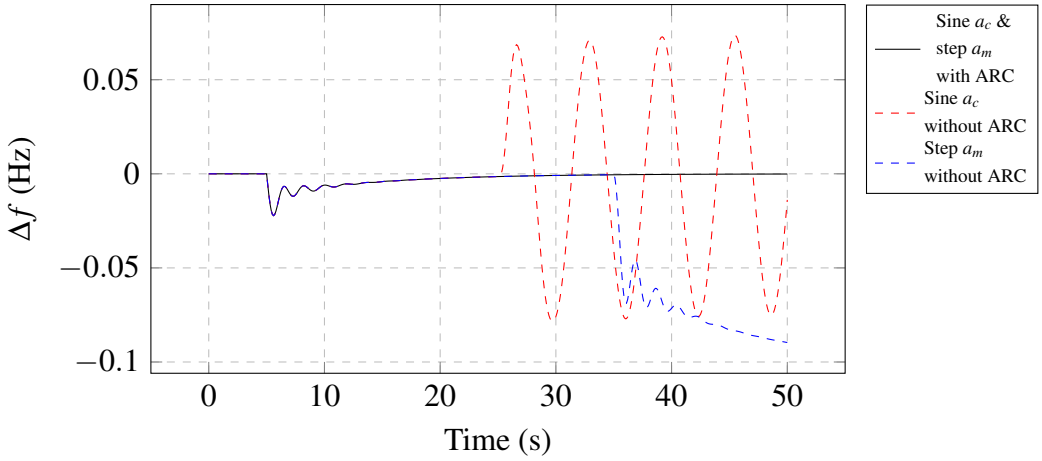
5.5.1.1 Use case analysis

According to Section 5.5, the proposed ARC strategy utilizes the output of the designed AE scheme in order to mitigate the impact of the launched cyberattacks. As these cyber defense layers are interdependent, it is reasonable to investigate their performance on the same grounds. Therefore, the evaluation of the introduced ARC scheme is conducted using the use cases developed for the AE scheme. These use cases have been already described in Section 5.4.1.1 and the reader may refer to the specific part of this study for more information on this topic.

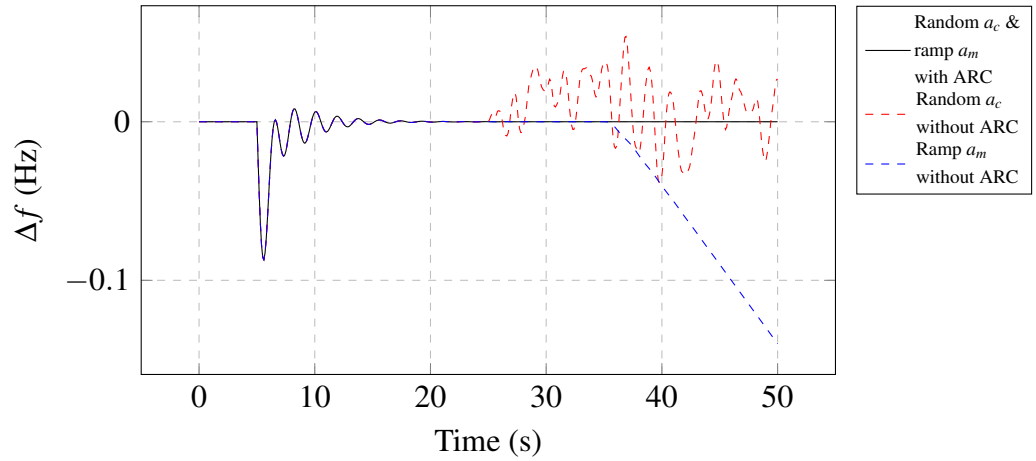
5.5.1.2 Performance analysis

The performance of the proposed ARC method is evaluated on the case studies defined in 5.5.1.1. For this experiment, the frequency responses of the implemented LFC systems are analyzed under the following two conditions: i) with the LFC utilizing the suggested ARC mechanism, and ii) with the suggested ARC being disabled and the LFC using other, existing control methods. For a better insight on the impact of cyberattacks against the system frequency, a single attack between a_c and a_m is simulated when the introduced ARC is inactive. The results are illustrated in Fig. 5.7. Particularly, Fig. 5.7a portrays the performance of the presented ARC in case study 1, where the black line illustrates the behavior of Δf_1 when the suggested ARC is enabled, the red line refers to Δf_1 response under a_c without using the suggested ARC and the blue line displays the Δf_1 response under a_m with the proposed ARC being disabled. The same information is provided in Fig. 5.7b for case study 2.

Based on the results of Fig. 5.7, the frequency responses of the LFC systems without the proposed ARC method start to deviate significantly from their nominal values at $t = 25$ sec and at $t = 35$ sec, when the a_c and the a_m are launched, respectively. On the contrary, when the proposed ARC method is active, the frequency responses deviate only in the event of the scheduled load disturbance at $t = 5$ sec and remain unaffected by the launched cyberattacks. This implies that the introduced control scheme can effectively mitigate the impact of FDIAs, allowing the LFC system to continue operating based on its primary specifications. Moreover, these experiments validate the scalability of the proposed ARC mechanism, as it is successfully applied to several case studies of varying complexity.



(a) Case study 1

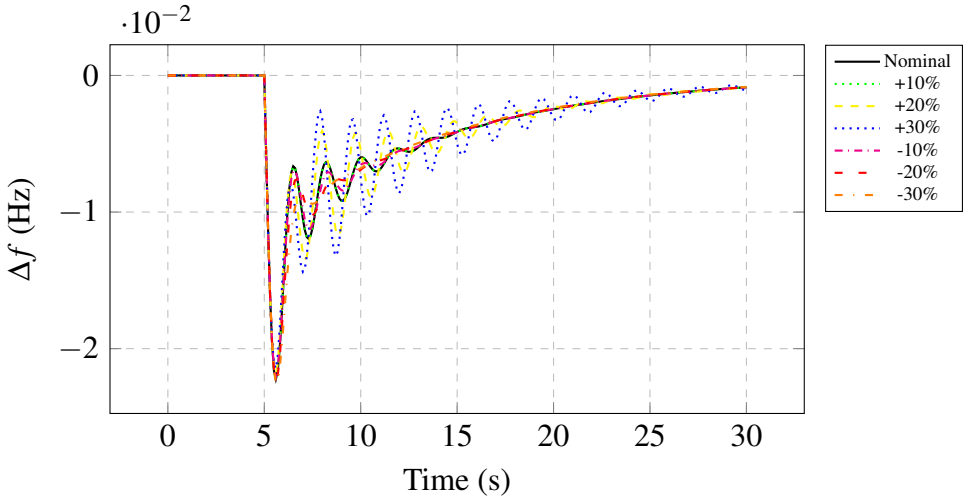


(b) Case study 2

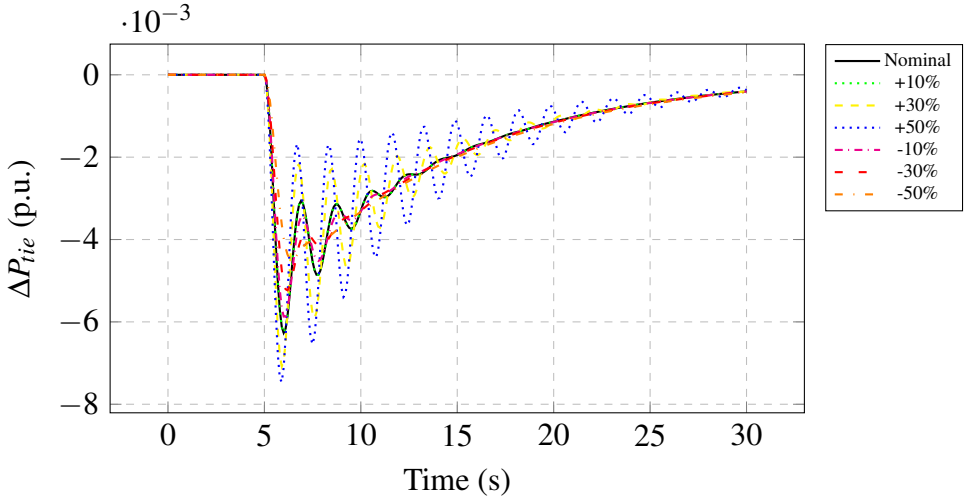
Figure 5.7 Performance of the proposed ARC scheme.

5.5.1.3 Sensitivity analysis on power system parameters

For feasibility reasons, it is necessary to investigate the robustness of a cyber defense mechanism against power system parameter uncertainties, according to Section 4.6.3. In this context, an analogous sensitivity analysis of the proposed ARC mechanism is performed in this subsection. This experiment adopts the methodology described in Section 4.6.3 and the reader may refer to the specific part of this study for more information. The implemented scenarios are the following:



(a) Uncertainties in T_{g_i} and T_{t_i}, $i = 1, 2$.



(b) Uncertainties in T_{12}.

Figure 5.8 Sensitivity analysis on power system parameter uncertainties of the proposed ARC scheme.

- a 10%, 20%, and 30% increase in the turbine and governor time constants of areas 1 and 2 ($T_{t_1}, T_{g_1}, T_{g_2}, T_{t_2}$, respectively),
- a 10%, 20%, and 30% decrease in the turbine and governor time constants of areas 1 and 2 ($T_{t_1}, T_{g_1}, T_{g_2}, T_{t_2}$, respectively),
- a 10%, 30% and 50% increase in the tie-line synchronizing coefficient between areas 1 and 2 (T_{12}),
- a 10%, 30% and 50% decrease in the tie-line synchronizing coefficient between areas 1 and 2 (T_{12}),

The results of the present sensitivity analysis are illustrated in Fig. 5.8. More specifically, Fig. 5.8a depicts the system frequency responses when uncertainties in T_{g_i} and T_{t_i} ($i = 1, 2$) occur and Fig. 5.8b portrays the tie-line power flow responses of the system in case of T_{12} uncertainties. These responses demonstrate minimal deviations in case of system parameter uncertainties compared to their standard behavior, when the nominal values of the system parameters are used. Furthermore, the FDIAs against control signals and measurements are still effectively mitigated even when the system parameters are miscalculated. Based on these results, it can be concluded that the suggested attack-resilient strategy is robust against possible inaccuracies in the power system parameters.

5.5.1.4 Software-in-the-loop simulation

So far, the assessment of the proposed ARC method has been performed within a software environment. Despite the multiple practical features considered in the modeling of LFC, software simulations cannot completely capture the real-time nature of power systems. To evaluate the performance of the presented ARC strategy in more realistic conditions, a Software-In-the-Loop (SIL) testbed has been implemented. SIL (or hardware-in-the-loop) is a real-time simulation technique that enables a highly detailed and accurate design of power systems. In SIL testing, the behavior of the physical system is emulated by a specialized hardware that can interact with external components, such as software applications and embedded systems. These external components encapsulate the algorithms proposed for optimizing the performance of the designed power system.

The architecture of the developed SIL testbed is demonstrated in Fig. 5.9 and described in what follows. The standard IEEE 39-bus system [131, 132], divided into three power areas, is implemented in a Real Time Digital Simulator (RTDS) infrastructure [133] to simulate the physical system. The frequency control of the IEEE 39-bus system is performed outside of RTDS as a standalone Python application, which can utilize the proposed ARC upon request.

The RTDS and the Python application communicate remotely through the DNP3 protocol. The events in the SIL simulation include a step load disturbance of 1% p.u. at $t = 50$ sec in area 1, a step FDIA of 1% p.u. against the control signal of area 3 at $t = 100$ sec and a sine FDIA of 10% p.u. amplitude against the frequency measurement of area 1 at $t = 120$ sec.

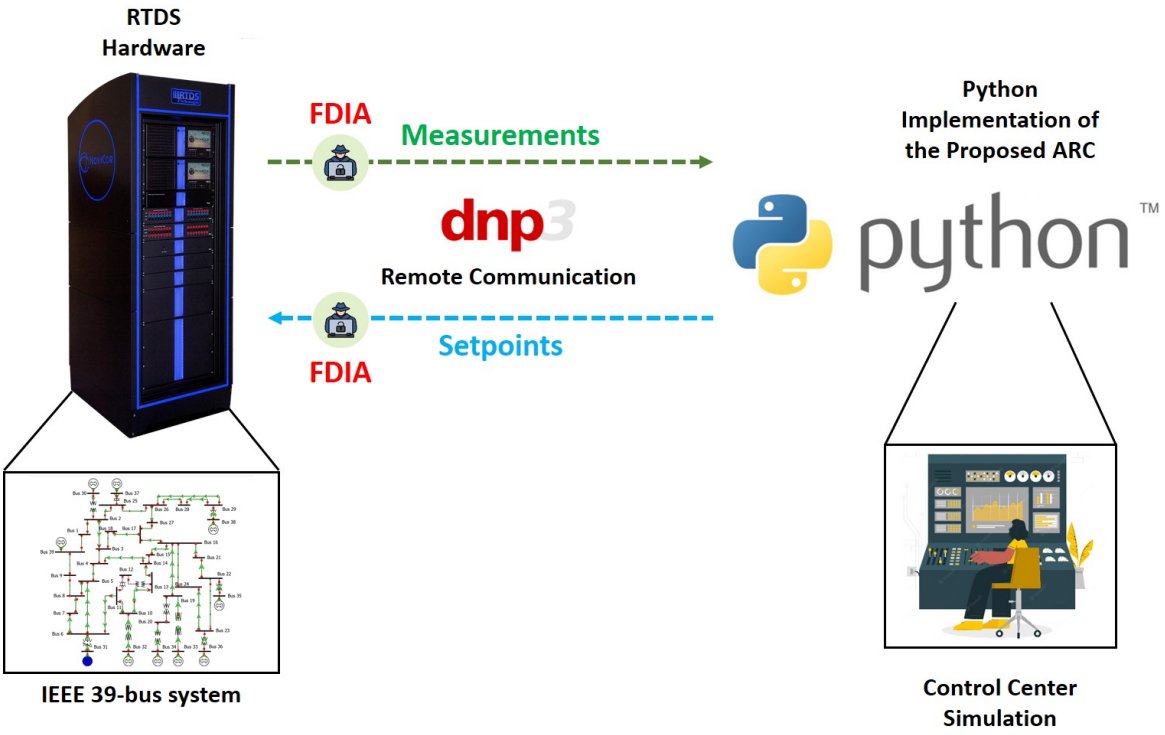


Figure 5.9 Implemented SIL testbed for the performance assessment of the proposed ARC scheme.

Fig. 5.10 illustrates the frequency response of the power system implemented in RTDS for the described simulation scenarios. More specifically, the red line corresponds to the scenario where the system faces the control signal FDIA and the load disturbance without using the proposed ARC strategy; the blue line characterizes the situation in which the measurement FDIA and the load disturbance take place with the introduced ARC scheme being disabled; finally, the black line represents the case where the load disturbance occurs and both measurement and control signal FDIAs are launched, while the presented ARC method is in operation. The results demonstrate that the frequency response of the system without any cyber defense measure experiences significant variations after the launch of FDIAs at 100 sec and 120 sec. On the contrary, the FDIAs against the system that utilizes the proposed ARC scheme have no impact on its frequency response. Therefore, it is confirmed that the introduced cyber defense method is still effective in real-world power systems. It is

also worth mentioning that the minor fluctuations of the frequency response throughout the SIL simulation are an expected phenomenon in realistic conditions.

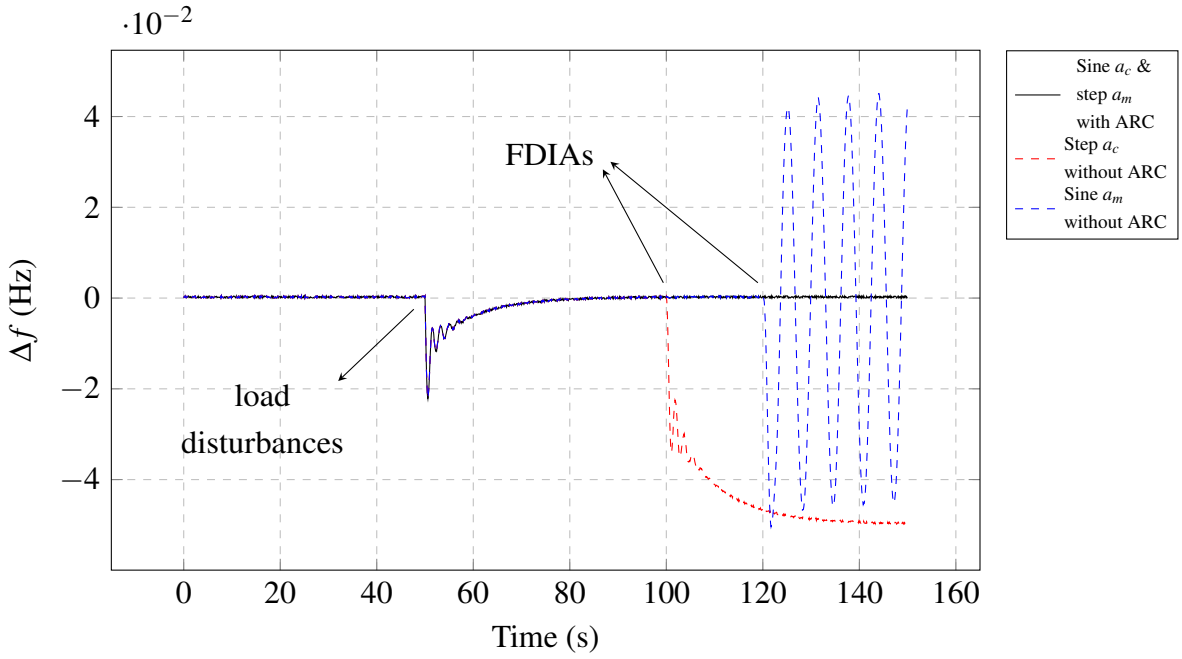


Figure 5.10 Performance assessment of the proposed ARC scheme in the SIL simulation.

5.5.1.5 Comparative study

To highlight the novelties of the proposed ARC mechanism, it is important to compare it with other related works from the literature. For this reason, a comparative analysis is conducted between the presented ARC method and several state-of-the-art methods, based on a set of selected quality features. The results of this analysis are demonstrated in Table 5.1; "✓" notation declares that the specific attack mitigation method for LFC meets the corresponding feature and "×" symbol implies that it does not. The considered set of quality features includes:

1. **Estimation**: the property of a method to provide full information about the launched cyberattacks;
2. **Global mitigation**: the attribute of a methodology to mitigate both measurement and control signal attacks;
3. **Decoupling**: the robustness of a method against external system disturbances;
4. **Nonlinearities**: determines if a defense mechanism takes the nonlinearities of LFC into account or not;

5. **Diverse tie-lines**: the applicability of a method to power systems with different types of tie-lines;

6. **RES**: declares whether the specified technique considers RES disturbances or not;

7. **Parameter uncertainties**: the sensitivity of an approach to power system parameter uncertainties;

8. **Time delays**: the robustness of a defensive strategy against network time delays;

9. **Scalability**: the applicability of an attack mitigation scheme to power systems of various sizes.

Table 5.1 Quality comparative analysis of the proposed ARC scheme.

Features \ **Methods**	[21]	[72]	[75]	[71]	[74]	**Proposed**
Estimation	×	✓	×	✓	✓	✓
Global mitigation	×	×	×	×	×	✓
Decoupling	×	✓	×	✓	✓	✓
Nonlinearities	×	×	✓	×	×	✓
Diverse tie-lines	×	×	×	×	×	✓
RES	×	×	×	×	×	✓
Parameter uncertainties	×	×	×	×	×	✓
Time delays	×	×	×	×	✓	✓
Scalability	×	×	×	×	✓	✓

Table 5.1 demonstrates the superiority of the proposed ARC methodology over various related works. Particularly, the introduced method can effectively approximate the magnitude of the launched FDIA signals and is resilient against external system disturbances, such as load variation and RES generation; both of these features are not met in [21] and [75]. Moreover, the suggested methodology can effectively eliminate both measurement and control signal attacks, while none of the benchmarked methods are capable of it. Regarding the LFC nonlinearities, very few research works [75] take them into consideration, apart from the proposed strategy. Finally, the feasibility of the existing methodologies to real-world electrical systems is not properly assessed, contrary to the introduced method. For example, none of the works under comparison is evaluated in power systems that use RES or

diverse types of interconnecting lines (HVDC, TCPS-equiped). Likewise, the sensitivity of a method against power systems parameters is not studied in any of the compared approaches. Additionally, the time delays due to network limitations along with the scalability over power systems of various sizes are only investigated in [74] and in the presented technique.

Chapter 6

Data-driven Attack Detection & Mitigation for LFC

This chapter describes the data-driven part of the framework developed in this thesis for identifying and mitigating cyberattacks against LFC. The analysis starts with the proposed autoencoder-based attack detection mechanism which can identify a wide range of FDIAs and distinguish them from other external disturbances. The autoencoder learns the healthy status of the system and then, its inputs and outputs are used to formulate the introduced cyberattack indicator for LFC. The evaluation of this method is performed through several experiments on diverse types of use cases. Next, the proposed data-driven attack mitigation technique follows, called *DAR-LFC*, which is based on a DNN architecture. The trained neural network model estimates the healthy control signals of LFC in an innovative way. Then, the system can temporarily use the approximated signal to regulate the system generators during cyberattacks. To verify the effectiveness of this method, DAR-LFC is applied to a series of use cases where various types of cyberattacks and external disturbance incidents are considered.

6.1 Autoencoder-based Attack Detection Method

6.1.1 Motivation

The core idea in attack (or anomaly, in general) detection methodologies is to design a baseline state within which the system is considered to be under normal operation. To achieve this, a metric is needed that will quantify the status of the investigated system in terms of cybersecurity. This metric is typically called *residual* and it can be acquired through the development of a model, e.g. mathematical, statistical or data-driven. The designed

model receives inputs from the actual system and computes the detection residual at each time step. Then, if the residual exceeds a specified threshold, it is considered that the system status is abnormal and a proper alarm is triggered; otherwise, the system is assumed to be in a healthy state and continues its operation. The selection and the design of the attack threshold, e.g. static or adaptive, is a key aspect towards the development of an effective attack detection methodology.

6.1.2 Algorithm inputs

Before selecting the model that will generate the attack detection residual, it is necessary to define its inputs. The applications that provide insights about the cyber resilience status of CPSs typically operate within the control center. Therefore, the inputs of these cyber defense mechanisms is a subset of the system variables and the generated control signals. For the case of LFC, the control is performed using the local frequency measurement of each area and the power flow readings of each tie-line. Several approaches, such as observer-based methodologies, require both the measurements and control signals of LFC to operate. On the other hand, data-driven attach detection mechanisms can achieve the same goal without using the control commands of this system. Consistent with its related works, the introduced algorithm is designed to operate using only the LFC measurements, specifically Δf_i and $\Delta P_{tie_{ij}}$, where $i, j = 1, 2, \ldots N$ and $i \neq j$.

Another critical aspect is to decide the number of the historical LFC measurements that will be used in the input vector of the proposed detection algorithm. In general, attack detection methods either operate with real-time measurements or store the last K values of the input vector variables and feed them to their models. For example, observer-based approaches do not utilize any past data and receive only real-time LFC readings as input. In data-driven methods, using a historical time window provides their models with an additional perspective on the training data and enhances the learning procedure. Thus, an input vector that includes previous information of its variables is selected for the proposed algorithm. The proper length of the time window can be determined by approximately calculating the K time steps that can adequately capture the impact of the external events on the input vector variables. Considering the analysis of this subsection, the input vector is modeled as:

$$x_t = \{\Delta f_i^1, \Delta f_i^2, \ldots, \Delta f_i^K, \Delta P_{tie_{ij}}^1, \Delta P_{tie_{ij}}^2, \ldots, \Delta P_{tie_{ij}}^K\}.$$

6.1.3 Utilized model

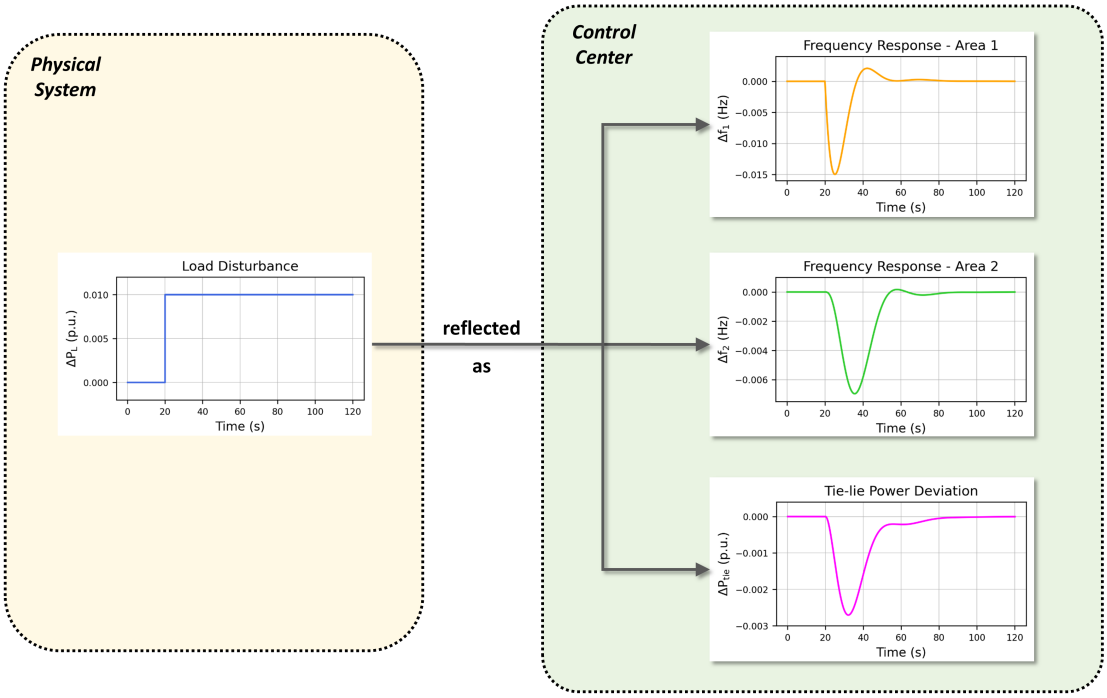


Figure 6.1 Impact of a step load disturbance on the measurements of the control center.

After defining the inputs of the proposed method, the next objective is to select a proper model that will produce the detection residual. In the field of AI, various data-driven algorithms have been developed for addressing the problem of anomaly detection, such as isolation forests, support vector machines (SVMs), autoencoders, etc. While simplistic data-driven algorithms demonstrate acceptable performance for anomaly detection, autoencoders are more appropriate for the LFC system and thus, they are selected as the model of this method. According to Section 2.2.2, a trained autoencoder can reconstruct the received input vector at its output with high accuracy. By training on normal data, this feature allows autoencoders to identify anomalies or outliers in new data that deviate from the learned patterns. The main reason of employing autoencoders in this part of the framework is the significant challenges met in identifying cyberattacks in LFC, such as the complex impact of cyberattacks on the system measurements and the distinction of digital threats from other external incidents. Therefore, the advanced data-driven model of the autoencoder, that is capable of recognizing the narrow underlying patterns of LFC data, is suitable for this case. Furthermore, autoencoders are implemented using deep neural networks, as explained in Section 2.2.2. The flexibility of DNNs enables the autoencoder to continuously learn new healthy states during its online operation, as it will be explained in the next subsection.

6.1.4 Proposed attack detection algorithm

In this subsection, the introduced data-driven attack detection mechanism for LFC is analyzed in detail. Before proceeding to this analysis, it is important to investigate the impact of the various external incidents on the power system for a better comprehension of the proposed method. To this end, assume that the LFC algorithm is applied to a two-area power system that faces load variations and cyberattacks. The operator of the control center can monitor the system status through the measurements of frequency and tie-line power flows. More specifically, if a step load increase occurs on the physical system, the control center operator views it as the damping oscillations of frequency and tie-line power flow measurements in Fig. 6.1. On the contrary, a step FDIA launched against the local frequency of area 1 is reflected to the control center as the deviations of frequency and tie-line power flow readings from their nominal values shown in Fig. 6.2. This paradigm highlights the different effects of cyberattacks and load disturbances on the LFC system.

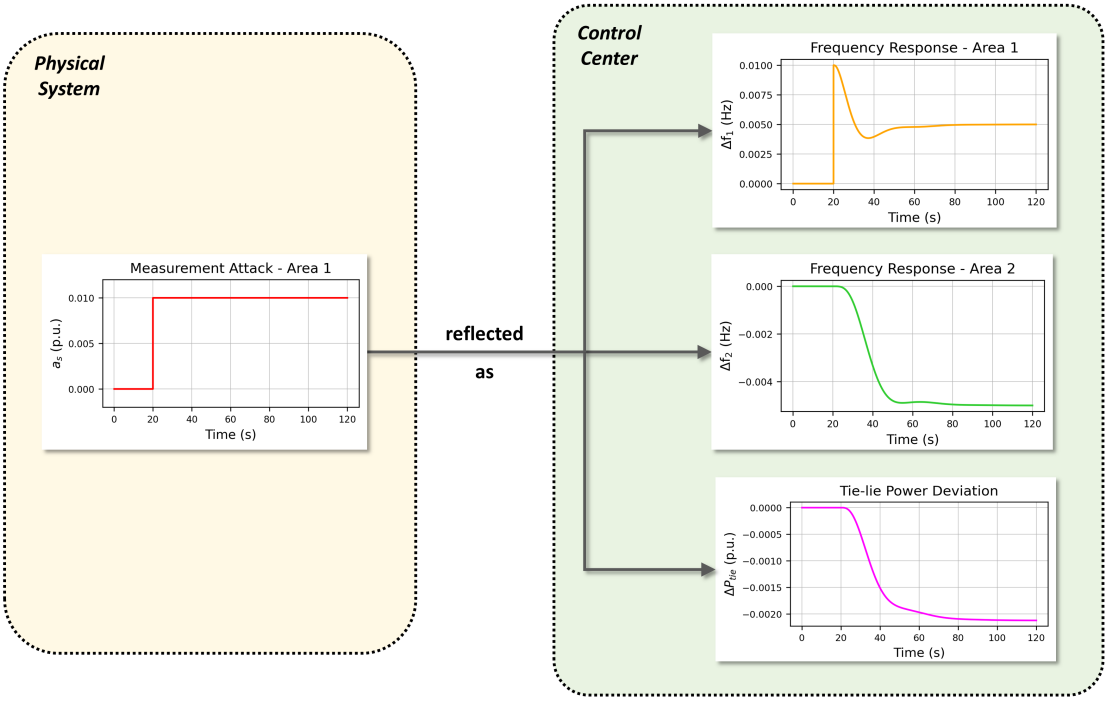


Figure 6.2 Impact of a step load disturbance on the measurements of the control center.

The distinct impact of cyberattacks on the remote measurements of LFC can be leveraged to detect potential malicious behavior in the system. This is achieved by modeling the normal status of LFC using the steady-state of the system, load variations and RES disturbances while considering cyberattacks as LFC anomalies. The employed autoencoder is then trained on a dataset of frequency and tie-line power flow measurements that reflect the healthy status

of LFC. After completing the training process, the autoencoder is deployed in the control center, where the LFC measurements are forward to it. If the autoencoder receives streams of frequency and tie-line power flow measurements that correspond to a load variation but have not been learned, it will replicate them with high accuracy, as illustrated in Fig. 6.3. On the other hand, if the autoencoder is given time-series of LFC measurements that describe the effect of a cyberattack (an event that the autoencoder has not seen during the training phase), the replication of its input will demonstrate significant error, similar to Fig. 6.4. This feature makes the trained autoencoder an effective indicator of cyberattacks for LFC.

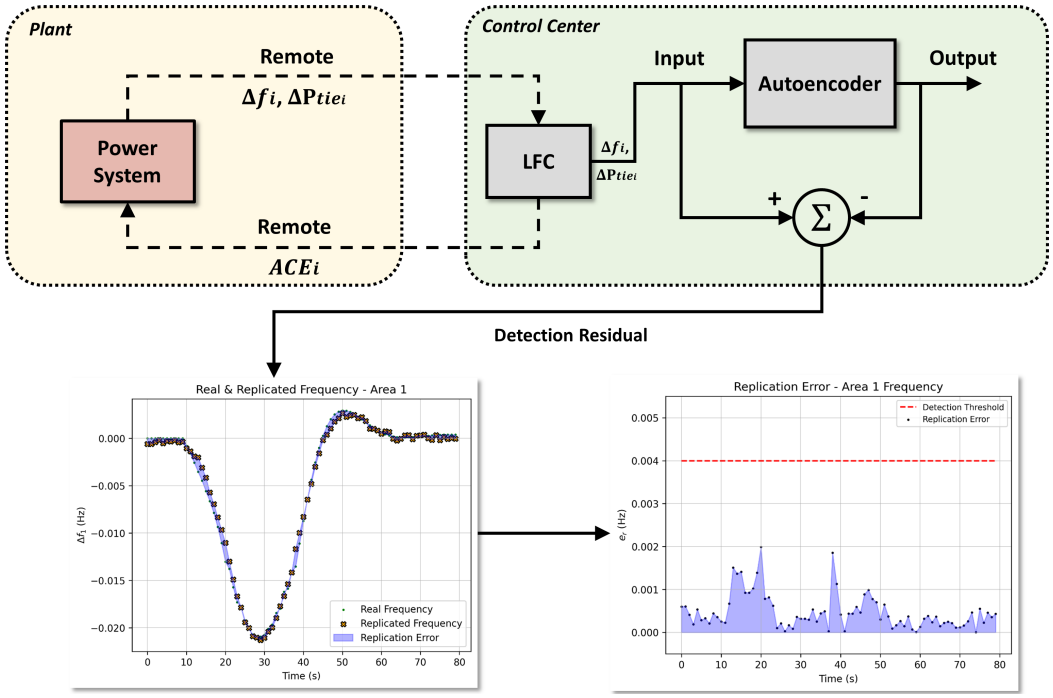


Figure 6.3 Diagram of the proposed data-driven attack detection method under normal conditions.

In the proposed attack detection method, the autoencoder is interchanged between three operational states: the offline training phase, the online operation phase and the online training phase. These states are utilized to provide the autoencoder with a basic knowledge about the normal LFC status, extract cyber resilience insights from it and maintain its robustness through continuous learning. A more detailed analysis of these phases is included in what follows:

1. **Offline training phase**: in this phase, the autoencoder is trained on streams of frequency and tie-line power flow measurements that correspond to load disturbances, RES variations and the steady-state of the system. The training is performed offline,

in an isolated computer environment outside the control center. These streams are composed of K consecutive points over the time of the training simulations. The maximum error obtained by the training samples is selected as the attack detection threshold t_d of the proposed method while the average of these error is used as the transition threshold t_t between the online operation and training phases.

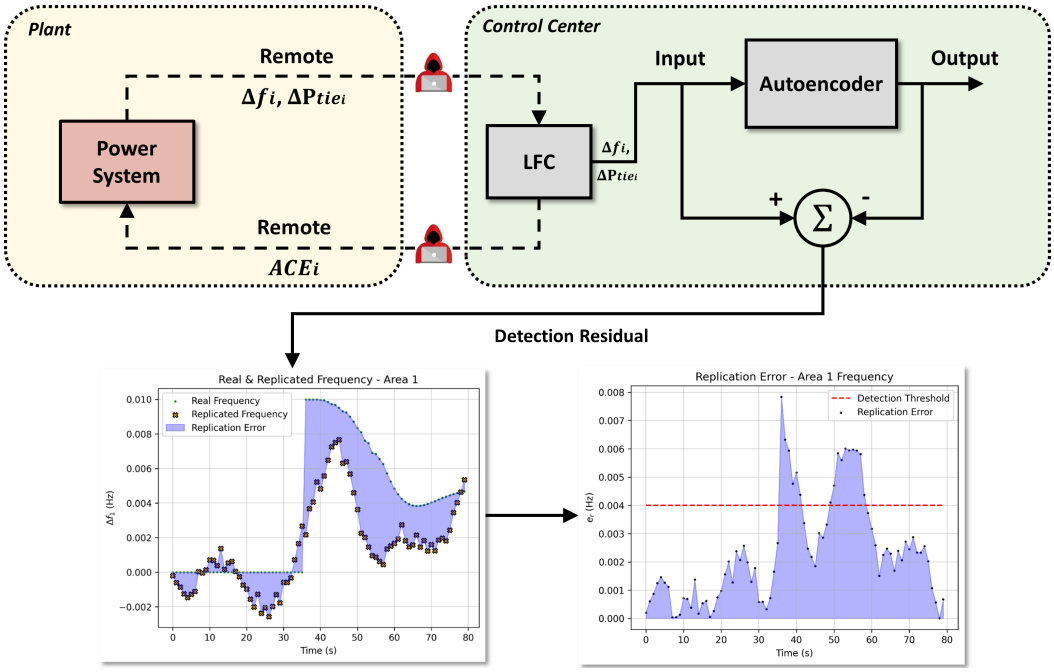


Figure 6.4 Diagram of the proposed data-driven attack detection method under cyberattacks.

2. **Online operation phase**: when the training process is completed, the autoencoder transits to its online operation phase, according to in Fig. 6.5. In this phase, the application of the utilized model is launched in the control center to generate the attack detection residual. At each time step of this mode, the autoencoder receives the last K LFC measurements as an input and produces a replica of them. If the autoencoder accuracy exceeds the t_d threshold, it is assumed that the LFC system is under attack; otherwise, LFC is considered to be in normal condition. The accuracy of the autoencoder is measured by its replication error, which is annotated as e_r and expresses the difference between the input measurements and the generated output.

3. **Online training phase**: if the autoencoder is in the online operation phase and its e_r ranges between t_t and t_d, it is assumed that the received normal status has not been learned. Therefore, the autoencoder transits to its online operation phase, as depicted in Fig. 6.5. In this state, a copy of the current autoencoder is created to be retrained

using the last K received measurements of LFC. When the online training phase is completed, the newly trained copy replaces the operating autoencoder and the system returns to its online operation phase. By continuously learning new normal conditions, the autoencoder stays updated and preserves its robustness against upcoming digital threats.

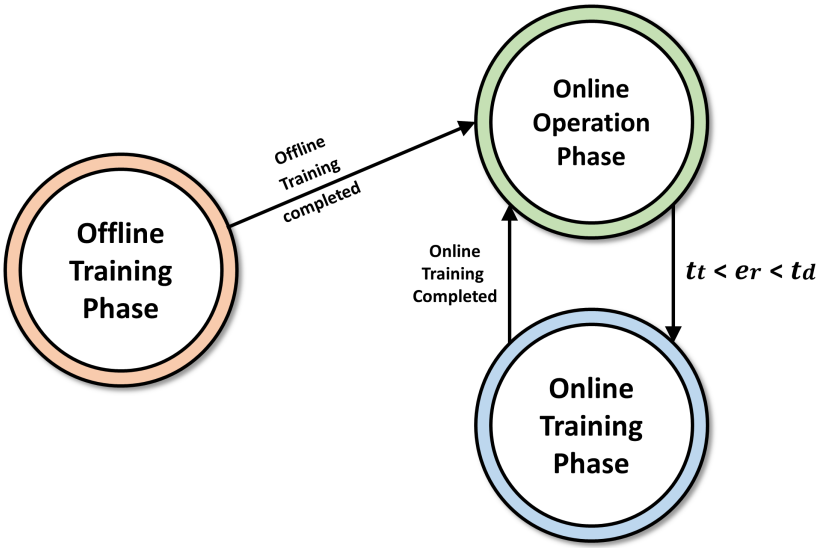


Figure 6.5 Autoencoder phases in the proposed data-driven attack detection method.

At this point, it is important to shed more insights on the decision process of the attack detection and state transition thresholds. After the final epoch of the training process, the replication errors produced by each sample of the training dataset have been minimized and can be acquired. If a stream of LFC measurements that reflects normal system status is fed to the autoencoder, the value of the generated e_r will vary between those obtained during the training process. Therefore, it is reasonable to set the maximum of these errors as the boundary that distinguishes normal disturbance events from cyberattacks in LFC. Furthermore, if the autoencoder generates an e_r that is close but below t_d, the corresponding normal condition of LFC might have not been learned. The average error obtained by the training process is considered an appropriate benchmark for unseen normal LFC conditions and thus, it is selected as the state transition threshold of this methodology.

6.1.5 Experimental results

6.1.5.1 Autoencoder training process

Before the deployment of the proposed cyberattack detection method for LFC, the utilized autoencoder needs to be trained on the healthy system measurements first. This training process involves the iterative adjustment of the model parameters based on the collected and preprocessed LFC data. The objective is to minimize the difference between the autoencoder output and the actual targets. For the training procedure of this work, the next, consecutive steps have been followed:

1. **Simulated events**: initially, the disturbance events that describe the normal dynamics of LFC have to be defined and simulated. Such disturbances are the load changes and the variations due to RES. The selection of these disturbance incidents should capture the LFC operation under diverse conditions to ensure that the model can perform well on unseen data, i.e. it can generalize. For this reason, load changes are simulated as step functions of various slopes and $\pm 35\%$ magnitudes, while RES disturbances are modeled according to Eq. (3.25)-(3.27) and occur throughout the simulation.

2. **Dataset collection**: after selecting the normal disturbance events, the LFC operation is simulated. Its duration is 4000 seconds and a single set of the autoencoder input vector is sampled every second. Therefore, the total samples of the dataset are $4000 \times n$, where n is the dimension of the input vector. The partitioning strategy of the dataset follows a 70-20-10 split, implying that 70% of the data is used for training, 20% for validation and 10% for testing.

3. **Data preprocessing**: the dataset acquired from the LFC simulation is generally regarded as clean, without any noisy or missing values. Thus, the only necessary data preprocessing technique is the data normalization for improving the performance and training stability of the model. Since the dataset features do not contain extreme outliers, the z-score method of $\mu = 0$ mean and $\sigma = 1$ standard deviation is applied to normalize the input data.

4. **Autoencoder architecture**: after thorough experimentation with the architecture of the autoencoder, optimal performance has been achieved using a DNN-based encoder comprised of three layers, each with 128, 64, and 32 units, respectively. Therefore, the dimension of the latent space is 32 while the structure of the decoder is the mirrored image of the encoder.

5. **Hyperparameters**: besides the model architecture, there are several other training hyperparameters that have to be properly configured to optimize the performance of the autoencoder. In this work, the accuracy of the utilized autoencoder has been maximized using 2000 epochs, 0.0001 learning rate, 64 batch size, the mean squared error (MSE) as the loss function and the Adam optimizer.

Table 6.1 Performance of various autoencoder implementations during training process.

Autoencoder implementation	Mean MSE	Maximum MSE	Training time (min)
DNN-based	0.0819	0.0011	8
CNN-based	0.0796	0.0098	47
RNN-based	0.0781	0.0092	76
LSTM-based	0.0763	0.0089	109

In the previously described procedure, Step 4 can be implemented using various neural network architectures, such as DNNs, Long Short-Term Memory (LSTMs), etc. The selection of the proper architecture depends on the application that the autoencoder is intended for. Typically, there is a trade-off between the complexity of the utilized neural network and the accuracy of the autoencoder. To find the proper architecture for the investigated problem, four types of autoencoders have been implemented: (i) the DNN-based, (ii) the Convolutional Neural Network-based, (iii) the Recurrent Neural Network-based and (iv) the LSTM-based autoencoders. The results from the training process of each autoencoder variant are shown in Table 6.1, where their mean MSE, their maximum MSE and their training time are included. Assuming that the MSE of each sample s is given by $MSE_s = \left(\sum_{k=1}^{K}(x_k - \tilde{x}_k)\right)/K$, where K is the total number of observations per sample, mean MSE is defined as $(\sum_{s=1}^{S} MSE_s)/S$, where S is the total number of samples, while maximum MSE is computed by $max(MSE_1, MSE_2, \ldots, MSE_S)$. The outcomes of this comparison verify the aforementioned, theoretical trade-off: the LSTM-based autoencoder has the best performance by demonstrating the minimum mean MSE. However, this variant provides a slight improvement in the model accuracy (7.34%) compared to the DNN-based implementation, while it significantly increases the training time (92.66%). Since a 0.0819 mean MSE is considered sufficient performance for the needs of the cyberattack detection in LFC, the DNN-based autoencoder is chosen due to its simplicity. Furthermore, this lightweight implementation does not introduce a large computational overhead into the learning phase of the proposed method, enhancing its stability at this stage.

6.1.5.2 Use cases description

For the evaluation of the proposed methodology, two use cases have been implemented, i.e. *Use Case A* and *Use Case B*. The operation of LFC in these scenarios is simulated using both a software environment and a Hardware-in-the-Loop testbed. Also, a wide range of external incidents has been considered, such as load changes, various FDIAs, etc. The diversity of the use cases ensures that the introduced method can scale effectively across different types of power systems and remain applicable in various conditions. The proposed data-driven approach has been implemented in Python, using the PyTorch library [134]. The detailed analysis of these use cases is provided in the remainder of this subsection:

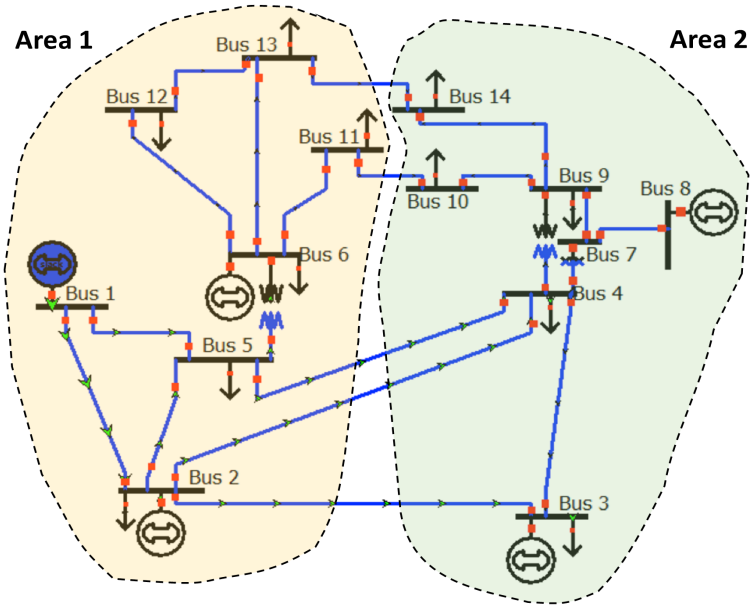


Figure 6.6 Diagram of the power system simulated in Use Case A.

- **Use Case A**: this baseline scenario for the evaluation of the proposed methodology is implemented within a software environment. More specifically, the operation of LFC is simulated in MATLAB/Simulink using the SFR model of the IEEE 14-bus system [135]. This power grid is separated into two distinct areas and its diagram is shown in Fig. 6.6. The Python module that applies the introduced attack detection technique is deployed on a computer node outside the MATLAB/Simulink environment and communicates with it over the TCP/IP protocol. The normal disturbance events considered in this use case are a 0.01 p.u. load increase at the 50th second of the simulation and generation variations due to RES that occur throughout the system operation. The simulated cyberattacks include a ramp FDIA of 0.01 p.u. slope that is injected into the tie-lie measurement between the 150th and 200th second of the

simulation, along with a step FDIA of 0.02 p.u. which is launched against the frequency measurement of area 1 at $t = 300$s and lasts for 50 seconds.

- **Use Case B**: to test the proposed method in more realistic conditions, a Hardware-in-the-Loop (HITL) testbed has been implemented in Use Case B. In this platform, which is illustrated in Fig. 6.7, the operation of a three-area, IEEE 118-bus system [136] is simulated on the RTDS hardware while its secondary frequency control is performed using a SEL Real-Time Automation Controller (RTAC) [137]. To act as the control center of this HITL simulation, RTAC receives measurements from the simulated power system and sends control commands back to it over the IEC 61850 protocol. RTAC also forwards the LFC measurements to the Python module that applies the proposed attack detection method via the TCP/IP protocol. The disturbance incidents include a 0.008 p.u. load decrease at $t = 130$s and generation variations from RES that happen during the whole system operation. Regarding the simulated cyber threats, a 0.01 p.u. sine FDIA is launched against the frequency measurement of area 1 at $t = 30$s until $t = 80$s, followed by a [-0.02, 0.02] p.u. random attack against the frequency measurement of area 3 between $220 - 260$s.

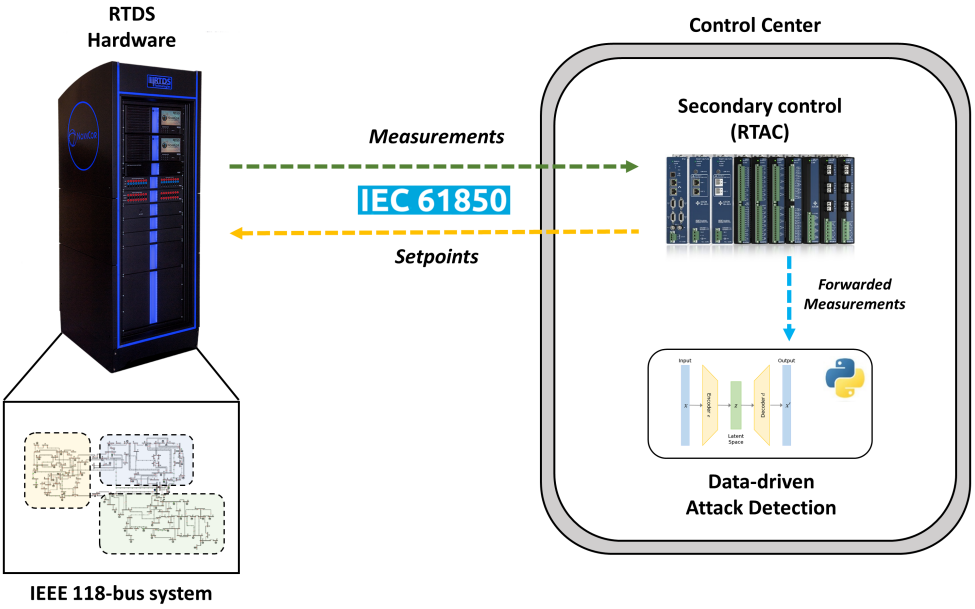


Figure 6.7 Diagram of the implemented HITL testbed in Use Case B.

6.1.5.3 Performance analysis

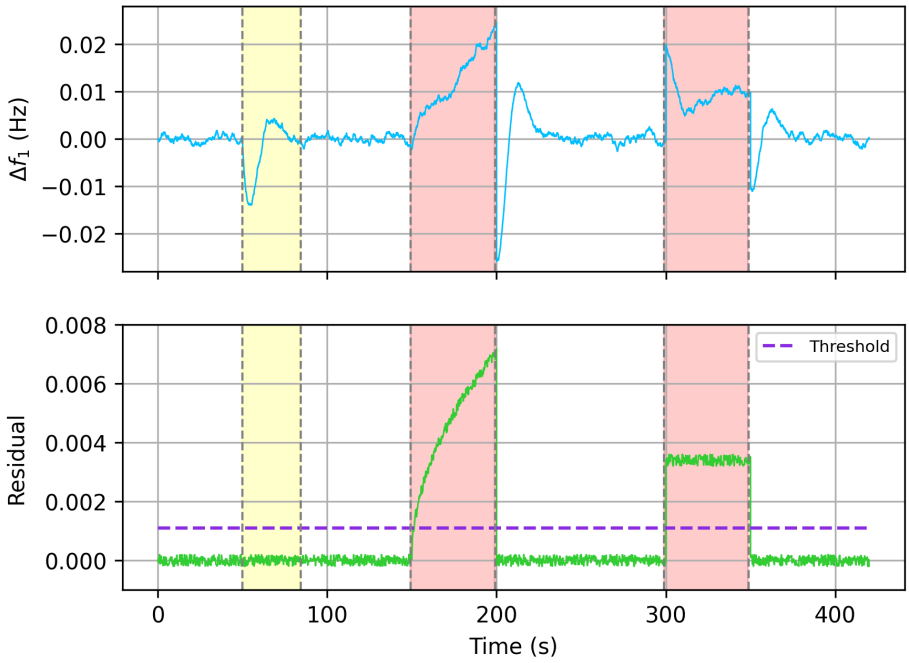
The proposed data-driven attack detection method for LFC is applied to Use Cases A and B to test its effectiveness. The results from this performance assessment are illustrated in Fig.

6.8. This Figure shows how the system frequency and the attack detection residual evolve over the simulation time of each use case. More specifically, the upper graph of Fig. 6.8a depicts the frequency response in the area 1 of Use Case A, while the lower graph shows the attack detection residual of Use Case A. Both of these variables are visualized in a common horizontal time axis. For Use Case B, the same information is depicted in Fig. 6.8b. The blue areas of Fig. 6.8 correspond to the time windows in which the system faces the transient effects of load disturbances while the red areas refer to the intervals in which cyberattacks are launched against the system.

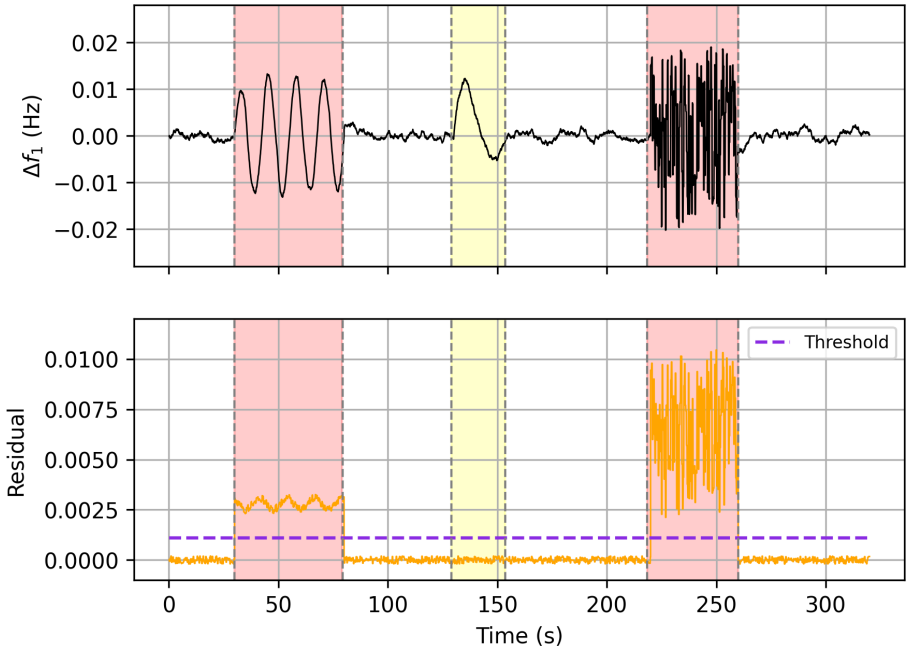
This performance assessment numerically verifies the effectiveness of the proposed data-driven attack detection mechanism. Particularly, the detection residual in Use Case A exceeds the selected predefined threshold during the ramp and step FDIAs (red areas in Fig. 6.8a) launched between $150-200$s and $300-350$s, respectively. On the contrary, the residual remains below the selected threshold during normal operation and under the transient effects of the load increase that starts at $t = 50$s (yellow area in Fig. 6.8a). This yields that the proposed detection strategy can successfully identify various types of FDIAs. Furthermore, this mechanism is capable of distinguishing if a frequency deviation is due to cyberattacks or another external incidents. Regarding Use Case B, similar conclusions are extracted from Fig. 6.8b. The effective performance of the presented methodology to various power systems ensures that it is scalable to large electrical grids and applicable to realistic conditions.

6.1.5.4 Sensitivity analysis on time delays

In real-world environments, various phenomena which are not captured by standard simulations can affect the system performance. Such phenomena involve latencies in the data exchange of the system caused by deficiencies in the communication mediums. To further investigate the applicability of the introduced methodology to realistic conditions, it is necessary to evaluate its sensitivity to these time delays. This sensitivity analysis is included in the present subsection. The analysis is conducted by deliberately injecting time delays ranging from 0.1 to 1 second into the communication channels that carry the measurements and control signals of the LFC systems implemented in Use Cases A and B. Fig. 6.9 presents the results of this analysis, where the detection residuals of each use case is illustrated under different amounts of time delays. The results show that the residuals in systems facing time delays exhibit similar behavior to those in latency-free systems, with minor shifts along the time axis. These slight latencies in the residuals are expected and they can be considered negligible for the scope of this application. Since time delays do not heavily impact the performance of the proposed cyber defense technique, its robustness against such phenomena is numerically verified.

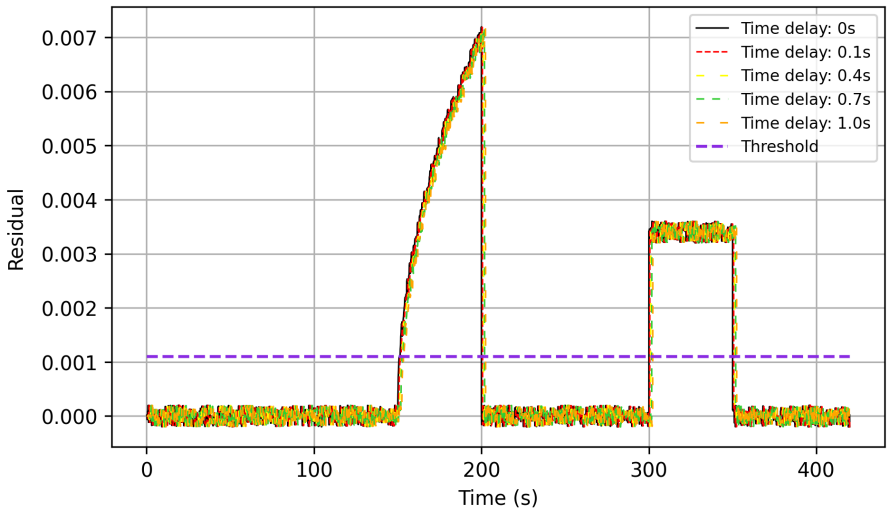


(a) Use Case A.

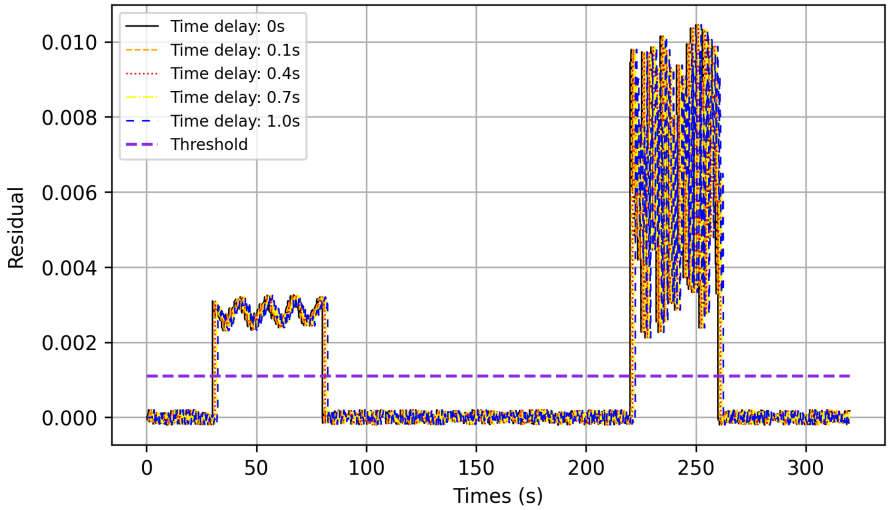


(b) Use Case B.

Figure 6.8 Performance of the proposed data-driven attack detection method.



(a) Use Case A.



(b) Use Case B.

Figure 6.9 Sensitivity analysis of the proposed data-driven attack detection method against times delays.

6.2 DNN-based Attack Recovery Mechanism

6.2.1 Motivation

According to [138], a natural attack response strategy for automation systems is the implementation of a backup control loop that is utilized under cyberattacks. This reserve control consists of a specially designed model that receives measurements from the investigated plant to produce an estimation of its healthy control signals. When a cyberattack is detected,

the original control system is temporarily discarded and replaced by this cyber defense mechanism which drives the protected plant with the generated control signal estimations. The spare control loop is activated for a short-term period, until the attack is mitigated and the normal functionality of the plant is fully restored. Inspired by this approach, a DNN-based attack recovery methodology is proposed in this thesis [139] that is particularly formulated for the case of LFC. A descriptive analysis of the presented cyber resilience technique is included in the remainder of this section.

6.2.2 Utilized model

As explained in Section 6.2.1, the objective of the proposed attack recovery mechanism is to produce an estimation of the control signals from the available measurements of the plant. Particularly, the introduced attack mitigation technique focuses on predicting the healthy ACE signals of one or more areas, as they form the basis of the LFC commands. This is a typical regression problem which involves the approximation of the underlying relationship between the LFC commands and one or more field measurements. To solve this task, a proper data-driven model is required. In the scientific field of AI, several data-driven algorithms have been developed to address regression problems. These models vary from simplistic ones, e.g. SVMs, to more sophisticated, such as LSTM networks, which demonstrate trade-offs between accuracy and complexity. To maintain an acceptable balance between these properties, a DNN architecture is considered appropriate to estimate the normal ACE commands and thus, it is selected as the model of this method.

6.2.3 Algorithm inputs

For large-scale, interconnected LFC systems, the ACE_i of area i is computed as the weighted sum of the frequency and area power export. In its general form, the ACE_i is computed as:

$$ACE_i = G\left(\Delta f_i, \quad \sum_{j=1, j\neq i}^{N} \Delta P_{ac_{ij}}, \quad \sum_{j=1, j\neq i}^{N} \Delta P_{dc_{ij}}\right) \tag{6.1}$$

where Δf_i is the frequency deviation in area i, $\Delta P_{ac_{ij}}$ represents the power deviation of the tie-lines between areas i and j if there are AC links between them, $\Delta P_{dc_{ij}}$ denotes the power deviation of the tie-lines between areas i and j if HVDC links between them exist [110], N is the total number of areas and G expresses the mathematical relationship between the aforementioned measurements and the ACE_i. When one or more frequency and tie-line power deviation measurements are missing or tampered, the ACE_i is miscalculated, leading

to malfunction of LFC and grid instability. One possible way to resolve this issue is to use alternative field measurements to estimate each ACE_i signal.

At this point, the goal is to find proper field measurements that will be used for the approximation of each ACE_i signal. According to subsection 3.1.4.2, the generators of an area that participate in the primary control of LFC use droop speed control for stable load division. Thus, the relationship between the local frequency and the generation of power plants in area i are inversely proportional to each other, depending on the type of the controllers and the dynamic constraints in the turbine-governor system. This relationship is mathematically expressed as $\Delta f_i = H(\Delta P_{G_i})$. Furthermore, the power export of an area is directly related with the generation of its local power plants. Based on the previous analysis, Eq. (6.1) can be written as:

$$ACE_i = G\left(\Delta H(\Delta P_{G_i}), \quad \sum_{j=1,j\neq i}^{N} \Delta P_{ac_{ij}}, \quad \sum_{j=1,j\neq i}^{N} \Delta P_{dc_{ij}}\right). \tag{6.2}$$

Eq. (6.2) indicates that it is reasonable to use the generation of a power plant as the input of the designed DNN in order to estimate each ACE_i, since there is an underlying relationship between them. The key aspect now is the selection of the proper local generator. The criterion for this decision is the similarity between the output of the selected generator and the ACE_i responses. Particularly, Fig. 6.10 shows the generation response of a power plant that participates in both primary and secondary control of the LFC ($\Delta P_{G_{sec}}$), the generation response of power plant that participates only in primary control ($\Delta P_{G_{pr}}$) and the ACE_i response to a 0.01 p.u. load increase. Based on this visualization, the field measurement that is closer to the oscillating ACE_i signal is the $\Delta P_{G_{pr}}$. Therefore, the output power of a generator that participates only in primary control, i.e. $\Delta P_{G_{pr}}$, is selected as the input of the DNN for the estimation of each ACE_i.

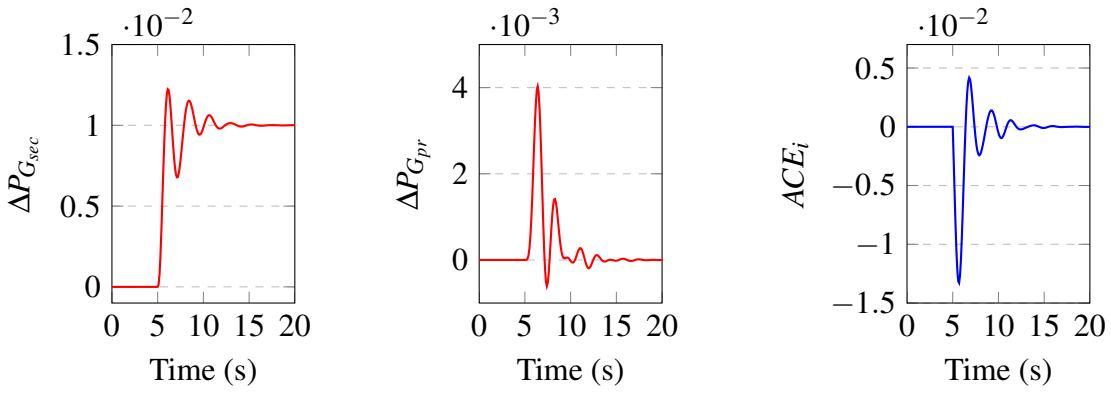


Figure 6.10 Generation and ACE responses to 0.01 p.u. load disturbance

6.2.4 Proposed attack recovery algorithm

In this subsection, the functionality of the proposed DNN-based attack mitigation strategy is analyzed in detail. This cyber resilience method is called DAR-LFC and its architecture is illustrated in Fig. 6.11. Before the activation of the DAR-LFC, a DNN is trained using the selected local generation measurements as the input data and the time-series of the ACEs in each area as the labeled data. After the training process, a spare telemetry system is installed (green dashed line in Fig. 6.11) for transmitting the output measurements from the selected local machine to the DNN. Then, the trained DNN is deployed in the control center ("DNN" module in Fig. 6.11) to estimate each normal ACE_i, i.e. $\widetilde{ACE}_i$, using the data that receives from the backup channel. Under normal conditions, generation is adjusted using the regular ACE_i, according to the original LFC specifications. When a cyberattack against LFC is identified ("Attack Detector" module in Fig. 6.11), the control center uses the $\widetilde{ACE}_i$ to regulate the generators that participate in the secondary control instead of the original ACE_i. Thus, the LFC operates with acceptable performance even in the presence of cyberattacks.

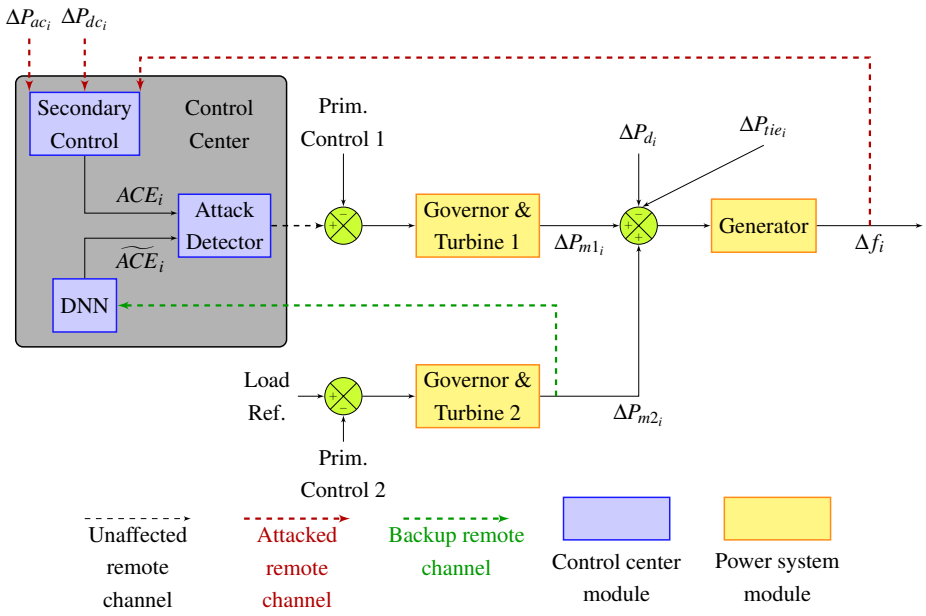


Figure 6.11 Schematic diagram of DAR-LFC for the i^{th} LFC area.

The DAR-LFC mechanism can estimate every ACE_i by using only one additional measurement, i.e. $\Delta P_{G_{pr}}$. This is one of the main advantages of the proposed methodology and it can be better demonstrated in large-scale, interconnected power grids, as it will be shown in the experiments. In these large systems, each ACE_i is computed using several measurements of frequency and tie-line power flows. If one or more of these measurements are tampered,

DAR-LFC can adequately estimated every ACE_i using a single input. Other similar works tackle this issue by approximating all the missing or affected measurements of the secondary control individually. For the approximation, these methods utilize algorithms like state estimation that require multiple inputs from the system.

Another benefit of DAR-LFC is its robustness against false positive cyberattack alarms. Sensitivity to these errors can lead to performance degradation and thus, it is a major issue when designing cyber defense strategies. If the utilized attack detector triggers an alarm that does not reflect to an actual cyber threat, the setpoints of the generators that participate in the secondary control will be adjusted by the estimated $\widetilde{ACE}_i$, as explained previously. However, the trained DNN can predict the healthy ACE_i signals with high accuracy. Since the LFC will be driven by an estimated $\widetilde{ACE}_i$ that is very close to the normal one, its operation will not be significantly affected. After the identification of the false positive flag, the LFC will use the original ACE_i signal and its functionality will be restored.

A weak point of DAR-LFC is the dependency of its additional communication channel to ICT. This makes it susceptible to cyber threats and opens another door for adversaries. However, the conducted literature review indicates that it is inevitable to design a data-driven attack mitigation mechanism for LFC without exposing the new parts of the system to cyber risks. For example, similar data-driven approaches [140], [141] require a historical database to feed their utilized models. In these works, it is implicitly assumed that these database servers are invulnerable to cyberattacks, which does not happen in practice [12]. Moreover, the collection, installation and maintenance of such large historical databases is far more expensive than the establishment of an extra telemetry system. Therefore, there is always a trade-off between the security that a data-driven model offers, the risks that introduces and its implementation costs.

Regarding the attack detection module of DAR-LFC, one possible solution is the utilization of the trained DNN. In this approach, the $\widetilde{ACE}_i$ generated by the deployed DNN is compared with the actual ACE_i and any significant deviation between them can be considered as a cyberattack. However, this approach cannot distinguish if a specific deviation is caused by a load change, cyberattack or other external disturbance. Moreover, it cannot identify the type of the cyberattack. Therefore, this technique has been intentionally dismissed. Instead, the autoencoder-based methodology described in Section 6.1 has been chosen for the attack identification module of DAR-LFC due to the benefits that brings. The assembly of these detection and mitigation mechanisms forms a solid data-driven framework that can significantly strengthen the cyber resilience of LFC against malicious activities.

6.2.5 Experimental results

In this Section, the results from the experimental tests of DAR-LFC are presented and discussed. This cyber resilience method is evaluated in two different LFC systems of varying complexity. The details of these power systems are described in the next subsection along with the simulated events (cyberattacks, disturbances, etc.) and their characteristics (time, magnitude, etc.). Then, the training process of the developed DNN is presented and the performance of the applied DAR-LFC is demonstrated. All the designed LFC systems and defence schemes are modeled in a MATLAB/Simulink [142] environment.

Table 6.2 Parameters of each power area

Parameter	Symbol	Value
Inertia constant	$2H_i$	0.1667 p.u. s
Damping coefficient	D_i	0.0083 p.u./Hz
Turbine time constant	T_{t_i}	0.3 s
Governor time constant	T_{g_i}	0.08 s
Governor regulating constant	R_i	2.4 Hz/p.u
Tie-line synchronize coefficient	T_{ij}	0.026 p.u./Hz
Frequency bias	β_i	0.425 p.u./Hz

6.2.5.1 Training of DNN

For the generation of the training dataset, several disturbances were simulated in order to make the neural network learn the underlying dynamics of the LFC. The diversity of these events assists the estimation model to perform well on data that it has not been trained on, i.e. to generalize. The testing and training datasets are composed of two vectors with 3600 datapoints, where each point of vector $\Delta P_{G_{pr}}$ should be mapped to the corresponding point of the vector ACE. Out of these 3600 datapoints, 70% were used for training and validation, and 30% were used for testing. The simulated training events are the following: the simulation starts at $t = 0$ sec, then at $t = 10$ sec a 5% step load increase occurs, at $t = 40$ sec a 10% step load decrease occurs, at $t = 80$ sec a 15% step load increase occurs, at $t = 120$ sec a 20% step load increase occurs and finally, at $t = 150$ sec the simulation ends.

For each use case a different neural network is trained. Each deployed neural network has a single input that corresponds to $\Delta P_{G_{pr}}$ and N outputs (N is the number of power system areas) that correspond to every ACE_i. For the optimization of the model performance, it is important to find the optimal architecture, i.e. finding the optimal hyper-parameters of each neural network, during the training phase. Hyper-parameters include the number of

hidden layers, number of neurons per layer, and the learning rate. For this purpose, the grid search technique [143] has been used as a systematic method of finding the optimal architecture of each neural network. The searching space and the selected, optimal values of the hyper-parameters are depicted in Table 6.3.

Table 6.3 Grid search values for hyper-parameter tuning

Hyper-parameter	**Search values**	**Optimal value**	
		Use case 1	Use case 2
Hidden layers (#)	1, 2, 5, 10	2	3
Neurons / layer (#)	10, 20, 50, 100	20	10
Optimizer	Adam, SGD	Adam	Adam
Learning rate	1e-2, 1e-3, 1e-4, 1e-5, 1e-6, 1e-7	0.0001	0.00001

The metric which is used to evaluate the performance of the proposed neural network is the MSE. Fig. 6.12 depicts the training progress of the neural network for use case 1. The improvement of this model is shown by the decrease in the MSE value for the testing data. In this way, it is verified that the proposed models can make a good estimation of the ACE of each area.

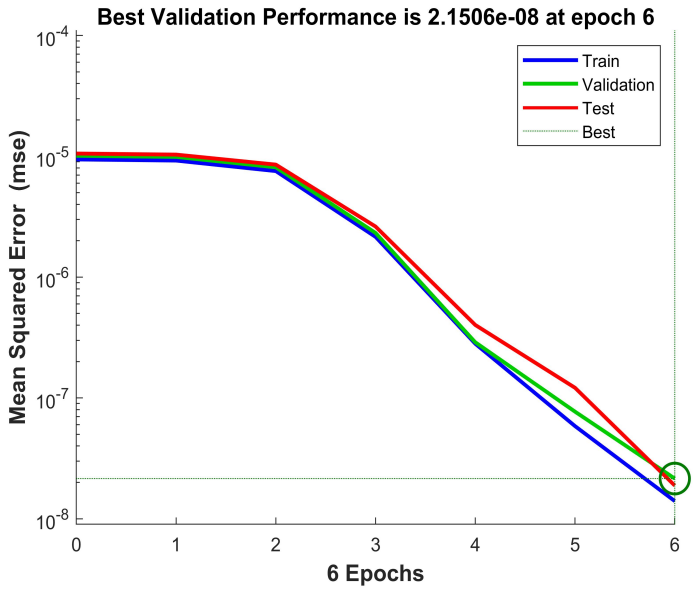


Figure 6.12 Performance validation of the proposed estimation model.

6.2.5.2 Evaluation of DAR-LFC

In this subsection, the experiments that verify the effectiveness of the proposed attack mitigation mechanism are presented. To illustrate the scalability of the introduced method, two different use cases of varying complexity are generated. Each of the following subsections is based on these use cases, where the impact of the considered cyberattacks against the LFC is shown, along with the restorative effects of the designed attack mitigation mechanism.

Use case 1: The power system of this use case is a single-area LFC whose parameters are given in Table 6.2. These parameters are tuned based on [144, 145]. It includes two generators with non-reheated turbines and the parameter $T_{ij} = 0$ since there is only one area. For the case of the DoS attack, a 10% p.u. step load disturbance at $t^1_{dist_1} = 5s$ is considered while there is no load change during FDIAs. The blue lines of Fig. 6.13 show that the frequency of this system remains within its preset value range when there is no attack and load disturbance; they also demonstrate that the frequency fluctuates until it reaches its nominal value in the event of a load disturbance but without attack.

For the DoS attack case, the attack is launched at $t^1_{dos} = 0s$ against the Δf. During this event, the ACE is missing and therefore, the LFC operates only with the primary control that results in a steady-state error, as shown in 6.13. For the FDIA cases, the scaling attack is launched at $t^1_{fdi_1} = 5s$ and the additive attack is launched at $t^1_{fdi_2} = 3s$, both against the Δf. The frequency is gradually increased under scaling attack and it constantly oscillates during additive attack as depicted in Fig. 6.13, due to the faulty ACE.

For these simulations, DoS attack is detected at $t^1_{det_1} = 1s$, scaling attack is detected at $t^1_{det_2} = 6.78s$ and additive attack is detected at $t^1_{det_3} = 5.93s$. After the detection, the DAR-LFC is activated and frequency evolves over time as illustrated in the green lines of Fig. 6.13. It is shown that using the estimated ACE in the LFC control loop, the frequency initially follows the changing trend caused by the cyberattack but, as soon as the attack is detected, it is gradually restored to its nominal value. The features of rise time, settling time, overshoot, etc. of the frequency response that results from the estimated ACE are different than the original one. However, these deviations are acceptable since the frequency is restored within a reasonable time. The green lines of Fig. 6.13 indicate that the proposed defence method is able to tackle the effects caused by the considered cyberattacks, highlighting its effectiveness.

Use case 2: In this case, the presented attack recovery mechanism is evaluated in a two-area LFC. In this way, the scalability of the proposed method is verified. The LFC parameters of each area are given in Table 1 [145]. Each area includes two generators with non-reheated turbines. The characteristics (type, time, magnitude) of the considered disturbances and

attacks are the same with the corresponding ones of use case 1, except that in this case the same attacks are launched against the ΔP_{tie} as well. All attacks affect area 1, i.e. Δf_1 and ΔP_{tie}.

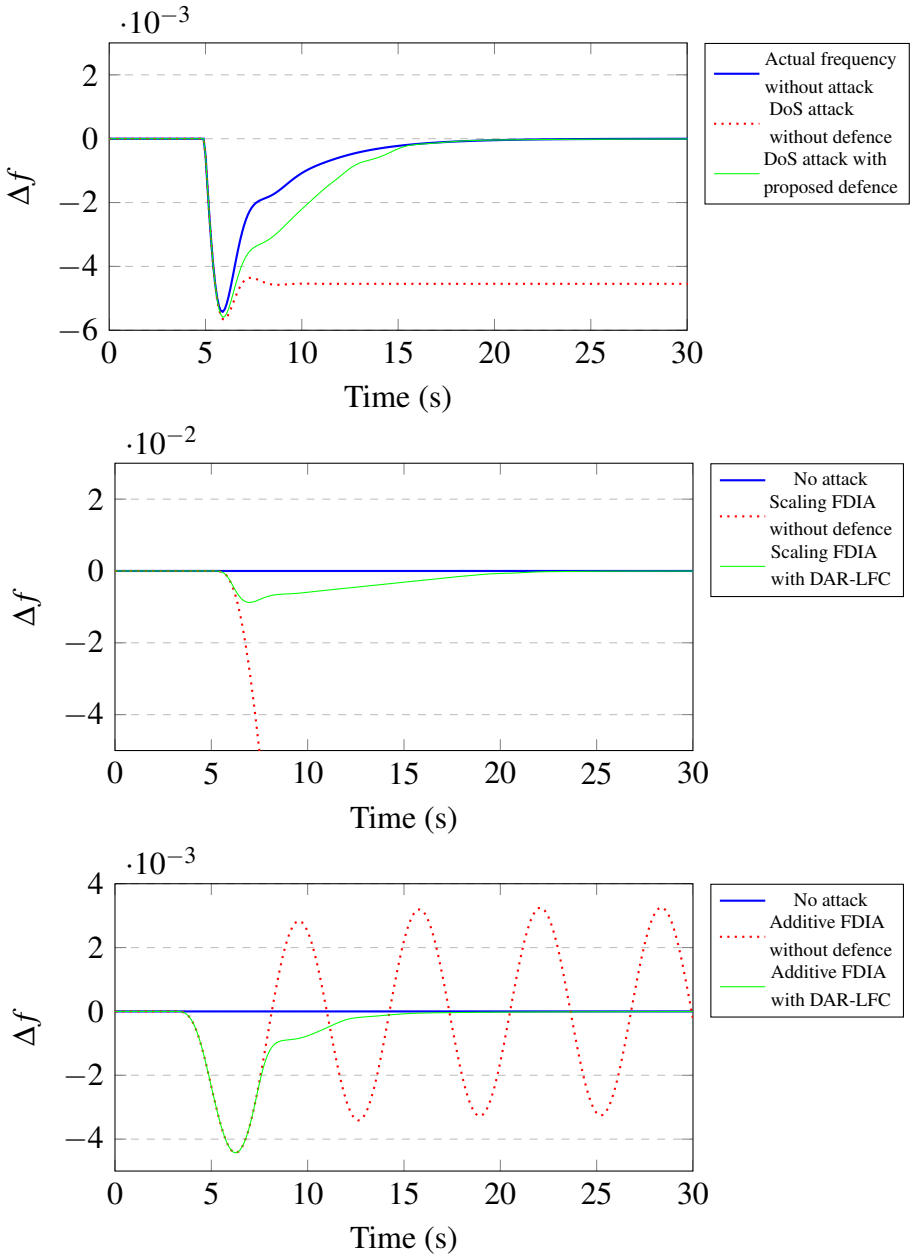


Figure 6.13 Performance evaluation - Use case 1

The effects of the considered cyberattacks are similar with the use case 1, besides the scaling attack where the frequency responses and tie-line power deviation are not gradually increasing (as in use case 1) but they are converging to a non-nominal value. This is

reasonable because a part of the impact of the cyberattacks is compensated by the secondary control of area 2, which is not affected by adversaries.

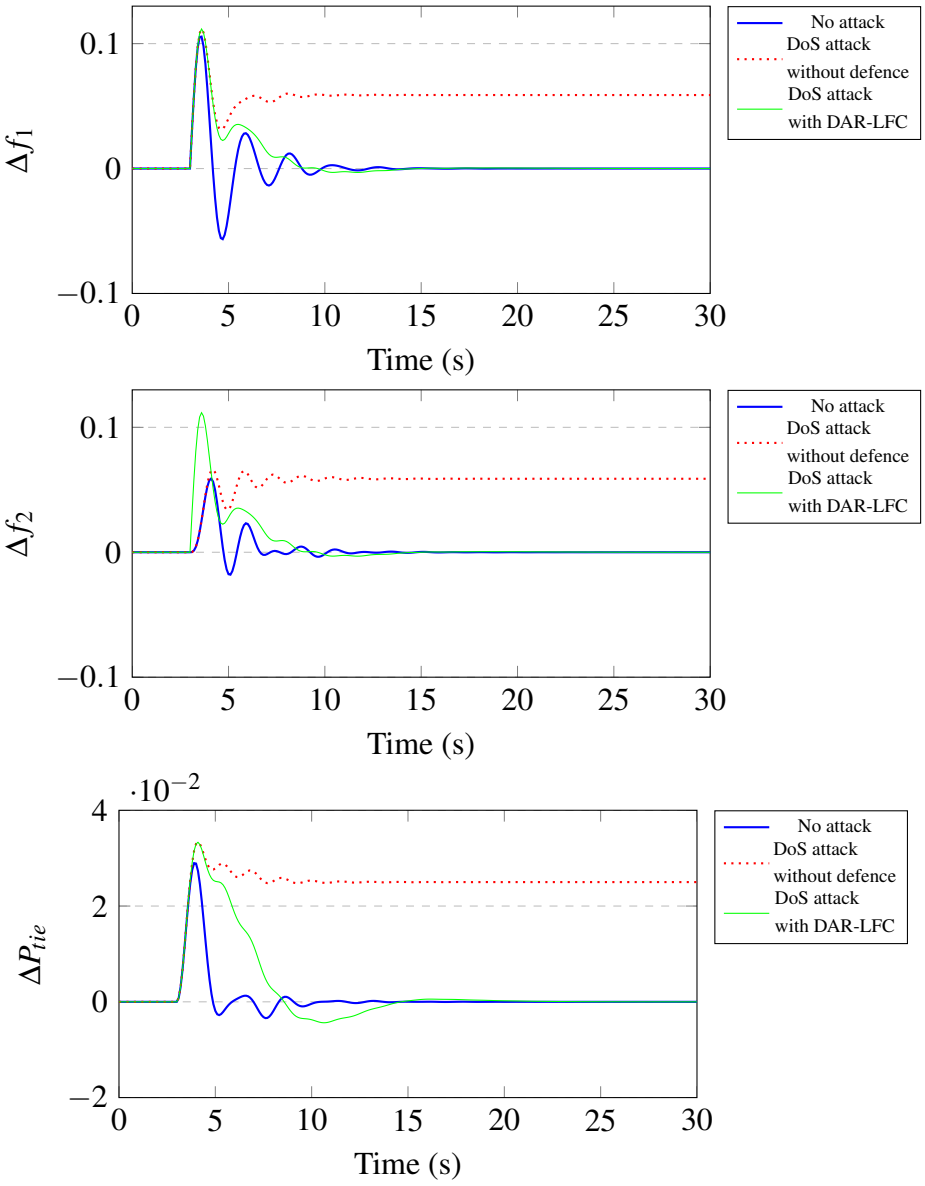


Figure 6.14 Performance evaluation - Use case 2 - DoS attack

The effectiveness of the proposed defence mechanism is illustrated in Fig. 6.14–6.16. The green lines of these figures demonstrate that the frequency response of each area and the tie-line power deviation are restored with minimum deviations in an acceptable time. The main advantage of DAR-LFC is highlighted in this Section; the neural network can estimate the ACE of each area using only an additional measurement, that is the generation response of a selected power plant that does not participate in secondary control of the LFC. Therefore,

the proposed method requires less inputs than other standard approximation methods, e.g. state estimation, in order to compute the affected measurements of the ACE_i.

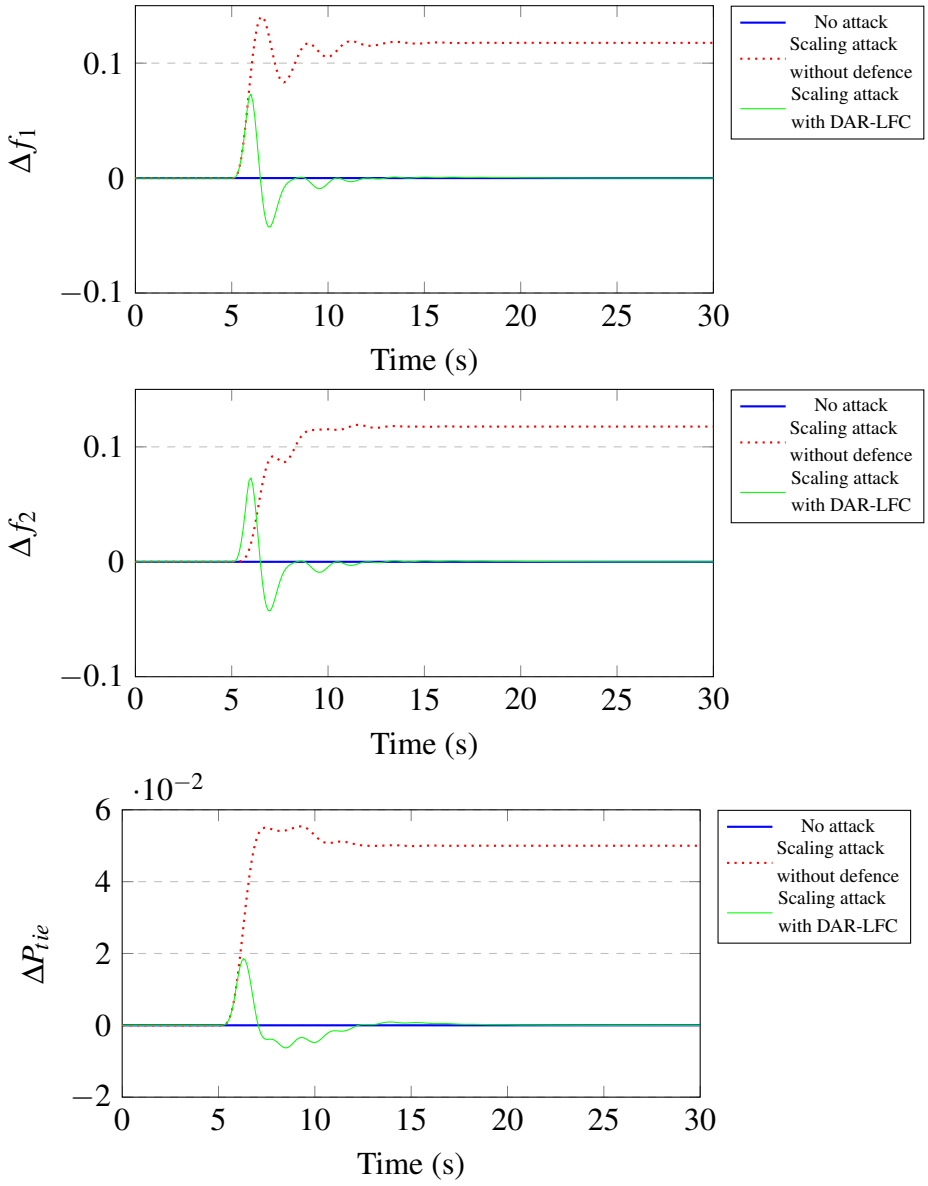


Figure 6.15 Performance evaluation - Use case 2 - Scaling attack

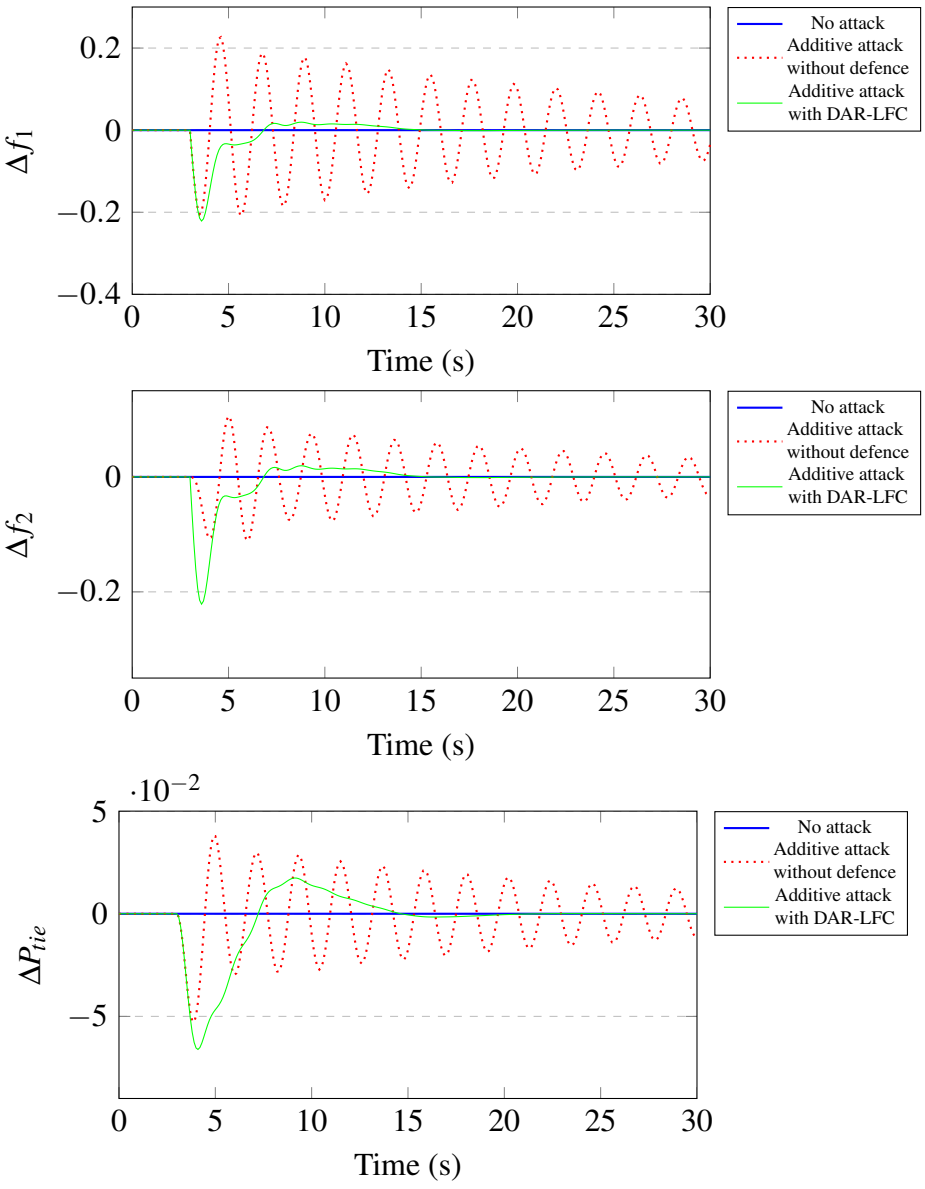


Figure 6.16 Performance evaluation - Use case 2 - Additive attack

Chapter 7

Conclusions & Future Work

In the present thesis, the cyber resilience of power systems frequency control and its enhancement has been thoroughly investigated. Inspired by the identified challenges, this thesis aims to bridge the existing research gaps and introduce significant innovations to the investigated research field. The design process and the experimental evaluation of the proposed methodologies have led to the discovery of various conclusive insights and revealed a series of directions for future research, which are presented in this final chapter.

7.1 Conclusions on Proposed Framework

In this Section, the conclusions about the introduced hybrid framework are thoroughly discussed. The analysis starts by comparing the two main categories of the proposed methodologies on various performance aspects. The results of this comparative study are then used to develop a set of configuration instructions for system operators. These instructions assist the assembly of a framework variant that is adjusted to the characteristics of the protected system.

7.1.1 Comparison between observer-based & data-driven approaches

This subsection offers a comparative analysis between observer-based & data-driven approaches for enhancing the cyber resilience of LFC. The conclusions have been drawn from the designing, implementation and testing of the various methodologies proposed in this thesis. The categories of the introduced methodologies are compared on several aspects that determine their overall performance. A detailed study of these aspects is presented in what follows:

- **Methodology Effectiveness**: as demonstrated by the experimental simulations, the proposed observer-based techniques are highly effective. However, these approaches are functional under specific conditions, due to their sensitivity in design matters such as model accuracy and threshold selection. While data-driven algorithms show similar results, their operation relies only on the historical system data. This information encapsulates more realistic system characteristics, such as nonlinearities, noisy signals, time delays, etc. Through their training process, these algorithms learn all potential operating points of the system and become model-independent. Therefore, data-driven techniques are universally effective, contrary to observer-based approaches.

- **Operational Requirements**: the observer-based category is a subset of model-based approaches and as a result, the algorithms of these techniques are based on mathematical models. If the LFC dynamics are effectively captured by the utilized models, these methods can achieve high levels of performance. However, electrical grids are complex systems that constantly evolve, which often leads to model inaccuracies. On the other hand, data-driven methodologies are robust against model uncertainties as they require only a descriptive dataset to operate. Nevertheless, the availability of the necessary data is not always guaranteed, which makes the data collection process a challenging task.

- **Computational Requirements**: once observer-based methodologies are mathematically established, their proposed algorithms are implemented either as software or hardware applications. The outputs of these algorithms are computed by numerically solving differential equations. The resulting formulas are typically low-order, ordinary, linear differential equations and thus, their solutions can be easily calculated. In data-driven approaches, the main computational bottleneck is caused by their training procedures. This is due to the complexity of the research problem, the sophisticated architecture of the utilized deep learning models and the need for online training of these algorithms. From the previous analysis, it is concluded that the proposed observer-based methods outperform the data-driven ones in terms of computational requirements.

For better comprehension, the conclusions drawn from the previous study are summarized in Table 7.1. This Table illustrates which cyber resilience category of LFC is the most superior in each of the aforementioned performance aspects. In this way, the benefits and the drawbacks of each LFC cyber resilience category can be easily extracted. The information included in Table 7.1 is used to establish the configuration guideline presented in subsection

7.1.2. This guideline assists system operators to select the most appropriate cyber resilience category for LFC in each layer of the proposed hybrid framework.

Table 7.1 Comparative study between observer-based and data-driven approaches for the cyber resilience enhancement of LFC.

	Methodology Performance	**Operational Requirements**	**Computational Requirements**
Observer-based	Conditionally High	Model	Low
Data-driven	Universally High	Data	High

7.1.2 Guidelines for framework configuration

The comparative study presented in subsection 7.1.1 can be used to extract a series of instructions regarding the configuration of the designed hybrid framework. These instructions serve as a guideline for system operators to determine which of the developed cyber resilience categories is the most appropriate at each layer of the framework. In this way, the introduced framework can be adjusted to the special characteristics of the LFC system that is applied to. For example, if immediate identification of cyberattacks is crucial in a power system and historical grid data are available, a data-driven method is highly recommended for the attack detection layer. Then, the observer-based approaches can be utilized for the attack estimation and the mitigation layers to balance the computational requirements of the framework. The extracted guidelines for the proper configuration of the proposed framework are summarized in what follows:

- if historical LFC data are available and their quality is acceptable, then the data-driven methods are appropriate for the investigated layer.
- if the quality of the available LFC system model is high, then observer-based methods fit better to the considered layer.
- if the top priority in a cyber resilience layer is its performance accuracy, data-driven methods are more suitable, assuming that historical LFC data are available.
- if the lightweight implementation of a cyber resilience layer is prioritized, the indicative category is the observer-based methods, considering the quality of the available LFC model.

7.2 Future Research

Despite the significant efforts of the present thesis to strengthen the cyber resilience of LFC, it is highly challenging to address every issue of this vast research field. As a result, there is still room for further development. To this end, this final section concludes the thesis with a series of recommendations for future research, based on the results of the present study and the challenges that have emerged throughout the research process. The suggestions for future research include:

- the design of a computational formula that will specify which cyber resilience category fits better at each framework layer. While the guidelines presented in subsection 7.1.2 provide a solid foundation for framework configuration, they fail to capture the full spectrum of conditions and challenges the LFC may encounter. By developing a specialized metric that can quantify the available LFC resources and the determined cybersecurity objectives, system operators would be better equipped to make informed decisions when configuring the framework. This formula would serve as a critical tool, offering a more nuanced and tailored approach towards enhancing the overall cyber resilience of LFC.

- the development of attack-resilient control strategies based on reinforcement learning techniques. Reinforcement learning is a machine learning paradigm where the models learn the desired behavior from their environments. In case of power system automation, the frequency controller, which acts as the agent of the reinforcement learning algorithm, aims to minimize the impact of cyberattacks within a dynamic LFC environment. This strategy does not require any additional control input and can mitigate attack patterns that has not encountered during the training phase. This introduces the generalization ability to the attack-resilient control mechanisms.

- the deployment of a digital twin to identify cyberattacks in LFC and recover it from such malicious activities. This technology is a digital representation of the power grid and its automation, allowing the system operators to simulate realistic situations of these infrastructures and obtain the results. Digital twin receives inputs from the physical system and then, reproduces the attack-free operation of the power grid. If the actual and the simulation results demonstrate specific patterns, then successfully launched cyberattack attack can be identified. If a cyberattack has been detected, the attack-free simulation results of the digital twin can replace the compromised signals of LFC to restore its functionality.

- the design of sophisticated cyberattacks using optimization algorithms or game-theoretic approaches. In case of power systems, these algorithms can produce optimal attack strategies against LFC, based on the specifications its operation, in order to stealthily cause the maximum damage to the power infrastructure. While these methodologies do not directly offer cybersecurity protection to the LFC system, they can be leveraged to understand the attack patterns that the adversaries follow. By studying these motifs, the system operators can develop counter-strategies that identify stealthy cyberattacsks and neutralize their impact on the system under protection.

Chapter 8

Extensive Summary in Greek

8.1 Κεφάλαιο 1

Στο κεφάλαιο 1 γίνεται μια εισαγωγική παρουσίαση των βασικών εννοιών με τις οποίες ασχολείται η παρούσα διατριβή. Άρχικά, ένας από τους λόγους για τους οποίους είναι απαραίτητη η ενεργειακή μετάβαση, είναι ότι παρατηρείται σε παγκόσμια κλίμακα μία διαρκώς αυξανόμενη ζήτηση για πιο ποιοτικές υπηρεσίες ηλεκτρικής ενέργειας. Αυτό με τη σειρά του επιτείνει την ανάγκη για πιο αξιόπιστα, ασφαλή και περιβαλλοντικά φιλικά συστήματα ηλεκτρικής ενέργειας. Για τον σκοπό αυτό, οι περισσότεροι διεθνείς ερευνητικοί και βιομηχανικοί οργανισμοί (π.χ. Η.Π.Α., Ε.Ε., Κίνα, Αυστραλία, κλπ.) που ασχολούνται με την ηλεκτρική ενέργεια, επικεντρώνουν τις προσπάθειές τους στο να κάνουν τα δίκτυα ηλεκτρικής ενέργειας πιο "έξυπνα" [3, 4]. Με αυτόν τον τρόπο, τα δίκτυα ηλεκτρικής ενέργειας μπορούν να προσαρμοστούν πιο αποτελεσματικά στις ανάγκες όλων των χρηστών τους, δηλαδή και των παραγωγών και των καταναλωτών ενέργειας.

Τα Ευφυή Δίκτυα Ηλεκτρικής Ενέργειας (SGs) είναι δίκτυα ηλεκτρικής ενέργειας που χρησιμοποιούν προηγμένες τεχνολογίες πληροφορικής και επικοινωνιών (ICT) όπως αισθητήρες, εφαρμογές λογισμικού, υπολογιστικά δίκτυα και ανάλυση δεδομένων για την παροχή αποτελεσματικών και βιώσιμων υπηρεσιών ενέργειας. Οι ICT διευκολύνουν την παρακολούθηση και τον έλεγχο του ηλεκτρικού δικτύου σε σύγκριση με τις συμβατικές υποδομές ενέργειας. Προσφέρουν καλύτερη εποπτεία της κατάστασης του δικτύου και ρυθμίζουν τη λειτουργία του με βέλτιστο τρόπο. Ενώ οι ICT προσφέρουν μια πληθώρα από πλεονεκτήματα στο δίκτυο ηλεκτρικής ενέργειας, ταυτόχρονα το εκθέτουν σε σοβαρούς κινδύνους κυβερνοασφάλειας [5, 6]. Ο ψηφιακός μετασχηματισμός του δικτύου ηλεκτρικής ισχύος δημιουργεί διάφορα τρωτά σημεία κυβερνοασφάλειας στο σύστημα, τα οποία με τη σειρά τους επιτρέπουν σε κακόβουλους χρήστες να εξαπολύσουν μια ευρεία γκάμα κυβερνοεπιθέσεων εναντίον τους.

Ένας σημαντικός δείκτης αξιοπιστίας και ασφάλειας των δικτύων ηλεκτρικής ενέργειας είναι η ανθεκτικότητα τους. Η ανθεκτικότητα είναι μία από τις πιο σημαντικές ιδιότητες του δικτύου ηλεκτρικής ενέργειας, καθώς εξασφαλίζει τη συνεχή παροχή ηλεκτρικής ισχύος προς του καταναλωτές. Σύμφωνα με το [19], η ανθεκτικότητα στα συστήματα ηλεκτρικής ενέργειας ορίζεται ως η ικανότητα ενός συστήματος να αντέχει, να απορροφά και να ανακάμπτει άμεσα από μια εξωγενές καταστροφικό γεγονός που χαρακτηρίζεται από υψηλή επίπτωση αλλά χαμηλή πιθανότητα. Καθώς τα ηλεκτρικά συστήματα εξελίσσονται με γοργούς ρυθμούς, νέοι τύποι ανεπιθύμητων γεγονότων επηρεάζουν την ανθεκτικότητά τους, όπως για παράδειγμα οι κυβερνοεπιθέσεις. Επομένως, είναι κρίσιμο να επανεξεταστεί η συνήθης έννοια της ανθεκτικότητας του ηλεκτρικού συστήματος προκειμένου να συμπεριληφθεί και η επίδραση αυτών των ανερχόμενων κινδύνων. Προς αυτό το σκοπό, ο ορισμός της ανθεκτικότητας που παρέχεται από το [19] επεκτείνεται στο [1] προκειμένου να περιλαμβάνει τον κυβερνοχώρο των Έξυπνων Δικτύων.

Ανάμεσα στους διάφορους μηχανισμούς ελέγχου ενέργειας που διευκολύνονται από τις ICT, ο έλεγχος φορτίου-συχνότητας (LFC) είναι ένας από τους πιο σημαντικούς. Ο ρόλος του LFC είναι η διατήρηση της ενεργειακής ισορροπίας μεταξύ παραγωγής και ζήτησης στα ηλεκτρικά δίκτυα για την πρόληψη οποιασδήποτε υποβάθμισης της απόδοσης του συστήματος. Ένα κλειδί δείκτης της ενεργειακής ισορροπίας είναι η απόκλιση της συχνότητας από την ονομαστική της τιμή, όπως φαίνεται στο Σχήμα 1.2. Για να διατηρηθεί το ισοζύγιο ισχύος στο σύστημα, το LFC λαμβάνει μετρήσεις συχνότητας από το δίκτυο ενέργειας, υπολογίζει το απαραίτητα σήμα ελέγχου και το στέλνει στις γεννήτριες για να ρυθμίσουν αναλόγως την παραγωγή τους. Εφόσον ορισμένες από τις κύριες λειτουργίες του LFC γίνονται με χρήση ICT, το συγκεκριμένο σύστημα ελέγχου κρίνεται ευάλωτο σε κυβερνοεπιθέσεις. Η σπουδαιότητα του LFC κάνει σημαντική την ανάπτυξη προηγμένων στρατηγικών κυβερνοάμυνας προκειμένου να διασφαλιστεί κυβερνοανθεκτικότητα του [27].

Μια σειρά από αποτελεσματικά εργαλεία κυβερνοάμυνας ενάντια στους ψηφιακούς κινδύνους που απειλούν το LFC είναι η ανίχνευση κυβερνοεπιθέσεων, η εντοπισμός κυβερνοεπιθέσεων, η εκτίμηση κυβερνοεπιθέσεων και οι ανθεκτικοί-σε-κυβερνοεπιθέσεις μηχανισμοί ελέγχου. Η ανίχνευση επιθέσεων προσδιορίζει εάν ένα σύστημα ηλεκτρικής ενέργειας έχει δεχθεί κυβερνοεπίθεση, καθώς επίσης και το πότε αυτή πραγματοποιήθηκε. Ο εντοπισμός επιθέσεων προσδιορίζει ποιά τμήματα του συστήματος (αισθητήρες, ελεγκτές κλπ.) έχουν δεχτεί κυβερνοεπίθεση. Η εκτίμηση επιθέσεων είναι ένα εργαλείο που εφαρμόζεται μετά την πραγματοποίηση μιας κυβερνοεπίθεσης στο σύστημα και παρέχει λεπτομερής πληροφορίες σχετικά με το είδος και την κυματομορφή της, όπως

είναι η ένταση, η συχνότητα, η μορφή κ.λπ. Η εκτίμηση επιθέσεων μπορεί επίσης να αποτελέσει το θεμέλιο λίθο για το σχεδιασμό ενός ανθεκτικού-σε-κυβερνοεπιθέσεις μηχανισμού ελέγχου για το **LFC**. Ο ανθεκτικός-σε-κυβερνοεπιθέσεις έλεγχος είναι ένας προηγμένος μηχανισμός που εξαλείφει τις επιπτώσεις των κυβερνοεπιθέσεων ενάντια στο εξεταζόμενο σύστημα ελέγχου, ώστε να διατηρηθεί η λειτουργικότητα του δικτύου ηλεκτρικής ενέργειας ακόμη κάτω από συνθήκες κακόβουλων δραστηριοτήτων. Κάθε μια από αυτές τις μεθοδολογίες μπορεί να εφαρμοστεί ακολουθιακά, με τη σειρά που παρουσιάστηκαν, προκειμένου να σχηματιστεί ένας πολυεπίπεδο μηχανισμό κυβερνοασφάλειας που μπορεί να αναγνωρίσει και να αντιμετωπίσει κακόβουλες δραστηριότητες εναντίον των συστημάτων ηλεκτρικής ισχύος.

Έπειτα από εκτενή βιβλιογραφική μελέτη των μεθόδων ανίχνευσης, εντοπισμού, εκτίμησης και εξάλειψης κυβερνοεπιθέσεων, χαρτογραφήθηκαν τα πλεονεκτήματα και οι περιορισμοί των προτεινόμενων εργασιών προκειμένου να αναγνωριστούν τα κενά στον συγκεκριμένο ερευνητικό τομέα. Η παρούσα διδακτορική διατριβή συμβάλει στην κάλυψη αυτών των ερευνητικών κενών με τον σχεδιασμό ενός καινοτόμου πλαισίου που συνδυάζει μεθόδους ανίχνευσης, εντοπισμού, εκτίμησης και εξάλειψης κυβερνοεπιθέσεων οι οποίες βασίζονται σε παρατηρητές ολίσθησης (**SMOs**) και μοντέλα μηχανικής μάθησης. Στα κεφάλαια που ακολουθούν, γίνεται διεξοδική παρουσίαση και ανάλυση των προτεινόμενων μεθόδων κυβερνοάμυνας προκειμένου να επιβεβαιωθεί θεωρητικά η ψηφιακή θωράκιση των συστημάτων ηλεκτρικής ενέργειας μέσα από αυτές. Η θεωρητική ανάλυση ακολουθείται πειραματική εφαρμογή των προτεινόμενων μεθόδων σε ρεαλιστικά συστήματα έτσι ώστε να επιβεβαιωθεί και αριθμητικά η αποτελεσματικότητά τους.

8.2 Κεφάλαιο 2

Στο κεφάλαιο 2 παρουσιάζονται οι βασικές αρχές κυβερνοασφάλειας των κυβερνοφυσικών συστημάτων (**CPSs**). Αρχικά, είναι σημαντικό να αναλυθεί συνοπτικά το τρίγωνο της κυβερνοασφάλειας (ή αλλιώς τριάδα **CIA**), το οποίο αποτελεί το πιο σημαντικό μοντέλο περιγραφής ενός μηχανισμού προστασίας υπολογιστικών συστημάτων. Η τριάδα **CIA** περιγράφει συνοπτικά τους τρεις βασικούς στόχους που πρέπει να πληρούνται για την προστασία ενός συστήματος **ICT**: την εχεμύθεια, την διαθεσιμότητα και την ακεραιότητα. Αυτοί οι τρεις στόχοι αποτελούν τα θεμέλια της κυβερνοασφάλειας και καθορίζουν τις βασικές αρχές που πρέπει να λαμβάνονται υπόψη κατά την ανάπτυξη και τη συντήρηση συστημάτων που χρησιμοποιούν **ICT**. Η τριάδα **CIA** αποτελεί επίσης ένα πλαίσιο ανάλυσης κινδύνων και λήψης αποφάσεων στον τομέα της κυβερνοασφάλειας. Παρέχει κατευθυντήριες γραμμές στους επαγγελματίες του τομέα ώστε να διατηρούν μια ισορροπία

ανάμεσα στους τρεις βασικούς στόχους με βάση τις ανάγκες και τις απειλές που αντιμετωπίζει ένας οργανισμός ή ένα σύστημα. Συνοπτικά, το τρίγωνο της κυβερνοασφάλειας αποτελείται από τα εξής στοιχεία:

- Ακεραιότητα: Αφορά τη διατήρηση της ακεραιότητας των πληροφοριών, προστατεύοντας τες από μη εξουσιοδοτημένες τροποποιήσεις. Ο στόχος είναι να παραμένουν τα δεδομένα του συστήματος διαρκώς ακέραια και ανέπαφα.

- Διαθεσιμότητα: Εστιάζει στη διασφάλιση της διαθεσιμότητας και της προσβασιμότητας των πληροφοριών και των πόρων όταν αυτές χρειάζονται. Επομένως, ο στόχος είναι η αντιμετώπιση απειλών που μπορεί να επηρεάσουν την πρόσβαση ή τη χρήση των πόρων.

- Εχεμύθεια: Αναφέρεται στη διατήρηση του απόρρητου χαρακτήρα των πληροφοριών. Ο στόχος είναι οι πληροφορίες που ανταλλάσσονται να παραμένουν κρυφές από μη εξουσιοδοτημένα άτομα ή συστήματα.

Στα κλασικά δίκτυα υπολογιστών, όπως τα δίκτυα βάσεων δεδομένων εταιρειών, δίκτυα διακομιστών ιστού κ.λπ., οι κυβερνοαπειλές στοχεύουν κυρίως στην παραβίαση του απορρήτου των δεδομένων και στη διακοπή των εξουσιοδοτημένων προσβάσεων στους υπολογιστικούς πόρους. Για παράδειγμα, οι κυβερνοεισβολείς συνήθως προσπαθούν να κλέψουν τις πληροφορίες που αποθηκεύονται στο δικτυακό σύστημα της βάσης δεδομένων ενός τραπεζικού ιδρύματος ή να διακόψουν την κανονική λειτουργία ενός διακομιστή ιστού για να απαιτήσουν λύτρα. Συνεπώς, οι κυβερνοαπειλές σε αυτήν την περίπτωση στοχεύουν να πλήξουν την εμπιστευτικότητα και τη διαθεσιμότητα των συστημάτων, αναφορικά με την τριάδα CIA.

Στην περίπτωση των σύγχρονων δικτύων ηλεκτρικής ενέργειας, τα μέρη που είναι τρωτά σε κυβερνοεπιθέσεις είναι τα συστήματα παρακολούθησης και ελέγχου. Αυτό συμβαίνει διότι τα συστήματα αυτά χρησιμοποιούν τηλεπικοινωνιακές υποδομές καθώς και εφαρμογές λογισμικού/υλικού υπολογιστών για την ανταλλαγή δεδομένων. Οι κυβερνοεισβολείς στοχεύουν είτε να τροποποιήσουν το περιεχόμενο των δεδομένων που ανταλλάσσονται είτε να εμποδίσουν την κανονική μεταφορά πληροφοριών στα συστήματα αυτοματισμού των SGs. Με αυτό τον τρόπο, οι κυβερνοεπιθέσεις πλήττουν την ευστάθεια των συστημάτων ηλεκτρικής ενέργειας. Επομένως, οι βασικοί στόχοι κυβερνοασφάλειας που απειλούνται στην περίπτωση των SGs είναι η ακεραιότητα και η διαθεσιμότητα, αναφορικά με την τριάδα CIA.

Για τον εντοπισμό των τρωτών (όσον αφορά την κυβερνοασφάλεια) σημείων ενός σύγχρονου δικτύου ηλεκτρικής ενέργειας, είναι χρήσιμο να αναλυθούν τα διάφορα μέρη

των συστημάτων τηλεμετρίας και απομακρυσμένου ελέγχου. Τα επιμέρους αυτά στοιχεία φαίνονται στο σχήμα 2.6 και περιγράφονται συνοπτικά παρακάτω:

- Αισθητήρες: Πρόκειται για συσκευές πεδίου που μετρούν περιοδικά σημαντικές μεταβλητές του φυσικού συστήματος. Η λειτουργία και η ρύθμιση τους πραγματοποιείται συνήθως πάνω σε εξειδικευμένο υλικό υπολογιστών και μέσω ενός λιτού περιβάλλοντος λογισμικού.

- Κανάλια Μέτρησης: Πρόκειται για κανάλια τηλεμετρίας που είναι υπεύθυνα για τη μεταφορά των μετρήσεων από τους αισθητήρες στο κέντρο ελέγχου. Η υλοποίησή τους εξαρτάται από τις εφαρμογές για τις οποίες σχεδιάστηκαν και την αρχιτεκτονική του χρησιμοποιούμενου πρωτοκόλλου επικοινωνίας.

- Κέντρο Ελέγχου: Αποτελεί τον "εγκέφαλο" ενός συστήματος αυτοματισμού. Το κέντρο ελέγχου λαμβάνει μετρήσεις από τους αισθητήρες και τις επεξεργάζεται αναλόγως για να σχηματίσει τις εντολές ελέγχου. Οι αλγόριθμοι που έχουν σχεδιαστεί για τον έλεγχο των ηλεκτρικών συστημάτων εκτελούνται μέσω εφαρμογών λογισμικού οι οποίες λειτουργούν εντός του κέντρο ελέγχου.

- Κανάλια Εντολών: Πρόκειται για κανάλια επικοινωνίας που είναι υπεύθυνα για τη μεταφορά των εντολών ελέγχου από το κέντρο ελέγχου προς το σύστημα ηλεκτρικής ενέργειας. Η υλοποίησή τους είναι παρόμοια με τα κανάλια μέτρησης.

- Ρυθμιστές: Πρόκειται για συσκευές που μετατρέπουν τις εντολές ελέγχου σε σήματα κατάλληλα για τη ρύθμιση των αντίστοιχων μεταβλητών του ηλεκτρικού συστήματος. Οι ενέργειες των ρυθμιστών υλοποιούνται συνήθως μέσω μηχανικών, υδραυλικών ή ηλεκτρονικών συσκευών.

Υπάρχουν διάφορα είδη κυβερνοεπιθέσεων που απειλούν ένα σύστημα SG. Ορισμένα από αυτά απαντώνται και στα κλασικά συστήματα ICT, ενώ κάποια άλλα εμφανίζονται μόνο στα CPSs. Στην παρούσα διατριβή αναλύονται τα πιο σημαντικά είδη κυβερνοεπιθέσεων εναντίον των CPSs. Ένα από αυτά τα είδη είναι η κυβερνοεπίθεση άρνησης παροχής υπηρεσιών (DoS). Πιο συγκεκριμένα, μια κυβερνοεπίθεση DoS κατά των SGs στοχεύει στη διακοπή της κανονικής λειτουργίας ή στην υπερφόρτωση των συστημάτων ICT που χρησιμοποιεί το δίκτυο ενέργειας. Αυτό έχει ως αποτέλεσμα, τα συστήματα ICT που είναι υπεύθυνα για την υλοποίηση των αλγορίθμων απομακρυσμένου ελέγχου να καθίστανται προσωρινά ή μόνιμα μη διαθέσιμα. Στο πλαίσιο των SGs, αυτό θα μπορούσε να διαταράξει την ευστάθεια του συστήματος, να προκαλέσει δυσκολίες στην παρακολούθηση του, καθώς και να δημιουργήσει μεγάλες οικονομικές ζημιές στα νοικοκυριά

και στους διαχειριστές του ηλεκτρικού συστήματος. Σε όρους του τριγώνου κυβερνοασφάλειας, οι κυβερνοεπιθέσεις DoS απειλούν κυρίως τη διαθεσιμότητα των δεδομένων που ανταλλάσσονται σε ένα SG.

Ένας άλλος σημαντικός τύπος κυβερνοεπιθέσεων που απαντώνται στα CPSs, είναι είναι οι κυβερνοεπιθέσεις έγχυσης χρονικών καθυστερήσεων (TDAs). Για να διατηρηθεί η ευστάθεια ενός συστήματος ηλεκτρικής ενέργειας, είναι απαραίτητο οι μετρήσεις και οι εντολές ελέγχου που ανταλλάσσονται να μεταφέρονται εγκαίρως. Η παρουσία μικρών χρονικών καθυστερήσεων λόγω φυσικών περιορισμών των ICT που χρησιμοποιούνται είναι φυσιολογικές και οι επιπτώσεις τους στα συστήματα αυτοματισμού είναι αμελητέες ή αντιμετωπίζονται εύκολα από τους μηχανισμούς ελέγχου. Ωστόσο, όταν αυτές οι χρονικές καθυστερήσεις είναι μεγάλες ή παρουσιάζουν συγκεκριμένα μοτίβα, η ευστάθεια ενός συστήματος SG μπορεί να επηρεαστεί σημαντικά. Αυτό μπορεί να οδηγήσει στην πλήρη απορρύθμιση του συστήματος ηλεκτρικής ενέργειας, καθώς στο να τεθεί το σύστημα πλήρως εκτός λειτουργίας. Σε όρους του τριγώνου κυβερνοασφάλειας, οι TDAs απειλούν κυρίως την ακεραιότητα των δεδομένων που ανταλλάσσονται σε ένα SG.

Ένας από τους πιο σημαντικούς τύπους κυβερνοεπιθέσεων που απαντάται κυρίως στα CPSs και όχι στα κλασικά συστήματα ICT, είναι οι κυβερνοεπιθέσεις έγχυσης ψευδών δεδομένων (FDIAs). Εάν κάποιος κυβερνοεισβολέας αποκτήσει πρόσβαση σε ένα μέρος του συστήματος αυτοματισμού του SG, μπορεί να αλλοιώσει το περιεχόμενο των πακέτων δικτύου που φέρουν τα δεδομένα μετρήσεων ή εντολών ελέγχου. Οι FDIAs μπορούν με διάφορους τρόπους να μεταβάλλουν το περιεχόμενο των πακέτων δικτύου, οι οποίοι περιγράφονται από μαθηματικές σχέσεις μικρής ή μεγάλης πολυπλοκότητας. Ανάλογα με τις προθέσεις των επιτιθέμενων, η δομή των FDIAs μπορεί να είναι τέτοια που είτε μυστικά να προκαλούν βλάβες σε διάφορα μέρη του συστήματος ενέργειας είτε να δημιουργούν εκτεταμένες διακοπές ρεύματος. Ως αποτέλεσμα, οι FDIAs μπορούν να διαταράξουν σημαντικά την ευστάθεια των σύγχρονων συστημάτων ηλεκτρικής ενέργειας και να οδηγήσουν σε σοβαρές οικονομικές απώλειες. Σε όρους του τριγώνου κυβερνοασφάλειας, οι FDIAs απειλούν κυρίως την ακεραιότητα των δεδομένων που ανταλλάσσονται σε ένα SG.

8.3 Κεφάλαιο 3

Στο κεφάλαιο 3 αναλύονται λεπτομερώς οι βασικές αρχές λειτουργίας του συστήματος LFC. Αρχικά, παρουσιάζεται η λειτουργία του συστήματος ελέγχου στροφών των γεννητριών παραγωγής ηλεκτρικής ενέργειας, το οποίο αποτελεί τη βάση του συστήματος ελέγχου συχνότητας των ηλεκτρικών δικτύων. Στη συνέχεια, περιγράφεται το σύστημα

LFC μέσω του μοντέλου απόκρισης συχνότητας που προκύπτει από τον συνδυασμό των επιμέρους μερών του συστήματος ελέγχου στροφών. Τέλος, σχεδιάζεται το μοντέλο χώρου-κατάστασης του LFC χρησιμοποιώντας τις διαφορικές-αλγεβρικές εξισώσεις που περιγράφουν τη δυναμική συμπεριφορά αυτού του συστήματος. Ο συγκεκριμένος τύπος μοντελοποίησης αποτελεί τη βάση των προτεινόμενων μηχανισμών κυβερνοάμυνας για το LFC, όπως θα φανεί στα επόμενα κεφάλαια.

Για την καλή λειτουργία των συστημάτων ηλεκτρικής ενέργειας είναι απαραίτητη η συνεχής διατήρηση του ισοζυγίου ισχύος εντός αποδεκτών ορίων. Το ισοζύγιο ισχύος επηρεάζεται συνήθως από τις μεταβολές φορτίου που διαρκώς συμβαίνουν στο δίκτυο. Επομένως, η παραγόμενη ηλεκτρική ενέργεια πρέπει να προσαρμόζεται συνεχώς στα επίπεδα της ζήτησης. Ένα σημαντικός δείκτης του ισοζυγίου ενέργειας είναι η συχνότητα του ηλεκτρικού δικτύου: κάθε απόκλιση της συχνότητας από την ονομαστική της τιμή υποδηλώνει ότι η ισορροπία μεταξύ παραγωγής και ζήτησης έχει διαταραχθεί. Το σύστημα το οποίο είναι υπεύθυνο για την αποκατάσταση του ισοζυγίου ενέργειας έπειτα από διαταραχές είναι ο μηχανισμός LFC. Το LFC λαμβάνει ως είσοδο μετρήσεις συχνότητας ώστε να ανιχνεύσει τυχόν αποκλίσεις από την ονομαστική της τιμή. Στη συνέχεια, ρυθμίζει κατάλληλα την έξοδο των γεννητριών του συστήματος ώστε η παραγόμενη ηλεκτρική ενέργεια να είναι στα ίδια επίπεδα με τη ζητούμενη.

Η βάση της λειτουργίας του LFC είναι το σύστημα ρύθμισης στροφών των γεννητριών. Τα βασικά μέρη αυτού του συστήματος είναι η γεννήτρια παραγωγής της ηλεκτρικής ενέργειας, η τουρμπίνα της γεννήτριας και ο ρυθμιστής της ταχύτητας. Όταν μεταβάλλεται το φορτίο που τροφοδοτεί η γεννήτρια, αλλάζει η ηλεκτρομαγνητική ροπή που ασκείται στον άξονα της. Αυτό οδηγεί στην επιτάχυνση ή επιβράδυνση της γεννήτριας, δηλαδή στην μεταβολή του ρυθμού περιστροφής της. Για να συνεχίσει η γεννήτρια να λειτουργεί με σταθερή ταχύτητα, θα πρέπει η παραγωγή της γεννήτριας να ακολουθήσει τη μεταβολή του φορτίου. Αυτό επιτυγχάνεται με την αλλαγή της μηχανικής ενέργειας που προσφέρει η τουρμπίνα στη γεννήτρια μέσω του ρυθμιστή των στροφών της.

Το σύστημα LFC αποτελείται από πολλαπλά επίπεδα ελέγχου, τα οποία είναι οργανωμένα μεταξύ τους με ιεραρχικό τρόπο. Κάθε ένα από αυτά τα επίπεδα επιτελεί μια διαφορετική λειτουργία για το LFC, όπως θα φανεί στη συνέχεια. Η παρούσα διατριβή επικεντρώνεται στο πρωτεύον και στο δευτερεύον επίπεδο ελέγχου, λόγω της σπουδαιότητας τους για το LFC. Οι αρχές λειτουργίας των επιπέδων αυτών περιγράφονται συνοπτικά ως εξής:

- Πρωτεύον έλεγχος: Αποτελεί το πρώτο επίπεδο ελέγχου της συχνότητας των ηλεκτρικών δικτύων και είναι υπεύθυνος για τη σταθεροποίηση της εντός αποδεκτών τιμών. Σε αυτό το επίπεδο ελέγχου, οι τοπικοί ρυθμιστές στροφών των γεννητρι-

ών ανιχνεύουν αυτόματα μια διαταραχή του ισοζυγίου ισχύος μέσω αποκλίσεων της συχνότητας από την ονομαστική της τιμή. Στη συνέχεια, οι έξοδοι των ρυθμιστών προσαρμόζονται έτσι ώστε να μεταβάλλουν κατάλληλα τις θέσεις των βαλβίδων της τουρμπίνων. Με αυτό τον τρόπο, οι ρυθμιστές στροφών αναγκάζουν την παραγόμενη ισχύ να προσαρμοστεί στα επίπεδα της ζήτησης. Παρόλο που αυτός ο τύπος ελέγχου σταθεροποιεί τη συχνότητα του συστήματος, δεν μπορεί να την επαναφέρει στην ονομαστική της τιμή, κι έτσι αφήνει ένα σφάλμα σε αυτή μετά τη λειτουργία του.

- Δευτερεύον/Συμπληρωματικός έλεγχος: Αποτελεί μια εφαρμογή του κέντρο ελέγχου ενέργειας του ηλεκτρικού δικτύου, γνωστή και ως αυτόματος έλεγχος παραγωγής (AGC), η οποία λαμβάνει χώρα αμέσως μετά τον πρωτεύοντα έλεγχο της συχνότητας. Σκοπός αυτού του συστήματος είναι να εξαλείψει τα προβλήματα που εισάγει ο πρωτεύον έλεγχος, δηλαδή να επαναφέρει τη συχνότητα στην ονομαστική της τιμή και να διατηρήσει τις ροές των γραμμών διασύνδεσης στα προγραμματισμένα τους επίπεδα (στην περίπτωση ηλεκτρικών συστημάτων πολλών περιοχών). Για να επιτευχθεί αυτό, ο δευτερεύον έλεγχος λαμβάνει αρχικά μετρήσεις της συχνότητας του δικτύου και των ροών ενέργειας των γραμμών διασύνδεσης μέσω τηλεμετρίας προκειμένου να σχηματίσει το σφάλμα ελέγχου περιοχής (ACE). Έπειτα, υπολογίζει το σήμα ελέγχου με βάση το ACE και το στέλνει ως απομακρυσμένη εντολή στην κατάλληλη είσοδο των ρυθμιστών στροφών των γεννητριών που συμμετέχουν στο συγκεκριμένο επίπεδο ελέγχου.

Για τον σχεδιασμό των μεθόδων κυβερνοάμυνας που προτείνονται για το LFC, είναι η απαραίτητη η μοντελοποίηση του συστήματος αυτού μέσω των εξισώσεων του χώρου-κατάστασης του. Αυτός ο τύπος μοντελοποίησης εκφράζει τη δυναμική συμπεριφορά του LFC ως ένα σύνολο μεταβλητών εισόδου, εξόδου και κατάστασης που σχετίζονται μεταξύ τους μέσω διαφορικών εξισώσεων πρώτης τάξης. Οι εξισώσεις αυτές περιγράφονται από τις μαθηματικές σχέσεις (3.17)-(3.33), οι οποίες προκύπτουν από τις συναρτήσεις μεταφοράς των υποσυστημάτων του LFC στο πεδίο της συχνότητας, όπως αυτές φαίνονται στο σχήμα 3.11.

8.4 Κεφάλαιο 4

Στο κεφάλαιο 4 γίνεται η παρουσίαση των προτεινόμενων μεθόδων ανίχνευσης και εντοπισμού κυβερνοεπιθέσεων στο σύστημα LFC των ηλεκτρικών δικτύων οι οποίες βασίζονται σε SMOs. Το κοινό σημείο ανάμεσα στις μεθόδους ανίχνευσης και εντοπισμού κυβερ-

νοεπιθέσεων που βασίζονται σε παρατηρητές είναι ότι τα σφάλματα εκτίμησης που προκύπτουν είναι ασυμπτωτικά ευσταθή μόνο κάτω από κανονικές συνθήκες, και όχι όταν το σύστημα αντιμετωπίζει κυβερνοεπιθέσεις. Αυτό εξασφαλίζει ότι τα σφάλματα εκτίμησης θα είναι πάντα μηδενικά εκτός εάν συμβεί μια κυβερνοεπίθεση, καθιστώντας τα αξιόπιστους δείκτες κυβερνοαπειλών. Παρόμοια φιλοσοφία ακολουθείται και στο σχεδιασμό των αντίστοιχων προτεινόμενων μεθόδων, οι οποίες παρουσιάζονται στη συνέχεια.

Στην αρχή αυτής της ενότητας γίνεται παρουσίαση του μοντέλου χώρου-κατάστασης του συστήματος **LFC**, όταν ο μηχανισμός αυτός υπόκειται σε κυβερνοεπιθέσεις. Το μοντέλο αυτό είναι απαραίτητο για την ανάπτυξη μεθόδων κυβερνοάμυνας που βασίζονται σε παρατηρητές. Προτού παρουσιαστούν οι προτεινόμενες μέθοδοι ανίχνευσης και εντοπισμού κυβερνοεπιθέσεων, πρέπει να θεμελιωθούν οι απαραίτητες μαθηματικές συνθήκες που επιτρέπουν το σχεδιασμό των παρατηρητών που χρησιμοποιούνται στις μεθόδους αυτές. Οι απαραίτητες αυτές μαθηματικές συνθήκες περιγράφονται στις υποθέσεις 1-4. Στη συνέχεια, του σύστημα **LFC** που περιγράφεται από την εξίσωση (4.1) διαχωρίζεται εικονικά στα υποσυστήματα (4.6) και (4.7) μέσω του προτεινόμενου μετασχηματισμού συντεταγμένων.

Μία από τις μεγαλύτερες προκλήσεις που αντιμετωπίζουν αυτές οι μέθοδοι είναι να σχεδιαστούν με τέτοιο τρόπο ώστε να μπορούν να διαχωρίσουν τις κυβερνοεπιθέσεις από άλλες είδη εξωγενών διαταραχών. Στις μεθόδους ανίχνευσης και εντοπισμού κυβερνοεπιθέσεων που προτείνονται στην παρούσα διατριβή, αυτός ο διαχωρισμός επιτυγχάνεται μέσω μετασχηματισμού συντεταγμένων του μοντέλου χώρου-κατάστασης του **LFC**, το οποίο βασίζεται σε προηγμένες τεχνικές παρατήρησης [117]. Μέσω του παραπάνω μετασχηματισμού, το αρχικό σύστημα διαχωρίζεται στο εικονικό υποσύστημα-I, το οποίο φέρει μόνο τις διαταραχές του συστήματος, και στο εικονικό υποσύστημα-II, το οποίο υπόκειται μόνο τις κυβερνοεπιθέσεις. Στη συνέχεια, ο παρατηρητής-I σχεδιάζεται για το υποσύστημα-I, ο οποίος επηρεάζεται μόνο από τις εξωγενείς διαταραχές του συστήματος, και τον παρατηρητή-II για το υποσύστημα-II, ο οποίος με τη σειρά του είναι ευαίσθητος μόνο σε κυβερνοεπιθέσεις.

Για την μέθοδο ανίχνευσης κυβερνοεπιθέσεων που προτείνεται στην παρούσα διατριβή, σχεδιάζεται ο **SMO** (4.8) για το (4.7) και ο κλειστός παρατηρητής (4.10) για το (4.6). Μετα τον σχεδιασμό των παραπάνω παρατηρητών, προκύπτουν οι δυναμικές εξισώσεις των σφαλμάτων εκτίμησης. Πιο συγκεκριμένα, η δυναμική εξίσωση (4.16) περιγράφει τη συμπεριφορά του σφάλματος εκτίμησης του σχεδιασμένου **SMO** και η δυναμική εξίσωση (4.17) τη συμπεριφορά του σφάλματος εκτίμησης του σχεδιασμένου κλειστού παρατηρητή.

Σύμφωνα με το θεώρημα (4.3.1) και την εξίσωση (4.17), το σφάλμα εκτίμησης του κλειστού παρατηρητή είναι ασυμπτωτικά ευσταθές υπό κανονικές συνθήκες. Αντίθετα,

όταν το σύστημα LFC υπόκειται σε κυβερνοεπιθέσεις, η ευστάθεια του επηρεάζεται από αυτές τις ανεπιθύμητες δραστηριότητες. Αυτό σημαίνει πως το σφάλμα εκτίμησης του κλειστού παρατηρητή συγκλίνει στο μηδέν υπό κανονικές συνθήκες ενώ γίνεται μη-μηδενικό όταν το σύστημα αυτό δέχεται κυβερνοεπιθέσεις Επιπλέον, παρατηρώντας την εξίσωση (4.17) συμπεραίνουμε ότι αυτή περιλαμβάνει μόνο το διάνυσμα κυβερνοεπίθεσης, ενώ δεν περιέχει καθόλου το διάνυσμα εξωγενών διαταραχών του συστήματος ή το σφάλμα εξόδου που εισάγει ο σχεδιασμένος SMO. Έτσι, το σφάλμα εκτίμησης κατάστασης του κλειστού παρατηρητή είναι ευαίσθητο μόνο στις κυβερνοεπιθέσεις και ανθεκτικό στις εξωγενείς διαταραχές του συστήματος.

Με βάση όλα τα παραπάνω, εάν το σφάλμα εκτίμησης του κλειστού παρατηρητή συγκλίνει στο μηδέν, το σύστημα βρίσκεται σε κανονική κατάσταση ενώ αν είναι μη-μηδενικό, το σύστημα είναι υπό κυβερνοεπίθεση. Αυτή η ιδιότητα του σφάλματος εκτίμησης του κλειστού παρατηρητή το καθιστά κατάλληλο δείκτη για την ανίχνευση κυβερνοεπιθέσεων ενάντια στο σύστημα LFC. Η μέθοδος ανίχνευσης επιθέσεων που προτείνεται στην παρούσα διατριβή μπορεί να συνοψιστεί ως εξής: υποθέτουμε ότι το $\|e_{h_3}\|$ επιλέγεται ως δείκτης ανίχνευσης, το ς_d αντιπροσωπεύει ένα προσαρμοστικό κατώφλι, το t_d είναι η χρονική στιγμή ανίχνευσης της επίθεσης και t_e είναι ο χρόνος που έχει παρέλθει από την προηγούμενη ανίχνευση επίθεσης μέχρι το t_d. Εάν $\|e_{h_3}\| \geq \varsigma_d$, τότε θεωρούμε ότι το σύστημα LFC έχει δεχτεί FDIA τη χρονική στιγμή t_d, διαφορετικά, το σύστημα βρίσκεται υπό κανονικές συνθήκες για το χρονικό διάστημα t_e.

Η μέθοδος ανίχνευσης κυβερνοεπιθέσεων που προτείνεται στην παρούσα διατριβή μπορεί να προσδιορίσει μόνο εάν και πότε το σύστημα LFC υπόκειται σε κυβερνοεπίθεση. Ωστόσο, δεν μπορεί να αναγνωρίσει ποιο σήμα ή ποια συσκευή που χρησιμοποιείται στο σύστημα LFC έχει δεχτεί κυβερνοεπίθεση, ειδικά σε σενάρια όπου το εν λόγω σύστημα αντιμετωπίζει πολλαπλές FDIAs ταυτόχρονα. Οι παραπάνω πληροφορίες μπορούν γίνουν διαθέσιμες χρησιμοποιώντας μεθόδους εντοπισμού κυβερνοεπιθέσεων. Στα πλαίσια της παρούσας διατριβής σχεδιάστηκε μια καινοτόμα μέθοδος εντοπισμού κυβερνοεπιθέσεων, η οποία και παρουσιάζεται στο υπόλοιπο αυτής της ενότητας.

Η βασική ιδέα της προτεινόμενης μεθόδου εντοπισμού κυβερνοεπιθέσεων είναι η εξής: προσδιορίζοντας ποια στοιχεία του διανύσματος επίθεσης a_m είναι μηδενικά ή και ποια όχι, ο εντοπισμός των μεταβλητών κατάστασης που έχουν επηρεαστεί από FDIAs μπορεί να πραγματοποιηθεί πολλαπλασιάζοντας το διάνυσμα επιθέσεων a_m με τον αντίστοιχο πίνακα κατανομής επιθέσεων D. Σε περίπτωση εντοπισμού μιας κυβερνοεπίθεσης από την προτεινόμενη μέθοδο, ένα κατάλληλα σχεδιασμένο σύστημα ειδοποίησης ενημερώνει τον χειριστή του συστήματος για τα ποιά σήματα που αλλοιωθεί από FDIAs μέσω ενός ψηφιακού συστήματος λογικής.

Πιο αναλυτικά, για την προτεινόμενη μέθοδο εντοπισμού κυβερνοεπιθέσεων σχεδιάστηκε ένα ειδικό ζεύγος SMOs για κάθε στοιχείο του διανύσματος επιθέσεων, σχηματίζοντας έτσι μία τράπεζα παρατηρητών αποτελούμενη από $2q$ δομικές μονάδες. Σε κάθε ζεύγος SMOs των δομικών μονάδων αυτής της τράπεζες, ο SMO-I σχεδιάζεται έτσι ώστε να επηρεάζεται από τις εξωγενείς διαταραχές του συστήματος LFC και ταυτόχρονα να είναι ανθεκτικός στις κυβερνοεπιθέσεις. Αντίθετα, ο SMO-II σχεδιάστηκε έτσι ώστε να επηρεάζεται μόνο από κυβερνοεπιθέσεις ενάντια στο σύστημα LFC ενω ταυτόχρονα να είναι ανθεκτικός στις εξωγενείς διαταραχές. Η ειδική ιδιότητα του SMO-II είναι ότι το σφάλμα εξόδου που αυτός εισάγει μοντελοποιείται με τρόπο που να είναι ισοδύναμος με το διάνυσμα όλων των στοιχείων επίθεσης εκτός από το στοιχείο στο οποίο αντιστοιχεί ο SMO-II, που συμβολίζεται με $\bar{a}_m^l$. Με αυτό τον τρόπο, οι δυναμικές εξισώσεις των σφαλμάτων εκτίμησης που προκύπτουν από τον κάθε SMO-II μπορούν να εξαλείψουν την επίδραση του $\bar{a}_m^l$ υπό συγκεκριμένες συνθήκες. Αυτό έχει ως αποτέλεσμα, ο SMO-II είναι ευαίσθητος μόνο στο στοιχείο του διανύσματος επίθεσης για το οποίο έχει σχεδιαστεί. Από όλα τα παραπάνω, συμπεραίνουμε ότι το σφάλμα εκτίμησης του SMO-II αποκλίνει από το μηδέν μόνο όταν το στοιχείο του διανύσματος επίθεσης για το οποίο έχει σχεδιαστεί είναι μη μηδενικό, καθιστώντας το ένα κατάλληλο δείκτη για τον εντοπισμό επιθέσεων.

8.5 Κεφάλαιο 5

Στο κεφάλαιο 5 παρουσιάζεται η προτεινόμενη μέθοδος εκτίμησης κυβερνοεπιθέσεων, καθώς και ο ανθεκτικός-στις-κυβερνοεπιθέσεις έλεγχος που έχει σχεδιαστεί στα πλαίσια της παρούσας διατριβής για τη ρύθμιση της συχνότητας των συστημάτων ηλεκτρικής ενέργειας. Οι μέθοδοι αυτές βασίζονται σε ένα ζευγάρι ειδικά σχεδιασμένων παρατηρητών κατάστασης. Όπως είναι γνωστό από τη θεωρία συστημάτων, ο σκοπός ενός παρατηρητή κατάστασης είναι να παρέχει μια εκτίμηση του διανύσματος κατάστασης ενός συστήματος για το οποίο έχει σχεδιαστεί. Μαθηματικά, η ύπαρξη ενός παρατηρητή κατάστασης είναι εξασφαλισμένη για ένα σύστημα όταν η διαφορά μεταξύ του πραγματικού και του εκτιμώμενου διανύσματος κατάστασης (δηλαδή το σφάλμα εκτίμησης) συγκλίνει στο μηδέν. Στα πλαίσια της παρούσας διατριβής, σχεδιάστηκαν κατάλληλοι παρατηρητές κατάστασης για το σύστημα LFC το οποίο υπόκειται σε κυβερνοεπιθέσεις. Συνεπώς, όταν το σφάλμα εκτίμησης που προκύπτει από τους υλοποιημένους παρατηρητές κατάστασης τείνει στο μηδέν, προκύπτει ένα σύνολο μαθηματικών σχέσεων που μπορούν να προσεγγίσουν τα αντίστοιχα διανύσματα κυβερνοεπίθεσης, όπως θα φανεί στη συνέχεια. Στο υπόλοιπο

αυτής της ενότητας παρουσιάζεται στη διαδικασία σχεδιασμού των προτεινόμενων παρατηρητών κατάστασης.

Όπως και στην περίπτωση των προτεινόμενων μεθόδων ανίχνευσής και εντοπισμού κυβερνοεπιθέσεων, αρχικά σχεδιάζεται το μοντέλου χώρου-κατάστασης του συστήματος LFC αυτό υπόκειται σε FDIAs. Πριν το σχεδιασμό των παρατηρητών κατάστασης, πρέπει πρώτα να θεμελιωθούν οι απαραίτητες μαθηματικές συνθήκες που εξασφαλίζουν την ύπαρξή τους [117]. Οι μαθηματικές αυτές συνθήκες εκφράζονται μέσα από μια σειρά υποθέσεων, οι οποίες παρουσιάζονται παρακάτω. Πιο συγκεκριμένα, η υπόθεση 5 επιτρέπει τον εικονικό διαχωρισμό του συστήματος (5.1) στο υποσύστημα-I και στο υποσύστημα-II. Το υποσύστημα-I (5.5) επηρεάζεται από τις FDIAs εναντίον των σημάτων ελέγχου και είναι ανθεκτικό στις FDIAs εναντίον των μετρήσεων, ενώ το υποσύστημα-II (5.4) επηρεάζεται από τις FDIAs εναντίον των μετρήσεων και είναι ανθεκτικό στις FDIAs εναντίον των σημάτων ελέγχου. Αυτός ο διαχωρισμός διευκολύνει τη διαδικασία σχεδιασμού των παρατηρητών, καθώς διαίρεί το αρχικό σύστημα σε πιο απλά, ισοδύναμα υποσυστήματα με λιγότερους όρους. Από την άλλη, οι υποθέσεις 6 και 7 είναι απαραίτητες για να αποδειχθεί ότι τα σφάλματα εκτίμησης των παρατηρητών που σχεδιάστηκαν είναι ασυμπτωτικά ευσταθή. Πιο συγκεκριμένα, οι υποθέσεις αυτές προσφέρουν χρήσιμες ανισότητες προκειμένου να φραχτούν οι συναρτήσεις Lyapunov που έχουν επιλεχθεί.

Για την προτεινόμενη μέθοδο εκτίμησης κυβερνοεπιθέσεων, σχεδιάζεται ο SMO (5.6) για το υποσύστημα-I (5.5) και ο παρατηρητής αγνώστου εισόδου (UIO) (5.8) για το υποσύστημα-II (5.4). Μετα τον σχεδιασμό των παραπάνω παρατηρητών, προκύπτουν η διαφορικές εξισώσεις (5.9) και (5.14) που περιγράφουν τη συμπεριφορά του σφάλματος εκτίμησης του υλοποιημένου SMO και του σχεδιασμένου UIO, αντίστοιχα. Οι τελευταίες αυτές εξισώσεις δείχνει ότι ο στόχος της διαδικασίας σχεδιασμού του παρατηρητών έχει επιτευχθεί: το σφάλμα εκτίμησης του SMO επηρεάζεται μόνο από FDIAs εναντίον των σημάτων ελέγχου ενώ το σφάλμα εκτίμησης του UIO επηρεάζεται μόνο από FDIAs εναντίον των μετρήσεων. Ωστόσο, τα σφάλματα εκτίμησης δεν έχουν ακόμη αποσυμπλεγχθεί πλήρως από τις εξωγενείς διαταραχές του LFC. Για να ελαχιστοποιηθεί η επίδραση των εξωγενών αυτών διαταραχών στα σφάλματα κατάστασης, χρησιμοποιείται η γνωστή H_∞ μέθοδος. Τα θεώρηματα 5.3.1 και 5.3.2 καθορίζουν τις προϋποθέσεις που πρέπει να πληρούνται ώστε να επιτευχθεί η επιθυμητή ευστάθεια των παρατηρητών που σχεδιάστηκαν. Μετά τον σχεδιασμό των παραπάνω παρατηρητών, το διάνυσμα των κυβερνοεπιθέσεων εναντίον των μετρήσεων προσεγγίζεται από την εξίσωση:

$$\hat{a}_m \approx \begin{bmatrix} 0 & I_q \end{bmatrix} \hat{\bar{\xi}}_2,$$

ενώ το διάνυσμα των κυβερνοεπιθέσεων εναντίον των σημάτων ελέγχου προσεγγίζεται από την εξίσωση:

$$\hat{a}_c \approx (\rho+\eta)\frac{B_1^T P_1 (C_1^{-1}\omega_1 - \hat{\xi}_1)}{\|B_1^T P_1 (C_1^{-1}\omega_1 - \hat{\xi}_1)\| + \delta}.$$

Οι παραπάνω μαθηματικές φόρμουλες εκτίμησης κυβερνοεπιθέσεων που προέκυψαν, μπορούν να χρησιμοποιηθούν για τον σχεδιασμό ενός ανθεκτικου-στις-κυβερνοεπιθέσεις μηχανισμού ελέγχου της συχνότητας των ηλεκτρικών δικτύων. Πιο συγκεκριμένα, για την εξάλειψη των FDIAs εναντίον των σημάτων ελέγχου θα πρέπει το διάνυσμα της εκτιμώμενης FDIA εναντίον των σημάτων ελέγχου $\hat{a}_c$ να ενσωματωθεί στο βρόχο ανάδρασης του LFC. Αυτό μπορεί να επιτευχθεί εύκολα προσθέτοντας το $\hat{a}_c$ ως μια επιπλέον είσοδο ελέγχου στο σύστημα LFC που υπόκειται σε κακόβουλες δραστηριότητες, προκειμένου να αντισταθμιστεί η επίδραση του αντίστοιχου διανύσματος κυβερνοεπιθέσεων. Η προσθήκη αυτής της νέας εισόδου ελέγχου δεν επηρεάζει το διάνυσμα εισόδου του συστήματος και κατα συνέπεια, δεν αλλοιώνει την ορθή λειτουργία του αρχικού ελέγχου συχνότητας. Επομένως, η ευστάθεια του LFC διατηρείται ακόμα και μετά την ενσωμάτωση του $\hat{a}_c$ στο βρόχο ανάδρασης. Η δομή του συστηματος LFC που δέχεται κυβερνοεπιθέσεις , μετά την προσθήκη της νέας εισόδου ελέγχου $\hat{a}_c$ μετατρέπεται στο ακόλουθο μοντέλο:

$$\begin{cases} \dot{x}(t) = Ax(t) + F\phi(x,t) + B\big(u(t) + a_c(t) - \hat{a}_c(t)\big) + Ed(t) \\ y(t) = Cx(t) + Da_m(t). \end{cases}$$

Για λόγους απλότητας, υποθέτουμε από εδώ και στο εξής πως οι FDIAs εναντίον των σημάτων ελέγχου εξαλείφονται άμεσα με την προσθήκη της νέας εισόδου ελέγχου $\hat{a}_c$ και η διαφορά μεταξύ του πραγματικού και των εκτιμώμενου διανύσματος επίθεσης είναι αμελητέα.

Παρόμοια με το προηγούμενο είδος κυβερνοεπιθέσεων, η εξάλειψη των FDIAs εναντίον των μετρήσεων μπορεί να επιτευχθεί με την ενσωμάτωση του σχεδιασμένου διανύσματος εκτίμησης τους στο σύστημα LFC. Η παραπάνω διαδικασία ενσωμάτωσης πραγματοποιείται ως εξής: το διάνυσμα εξόδου του συστήματος LFC θα τροποποιηθεί αφαιρώντας από αυτό το εκτιμώμενο διάνυσμα επίθεσης $\hat{a}_m$, αφού σε αυτό περιλαμβάνεται και το διάνυσμα των FDIAs ενάντια σε μετρήσεις. Ωστόσο, αυτή η προσθήκη στο βρόχο ανάδρασης του LFC επηρεάζει τον αρχικό μηχανισμό ελέγχου συχνότητας, αντίθετα με το προηγούμενο σύστημα αντιστάθμισης κυβερνοεπιθέσεων. Αυτό οφείλεται στο ότι ο αρχικός μηχανισμός ρύθμισης της συχνότητας λειτουργεί με ελεγκτή ανάδραση-εξόδου και άρα, η είσοδος και η έξοδος του LFC συνδέονται τη σχέση $u(t) = -Ky(t)$. Εφόσον το διάνυσμα εξόδου τροποποιείται από την προτεινόμενη προσθήκη, θα τρο-

ποποιηθεί και το διάνυσμα εισόδου, επηρεάζοντας έτσι την εξίσωση του διανύσματος κατάστασης. Επομένως, είναι απαραίτητο να αποδειχθεί θεωρητικά η αποτελεσματικότητα του νέου, ανθεκτικού-στις-κυβερνοεπιθέσεις μηχανισμού ελέγχου τόσο απέναντι στις εξωγενείς διαταραχές όσο και απέναντι στις FDIAs που στοχεύουν τις μετρήσεις. Η ανάλυση ευστάθειας του προτεινόμενου αυτού ελέγχου, που περιγράφεται στην ενότητα 5.5, αποδεικνύει μαθηματικά πως ο μηχανισμός αυτός πληροί τις προδιαγραφές που είναι απαραίτητες για την ενδεδειγμένη λειτουργία του. Έτσι, ο προτεινόμενος ανθεκτικός-στις-κυβερνοεπιθέσεις έλεγχος για το LFC περιγράφεται από το ακόλουθο μοντέλο:

$$\begin{cases} \dot{x}(t) = Ax(t) + F\phi(x,t) + B\big(u(t) + a_c(t) - \hat{a}_c(t)\big) + Ed(t) \\ y(t) = Cx(t) + D\big(a_m(t) - \hat{a}_m(t)\big). \end{cases}$$

8.6 Κεφάλαιο 6

Στο Κεφάλαιο 6 περιγράφεται το μέρος του προτεινόμενου πλαισίου ενίσχυσης της κυβερνοανθεκτικότητας του LFC το οποίο βασίζεται σε μεθόδους μηχανικής μάθησης. Η ανάλυση ξεκινά με τον αντίστοιχο μηχανισμό ανίχνευσης επιθέσεων ενάντια στο LFC. Η βασική ιδέα στις μεθοδολογίες ανίχνευσης κυβερνοεπιθέσεων είναι η οριοθέτηση της κανονικής κατάστασης, εντός της οποίας το σύστημα θεωρείται ότι λειτουργεί φυσιολογικά. Για να επιτευχθεί αυτός ο σκοπός, χρειάζεται μια μετρική ανίχνευσης που θα ποσοτικοποιεί την κατάσταση του εξεταζόμενου συστήματος σε όρους κυβερνοασφάλειας. Αυτή η μετρική ανίχνευσης υλοποιείται με την ανάπτυξη ενός κατάλληλου μοντέλου, χρησιμοποιώντας είτε μαθηματικές εξισώσεις είτε δεδομένα. Το μοντέλο που θα σχεδιαστεί λαμβάνει εισόδους από το πραγματικό σύστημα και υπολογίζει τη μετρική ανίχνευσης για κάθε χρονική στιγμή. Αν η μετρική ανίχνευσης υπερβεί ένα συγκεκριμένο κατώφλι, το σύστημα θεωρείται πως έχει δεχτεί κυβερνοεπίθεση, διαφορετικά η λειτουργία του συστήματος συνεχίζεται κανονικά.

Για το σχεδιασμό αυτού του μηχανισμού ανίχνευσης κυβερνοεπιθέσεων, θα πρέπει πρώτα να οριστούν οι είσοδοι της μεθόδου. Οι εφαρμογές που παρέχουν πληροφορίες για την κυβερνοανθεκτικότητα των CPSs συνήθως λειτουργούν εντός του κέντρου ελέγχου. Επομένως, οι είσοδοι τους είναι ένα υποσύνολο των μεταβλητών κατάστασης του συστήματος. Στην περίπτωση του LFC, ο έλεγχος πραγματοποιείται χρησιμοποιώντας την τοπική μέτρηση συχνότητας κάθε περιοχής και τις μετρήσεις ροής ισχύος κάθε διασυνδετικής γραμμής. Οι μηχανισμοί ανίχνευσης επιθέσεων που βασίζονται σε δεδομένα συνήθως λειτουργούν χρησιμοποιώντας αποκλειστικά τις μετρήσεις του συστήματος. Επιπλέον, η διαδικασία εκπαίδευσης των μεθόδων που βασίζονται σε μοντέλα μηχανικής

μάθησης βελτιώνεται σημαντικά με χρήση ιστορικών δεδομένων έναντι δεδομένων πραγματικού χρόνου. Βάσει της παραπάνω ανάλυσης, ο προτεινομενος αλγόριθμος ανίχνευσης επιθέσεων είναι σχεδιασμένος να λειτουργεί χρησιμοποιώντας τις ιστορικές μετρήσεις του LFC.

Μετά τον ορισμό των εισόδων της προτεινόμενης μεθόδου, ο επόμενος στόχος είναι να σχεδιαστεί το κατάλληλο μοντέλο που θα παράγει τη μετρική ανίχνευσης. Η ανάλυση που πραγματοποιήθηκε στα πλαίσια αυτής της διατριβής έδειξε πως οι αυτοκωδικοποιητές είναι το μοντέλο μηχανικής μάθησης που ταιριάζει περισσότερο στα χαρακτηριστικά του LFC. Ένας κατάλληλα εκπαιδευμένος αυτοκωδικοποιητής μπορεί να αναπαράγει στην έξοδό του το διάνυσμα που λαμβάνει στην εισόδου του με υψηλή ακρίβεια. Έτσι, η εκπαίδευση του αυτοκωδικοποιητή σε υγιή δεδομένα του επιτρέπει να αναγνωρίζει ανωμαλίες ή αποκλίσεις των νέων δεδομένων που λαμβάνονται από τα πρότυπα που έχει μάθει. Επιπλέον, ο προτεινόμενος αυτοκωδικοποιητής υλοποιείται με βαθιά νευρωνικά δίκτυα (DNNs), η ευελιξία των οποίων του επιτρέπει να μαθαίνει συνεχώς νέες υγιείς καταστάσεις, ακόμα και κατά τη διάρκεια της λειτουργίας του.

Προκειμένου να ανιχνεύσει κυβερνοεπιθέσεις ενάντια στο LFC, ο προτεινόμενος αυτοκωδικοποιητής εκπαιδεύεται πρώτα σε ένα σύνολο δεδομένων με μετρήσεις συχνότητας και ροής ισχύος των διασυνδετικών γραμμών που αντανακλούν την υγιή κατάσταση του συστήματος, π.χ. καταστάσεις ευστάθειας, μεταβολές φορτίου ή διαταραχές λόγω ανανεώσιμων πηγών ενέργειας, κλπ. Μετά την εκπαίδευση του, ο αυτοκωδικοποιητής εγκαθίσταται στο κέντρο ελέγχου, όπου οι μετρήσεις του LFC προωθούνται σε αυτόν. Αν ο αυτοκωδικοποιητής λάβει μετρήσεις συχνότητας και ροής ισχύος των διασυνδετικών γραμμών που αντιστοιχούν σε υγιή κατάσταση αλλά δεν τις έχει μάθει κατά την εκπαίδευση του, θα μπορέσει να τις αναπαράγει με υψηλή ακρίβεια. Από την άλλη πλευρά, αν ο αυτοκωδικοποιητής λάβει δεδομένα μετρήσεων LFC που αντιστοιχούν σε κυβερνοεπίθεση (ένα συμβάν που ο αυτοκωδικοποιητής δεν έχει μάθει κατά τη φάση εκπαίδευσης), η αναπαραγωγή της εισόδου του θα παρουσιάσει σημαντικό σφάλμα. Αυτό το χαρακτηριστικό του προτεινόμενου αυτοκωδικοποιητή τον καθιστά έναν αποτελεσματικό δείκτη κυβερνοεπιθέσεων για το LFC.

Στην προτεινόμενη μέθοδο ανίχνευσης επιθέσεων, ο υλοποιημένος αυτοκωδικοποιητής εναλλάσσεται μεταξύ τριών λειτουργικών καταστάσεων. Αρχικά, ο αυτοκωδικοποιητής βρίσκεται στη φάση εκπαίδευσης εκτός λειτουργίας, όπου μαθαίνει τις υγιείς κατάστασεις του συστήματος σε ένα απομονωμένο υπολογιστικό περιβάλλον. Μετά τη διαδικασίας εκπαίδευσης, ο αυτοκωδικοποιητής μεταβαίνει στη φάση λειτουργίας σε πραγματικό χρόνο, όπου εγκαθίσταται στο κέντρο ελέγχου και ξεκινά να λειτουργεί. Σε κάθε χρονικό στιγμή, το σφάλμα αναπαραγωγής εισόδου του αυτοκωδικοποιητή καθορίζει αν το LFC

έχει δεχτεί κυβερνοεπίθεση ή όχι. Αν ο αυτοκωδικοποιητής λάβει στη φάση λειτουργίας σε πραγματικό χρόνο μία κανονική κατάσταση που δεν έχει μάθει, τότε μεταβαίνει στη φάση εκπαίδευσης σε πραγματικό χρόνο. Σε αυτή την κατάσταση, ο αυτοκωδικοποιητής επαναεκπαιδεύεται με την τελευταία κανονική κατάσταση που δέχτηκε, ώσπου να επιστρέψει στη φάση λειτουργίας σε πραγματικό χρόνο. Αυτή η εναλλαγή λειτουργικών καταστάσεων κάνει τον αυτοκωδικοποιητή να παραμένει συνεχώς ενημερωμένος απέναντι σε επερχόμενες ψηφιακές απειλές.

Η επόμενη μέθοδος του προτεινόμενου πλαισίου που βασίζεται σε αλγορίθμους μηχανικής μάθησης έχει ως στόχο την εξάλείψη των κυβερνοεπιθέσεων ενάντια στο LFC. Σύμφωνα με το [138], μια φυσική στρατηγική αντιμετώπισης κυβερνοεπιθέσεων στα CPSs είναι η προσθήκη ενός εφεδρικού βρόχου ελέγχου. Αυτός ο εφεδρικός έλεγχος αποτελείται από ένα ειδικά σχεδιασμένο μοντέλο που λαμβάνει μετρήσεις από το υπό προστασία CPS ώστε να παράγει μια εκτίμηση των υγιών σημάτων ελέγχου του. Όταν ανιχνεύεται μια κυβερνοεπίθεση, το αρχικό σύστημα ελέγχου τίθεται προσωρινά εκτός λειτουργίας και το προστατευμένο σύστημα ελέγχεται από τα εκτιμώμενα σήματα του εφεδρικού βρόχου. Ο εφεδρικός βρόχος ελέγχου ενεργοποιείται για μικρό χρονικό διάστημα, μέχρις ότου να εξαλειφθεί η επίθεση και να αποκατασταθεί πλήρως η λειτουργία του CPS. Με βάση αυτή την προσέγγιση, σχεδιάστηκε σε αυτή τη διατριβή το DAR-LFC, μια μεθοδολογία επαναφοράς του LFC από κυβερνοεπιθέσεις μέ χρήση μοντέλων μηχανικής μάθησης.

Ένα σημαντικό στοιχείο στη σχεδίαση του DAR-LFC είναι η επιλογή του μοντέλου που θα παράγει τα εκτιμώμενα υγιή σήματα ελέγχου. Η προσέγγιση των εντολών ελέγχου του LFC μέσω μιας ή περισσότερων μετρήσεων πεδίου αποτελεί ένα τυπικό πρόβλημα παλινδρόμησης. Τα προβλήματα παλινδρόμησης μπορούν να αντιμετωπιστούν αποδοτικά από διάφορους αλγόριθμοι μηχανικής μάθησης. Οι αρχιτεκτονικές αυτών των μοντέλων κυμαίνονται από σχετικά απλές, π.χ. μηχανές διανυσμάτων υποστήριξης (SVMs), έως πολύ σύνθετες, π.χ. δίκτυα μακράς βραχυπρόθεσμης μνήμης (LSTMs). Για την διατήρηση μιας αποδεκτής ισορροπίας μεταξύ ακρίβειας και πολυπλοκότητας του επιλεγμένου μοντέλου, η αρχιτεκτονική των DNNs κρίθηκε ως η πλέον κατάλληλη για την εκτίμηση των υγιών εντολών ελέγχου του LFC και έτσι, επιλέχθηκε ως μοντέλο του DAR-LFC.

Ένα άλλο σημαντικό στοιχείο για τη σχεδίαση του DAR-LFC, είναι η επιλογή των σημάτων που θα δοθούν ως είσοδοι στο DNN. Εφόσον οι μετρήσεις της συχνότητας και της ισχύος των γραμμών διασύνδεσης αλλοιώνονται κατά τη διάρκεια μιας κυβερνοεπίθεσης στο LFC, χρειάζεται να βρεθούν κατάλληλες εναλλακτικες μετρήσεις πεδίου για την εκτίμηση του κάθε σήματος ελέγχου ACE_i. Παρατηρώντας την εξίσωση (6.1) και τον τύπο του ελεγκτή στροφών των γεννητριών, διαπιστώνεται ότι το εκάστοτε ACE_i σχετίζεται με την παραγωγή των τοπικών γεννητριών. Έτσι, η ισχύς που παράγεται από

τις τοπικές γεννήτριες αποτελεί κατάλληλη είσοδο για το DNN ώστε να εκτιμήσει με ακρίβεια το ACE_i. Το βασικό σημείο τώρα είναι η επιλογή της παραγόμενης ισχύς από την κατάλληλη τοπική γεννήτρια. Σύμφωνα με το Σχήμα 6.10, η τοπική παραγωγή που είναι πιο κοντά στο ACE_i είναι η $\Delta P_{G_{pr}}$. Επομένως, η έξοδος ισχύος μιας γεννήτριας που συμμετέχει μόνο στον πρωτογενή έλεγχο επιλέγεται ως είσοδος στο DNN.

Το διάγραμμα του μηχανισμού άμυνας DAR-LFC απεικονίζεται στο Σχήμα 6.11, στο οποίο φαίνεται αναλυτικά και η λειτουργία του. Αρχικά, το DNN εκπαιδεύεται χρησιμοποιώντας ως είσοδο τις επιλεγμένες μετρήσεις τοπικής παραγωγής και ως έξοδο τις χρονοσειρές των ACE_i της κάθε περιοχής. Μετά την εκπαίδευση, εγκαθίσταται ένα εφεδρικό σύστημα τηλεμετρίας για τη μετάδοση της παραγόμενης ισχύς από την επιλεγμένη τοπική γεννήτρια στο DNN. Στη συνέχεια, το εκπαιδευμένο DNN εγκαθίσταται στο κέντρο ελέγχου για να παράγει εκτιμήσεις των υγιούς ACE_i ($\widetilde{ACE}_i$) χρησιμοποιώντας τα δεδομένα που λαμβάνει από το εφεδρικό κανάλι επικοινωνίας. Έτσι, όταν ανιχνεύεται μια κυβερνοεπίθεση κατά του LFC, το κέντρο ελέγχου χρησιμοποιεί το $\widetilde{ACE}_i$ για να ρυθμίσει τις γεννήτριες που συμμετέχουν στον δευτερογενή έλεγχο αντί για το αρχικό ACE_i. Με αυτό τον τρόπο, το LFC μπορεί να λειτουργεί σύμφωνα με τις προβλεπόμενες προδιαγραφές του ακόμα και όταν έχει δεχτεί κυβερνοεπιθέση.

8.7 Κεφάλαιο 7

Στο τελευταίο αυτό κεφάλαιο πραγματοποιείται μια ολοκληρωμένη ανασκόπηση της διατριβής, αναλύοντας τα πιο σημαντικά σημεία του ερευνητικού προβλήματος που εξετάστηκε και των μεθόδων που χρησιμοποιήθηκαν για την αντιμετώπισή του. Παράλληλα, παρουσιάζονται συνοπτικά τα κυριότερα συμπεράσματα που προέκυψαν από την ανάπτυξη των προτεινόμενων μεθόδων και από τα αποτελέσματα των πειραματικών τους εφαρμογών. Τέλος, χαρτογραφούνται τα διάφορα ανοιχτά ερευνητικά θέματα για ενδεχόμενες μελλοντικές εργασίες, τα οποία μπορεί είτε να συνεχίσουν την τρέχουσες ερευνητικές δραστηριότητες, είτε να αξιοποιήσουν τη γνώση που αποκτήθηκε σε νέες εφαρμογές ή και σε διαφορετικά επιστημονικά πεδία.

Bibliography

[1] A. D. Syrmakesis, C. Alcaraz, and N. D. Hatziargyriou, "Classifying resilience approaches for protecting smart grids against cyber threats," *International Journal of Information Security*, pp. 1–22, 2022.

[2] S. Nazari, "The unknown input observer and its advantages with examples," *arXiv preprint arXiv:1504.07300*, 2015.

[3] U.S. Department of Energy, Grid Modernization and the Smart Grid. [Online].

[4] European Commission, European Technology Platform SmartGrids; Vision and Strategy for Europe's Electricity Networks of the Future. [Online].

[5] U.S. Department of Energy, Cybersecurity. [Online].

[6] European Union Agency for Cybersecurity (ENISA), Smart Grids. [Online].

[7] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, "Smart grid technologies: Communication technologies and standards," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 4, pp. 529–539, 2011.

[8] M. Z. Gunduz and R. Das, "Cyber-security on smart grid: Threats and potential solutions," *Computer networks*, vol. 169, p. 107094, 2020.

[9] C. Alcaraz and J. Lopez, "Analysis of requirements for critical control systems," *International Journal of Critical Infrastructure Protection (IJCIP)*, vol. 5, p. 137–145, 2012 2012.

[10] D. U. Case, "Analysis of the cyber attack on the Ukrainian power grid," *Electricity Information Sharing and Analysis Center (E-ISAC)*, vol. 388, 2016.

[11] N. Falliere, L. O. Murchu, and E. Chien, "W32. stuxnet dossier," *White paper, Symantec Corp., Security Response*, vol. 5, no. 6, p. 29, 2011.

[12] S. Karnouskos, "Stuxnet worm impact on industrial cyber-physical system security," in *IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society*, pp. 4490–4494, 2011.

[13] N. I. of Standards and Technology, "Framework for improving critical infrastructure cybersecurity," tech. rep., NIST, 2018.

[14] V. Y. Pillitteri and T. L. Brewer, "Guidelines for smart grid cybersecurity," tech. rep., 2014.

[15] EPRI, "Enhancing distribution resiliency: Opportunities for applying innovative technologies," Tech. Rep. 1026889, Palo Alto, CA, USA, Jan. 2013.

[16] N. Council, "Critical infrastructure resilience final report and recommendations," tech. rep., Nat. Infrastruct. Advisory Council, Washington, DC, USA, 2009.

[17] S. I. R. T. Force, "Severe impact resilience: Considerations and recommendations," tech. rep., NERC, Atlanta, GA, USA, May 2012.

[18] M. Chaudry, P. Ekins, K. Ramachandran, A. Shakoor, J. Skea, G. Strbac, X. Wang, and J. Whitaker, "Building a resilient UK energy system," Tech. Rep. UKERC/WP/ES/2009/023, UK Energy Res. Center, London, U.K, Apr. 2011.

[19] M. Panteli and P. Mancarella, "The Grid: Stronger, Bigger, Smarter?: Presenting a Conceptual Framework of Power System Resilience," *IEEE Power and Energy Magazine*, vol. 13, no. 3, pp. 58–66, 2015.

[20] A. Ameli, A. Hooshyar, E. F. El-Saadany, and A. M. Youssef, "Attack Detection and Identification for Automatic Generation Control Systems," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4760–4774, 2018.

[21] S. Sridhar and M. Govindarasu, "Model-Based Attack Detection and Mitigation for Automatic Generation Control," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 580–591, 2014.

[22] S. East, J. Butts, M. Papa, and S. Shenoi, "A Taxonomy of Attacks on the DNP3 Protocol," in *International Conference on Critical Infrastructure Protection*, pp. 67–81, Springer, 2009.

[23] M. J. Rice, C. A. Bonebrake, G. K. Dayley, and L. J. Becker, "Secure ICCP Final Report," 6 2017.

[24] S. Liu, X. P. Liu, and A. El Saddik, "Denial-of-service (dos) attacks on load frequency control in smart grids," in *2013 IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–6, 2013.

[25] A. Sargolzaei, K. K. Yen, and M. N. Abdelghani, "Preventing Time-Delay Switch Attack on Load Frequency Control in Distributed Power Systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1176–1185, 2016.

[26] T. Huang, B. Satchidanandan, P. R. Kumar, and L. Xie, "An Online Detection Framework for Cyber Attacks on Automatic Generation Control," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6816–6827, 2018.

[27] I. Zografopoulos, J. Ospina, X. Liu, and C. Konstantinou, "Cyber-Physical Energy Systems Security: Threat Modeling, Risk Assessment, Resources, Metrics, and Case Studies," *IEEE Access*, vol. 9, pp. 29775–29818, 2021.

[28] Z. Zheng, S. Jin, R. Bettati, and A. L. N. Reddy, "Securing cyber-physical systems with adaptive commensurate response," in *2017 IEEE Conference on Communications and Network Security (CNS)*, pp. 1–6, Oct 2017.

[29] L. F. Cómbita, J. Giraldo, A. A. Cárdenas, and N. Quijano, "Response and reconfiguration of cyber-physical control systems: A survey," in *2015 IEEE 2nd Colombian Conference on Automatic Control (CCAC)*, pp. 1–6, 2015.

[30] S. Gholami, S. Saha, and M. Aldeen, "A cyber attack resilient control for distributed energy resources," in *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pp. 1–6, Sep. 2017.

[31] K. Paridari, N. O'Mahony, A. El-Din Mady, R. Chabukswar, M. Boubekeur, and H. Sandberg, "A Framework for Attack-Resilient Industrial Control Systems: Attack Detection and Controller Reconfiguration," *Proceedings of the IEEE*, vol. 106, no. 1, pp. 113–128, 2018.

[32] X. Jia, T. Lv, B. Li, H. Li, and B. Liu, "Decentralized Secure Load Frequency Control for Multi-Area Power Systems Under Complex Cyber Attacks," *IEEE Transactions on Smart Grid*, pp. 1–1, 2024.

[33] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure Estimation and Control for Cyber-Physical Systems Under Adversarial Attacks," *IEEE Transactions on Automatic Control*, vol. 59, pp. 1454–1467, June 2014.

[34] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. J. Pappas, "Robustness of attack-resilient state estimators," in *2014 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*, pp. 163–174, April 2014.

[35] N. Bezzo, J. Weimer, M. Pajic, O. Sokolsky, G. J. Pappas, and I. Lee, "Attack resilient state estimation for autonomous robotic systems," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3692–3698, Sep. 2014.

[36] R. Hewett, S. Rudrapattana, and P. Kijsanayothin, "Cyber-Security Analysis of Smart Grid SCADA Systems with Game Models," in *Proceedings of the 9th Annual Cyber and Information Security Research Conference*, CISR '14, (New York, NY, USA), p. 109–112, Association for Computing Machinery, 2014.

[37] J. E. Rubio, C. Alcaraz, and J. Lopez, "Game theory-based approach for defense against apts," in *18th International Conference on Applied Cryptography and Network Security (ACNS'20)*, vol. 12147, pp. 297–320, Springer, Springer, 10/2020 2020.

[38] R. Hewett, S. Rudrapattana, and P. Kijsanayothin, "Smart Grid security: Deriving informed decisions from cyber attack game analysis," in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 946–951, 2014.

[39] P. Srikantha and D. Kundur, "A DER Attack-Mitigation Differential Game for Smart Grid Security Analysis," *IEEE Transactions on Smart Grid*, vol. 7, pp. 1476–1485, May 2016.

[40] Y. Li, L. Shi, P. Cheng, J. Chen, and D. E. Quevedo, "Jamming Attacks on Remote State Estimation in Cyber-Physical Systems: A Game-Theoretic Approach," *IEEE Transactions on Automatic Control*, vol. 60, pp. 2831–2836, Oct 2015.

[41] R. Deng, G. Xiao, and R. Lu, "Defending Against False Data Injection Attacks on Power System State Estimation," *IEEE Transactions on Industrial Informatics*, vol. 13, pp. 198–207, Feb 2017.

[42] S. Ghosh and C. Konstantinou, "A Bi-Level Differential Game-Based Load Frequency Control With Cyber-Physical Security," *IEEE Transactions on Smart Grid*, pp. 1–1, 2024.

[43] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.

[44] H. Jia, Y. Gai, and H. Zheng, "Network Recovery for Large-scale Failures in Smart Grid by Reinforcement Learning," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, pp. 2658–2663, Dec 2018.

[45] Y. Zhang, J. Wu, Z. Chen, Y. Huang, and Z. Zheng, "Sequential Node/Link Recovery Strategy of Power Grids Based on Q-Learning Approach," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, May 2019.

[46] F. Wei, Z. Wan, and H. He, "Cyber-Attack Recovery Strategy for Smart Grid Based on Deep Reinforcement Learning," *IEEE Transactions on Smart Grid*, pp. 1–1, 2019.

[47] J. Niu, Z. Ming, M. Qiu, H. Su, Z. Gu, and X. Qin, "Defending jamming attack in wide-area monitoring system for smart grid," *Telecommunication Systems*, vol. 60, no. 1, pp. 159–167, 2015.

[48] Cárdenas, Alvaro A. and Amin, Saurabh and Lin, Zong-Syun and Huang, Yu-Lun and Huang, Chi-Yen and Sastry, Shankar, "Attacks Against Process Control Systems: Risk Assessment, Detection, and Response," in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, ASIACCS '11, (New York, NY, USA), pp. 355–366, ACM, 2011.

[49] N. L. Ricker, "Model predictive control of a continuous, nonlinear, two-phase reactor," *Journal of Process Control*, vol. 3, no. 2, pp. 109–123, 1993.

[50] A. F. Murillo Piedrahita, V. Gaur, J. Giraldo, A. A. Cárdenas, and S. J. Rueda, "Leveraging Software-Defined Networking for Incident Response in Industrial Control Systems," *IEEE Software*, vol. 35, pp. 44–50, January 2018.

[51] A. Belmonte Martin, L. Marinos, E. Rekleitis, G. Spanoudakis, and N. Petroulakis, "Threat landscape and good practice guide for software defined networks/5g," 2015.

[52] D. Antonioli and N. Tippenhauer, "MiniCPS: A toolkit for security research on CPS networks," in *Proceedings of the First ACM workshop on cyber-physical systems-security and/or privacy*, pp. 91–100, 2015.

[53] R. Tan, H. H. Nguyen, E. Y. S. Foo, D. K. Y. Yau, Z. Kalbarczyk, R. K. Iyer, and H. B. Gooi, "Modeling and Mitigating Impact of False Data Injection Attacks on Automatic Generation Control," *IEEE Transactions on Information Forensics and Security*, vol. 12, pp. 1609–1624, July 2017.

[54] PowerWorld, 2016. [Online].

[55] S. D. Roy and S. Debbarma, "Detection and Mitigation of Cyber-Attacks on AGC Systems of Low Inertia Power Grid," *IEEE Systems Journal*, vol. 14, no. 2, pp. 2023–2031, 2020.

[56] M. Khalaf, A. Youssef, and E. El-Saadany, "Joint Detection and Mitigation of False Data Injection Attacks in AGC Systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 4985–4995, 2019.

[57] S. Alhalali, C. Nielsen, and R. El-Shatshat, "Mitigation of cyber-physical attacks in multi-area automatic generation control," *International Journal of Electrical Power & Energy Systems*, vol. 112, pp. 362–369, 2019.

[58] G. Wang, C. Wang, Q. Hao, and M. Shahidehpour, "Load Frequency Control Method for Cyber-Physical Power Systems With 100% Renewable Energy," *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 4684–4698, 2024.

[59] X. Wang, X. Luo, X. Pan, and X. Guan, "Detection and Location of Bias Load Injection Attack in Smart Grid via Robust Adaptive Observer," *IEEE Systems Journal*, vol. 14, no. 3, pp. 4454–4465, 2020.

[60] X. Wang, X. Luo, M. Zhang, Z. Jiang, and X. Guan, "Detection and Isolation of False Data Injection Attacks in Smart Grid via Unknown Input Interval Observer," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3214–3229, 2020.

[61] Z. Zhang, M. Easley, M. Hosseinzadehtaher, G. Amariucai, M. B. Shadmand, and H. Abu-Rub, "An Observer Based Intrusion Detection Framework for Smart Inverters at the Grid-Edge," in *2020 IEEE Energy Conversion Congress and Exposition (ECCE)*, pp. 1957–1962, 2020.

[62] K. Xiahou, Y. Liu, and Q. H. Wu, "Decentralized Detection and Mitigation of Multiple False Data Injection Attacks in Multiarea Power Systems," *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, vol. 3, no. 1, pp. 101–112, 2022.

[63] A. O. Aluko, D. G. Dorrell, and E. E. Ojo, "Observer-Based Detection and Mitigation Scheme for Isolated Microgrid Under False Data Injection Attack," in *2021 IEEE Southern Power Electronics Conference (SPEC)*, pp. 1–6, 2021.

[64] S. Tan, P. Xie, J. M. Guerrero, J. C. Vasquez, and R. Han, "Cyberattack Detection for Converter-Based Distributed dc Microgrids: Observer-Based Approaches," *IEEE Industrial Electronics Magazine*, vol. 16, no. 3, pp. 67–77, 2022.

[65] J. Yang, Q. Zhong, K. Shi, and S. Zhong, "Co-Design of Observer-Based Fault Detection Filter and Dynamic Event-Triggered Controller for Wind Power System Under Dual Alterable DoS Attacks," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1270–1284, 2022.

[66] S. Zhao, J. Xia, R. Deng, P. Cheng, and Q. Yang, "Adaptive Observer-Based Resilient Control Strategy for Wind Turbines Against Time-Delay Attacks on Rotor Speed Sensor Measurement," *IEEE Transactions on Sustainable Energy*, vol. 14, no. 3, pp. 1807–1821, 2023.

[67] A. Sargolzaei, K. Yazdani, A. Abbaspour, C. D. Crane III, and W. E. Dixon, "Detection and Mitigation of False Data Injection Attacks in Networked Control Systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4281–4292, 2020.

[68] S. Nateghi, Y. Shtessel, and C. Edwards, "Resilient control of cyber-physical systems under sensor and actuator attacks driven by adaptive sliding mode observer," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 15, pp. 7425–7443, 2021.

[69] A.-Y. Lu and G.-H. Yang, "Observer-Based Control for Cyber-Physical Systems Under Denial-of-Service With a Decentralized Event-Triggered Scheme," *IEEE Transactions on Cybernetics*, vol. 50, no. 12, pp. 4886–4895, 2020.

[70] J. Ye and X. Yu, "Detection and Estimation of False Data Injection Attacks for Load Frequency Control Systems," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 4, pp. 861–870, 2022.

[71] A. Abbaspour, A. Sargolzaei, P. Forouzannezhad, K. K. Yen, and A. I. Sarwat, "Resilient Control Design for Load Frequency Control System Under False Data Injection Attacks," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 9, pp. 7951–7962, 2020.

[72] X. Chen, S. Hu, Y. Li, D. Yue, C. Dou, and L. Ding, "Co-Estimation of State and FDI Attacks and Attack Compensation Control for Multi-Area Load Frequency Control Systems Under FDI and DoS Attacks," *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 2357–2368, 2022.

[73] H. H. Alhelou and P. Cuffe, "A Dynamic-State-Estimator-Based Tolerance Control Method Against Cyberattack and Erroneous Measured Data for Power Systems," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4990–4999, 2022.

[74] C. Chen, Y. Chen, J. Zhao, K. Zhang, M. Ni, and B. Ren, "Data-Driven Resilient Automatic Generation Control Against False Data Injection Attacks," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8092–8101, 2021.

[75] A. Ayad, M. Khalaf, M. Salama, and E. F. El-Saadany, "Mitigation of false data injection attacks on automatic generation control considering nonlinearities," *Electric Power Systems Research*, vol. 209, p. 107958, 2022.

[76] Y. Li, P. Zhang, and L. Ma, "Denial of service attack and defense method on load frequency control system," *Journal of the Franklin Institute*, vol. 356, no. 15, pp. 8625–8645, 2019.

[77] H. Bevrani, "Robust power system frequency control," 2014.

[78] J. Wei and G. J. Mendis, "A deep learning-based cyber-physical strategy to mitigate false data injection attack in smart grids," in *2016 Joint Workshop on Cyber- Physical Security and Resilience in Smart Grids (CPSR-SG)*, pp. 1–6, April 2016.

[79] Y. He, G. J. Mendis, and J. Wei, "Real-Time Detection of False Data Injection Attacks in Smart Grid: A Deep Learning-Based Intelligent Mechanism," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505–2516, 2017.

[80] K. Liao and Y. Xu, "A Robust Load Frequency Control Scheme for Power Systems Based on Second-Order Sliding Mode and Extended Disturbance Observer," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3076–3086, 2018.

[81] G. Anagnostou, F. Boem, S. Kuenzel, B. C. Pal, and T. Parisini, "Observer-Based Anomaly Detection of Synchronous Generators for Power Systems Monitoring," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4228–4237, 2018.

[82] M. Zeitz, "The extended luenberger observer for nonlinear systems," *Systems & Control Letters*, vol. 9, no. 2, pp. 149–156, 1987.

[83] Y. Guan and M. Saif, "A novel approach to the design of unknown input observers," *IEEE Transactions on Automatic Control*, vol. 36, no. 5, pp. 632–635, 1991.

[84] M. C. Turner and D. G. Bates, "Mathematical methods for robust and nonlinear control," *Book series on control systems. Springer, Berlin*, 2007.

[85] S. K. Spurgeon, "Sliding mode observers: a survey," *International Journal of Systems Science*, vol. 39, no. 8, pp. 751–764, 2008.

[86] C. Edwards, S. K. Spurgeon, C. P. Tan, and N. Patel, "Sliding-mode observers," *Mathematical Methods for Robust and Nonlinear Control: EPSRC Summer School*, pp. 221–242, 2007.

[87] E. Alpaydin, *Machine learning*. MIT press, 2021.

[88] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436–44, 05 2015.

[89] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[90] D. E. Rumelhart and J. L. McClelland, *Learning Internal Representations by Error Propagation*, pp. 318–362. 1987.

[91] A. Ng *et al.*, "Sparse autoencoder," *CS294A Lecture notes*, vol. 72, no. 2011, pp. 1–19, 2011.

[92] A. D. Syrmakesis and N. D. Hatziargyriou, "Cyber resilience methods for smart grids against false data injection attacks: categorization, review and future directions," *Frontiers in Smart Grids*, vol. 3, 2024.

[93] A. Huseinović, S. Mrdović, K. Bicakci, and S. Uludag, "A Survey of Denial-of-Service Attacks and Solutions in the Smart Grid," *IEEE Access*, vol. 8, pp. 177447–177470, 2020.

[94] Y. Li, R. Huang, and L. Ma, "Hierarchical-Attention-Based Defense Method for Load Frequency Control System Against DoS Attack," *IEEE Internet of Things Journal*, vol. 8, no. 20, pp. 15522–15530, 2021.

[95] X. Lou, C. Tran, R. Tan, D. K. Y. Yau, Z. T. Kalbarczyk, A. K. Banerjee, and P. Ganesh, "Assessing and Mitigating Impact of Time Delay Attack: Case Studies for Power Grid Controls," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 1, pp. 141–155, 2020.

[96] X.-C. Shangguan, Y. He, C.-K. Zhang, W. Yao, Y. Zhao, L. Jiang, and M. Wu, "Resilient Load Frequency Control of Power Systems to Compensate Random Time Delays and Time-Delay Attacks," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 5, pp. 5115–5128, 2023.

[97] Y. Wu, J. Weng, B. Qiu, Z. Wei, F. Qian, and R. H. Deng, "Random Delay Attack and Its Applications on Load Frequency Control of Power Systems," in *2019 IEEE Conference on Dependable and Secure Computing (DSC)*, pp. 1–8, 2019.

[98] A. Sargolzaei, K. Yen, and M. Abdelghani, "Delayed inputs attack on load frequency control in smart grid," in *ISGT 2014*, pp. 1–5, 2014.

[99] A. Sargolzaei, K. K. Yen, and M. Abdelghani, "Time-delay switch attack on load frequency control in smart grid," *Advances in Communication Technology*, vol. 5, pp. 55–64, 2013.

[100] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A Review of False Data Injection Attacks Against Modern Power Systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, 2017.

[101] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," vol. 14, jun 2011.

[102] A. S. Musleh, G. Chen, Z. Y. Dong, C. Wang, and S. Chen, "Attack Detection in Automatic Generation Control Systems using LSTM-Based Stacked Autoencoders," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 153–165, 2023.

[103] A. D. Syrmakesis, H. H. Alhelou, and N. D. Hatziargyriou, "A Novel Cyberattack-Resilient Frequency Control Method for Interconnected Power Systems Using SMO-based Attack Estimation," *IEEE Transactions on Power Systems*, pp. 1–13, 2023.

[104] X. Zhao, Z. Ma, X. Shi, and S. Zou, "Attack Detection and Mitigation Scheme of Load Frequency Control Systems Against False Data Injection Attacks," *IEEE Transactions on Industrial Informatics*, pp. 1–11, 2024.

[105] P. Anderson and M. Mirheydar, "A low-order system frequency response model," *IEEE Transactions on Power Systems*, vol. 5, no. 3, pp. 720–729, 1990.

[106] P. Kundur and N. Balu, *Power System Stability and Control*. EPRI power system engineering series, McGraw-Hill, 1994.

[107] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power generation, operation, and control*. John Wiley & Sons, 2013.

[108] F. Caliskan and I. Genc, "A Robust Fault Detection and Isolation Method in Load Frequency Control Loops," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1756–1767, 2008.

[109] A. M. A. Soliman, M. Bahaa Eldin, and M. Ahmed Mehanna, "Application of WOA Tuned Type-2 FLC for LFC of Two Area Power System With RFB and Solar Park Considering TCPS in Interline," *IEEE Access*, vol. 10, pp. 112007–112018, 2022.

[110] T. N. Pham, H. Trinh, and L. V. Hien, "Load Frequency Control of Power Systems With Electric Vehicles and Diverse Transmission Links Using Distributed Functional Observers," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 238–252, 2016.

[111] R. Abraham, D. Das, and A. Patra, "Effect of TCPS on oscillations in tie-power and area frequencies in an interconnected hydrothermal power system," *IET Generation, Transmission & Distribution*, vol. 1, pp. 632–639(7), July 2007.

[112] N. Pathak, A. Verma, T. S. Bhatti, and I. Nasiruddin, "Modeling of HVDC Tie Links and Their Utilization in AGC/LFC Operations of Multiarea Power Systems," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 3, pp. 2185–2197, 2019.

[113] Y. Dai, J. Phulpin, A. Sarlette, and D. Ernst, "Coordinated primary frequency control among non-synchronous systems connected by a multi-terminal high-voltage direct current grid," *IET Generation, Transmission & Distribution*, vol. 6, pp. 99–108(9), February 2012.

[114] M. Khudhair, M. Ragab, K. M. AboRas, and N. H. Abbasy, "Robust control of frequency variations for a multi-area power system in smart grid using a newly wild horse optimized combination of PIDD2 and PD controllers," *Sustainability*, vol. 14, no. 13, p. 8223, 2022.

[115] H. Haes Alhelou, M. E. H. Golshan, and N. D. Hatziargyriou, "A Decentralized Functional Observer Based Optimal LFC Considering Unknown Inputs, Uncertainties, and Cyber-Attacks," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4408–4417, 2020.

[116] H. Haes Alhelou, M. E. Hamedani Golshan, and M. Hajiakbari Fini, "Wind driven optimization algorithm application to load frequency control in interconnected power systems considering GRC and GDB nonlinearities," *Electric Power Components and Systems*, vol. 46, no. 11-12, pp. 1223–1238, 2018.

[117] A. D. Syrmakesis, H. H. Alhelou, and N. D. Hatziargyriou, "Novel SMO-Based Detection and Isolation of False Data Injection Attacks against Frequency Control Systems," *IEEE Transactions on Power Systems*, pp. 1–13, 2023.

[118] M. CORLESS and J. TU, "State and Input Estimation for a Class of Uncertain Systems," *Automatica*, vol. 34, no. 6, pp. 757–764, 1998.

[119] S. Hui and S. Żak, "Observer design for systems with unknown inputs," *International Journal of Applied Mathematics and Computer Science*, vol. 15, no. 4, pp. 431–446, 2005.

[120] W. Chen and F. N. Chowdhury, "A synthesized design of sliding-mode and luenberger observers for early detection of incipient faults," *International Journal of Adaptive Control and Signal Processing*, vol. 24, no. 12, pp. 1021–1035, 2010.

[121] H. Yang, S. Liu, and C. Fang, "Model-Based Secure Load Frequency Control of Smart Grids Against Data Integrity Attack," *IEEE Access*, vol. 8, pp. 159672–159682, 2020.

[122] W. Bi, K. Zhang, K. Yuan, Y. Wang, C. Chen, and K. Wang, "Observer-Based Attack Detection and Mitigation for Load Frequency Control System," in *2019 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–5, 2019.

[123] R. E. Larson, W. F. Tinney, and J. Peschon, "State Estimation in Power Systems Part I: Theory and Feasibility," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 3, pp. 345–352, 1970.

[124] M. Liu, C. Zhao, Z. Zhang, and R. Deng, "Explicit Analysis on Effectiveness and Hiddenness of Moving Target Defense in AC Power Systems," *IEEE Transactions on Power Systems*, pp. 1–1, 2022.

[125] C. Chen, K. Zhang, K. Yuan, L. Zhu, and M. Qian, "Novel Detection Scheme Design Considering Cyber Attacks on Load Frequency Control," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 5, pp. 1932–1941, 2018.

[126] A. Ameli, A. Hooshyar, A. H. Yazdavar, E. F. El-Saadany, and A. Youssef, "Attack Detection for Load Frequency Control Systems Using Stochastic Unknown Input Estimators," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 10, pp. 2575–2590, 2018.

[127] M. Ghiasi, M. Dehghani, T. Niknam, A. Kavousi-Fard, P. Siano, and H. H. Alhelou, "Cyber-attack detection and cyber-security enhancement in smart dc-microgrid based on blockchain technology and hilbert huang transform," *IEEE Access*, vol. 9, pp. 29429–29440, 2021.

[128] C. P. Tan and C. Edwards, "Sliding mode observers for robust fault detection & reconstruction," *IFAC Proceedings Volumes*, vol. 35, no. 1, pp. 347–352, 2002. 15th IFAC World Congress.

[129] C. P. Tan and C. Edwards, "Sliding mode observers for robust detection and reconstruction of actuator and sensor faults," *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal*, vol. 13, no. 5, pp. 443–463, 2003.

[130] C. Edwards and S. Spurgeon, *Sliding mode control: theory and applications*. Crc Press, 1998.

[131] T. Athay, R. Podmore, and S. Virmani, "A Practical Method for the Direct Analysis of Transient Stability," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-98, no. 2, pp. 573–584, 1979.

[132] A. Pai, *Energy function analysis for power system stability*. Springer Science & Business Media, 1989.

[133] "RTDS Technology Inc.." https://www.rtds.com/. Accessed: 2023-09-26.

[134] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.

[135] J. A. Boudreaux, *Design, simulation, and construction of an IEEE 14-bus power system*. Louisiana State University and Agricultural & Mechanical College, 2018.

[136] A. A. Saleh, T. Senjyu, S. Alkhalaf, M. A. Alotaibi, and A. M. Hemeida, "Water cycle algorithm for probabilistic planning of renewable energy resource, considering different load models," *Energies*, vol. 13, no. 21, p. 5800, 2020.

[137] D. Kite and R. Jenkins, "Automating Protection System Monitoring and Verification With the SEL RTAC [White Paper]," 2016.

[138] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response," in *Proceedings of the 6th ACM symposium on information, computer and communications security*, pp. 355–366, 2011.

[139] A. D. Syrmakesis, C. Alcaraz, and N. D. Hatziargyriou, "DAR-LFC: A data-driven attack recovery mechanism for Load Frequency Control," *International Journal of Critical Infrastructure Protection*, vol. 45, p. 100678, 2024.

[140] Y. Li, P. Zhang, and L. Ma, "Denial of service attack and defense method on load frequency control system," *Journal of the Franklin Institute*, vol. 356, no. 15, pp. 8625–8645, 2019.

[141] Y. Li, R. Huang, and L. Ma, "False Data Injection Attack and Defense Method on Load Frequency Control," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2910–2919, 2021.

[142] MATLAB, *version 9.4.0.813654 (R2018a)*. Natick, Massachusetts: The MathWorks Inc., 2018.

[143] I. Bello, B. Zoph, V. Vasudevan, and Q. V. Le, "Neural optimizer search with reinforcement learning," in *International Conference on Machine Learning*, pp. 459–468, PMLR, 2017.

[144] A. D. Syrmakesis, H. H. Alhelou, and N. D. Hatziargyriou, "A Novel Cyber Resilience Method for Frequency Control in Power Systems considering Nonlinearities and Practical Challenges," *IEEE Transactions on Industry Applications*, pp. 1–13, 2023.

[145] H. Haes Alhelou, M. Hamedani Golshan, and J. Askari-Marnani, "Robust sensor fault detection and isolation scheme for interconnected smart power systems in presence of RER and EVs using unknown input observer," *International Journal of Electrical Power & Energy Systems*, vol. 99, pp. 682–694, 2018.

[146] X.-G. Yan and C. Edwards, "Nonlinear robust fault reconstruction and estimation using a sliding mode observer," *Automatica*, vol. 43, no. 9, pp. 1605–1614, 2007.

[147] V. I. Utkin, *Sliding modes in control and optimization*. Springer Science & Business Media, 2013.

[148] H. Haes Alhelou, M. E. Hamedani Golshan, and N. D. Hatziargyriou, "Deterministic Dynamic State Estimation-Based Optimal LFC for Interconnected Power Systems Using Unknown Input Observer," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1582–1592, 2020.

Appendix A

Author's Publications

International Peer Reviewed Journals

- **A. D. Syrmakesis**, H. H. Alhelou and N. D. Hatziargyriou, "Novel SMO-Based Detection and Isolation of False Data Injection Attacks Against Frequency Control Systems," in IEEE Transactions on Power Systems, vol. 39, no. 1, pp. 1434-1446, Jan. 2024, doi: 10.1109/TPWRS.2023.3242015.

- **A. D. Syrmakesis**, H. H. Alhelou and N. D. Hatziargyriou, "A Novel Cyberattack-Resilient Frequency Control Method for Interconnected Power Systems Using SMO-Based Attack Estimation," in IEEE Transactions on Power Systems, vol. 39, no. 4, pp. 5672-5686, July 2024, doi: 10.1109/TPWRS.2023.3340744.

- **A. D. Syrmakesis**, H. H. Alhelou and N. D. Hatziargyriou, "A Novel Cyber Resilience Method for Frequency Control in Power Systems Considering Nonlinearities and Practical Challenges," in IEEE Transactions on Industry Applications, vol. 60, no. 2, pp. 2176-2190, March-April 2024, doi: 10.1109/TIA.2023.3332702.

- **Andrew D. Syrmakesis**, Cristina Alcaraz, Nikos D. Hatziargyriou, "DAR-LFC: A data-driven attack recovery mechanism for Load Frequency Control", in International Journal of Critical Infrastructure Protection, Volume 45, 2024, 100678, ISSN 1874-5482, https://doi.org/10.1016/j.ijcip.2024.100678.

- **A. D. Syrmakesis**, C. Alcaraz, and N. D. Hatziargyriou, "Classifying resilience approaches for protecting smart grids against cyber threats," in International Journal of Information Security, pp. 1–22, 2022.

- **A. D. Syrmakesis** and N. D. Hatziargyriou, "Cyber resilience methods for smart grids against false data injection attacks: categorization, review and future directions," Frontiers in Smart Grids, vol. 3, 2024.

International Conferences

- Dimitropoulos V., **Syrmakesis A.D.**, Hatziargyriou N., "DRL^{2}FC: An Attack-Resilient Controller for Automatic Generation Control Based on Deep Reinforcement Learning," submitted to the 14th Mediterranean Conference on Power Generation Transmission, Distribution and Energy Conversion.

Appendix B

Theorem Proofs

Proof of Lemma 2.

Proof. If and only if:

$$rank\begin{bmatrix} sI - A_0 \\ C_0 \end{bmatrix} = rank\begin{bmatrix} sI - A_4 & \underline{0} \\ -C_4 & sI \\ \underline{0} & I \end{bmatrix} = n + p - 2r \tag{B.1}$$

the pair (A_0, C_0) is observable for every $s \in \mathbb{C}$, based on the Popov–Belevitch–Hautus test that follows.

If $s = 0$, it is obtained:

$$rank\begin{bmatrix} sI - A_4 & \underline{0} \\ -C_4 & sI \\ \underline{0} & I \end{bmatrix} = rank\begin{bmatrix} -A_4 \\ -C4 \end{bmatrix} + p - r.$$

From Lemma 1, the (A_4, C_4) is detectable and therefore:

$$rank\begin{bmatrix} sI - A_4 \\ -C_4 \end{bmatrix} = n - r \; \forall \; s \in \mathbb{C}.$$

Therefore, the rank test (B.1) holds when $s = 0$.

Furthermore, since (A_4, C_4) is detectable, when $s \neq 0$, we have:

$$\begin{bmatrix} sI - A_4 & 0 \\ -C_4 & sI \\ 0 & I \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \underline{0} \Rightarrow (a_1, a_2) = (0, 0).$$

This indicates that the columns of the above matrix are linearly independent and its rank is $n+p-2r$.

This completes the proof. □

Proof of Theorem 4.3.1.

Proof. Assume that $V_1(e_1) = e_1^T P_1 e_1$ and $V_0(e_0) = e_0^T P_0 e_0$. The function $V(e_1, e_0) = V_1(e_1) + V_0(e_0)$ is selected as the Lyapunov candidate. For the derivative of V_1, it is obtained:

$$\dot{V}_1 = e_1^T (A_1^{s^T} P_1 + P_1 A_1^s) e_1 + 2e_1^T P_1 \bar{A}_2 e_0 + 2e_1^T P_1 E_1 d + 2e_1^T P_1 (F_1 \phi(G^{-1} g, t) - F_1 \phi(G^{-1} \hat{g}, t)) - \\ - 2e_1^T P_1 v_1.$$

Based on [146], the inequality $2X^T Y \leq \frac{1}{\alpha} X^T X + \alpha Y^T Y$ is met for any positive scalar α. Thus:

$$\begin{aligned} \dot{V}_1 \leq\ & e_1^T (A_1^{s^T} P_1 + P_1 A_1^s) e_1 + 2e_1^T P_1 \bar{A}_2 e_0 + 2e_1^T P_1 E_1 d + \\ & + \alpha_1 \left(F_1 \phi(G^{-1} g, t) - F_1 \phi(F^{-1} \hat{g}, t)\right)^T \left(F_1 \phi(G^{-1} g, t) - F_1 \phi(F^{-1} \hat{g}, t)\right) + \\ & + \frac{1}{\alpha_1} e_1^T P_1 P_1^T e_1 - 2e_1^T P_1 v_1. \end{aligned} \tag{B.2}$$

Since $\hat{\zeta} := [(C_1^{-1}\omega_1)^T, (\hat{\zeta}_0)^T]^T$, it is true that $\zeta - \hat{\zeta} = \begin{bmatrix} 0 \\ e_2 \end{bmatrix}$ before the launch of FDIAs. This yields that $\|G^{-1} g - G^{-1} \hat{g}\| = \|G^{-1} e_2\| \leq \|G^{-1} e_0\|$. Hence:

$$\begin{aligned} &\|F_1 \phi(G^{-1} g, t) - F_1 \phi(G^{-1} \hat{g}, t)\| \leq \|F_1\| \mathscr{L}_\phi \|G^{-1}\| \|e_0\|, \text{ and} \\ &\|F_2 \phi(G^{-1} g, t) - F_2 \phi(G^{-1} \hat{g}, t)\| \leq \|F_2\| \mathscr{L}_\phi \|G^{-1}\| \|e_0\|. \end{aligned}$$

Furthermore, from the definition of v_1 it is derived that:

$$e_1^T P_1 v_1 = (\|E_1\| \xi + \eta_1) \|P_1 e_1\|.$$

Then, inequality (B.2) can be simplified as:

$$\begin{aligned} \dot{V}_1 \leq\ & e_1^T (A_1^{s^T} P_1 + P_1 A_1^s) e_1 + 2e_1^T P_1 \bar{A}_2 e_0 + \frac{1}{\alpha_1} e_1^T P_1 P_1 e_1 + \alpha_1 \|F_1\|^2 \mathscr{L}_\phi^2 \|G^{-1}\|^2 \|e_0\|^2 + \\ & + 2\|E_1\| \xi \|P_1 e_1\| - 2(\|E_1\| \xi + \eta_1) \|P_1 e_1\| \\ \leq\ & e_1^T (A_1^{s^T} P_1 + P_1 A_1^s) e_1 + 2e_1^T P_1 \bar{A}_2 e_0 + \frac{1}{\alpha_1} e_1^T P_1 P_1 e_1 + \alpha_1 \|F_1\|^2 \mathscr{L}_\phi^2 \|G^{-1}\|^2 \|e_0\|^2. \end{aligned} \tag{B.3}$$

In a similar manner, for the derivative of V_0 it can be proven that:

$$\begin{aligned}\dot{V}_0 &= e_0^T\big((A_0-L_0C_0)^TP_0+P_0(A_0-L_0C_0)\big)e_0+2e_0^TP_0(\bar{F}_2\phi(G^{-1}g,t)-\bar{F}_2\phi(G^{-1}\hat{g},t))\\ &\leq e_0^T\big((A_0-L_0C_0)^TP_0+P_0(A_0-L_0C_0)\big)e_0+\frac{1}{\alpha_0}e_0^TP_0P_0e_0+\alpha_0\|F_2\|^2\mathscr{L}_\phi^2\|G^{-1}\|^2\|e_0\|^2.\end{aligned} \tag{B.4}$$

The combination of (B.3) and (B.4) indicates that:

$$\dot{V}=\dot{V}_1+\dot{V}_0\leq\begin{bmatrix}e_1\\e_0\end{bmatrix}^T\Lambda\begin{bmatrix}e_1\\e_0\end{bmatrix}.$$

If there are matrices $P_1=P_1^T>0$, $P_0=P_0^T$, $A_1^s<0$ and L_0 and positive scalars α_0 and α_1 so that inequality (4.13) is true, then $\dot{V}<0$ for any $e\neq 0$, where $e=\begin{bmatrix}e_1\\e_0\end{bmatrix}$. This proves the stability of the error dynamics (4.11) and (4.12) and completes the proof. □

Proof of Theorem 4.3.2.

Proof. Consider function $V_1=e_1^TP_1e_1$ as the Lyapunov candidate. The time derivative of V_1 is:

$$\begin{aligned}\dot{V}_1 = {} & e_1^T(A_1^{sT}P_1+P_1A_1^s)e_1+2e_1^TP_1\bar{A}_2e_0+2e_1^TP_1E_1d+2e_1^TP_1(F_1\phi(G^{-1}g,t)-F_1\phi(G^{-1}\hat{g},t))-\\ & -2e_1^TP_1v_1.\end{aligned}$$

Since A_1^s is a stable matrix, the term $A_1^{sT}P_1+P_1A_1^s<0$ by definition. From Eq. (4.9), it is obtained:

$$\begin{aligned}\dot{V}_1 &\leq 2\|P_1e_1\|\big((\|\bar{A}_2\|+\|F_1\|\mathscr{L}_\phi\|G^{-1}\|)\|e_0\|-\eta_1\big)\\ &\leq -2\eta_2\|P_1e_1\|\\ &\leq -2\eta_2\sqrt{\lambda_{min}(P_1)}\sqrt{V_1}.\end{aligned}$$

This indicates that the reachability condition [147] is met and a sliding motion is achieved and preserved after a finite time frame.

This completes the proof.

□

Proof of Theorem 4.4.1.

Proof. Let $V_1^l = {e_1^l}^T P_1 e_1^l$ and $V_0^l = {e_0^l}^T P_0 e_0^l$. The selected Lyapunov candidate function is $V^l = V_1^l + V_0^l$. The time derivative of V_1^l is:

$$\begin{aligned}\dot{V}_1^l &= {e_1^l}^T (A_1^{s^T} P_1 + P_1 A_1^s) e_1^l + 2{e_1^l}^T P_1 \bar{A}_2 e_0^l + 2{e_1^l}^T P_1 E_1 d + \\ &\quad + 2{e_1^l}^T P_1 \big(F_1\phi(G^{-1}g,t) - F_1\phi(G^{-1}\hat{g}^l,t)\big) - 2{e_1^l}^T P_1 v_1 \\ &\leq {e_1^l}^T \Pi_1 e_1^l + 2{e_1^l}^T P_1 \bar{A}_2 e_0^l + \frac{1}{\alpha_1} {e_1^l}^T P_1 P_1 e_1^l + \alpha_1 \mathscr{L}_{\phi_1}^2 \|G^{-1}\|^2 \|e_0^l\|^2.\end{aligned}$$

If $a_m^l = 0$, the error dynamics (4.21) are converted into the following:

$$\dot{e}_0^l = (A_0 - L_0 C_0) e_0^l + \bar{F}_2\phi(G^{-1}g,t) - \bar{F}_2\phi(G^{-1}\hat{g}^l,t) + \bar{D}_0^l(\bar{a}_m^l - v_2^l).$$

If (4.22) is satisfied, the derivative of V_0^l becomes:

$$\begin{aligned}\dot{V}_0^l &= {e_0^l}^T \Pi_0 e_0^l + 2{e_0^l}^T P_0 \big(\bar{F}_2\phi(G^{-1}g,t) - \bar{F}_2\phi(G^{-1}\hat{g}^l,t)\big) + 2{e_0^l}^T P_0 \bar{D}_0^l(\bar{a}_m^l - v_2^l) \\ &\leq {e_0^l}^T \Pi_0 e_0^l + 2{e_0^l}^T P_0 \big(\bar{F}_2\phi(G^{-1}g,t) - \bar{F}_2\phi(G^{-1}\hat{g}^l,t)\big) - 2\eta_3 \|\bar{F}_0^l e_{\omega_3}^l\| \\ &\leq {e_0^l}^T \Pi_0 e_0^l + 2{e_0^l}^T P_0 \big(\bar{F}_2\phi(G^{-1}g,t) - \bar{F}_2\phi(G^{-1}\hat{g}^l,t)\big) \\ &\leq {e_0^l}^T \Pi_0 e_0^l + \frac{1}{\alpha_0} {e_0^l}^T P_0 P_0 e_0^l + \alpha_0 \mathscr{L}_{\phi_2}^2 \|G^{-1}\|^2 \|e_0^l\|^2.\end{aligned}$$

If there are matrices $L_0, A_1^s < 0$, $P_0 = P_0^T > 0$, $P_1 = P_1^T > 0$ and F_0, and scalars $\alpha_0 > 0$ and $\alpha_1 > 0$ so that inequality (4.23) holds true, then we have:

$$\dot{V}^l = \dot{V}_1^l + \dot{V}_0^l < 0.$$

This indicates that if $a_m^l = 0$, then $\lim_{t\to\infty} e_0^l = 0$, despite the successful launch of FDIAs $a_m^j \neq 0,\ j \in \{1,2,\ldots,q\} \setminus \{l\}$. On the other hand, if $a_m^l, \neq 0$, then $\lim_{t\to\infty} e_0^l \neq 0$ since D_0 is of full column rank and the term $\bar{D}_0^l(\bar{a}_m^l - v_2^l)$ in (4.21) cannot eliminate the $D_0^l a_m^l$.

This completes the proof. □

Proof of Theorem 5.3.1.

Proof. The considered Lyapunov function is:

$$V = V_1 + V_2 + V_3 + V_4,$$

where $V_1 = e_1^T P_1 e_1$, $V_2 = \bar{e}_2^T P_2 \bar{e}_2$, $V_3 = \frac{e_{k_1}^2}{2l_{k_1}}$, $V_4 = \frac{e_{k_2}^2}{2l_{k_2}}$, $e_{k_1} = k_1 - \hat{k}_1$ and $e_{k_2} = k_2 - \hat{k}_2$. k_1 and k_2 are two positive constants that can be computed using (B.7).

For the time derivative of V_1, we have:

$$\dot{V}_1 = e_1^T(A_1^{s^T}P_1 + P_1A_1^s)e_1 + 2e_1^TP_1\bar{A}_2\bar{e}_2 + 2e_1^TP_1E_1d + 2e_1^TP_1F_1\left(\phi(T^{-1}\zeta,t) - \phi(T^{-1}\hat{\zeta},t)\right) + \\ + 2e_1^TP_1B_1(a_c - v) - \hat{k}_1\|F_1^TP_1e_1\|^2.$$

From Assumption 6 and $\hat{\zeta} := col(C_1^{-1}S_1y, \begin{bmatrix} I_{n-m} & 0 \end{bmatrix}\hat{\bar{\zeta}}_2)$, it is inferred that:

$$\|\phi(T^{-1}\zeta,t) - \phi(T^{-1}\hat{\zeta},t)\| \leq \mathscr{L}_\phi\|T^{-1}\|\|\bar{e}_2\|.$$

Based on [146], the inequality $2X^TY \leq \frac{1}{\alpha}X^TX + \alpha Y^TY$ holds true for any scalar $\alpha > 0$, thus:

$$\begin{aligned}\dot{V}_1 &\leq e_1^T(A_1^{s^T}P_1 + P_1A_1^s)e_1 + 2e_1^TP_1\bar{A}_2\bar{e}_2 + 2e_1^TP_1E_1d + \frac{1}{\alpha_1}e_1^TP_1F_1F_1^TP_1e_1 + 2e_1^TP_1B_1(a_c - v) + \\ &+ \alpha_1\left(\phi(T^{-1}\zeta,t) - \phi(T^{-1}\hat{\zeta},t)\right)^T\left(\phi(T^{-1}\zeta,t) - \phi(T^{-1}\hat{\zeta},t)\right) - \hat{k}_1\|F_1^TP_1e_1\|^2 \\ &\leq e_1^T(A_1^{s^T}P_1 + P_1A_1^s + \frac{1}{\alpha_1}P_1F_1F_1^TP_1)e_1 + 2e_1^TP_1\bar{A}_2\bar{e}_2 + 2e_1^TP_1E_1d + \alpha_1\mathscr{L}_\phi^2\|T^{-1}\|^2\|\bar{e}_2\|^2 + \\ &+ 2e_1^TP_1B_1(a_c - v) - \hat{k}_1\|F_1^TP_1e_1\|^2.\end{aligned}$$

Using (5.7), it can be proven that:

$$e_1^TP_1B_1(a_c - v) = e_1^TP_1B_1a_c - (\rho + \eta)\frac{\|B_1^TP_1e_1\|^2}{\|B_1^TP_1e_1\|} \leq -\eta\|B_1^TP_1e_1\| < 0.$$

Therefore:

$$\dot{V}_1 \leq e_1^T(A_1^{s^T}P_1 + P_1A_1^s)e_1 + \frac{1}{\alpha_1}e_1^TP_1F_1F_1^TP_1e_1 + \alpha_1\mathscr{L}_\phi^2\|T^{-1}\|^2\|\bar{e}_2\|^2 + 2e_1^TP_1\bar{A}_2\bar{e}_2 + 2e_1^TP_1E_1d - \\ - \hat{k}_1\|F_1^TP_1e_1\|^2.$$

Selecting $\alpha_1 = \frac{1}{\mathscr{L}_\phi^2\|T^{-1}\|^2}$, it follows that:

$$\dot{V}_1 \leq e_1^T(A_1^{s^T}P_1 + P_1A_1^s)e_1 + 2e_1^TP_1\bar{A}_2\bar{e}_2 + 2e_1^TP_1E_1d \ + (\mathscr{L}_\phi^2\|T^{-1}\|^2 - \hat{k}_1)\|F_1^TP_1e_1\|^2 + \|\bar{e}_2\|^2. \tag{B.5}$$

From (5.15), the time derivative of V_2 is determined as follows:

$$\begin{aligned}\dot{V}_2 &= \bar{e}_2^T(P_2F_0 + F_0^TP_2)\bar{e}_2 + 2\bar{e}_2^TP_2M_0\bar{E}_2d - \hat{k}_2\bar{e}_2^TP_2M_0\bar{F}_2H_0\bar{C}_4\bar{e}_2 + \\ &\quad + 2\bar{e}_2^TP_2M_0\bar{F}_2\big(\phi(T^{-1}\zeta,t) - \phi(T^{-1}\hat{\zeta},t)\big) \\ &\leq \bar{e}_2^T(P_2F_0 + F_0^TP_2)\bar{e}_2 + \frac{1}{\alpha_2}\bar{e}_2^TP_2M_0\bar{F}_2\bar{F}_2^TM_0^TP_2\bar{e}_2 + \alpha_2\mathscr{L}_\phi^2\|T^{-1}\|^2\|\bar{e}_2\|^2I_{n+q-m} + 2\bar{e}_2^TP_2M_0\bar{E}_2d - \\ &\quad - \hat{k}_2\|\bar{F}_2^TM_0^TP_2\bar{e}_2\|^2.\end{aligned}$$

If $\alpha_2 = \frac{1}{\mathscr{L}_\phi^2\|T^{-1}\|^2}$, then:

$$\dot{V}_2 \leq \ \bar{e}_2^T(P_2F_0 + F_0^TP_2)\bar{e}_2 + 2\bar{e}_2^TP_2M_0\bar{E}_2d + \|\bar{e}_2\|^2 \ + (\mathscr{L}_\phi^2\|T^{-1}\|^2 - \hat{k}_2)\|\bar{F}_2^TM_0^TP_2\bar{e}_2\|^2. \tag{B.6}$$

By defining:

$$k_1 = k_2 = \mathscr{L}_\phi^2\|T^{-1}\|^2, \tag{B.7}$$

the derivatives of V_3 and V_4 with respect to time are:

$$\dot{V}_3 = \frac{2e_{k_1}\dot{e}_{k_1}}{2l_{k_1}} = e_{k_1}\frac{\dot{k}_1 - \dot{\hat{k}}_1}{l_{k_1}} = e_{k_1}\frac{-l_{k_1}\|F_1^TP_1e_1\|^2}{l_{k_1}} = -e_{k_1}\|F_1^TP_1e_1\|^2 \text{ and} \tag{B.8}$$

$$\dot{V}_4 = \frac{2e_{k_2}\dot{e}_{k_2}}{2l_{k_2}} = e_{k_2}\frac{\dot{k}_2 - \dot{\hat{k}}_2}{l_{k_2}} = e_{k_2}\frac{-l_{k_2}\|H_0(\omega_2 - \hat{\omega}_2)\|^2}{l_{k_2}} = -e_{k_2}\|\bar{F}_2^TM_0^TP_2\bar{e}_2\|^2. \tag{B.9}$$

From (B.5), (B.6), (B.8) and (B.9), the time derivative of V can be obtained as:

$$\dot{V} = \dot{V}_1 + \dot{V}_2 + \dot{V}_3 + \dot{V}_4 \leq \begin{bmatrix} e_1 \\ \bar{e}_2 \end{bmatrix}^T \begin{bmatrix} \Pi_1 & P_1\bar{A}_2 \\ \bar{A}_2^TP_1 & \Pi_2 \end{bmatrix} \begin{bmatrix} e_1 \\ \bar{e}_2 \end{bmatrix} + 2e_1^TP_1E_1d + 2\bar{e}_2^TP_2M_0\bar{E}_2d.$$

For the case where $d = 0$, if (5.15) is feasibly solvable, then $\begin{bmatrix} \Pi_1 & P_1\bar{A}_2 \\ \bar{A}_2^TP_1 & \Pi_2 \end{bmatrix} < 0$ and thus, $\dot{V} < 0$. This indicates that $\lim_{t\to\infty} e(t) = 0$ and therefore, the error dynamics are asymptotically stable.

When $d \neq 0$, to make the proposed observers robust against disturbances d in L_2 sense, we define:

$$V_0 = \dot{V} + r^Tr - \mu d^Td.$$

The satisfaction of (5.15), infers that:

$$V_0 \leq \begin{bmatrix} e_1 \\ \bar{e}_2 \\ d \end{bmatrix}^T \Lambda \begin{bmatrix} e_1 \\ \bar{e}_2 \\ d \end{bmatrix} < 0$$

Then, under zero initial conditions, we have:

$$\begin{aligned} &\int_0^\infty (\|r\|^2 - \mu\|d\|^2)\,dt = \int_0^\infty (\|r\|^2 - \mu\|d\|^2 + \dot{V})\,dt - \int_0^\infty \dot{V}\,dt = \\ &= \int_0^\infty (\|r\|^2 - \mu\|d\|^2 \ + \dot{V})\,dt - V(\infty) + V(0) \\ &\leq \int_0^T V_0\,dt < 0 \end{aligned}$$

which implies that:

$$\int_0^T (r^T r)\,dt \leq \mu \int_0^T (d^T d)\,dt \Rightarrow \|r\|_{\mathscr{L}_2} \leq \sqrt{\mu}\|d\|_{\mathscr{L}_2}.$$

This completes the proof. □

Proof of Theorem 5.3.2.

Proof. For the Lyapunov candidate function $V_1 = e_1^T P_1 e_1$, we have:

$$\begin{aligned} \dot{V}_1 &= e_1^T (A_1^{sT} P_1 + P_1 A_1^s) e_1 + 2e_1^T P_1 \bar{A}_2 \bar{e}_2 + 2e_1^T P_1 E_1 d + 2e_1^T P_1 F_1 \left(\phi(T^{-1}\zeta, t) - \phi(T^{-1}\hat{\zeta}, t)\right) + \\ &\quad + 2e_1^T P_1 B_1 (a_c - v) - \hat{k}_1 \|F_1^T P_1 e_1\|^2 \\ &\leq e_1^T (A_1^{sT} P_1 + P_1 A_1^s) e_1 + 2e_1^T P_1 \bar{A}_2 \bar{e}_2 + 2e_1^T P_1 E_1 d + 2e_1^T P_1 F_1 \left(\phi(T^{-1}\zeta, t) - \phi(T^{-1}\hat{\zeta}, t)\right) + \\ &\quad + 2e_1^T P_1 B_1 (a_c - v). \end{aligned} \tag{B.10}$$

Since A_1^s is a Hurwiz matrix by definition, it can be easily concluded that $A_1^{sT} P_1 + P_1 A_1^s < 0$. Therefore, combining (5.7), the Cauchy-Schwartz inequality and the existence of B_1^{-1}, (B.10) becomes:

$$\begin{aligned} \dot{V}_1 &\leq 2\|P_1 e_1\| \left(\|\bar{A}_2\|\|\bar{e}_2\| + \mathscr{L}_\phi \|W_1\|\|T^{-1}\|\|\bar{e}_2\| + \|E_1\|\xi\right) - 2\eta\|B_1^T P_1 e_1\| \\ &\leq 2\|B_1^T P_1 e_1\| \left(\|B_1^{-1}\|(\|\bar{A}_2\|\varepsilon + \mathscr{L}_\phi \|W_1\|\|T^{-1}\|\varepsilon + \|E_1\|\xi) - \eta_1\right) \end{aligned}$$

This implies that the reachability condition [147] holds true, resulting in an ideal and permanent sliding motion within a finite period.

This completes the proof. □

Appendix C

Parameter Values

The power system parameter values of each area i that are used for the simulations, are [148]: $2H_i = 0.1667$ p.u. s, $D_i = 0.0083$ p.u./Hz, $T_{t_i} = 0.3$ s, $T_{g_i} = 0.08$ s, $R_i = 2.4$ Hz/p.u., $T_{ij} = T_{ji} = 0.026$ p.u./Hz, $\beta_i = 0.425$ p.u./Hz, $K_{ij} = 1$ p.u./Hz, $T_{dc_i} = 0.2$ s, $K_{s_{ij}} = 1$ s, $T_{s_{ij}} = 0.1$ s.

The values of the LFC nonlinearities for each area i, used in the simulations, are: $P_{GDB} = 1‰$ p.u., $P_{GRC} = 10\%$ p.u./min, $\tau_d = 1$ s.

For case study 1, the resulting SMO matrices are the following:

$$P_1 = \begin{bmatrix} 0.045548 & 0.007735 \\ 0.007735 & 0.045548 \end{bmatrix}$$

$$L_0 = \begin{bmatrix} -54370.083 & -18112.592 & -20166.878 \\ -81.039 & -25.238 & -41.517 \\ 328.126 & 104.133 & 155.328 \\ -330.883 & -99.256 & -195.751 \\ 69968.844 & 19977.54 & 47839.577 \\ 11.115 & 3.58 & 5.397 \\ 493.482 & 147.434 & 293.969 \\ 53453.184 & 17831.136 & 19669.277 \\ -69284.891 & -19772.1 & -47439.241 \\ -422.065 & -134.207 & -198.96 \end{bmatrix}$$

$$N_0 = \begin{bmatrix} 1 & 0.033 & -585.812 \\ 0 & 0.748 & -2.035 \\ 0 & -0.465 & 6.947 \end{bmatrix}$$

$$M_0 = \begin{bmatrix} 328.836 & 114.433 & 89.826 \\ 0.515 & 0.176 & 0.232 \\ -1.915 & -0.57 & -0.741 \end{bmatrix}$$